FOUNDATIONS OF PHARMACOKINETICS

ALDO RESCIGNO

FOUNDATIONS OF PHARMACOKINETICS

This page intentionally left blank

FOUNDATIONS OF PHARMACOKINETICS

ALDO RESCIGNO

University of Minnesota
Minneapolis, Minnesota

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-306-47924-9
Print ISBN: 0-306-47704-1

Created in the United States of America

Visit Kluwer Online at: http://kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

To my wife Luisa

This page intentionally left blank

Nec sum animi dubius, verbis ea vincere magnum
quam sit et angustis hunc addere rebus honorem.

Virgil, Georgicon, III, 289.

This page intentionally left blank

PREFACE

This book has its origin in my experience as a teacher of pharmacokinetics in many universities in four different continents.

It was not my intention to write a popular book; what distinguishes this one from many others on the same subject is its large use of algebra and calculus. For this I make no apologies; in fact a serious study of pharmacokinetics without the help of mathematics is, in my opinion, impossible. The exact definition of many pharmacokinetic quantities, even the most common, and the correct use of many equations, even the most simple, requires the constant use of mathematical language.

On the other hand I have made a considerable effort to use only elementary algebra and elementary calculus, as commonly taught in most introductory university courses. For the few exceptions, when less common mathematical concepts were needed, I have supplied the necessary explanations in four appendices.

The first three chapters are a general introduction to the scientific method.

Chapters 4 to 12 show different specific methods to deal with pharmacokinetic problems. There is considerable overlap among those chapters; this is intentional and its purpose is to convince the reader that every problem can be solved in more than one way, including ways that were not mentioned in this book and that intelligent readers can find for their own pleasure.

Chapters 13 to 17 show how different parameters of importance in pharmacokinetics can be exactly defined and measured.

The four appendices deal with a few mathematical concepts that are less frequently taught in introductory courses on algebra and calculus. The interested reader can study them in detail, or simply use their results as indicated by the cross-references.

I owe a big debt of gratitude to my late friend Giorgio Segre who introduced me to the study of pharmacokinetics. Many years ago we wrote together a monograph on "Drug and Tracer Kinetics", now out of print; this volume is, in a sense, the continuation and conclusion of that monograph.

For the preparation of this book I had no grants or financial aid on any sort, just the help of a few friends, most notably Dr. James S. Beck of Calgary, Alberta, and Dr. Marta Farolfi of Solarolo, Italy; to them go all my thanks.

This page intentionally left blank

CONTENTS

1. INTRODUCTION

1.1. DEFINITION OF PHARMACOKINETICS

The term "Pharmacokinetics" was first used by Dost [1] to mean the quantitative study of absorption, distribution, metabolism, and elimination of drugs, but pharmacologists have always been aware that drugs are absorbed, distributed, metabolized, and eliminated from the organism, and that the rates of absorption, distribution, metabolism, and elimination are fundamental in determining the effects on the organism they are administered to. In a sense, pharmacokinetics has always been a part of pharmacology; but it began to be considered a new discipline when more sophisticated methods were introduced to study the kinetic properties of drugs. On one side analytical chemistry and physical chemistry make now possible the detection and measurement of drugs in various parts of the organism in very low concentrations, and on the other side applied mathematics, statistics and physics make possible the systematic organization of the observed facts in a wide variety of circumstances.

1.2. OBJECTIVES OF PHARMACOKINETICS

Three different objectives can be observed in pharmacokinetics: a prescriptive one, which we can call passive pharmacokinetics, a descriptive one, which we can call active pharmacokinetics, and finally a predictive one, which we can call creative pharmacokinetics [2].

1.2.1. Passive Pharmacokinetics

By passive pharmacokinetics I mean the use of known pharmacokinetic concepts and previously determined pharmacokinetic parameters (see chapter 3) for the prediction of absorption, distribution, metabolism, and elimination of drugs administered in a given way (time or formulation) to living beings. The use of these known concepts and parameters give to the clinical pharmacokineticist the possibility to achieve a clinical goal. By using known facts about pharmacokinetics we can achieve a clinical goal (a particular concentration of a drug in a particular part of the body) by appropriate choice of dose, mode and site of administration, and of formulation of the drug and its vehicle among the available ones. This is a technological application—both common and important.

1.2.2. Active Pharmacokinetics

By active pharmacokinetics I mean the determination of pharmacokinetic parameters of new drugs or new formulations for their best utilization.

Each new molecule and new preparation of pharmacological interest need to be investigated for its pharmacokinetic properties; this is the routine job of the pharmacokineticist, and will never end as long as new substances will be added to the list of useful or potentially useful drugs, or new formulations are requested or proposed.

1.2.3. Creative Pharmacokinetics

I call creative pharmacokinetics, for lack of a better term, the study of new methods for the determination of pharmacokinetic parameters, and the critical study of old methods in order to evaluate their validity and applicability to the investigation of new drugs and new formulations.

An example of creative pharmacokinetics is the constant work of FDA's Committees about generic drugs, to establish guidelines for BA and BE determination studies.

1.2.4. Rhetorical Pharmacokinetics

A fourth aspect of pharmacokinetics can be added, that doesn't correspond to any of the objectives stated above. We can call it *rhetorical pharmacokinetics*: it is neither passive, active, or creative pharmacokinetics. Its objective is not to make predictions, neither explanations; it is a form of academic gamesmanship and as such is not a desirable activity. I shall show some examples of rhetorical pharmacokinetics in the next chapter.

1.3. METHODS OF PHARMACOKINETICS

As any other experimental science, pharmacokinetics has two inseparable components: the observation of facts, and the interpretation of the observations. I shall call the first component, *descriptive pharmacokinetics,* and the second one, *analytical pharmacokinetics*; obviously these two components are strictly interwoven: we can interpret only facts that can be observed, and we observe only facts that we think can be interpreted.

The strength of pharmacokinetics is due to its constant use of the *experimental method.* The experimental method is not just the acceptance of observed facts, but the acceptance of statements about facts that can be verified by experimental observations.

I will show with more details in the next chapter the meaning of the relationship

Observed facts ↔ Statement about facts.

1.4. FUTURE OF PHARMACOKINETICS

As all sciences, pharmacokinetics may progress in three different fields, namely the solution of new problems, the elaboration of new methods, and the development of new symbolism.

I have shown in section 1.1 that the investigation of pharmacokinetic properties needs the help of several other sciences, for instance analytical chemistry, physical biochemistry, numerical analysis. The methods of these sciences will progress concurrently making the solution of particular problems easier or more precise.

The third aspect in the progress of all sciences is often overlooked; it consists in the development of a new or an improved symbolism. Algebra never developed in a substantial way in Greece by reason of lack of an appropriate notation, while the Chinese were able to solve algebraic equations several centuries before the Europeans did, just because they had efficient symbols for them. In this respect the science of pharmacokinetics is in its state of infancy. Some efforts were made to develop a unified nomenclature, notably the proposals by Rowland and Tucker [3] and by Rescigno et al. [4], but besides the fact that they have not yet reached a general acceptance, those proposals are more on the line of abbreviations than symbols. For instance *AUC* (Area Under the Curve) means the integral $\int_0^\infty c(t)dt$, where $c(t)$ is the concentration of a drug in the plasma following a bolus injection at time t; but it is not always clear whether the drug was injected intravenously or intramuscularly, and whether the dose injected was a unit dose or not. Furthermore, there are many properties connected to the integral above that could be used directly in the description of the fate of a drug in vivo. For instance, the ratio of the *AUC*'*s* measured in two different points of an organism, irrespective of the dose and the mode of injection of the drug, is an invariant quantity for a linear system [5-8]. In addition to the advantage of being dimensionless, this parameter has an interesting property, namely it can be used as the element of an algebra to describe the connectivity of the organism.

Examples of this sort could be multiplied at will; each of them represents a theoretical concept that could lead to important practical developments of pharmacology. In general we could say that the present trend in pharmacokinetics is mostly to move away from descriptive models and toward interpretative models. To this end it is necessary to pay more attention to the mathematical methods necessary to transform the hypotheses of a physical and physiological character into differential or integral equations.

The numerical solution of those equations is a separate problem; it is an important one but not crucial as it was a few years ago, thanks to the very efficient hardware and software available today.

Much more important is the logical approach to model building, namely to the problem of determining the minimum number of hypotheses necessary for the explanation of observed phenomena, and to planning the experiments in the most efficient way in order to verify the validity of the hypotheses postulated.

Since in many practical cases (e.g., carcinogenic risk assessment), human inferences are made from experiments on animals, it is imperative that through a combination of biological facts and mathematical theory, appropriate methods of extrapolation from animal to human be developed.

1.5. REFERENCES

1. F. H. Dost, Der Blutspiegel. *Kinetik der Konzentrationsverläufe in der Kreislaufflüssigkeit* (Thieme, Leipzig, 1953).
2. A. Rescigno, Foundations of Pharmacokinetics, *Pharmacol. Res.* **42,** 527-38 (2000).
3. M. Rowland and G. Tucker, Symbols in Pharmacokinetics, *J. Pharmacokin. Biopharm.* **8,** 497-507 (1980).
4. A. Rescigno, A. K. Thakur, A. B. Brill and G. Mariani, Tracer Kinetics: A Proposal for Unified Symbols and Nomenclature, *Phys. Med. Biol.* **35,** 449-65 (1990).
5. A. Rescigno, On Transfer Times in Tracer Experiments, *J. Theoret. Biol.* **39,** 9-27 (1973).
6. A. Rescigno and L. D. Michels, On Dispersion in Tracer Experiments, *J. Theoret. Biol.* **41,** 451 -60 (1973).
7. A. Rescigno and L. D. Michels, Compartment Modeling from Tracer Experiments, *Bull. Math. Biol.* **35,** 245-57 (1973).
8. A. Rescigno and B. M. Bocchialini, Pharmacokinetics: Unfolding of a Concept, in: *New Trends in Pharmacokinetics,* edited by A. Rescigno and A. K. Thakur (Plenum Press, New York, 1991), pp. 1-26.

This page intentionally left blank

2. MODELS

2.1. THE EXPERIMENTAL METHOD

I have shown in the first chapter that the strength of Pharmacokinetics is due to its constant use of the *experimental method.* The experimental method is not just the acceptance of observed facts, but the acceptance of statements about facts that can be verified by experimental observations. The problem thus consists in selecting the facts that are to be observed; but how do we know which facts are worth observing? We do not proceed in a vacuum, but we observe only what we expect to lead to logical statements, and we use logical statements to decide which facts are worth observing. Now if any logical statement must be justified by an experiment, and each experiment by a previous logical statement, we end up to an infinite regression. If we want to avoid this infinite regression, we must start with some primitive statements: we shall call them *axioms.* I will explain the meaning of this term in section 2.4.

At this point I must introduce some definitions [1].

The system to be studied shall be called the *primary system.*

What is used by the investigator to study the primary system, by whatever means, shall be called a *secondary system.* I shall consider now a further subdivision of secondary systems.

A *simulator* is a secondary system that mimics certain aspects of the behavior of the primary system.

A *model* is a secondary system created to confirm hypotheses on the primary system.

Modulating the data means modifying them in accordance with a certain criterion; this transformation may involve a change in their information content.

Coding is a particular case of modulation in which no information is added or lost.

2.2. CODING AND MODULATION OF EXPERIMENTAL DATA

The data collected through experimental observations may be presented in the form of graphs or tables. As an example consider Table I, where I have listed the value of the quantity $x(t)$ observed at three different times in a hypothetical primary system.

Table I. Data from a hypothetical primary system

t	1	2	3
$x(t)$	2	3	4

The same data can also be represented by a graph, as in Fig. 1, or concisely by the formula:

$$x(t) = t + 1, t = 1, 2, 3. \quad (1)$$

Whether we use a graph, or a table, or a formula like the one above, the information content is exactly the same; those are examples of coding.

Other examples of coding are changes of coordinates, for instance from linear to logarithmic or semilogarithmic.

Other transformations of the data are possible. For instance we can redraw the graph of Fig. 1 with a straight line (Fig. 2), or we can fit the data of Table I with the expression

$$x(t) = t + 1, t \geq 0. \quad (2)$$

These last two transformations are examples of modulations in which there is a change of the information content of the original data; in fact here we loose the knowledge of which points were actually measured, and we add values at points where measurements were not actually taken.

This particular example of modulation, i.e., fitting the intermediate points with a straight line, may seem obvious, but it is not. The polynomial function

$$x(t) = t^3 - 6t^2 + 12t - 5$$

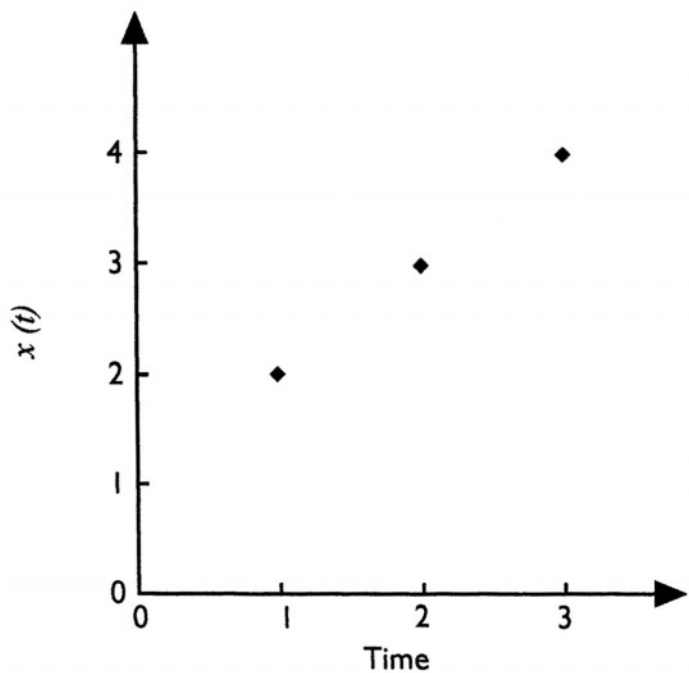

Fig. 1. The three points of Table I.

and the trigonometric function

$$x(t) = 3 - \sin(\pi \cdot t/2)$$

both fit the data as well as the linear function (2); infinitely many other functions can provide an equally good fit. The choice of the specific function implies a different hypothesis made on the primary system.

So modulation includes data fitting, i.e., computation of parameters of an equation of a given type, such that the divergence between the data from the primary system and the data from the secondary system is minimal according to a specified criterion.

The important point is that coding is, in a sense, neutral with regard to the primary system, while data fitting (modulation) implies some assumptions that need to be justified and hypothesis that need to be confirmed.

2.3. SIMULATION

In the definition of simulation I implied that its purpose is duplication of some aspects of the behavior of the primary system; this is done without regard to mechanism. An artificial arm for instance simulates the mechanical behavior of a human arm by very different mechanisms. A mathematical equation may predict the temporal variation of the concentration of a drug in treatment planning, for a limited range of its values and without its parameters having any specific meaning. Simulation may thus be done for a very useful purpose, but it must be kept in mind that a secondary system may mimic a primary system in a very satisfactory manner, but using a mechanism that has nothing, or very little, in common with the primary system.

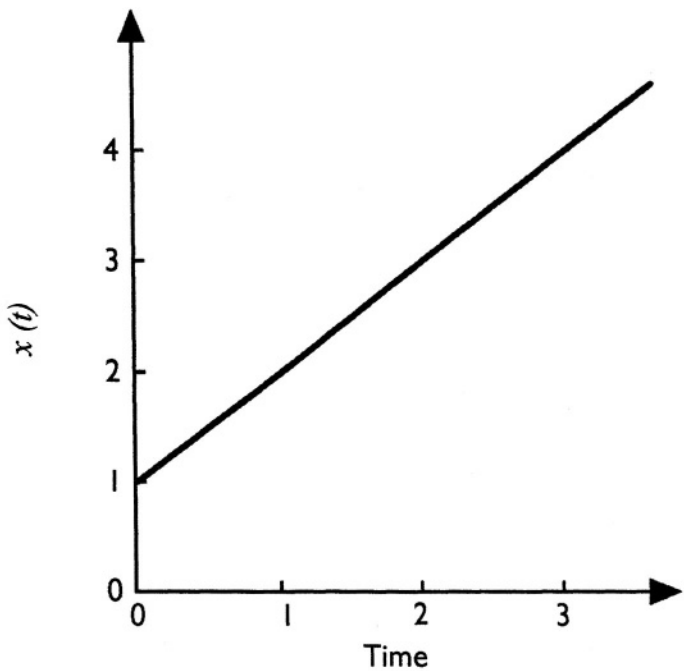

Fig. 2. The straight line of equation (2)

Powerful computers and sophisticated programs are available for modulating experimental data and making them fit rapidly a sum of exponentials, or a sum of trigonometric functions, or many other mathematical functions with an arbitrary accuracy; such an effort will always succeed and is, therefore, useful only as a means of imitation of the behavior of the primary system.

For instance consider the data in the first two columns of Table II; they can easily be fitted by a sum of three exponentials,

$$x(t) = 0.7616\, e^{-0.6749\, t} + 0.1815\, e^{-0.1412\, t} + 0.0569\, e^{-0.0219\, t}, \tag{3}$$

but also by a hyperbola,

$$x(t) = \frac{1.0006}{t + 1.0006}. \tag{4}$$

The third and fourth columns show the values computed from identities (3) and (4), and the last two columns the corresponding deviations of the computed from the measured data. In both cases the fitting appears to be very good. The parameters of identity (3) have only one thing in common with the parameters of identity (4): i.e., both provide a good fit when used in their respective places. But, if the parameters of identity (3) do have a biological meaning, then what biological meaning can the parameters of identity (4) possibly have? And vice versa.

Fig. 3 shows that the fitting of identities (3) and (4) is good only at the points where the experimental data were taken, but some discrepancies occur at other points. Only additional experiments can tell which model is more appropriate.

The procedure of curve fitting, inappropriately called "modeling the data", may be very useful for simulation. However we must resist the temptation of giving the results of those computations a meaning besides the phenomenological one.

A classic example of simulation is the Almagest; in it Ptolemy (~100-178 a.D.) used abstract mathematical functions (cycles and epicycles) to describe the movements of all known celestial bodies, obtaining a remarkably good fit; but the good fit did not help in interpreting the data or in predicting future observations. On the contrary Aristarchus of Samos (~310-230 b.C.) first, then more consistently Copernicus (1473-1545), put forward the heliocentric hypothesis, thereby basing their computations not on the mere fit-

Table II. Data fitted by Identities (3) and (4)

Experimental Data		Fitted Data		Deviations	
Time	Conc.	Exponentials	Hyperbola	Exponentials	Hyperbola
0	1	1.0000	1.0000	0.00	0.00
5	.1668	0.1667	0.1667	$+1.37\cdot10^{-4}$	$+5.00\cdot10^{-5}$
10	.0909	0.0908	0.0910	$+7.49\cdot10^{-5}$	$-5.87\cdot10^{-5}$
15	.0628	0.0628	0.0625	$-2.81\cdot10^{-5}$	$+2.65\cdot10^{-4}$
20	.0475	0.0475	0.0476	$+4.63\cdot10^{-6}$	$-1.46\cdot10^{-4}$
25	.0386	0.0382	0.0385	$+3.71\cdot10^{-4}$	$+1.16\cdot10^{-4}$
30	.0319	0.0321	0.0323	$-2.23\cdot10^{-4}$	$-3.77\cdot10^{-4}$
35	.0278	0.0277	0.0278	$+6.62\cdot10^{-5}$	$+6.02\cdot10^{-6}$
40	.0243	0.0243	0.0244	$-3.54\cdot10^{-5}$	$-1.05\cdot10^{-4}$
45	.0217	0.0216	0.0218	$+1.46\cdot10^{-4}$	$-5.19\cdot10^{-5}$

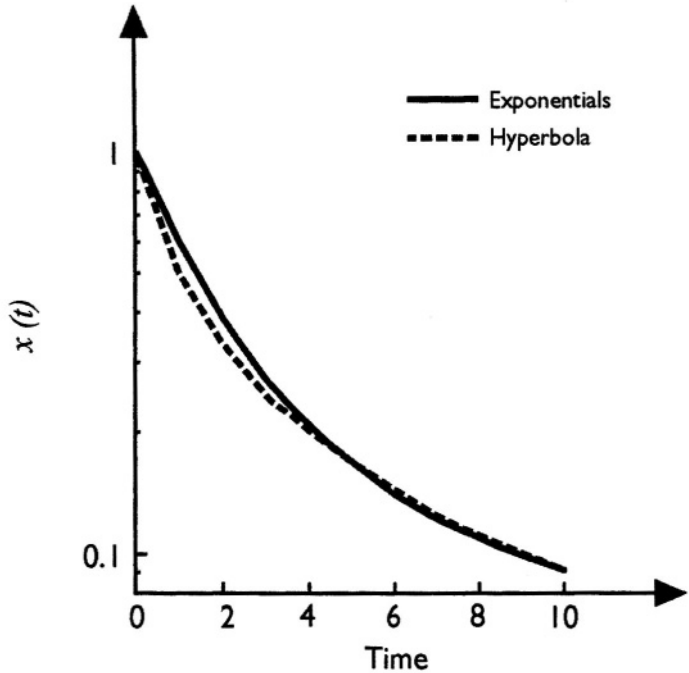

Fig. 3. A plot of identities (3) and (4) for $t \leq 10$.

ting of data (in fact Copernicus fitting was not substantially better than Ptolemy's), but on a physical basis [2].

2.3.1. Empirical analysis.

Some authors distinguish two methods for the analysis of scientific data: "empirical" and "model-based". In short, the distinction implies that, when reasonable hypotheses cannot be made on the physical or physiological process under observation, then it may be useful to fit the experimental data to a set of abstract mathematical functions, "functions that do not necessarily have any physical basis" [3]. The truth is that to elaborate any data one must have a model. If the model is based on some physical or physiological hypotheses, the elaboration of the data may confirm or falsify the hypotheses. If the model is simply a fitting program not based on any physical or physiological hypothesis, the fitting, per se, can never lead to confirmation or falsification, because there is nothing to confirm or to falsify [1].

Many models that are called "empirical" in reality are based on some physical hypothesis; statistical models, for instance, require that the data be a random sample drawn from a population with a certain probability distribution. But any probability distribution implies a specific hypothesis on the origin of the data. On the other hand, fitting the data to cubic splines, for instance, may help in putting the experimental results in a more aesthetic form, but does not add anything to our scientific knowledge. New knowledge is acquired only through the process of confirmation and rejection, in the best Galilean tradition.

"Model-free analysis of data" is a strange term used by some authors [4] to signify that they were unable to make any hypothesis on the system under observation, and they

expected a computer to do the job in their stead [5]. The result may be considered at most a simulator, its value will be limited to the description of the data. This is the classic example of rhetorical pharmacokinetics.

2.4. CONFIRMATION

In contrast to simulation, a model is a secondary system the purpose of which is to verify whether some hypotheses made on the primary system are valid. In other words, the purpose of a model is that of scientific confirmation.

Any scientist, whether a theorist or an experimentalist, pharmacokineticists included, when observing the object of his study (the *primary system*) has always present in his mind a set of statements, which he calls a *model.* The statements constituting a model are of two kinds: *axioms* and *hypotheses.* The distinction between axioms and hypotheses is apparently psychological, but it is a fundamental one: axioms are statements that we accept as true and are not to be doubted; hypotheses are statements accepted on a temporary bases, but subject to confirmation or rejection.

In short, when choosing a model the investigator knows some facts and wants to discover other facts; what the investigator presumes to know is an *axiom,* what he wants to check is a *hypothesis.* Now in this context what is the rationale for doing an experiment? The inescapable conceptual structure of any scientific investigation is as follow (Fig. 4):

a. Assumption (conscious or unconscious) of axioms;

b. Choice of hypotheses;

c. Specification and construction of a model (mathematical, mechanical, ...) incorporating those axioms and hypotheses;

d. Observation of the model (collection of data from equations or a device);

e. Observation of the primary system;

f. Comparison of data from model and primary system;

g. Conclusion (confirmation or rejection of hypotheses).

If the hypotheses are rejected, we make different hypotheses, i.e. we change the model and start the said process again; if the hypotheses are confirmed, we accept them.

We must not be misled by the term *confirmation.* Confirmation may occur when evidence is realized which is consistent with a hypothesis, but, as Popper [6] has repeatedly admonished, a hypothesis cannot possibly be proven true. We can, however, clearly disprove hypotheses. This is not so despairing as reducing an infinity of possibilities to an infinite set containing one less possibility, for we always test hypotheses within some theoretical framework, and generally that theoretical construct is valid for some purpose.

The nature and process of confirmation can be illustrated with an example. Suppose we are interested in insulin (H) binding by rat fat cell receptors (R) to form a hormone-receptor complex (C). Our hypothesis is that the only reaction taking place is $H + R \Leftrightarrow C$. Then the axiom of mass action leads us to equation

$$\frac{dC}{dt} = k_a[H][R] - k_d[C]$$

where k_a and k_d are the association and dissociation constants, respectively; the axiom of conservation of mass gives the equations

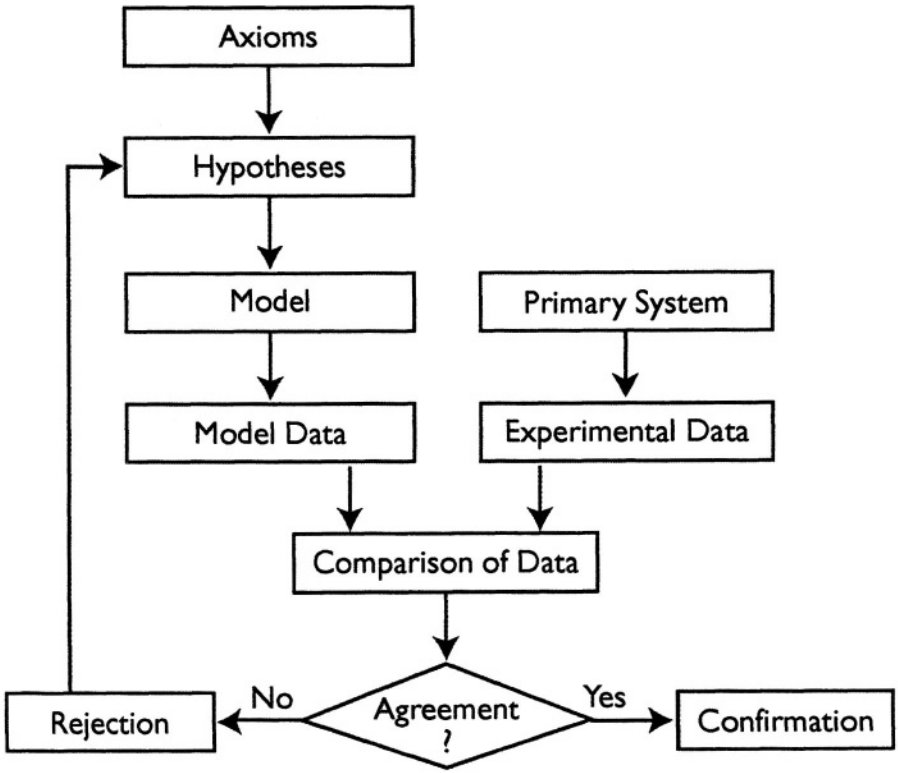

Fig. 4. Flow chart of the process of Confirmation or Rejection

$$[H]+[C]=h$$
$$[R]+[C]=r$$

where h and r are the constituent concentrations of insulin and of receptors, considered constant. If we eliminate the variables $[H]$ and $[R]$ from these three equations we get

$$\frac{dC}{dt}=k_a\big(h-[C]\big)\big(r-[C]\big)-k_d[C]. \tag{5}$$

In an actual experiment we can measure $[C]$ at different times; the experimental results may be represented in Table III. With these data we may determine, by regression analysis, the values of the two rate constants; the expected values are $k_a = 2\cdot10^{17}$ $\mathrm{l\cdot pmol^{-1}\cdot min^{-1}}$ and $k_d = 0.05\ \mathrm{min^{-1}}$. Fig. 5 shows a plot of the measured values given in Table III together with the values computed using equation (5) and the estimated rate constants. There is clearly a contradiction between the primary and the secondary system. We must discard the present hypothesis.

Table III. Hormone-Receptor Complex concentration *versus* time

t (minutes)	1	10	20	30	50	80	120	200
$[C]$ (picomolar)	0.1	4.8	5.6	5.5	4.7	3.5	2.0	0.6

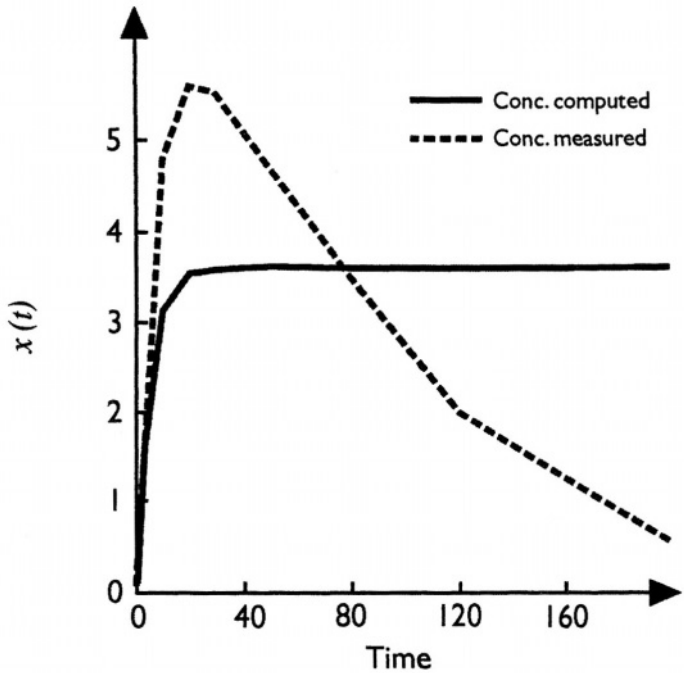

Fig. 5. Graphical representation of the data from Table III and the data computed from equation (5).

The data from the primary system suggest a loss of one of the species from the binding equation; this suggests one of these possibilities:

Model 2: $H^* \leftarrow H + R \Leftrightarrow C$,

Model 3: $R^* \leftarrow R + H \Leftrightarrow C$,

Model 4: $R + H \Leftrightarrow C \rightarrow C^*$.

Now the data from either models of loss of reagent or model of loss of complex from the site can be made to duplicate the data of the primary system by choosing appropriate values for the rate of loss of H or R or C. The choice between those models must be based on biological, not mathematical, considerations.

2.5. INFORMATION

All the above discussion can be summarized in just two propositions with respect to primary system:

1. Information comes from observation,
2. Mathematical and graphical models decrease information.

The first proposition is obvious to any experimental scientist; the second proposition may be a little puzzling at first. This subject has been discussed at length by Rescigno and Beck [1], but I shall give here some justification for it.

The *amount of information* is defined as the logarithm of $1/p$, where p is the expectation of a particular experimental result, before the experiment is done.

Given proposition 1, we can increase our scientific knowledge only by observation. But our time and abilities being limited, our observations must be planned in such a way that only relevant data are collected. Thus we observe only quantities that we consider important, and we discard as "experimental errors" anything that deviates from what we thought should have been the correct or the most likely result. These expectations of what is relevant to our hypotheses and what is not a deviant result are determined from the behavior of the model. Without a model we have no expectations: we have no basis for choosing what to observe or which results are useful. In such a conceptual vacuum we know nothing about the behavior of a given primary system. Any experiment can generate any result, as far as expectations go, and there are an infinite number of possible expected results with an infinitesimal probability, p. Any observation can generate an infinite amount of information $\log(1/p)$.

In contrast with such an unrealistic situation, we always have a model in mind, though some investigators do not explicitly acknowledge it. If we do have a model, we expect particular results, and we look only for them. Whether the hypotheses will be confirmed or not, the information gained from the experiments is smaller in amount than without a model or with a less specific model because the possible outcomes were more expected. Also, we are ready to discard as "random error" or "irrelevant result" anything that does not conform to our expectations, i.e., to our model.

There is an old appealing, but simplistic notion that a scientist must always be completely free from all preconceptions; we could call it "the dogma of the immaculate conception". The History of Science is full of examples of great discoveries missed because the observer was not expecting a particular result, or of great discoveries due to unconfessed or unconscious guiding presuppositions a scientist adopted without being forced to do so by either data or current theory [7].

To conclude this section, observe that step f in the section "Confirmation" includes modulation of data. Perhaps the most common modulation is computation of a mean. When we compute a mean, we discard some, possibly a large number of, individual data. That is, we reduce the amount of information, but we preserve a datum that we consider important, i.e., with an information that is more relevant for the confirmation or the rejection of the model. It is always the model that guides us and that enables us to evaluate which results are important and which are not.

2.6. VALUE AND STRENGTH OF A MODEL

From what I have said so far, it must be clear that a secondary system, be it a simulator or a model, implies some hypotheses. The difference is that with a simulator the hypotheses are accepted *a priori* and are used to duplicate the expected behavior of the primary system in situations that have not been observed directly, while with a model the hypotheses are the issue to be checked *a posteriori.* This distinction is a fundamental one and has been stressed very clearly by Zierler [8] in the context of compartmental analysis.

We can loosely define the *reliability* of a simulator by saying that a more reliable simulator mimics the behavior, or some relevant aspects of the behavior, of the primary system more closely. On the other hand, a model cannot be judged solely by the closeness of the data observed on the primary and on the secondary system.

A model must be judged from three different points of view: retrodiction, prediction, and understanding. Retrodiction is simply recalling what happened; this means that the model must conform with the original data from the primary system, i.e., that the hy-

potheses are consistent with our experimental knowledge. Although this condition is necessary, it is far from sufficient.

The model must be predictive; that is, it must tell us what will happen in future experiments. If a given model leads us to a new experiment, and this experiment confirms the hypotheses, we attribute a larger value to the model; but this temporal distinction between prediction and retrodiction is important only in the context of discovery, not in the context of explanation, and its value is only psychological. A deeper distinction between prediction and retrodiction is a categorical one, i.e., one referring not to experiments made at different times but to experiments made in different contexts.

I refer to the predictive property of a model in different contexts as its *strength.* The strength of a model can be illustrated with an example.

Consider the Volterra predator-prey equations

$$\begin{aligned} dN_1/dt &= (+e_1 - g_1 N_2) N_1 \\ dN_2/dt &= (-e_2 + g_2 N_1) N_2 \end{aligned} \tag{6}$$

where N_1 is the number of prey, N_2 the number of predators, e_1 and e_2 their coefficients of increase in the absence of interaction, g_1 and g_2 the coefficients of interaction. As shown by Volterra [9], the model equations predict oscillations with periods

$$T = 2\pi / \sqrt{e_1 e_2}\,. \tag{7}$$

The coefficients e_1 and e_2 can be measured in separate experiments with only the prey or only the predators present, i.e., in the absence of interaction; then with those values the period T can be computed using identity (7). If this value of T is consistent with the value of T measured in an experiment with the two species interacting, the hypotheses included in equations (6) are confirmed.

The third aspect of a model, understanding, is more difficult to define. I leave that to the philosophers and simply point out the significance of understanding in scientific investigations.

Understanding is what we apply in answering questions such as the following: "Have we defined the primary system optimally for our purpose?" The model and how it has performed in the process of confirmation helps us decide and then helps us refine the definition. Where, when, and what do we need to observe in order to test a particular hypothesis? The model is the basis on which we answer this question. This latter question and its answer constitute experimental design in laboratory and clinical studies. The model leads us to useful questions, suggests how we can answer them, enables us to avoid gathering useless data and helps us to anticipate where the possible results would lead us.

Finally, understanding includes some ideas of how the primary system fits into the universe, for in the ultimate analysis the universe—the grand primary system—is constituted conceptually of all the primary systems we investigate, collectively interacting. Our models require of us that we take note of where we have cut the connection to the rest of the universe in constructing them and that we consider the possible effects of such artificial isolation.

The *value* of a model is related to these three points of view. In the context of understanding, a model has minimal value if it leads to no new questions, if it does not help us to decide how to answer questions, and if it does not give perspective on where and how

the primary system fits into the world. It has a higher value the more it helps in further investigations, the more it helps us to integrate our concepts of primary systems into a wider, deeper and more effective concept of our world.

The result of confirming a model is a theorem or law. Theorems and laws have a value within the context of a model, never in the absolute. In physics, there are laws for the perfect gas (for instance, Boyle's law), other laws for gases formed by spherical particles (for instance, the equation of Van der Waals), and so forth. The laws valid for the model of classical mechanics are not valid for the model of relativistic mechanics. If in physics we often forget that a law is subordinate to a specific model, it is because these models are in general very strong, agreeing with a large number of real systems.

This generality is less frequent in biology. This fact does not detract from the utility of models in biology; on the contrary, it makes it necessary to formulate models more often, more explicitly, and more carefully.

As observed by Volterra [9], the value of T given by (7) is valid only for small oscillations. This presents another opportunity to stress the fact that all conclusions are valid within a specified model.

In contrast, the fitting of data *per se* can never lead to falsification, for there is nothing to be falsified, nothing to be tested. And of course there is nothing to be tested in an experiment of a different kind. There is nothing to be assigned strength to as defined here. The frequency of this kind of report in scientific literature compels further emphasis: a "model of data" has a strength of zero and value only as a simulator; that is, it retrodicts and (temporarily) predicts but contributes nothing to understanding.

2.7. REFERENCES

1. A. Rescigno and J. S. Beck, The Use and Abuse of Models, *J. Pharmacokin. Biopharm.* 15, 327-40 (1987).
2. A. Rescigno, Compartmental Analysis Revisited, *Pharmacol Res.* 39, 471-8 (1999).
3. J. J. DiStefano and E. M. Landaw, Multiexponential, multicompartmental, and noncompartmental modeling. I. Methodological limitations and physiological interpretations, *Am. J. Physiol.* 246, R651-64 (1984).
4. V. Guardabasso, P. J. Munson and D. Rodbard, A versatile method for simultaneous analysis of families of curves, *FASEB J. 2,* 209-15 (1988).
5. J. S. Beck and A. Rescigno, Simultaneous Analysis of Families of Curves, *FASEB J.* 3, 2113 (1989).
6. K. R. Popper, *The Logic of Scientific Discovery* (Basic Books, New York, 1959).
7. G. Holton, *Einstein, History, and Other Passions* (American Institute of Physics, Woodbury, NY, 1995).
8. K. Zierler, A Critique of Compartmental Analysis, *Ann. Rev. Biophys. Bioeng.* **10**, 531-62 (1981).
9. V. Volterra, *Variazioni e fluttuazioni del numero di individui in specie animali conviventi* (Report No. CXXXI, Memorie del R. Comitato Talassografico Italiano, Roma, 1927).

This page intentionally left blank

3. PHYSICAL QUANTITIES

3.1. INVARIANT QUANTITIES

In mathematics we learn that two pure numbers can always be added, subtracted, multiplied, divided, with just few restrictions; for instance a divisor cannot be zero. In physics things are quite different; physical quantities have dimensions, and we learn that we can add or subtract two quantities only if they have the same dimensions; multiplication and division of quantities are possible only if the result is an *invariant* quantity, i.e., a quantity whose value does not depend, within the limits of validity of some specific hypotheses, on the particular experimental conditions.

For instance a mass divided by a volume is a concentration, a quantity certainly invariant because we know that, in a well stirred fluid, the ratio *Mass/Volume* is constant for a wide range of sample sizes. In other words, the observation that the above quotient is constant is the operative definition of a new physical quantity, the concentration [1].

Unfortunately there are a few examples in the pharmacokinetic literature where new quantities are introduced by means of a product or a quotient without first checking whether or not the result of that operation is an invariant quantity, i.e., without a proper operational definition. Later in this book I shall try to indicate, for each pharmacokinetic quantity, its dimension and its correct definition.

3.2. EXTENSIVE AND INTENSIVE QUANTITIES

All physical quantities are classified as extensive or intensive. An extensive quantity is a quantity that is dependent upon the system extent. An intensive quantity is a quantity that is not dependent upon the system extent. All these quantities may be functions of time.

An alternative definition is that a quantity is extensive if

"Measure of A + Measure of B = Measure of $(A + B)$".

A quantity is intensive if

"Measure of $\mathbf{A} \leq$ Measure of $\mathbf{(A + B)} \leq$ Measure of B",

or

"Measure of $\mathbf{A} \geq$ Measure of $\mathbf{(A + B)} \geq$ Measure of B".

From this definition of extensive quantity, it follows that the quotient of two extensive quantities is always an intensive quantity. In other words, if x and y are two extensive quantities, then $x/y = z$ is an intensive quantity.

For instance, if M is mass and V volume, both extensive quantities, their ratio $D = M/V$ is density, an intensive quantity; in fact, if the subscripts refer to two separate objects A and B and to their combination,

$$D_{a+b} = \frac{M_a + M_b}{V_a + V_b} = \frac{M_{a+b}}{V_{a+b}};$$

from this identity we can compute

$$D_{a+b} - D_a = \frac{M_a + M_b}{V_a + V_b} - \frac{M_a}{V_a} = \frac{M_b V_a - M_a V_b}{V_a V_{a+b}}$$

and

$$D_{a+b} - D_b = \frac{M_a + M_b}{V_a + V_b} - \frac{M_b}{V_b} = \frac{M_a V_b - M_b V_a}{V_b V_{a+b}},$$

thence

$$D_a \leq D_{a+b} \leq D_b \text{ if } M_a V_b \leq M_b V_a$$

and

$$D_a \geq D_{a+b} \geq D_b \text{ if } M_a V_b \geq M_b V_a.$$

Only extensive quantities can be measured directly; in fact a direct measurement requires the interaction between a measuring subject and a measured object, and this last one must be present in a finite extent. An intensive quantity on the contrary can be determined indirectly as the result of a relationship between two extensive quantities. Two examples will better illustrate this concept.

From a solution we can measure the volume V of a sample and separately the amount M of solute contained in that volume; they are both extensive quantities and are dependent upon the chosen sample. Now we can compute the ratio M/V, but this ratio is invariant only if the solution is homogeneous; we can call it the "average concentration" in that particular sample. The correct value of the concentration C, an intensive quantity, is given by

$$C = \frac{dM}{dV},$$

i.e., the ratio of mass and volume of an infinitesimally small sample, and may vary according to where the sample is taken, if the solution is not homogeneous.

Now suppose that we have determined by other means the concentration of a solution; the amount of solute M contained in a certain volume V of solution should be given by multiplying the concentration by the volume, but this straight product will give the correct value of M only if the solution is homogeneous; if not, remembering that C is an intensive quantity, valid at one point only, we should write

$$dM = C \cdot dV,$$

where dM is the infinitesimal amount of solute contained in the infinitesimal volume dV of solution; by integration

$$M = \int_V C \cdot dV.$$

Summing up, the derivation of an intensive quantity from an extensive one requires a differentiation; the derivation of an extensive quantity from an intensive one requires an integration.

3.3. CONSERVED QUANTITIES

There are many conservation principles in elementary physics: the conservation of matter, of energy, of linear momentum, of angular momentum, of charge, and so forth; we see the first one applied in almost all pharmacokinetic equations, sometimes disguised with other names.

Any quantity that is conserved is extensive, but the converse is not necessarily true. In more precise terms, if E is an extensive and I an intensive quantity, their product $P = E \cdot I$ may be a conserved quantity, and all conserved quantities can be represented as the product of an extensive and an intensive quantity. The classic example of such products is given by Newton in the first two definitions of his Principia [2]:

> Def. I. Quantitas Materiae est mensura eiusdem orta ex illius Densitatae et Magnitudine conjunctim. (Mass = Density × Volume.)
>
> Def. II. Quantitas motus est mensura eiusdem orta ex Velocitatae et quantitate Materiae conjunctim. (Momentum = Velocity × Mass.)

Some examples of products of this form will be found several times in the second part of this book.

3.4. PARAMETERS

In mathematics a parameter is a fixed quantity that determines the behavior of a function. In the experimental sciences a parameter is an observable quantity that remains constant for every definable state of a system.

For each parameter mentioned in this book I will list a definition, a number of properties, and some methods for its determination; so it may be appropriate to spend a few lines to explain the meaning of these last terms.

A Definition is a statement that substitutes a new term to some previously known terms, a Property is a statement about previously defined subjects, and a Determination is a process used for the evaluation of a physical quantity.

Unfortunately, some authors do not make a clear distinction between *definition, property* and *determination.* For an example of the confusion among those three terms, see for instance "Fundamental Concepts in Pharmacokinetics" [3] under the entry "Definition."

Another source of confusion is generated by the terms *model-independent parameter* and *model-free parameter,* used by some authors in various contexts and with different meanings; sometimes they mean "a parameter not based on a compartmental model" [4], sometimes a parameter that "can be evaluated without an explicit pharmacokinetic model but under basic assumptions" [5], sometimes even a parameter computed "where no

mathematical formulation is known and when the available data do not permit derivation of such a model" [6].

It is important to separate the parameters used in pharmacokinetics into three different categories, namely *pharmacokinetic parameters, model parameters,* and *incidental parameters.*

3.4.1. Pharmacokinetic parameters

A pharmacokinetic parameter is a quantity that depends on intrinsic properties of a drug and of the biological system it interacts with, but not on any specific hypotheses made on the system. Examples of pharmacokinetic parameters are *clearance, turnover time, volume of distribution, turnover number, permanence time, yield, residence time.* They are connected by the relationships

$$\text{Clearance} \times \text{Turnover time} = \text{Volume of distribution},$$

$$\text{Turnover number} \times \text{Turnover time} = \text{Permanence time},$$

$$\text{Yield} \times \text{Permanence time} = \text{Residence time}.$$

We may chose four of them as fundamental parameters, and define the other three from them. For instance, if we choose as fundamental parameters Clearance, Turnover time, Residence time, and Yield, then the derived parameters are

$$\text{Volume} = \text{Clearance} \times \text{Turnover time},$$

$$\text{Permanence time} = \text{Residence time} \div \text{Yield},$$

$$\text{Turnover number} = \text{Permanence time} \div \text{Turnover time}.$$

The problem of defining pharmacokinetic parameters independently of a particular model has been dealt with by several authors, but seldom with convincing results. It is important to be aware that even if a parameter may be defined independently of a particular model, its determination is not always independent from the model; the second part of this book will make this point more evident.

3.4.2. Model Parameters

A model parameter is a quantity that depends on intrinsic properties of a drug and of the biological system it interacts with, and also on hypotheses made on that system.

For instance the terminal half-life is a model parameter because it depends on the number of exponentials used to fit the data.

3.4.3. Incidental Parameter

An incidental parameter is a quantity that describes the result of an experiment, but is not an invariant quantity under a reasonable set of experimental conditions. For instance, the integral $\int_0^\infty c(t)\mathrm{d}t$ is not an invariant quantity because it depends upon the site of administration, the mode of administration, and the dose; the ratio $\int_0^\infty c(t)\mathrm{d}t / D$, on the contrary, may be considered invariant because, in many cases and within a certain range, it does not depend upon the dose D.

Another example of incidental parameter is the ratio $\int_0^\infty t\,c(t)\mathrm{d}t \,/\, \int_0^\infty c(t)\mathrm{d}t$, inappropriately called by Yamaoka et al. [7] *mean residence time*; it depends upon the time of administration, the site of administration, and the mode of administration; it should more properly be called *time of exit* [8, 9].

3.4.4. Macroparameters and Microparameters

The terms *macroparameters* and *microparameters* are sometimes used in the literature, though their meaning is not always very clear.

In short, when assuming a linear compartmental model with time-invariant parameters, one writes the function $c(t)/D$ as a sum of n exponential terms; this function is fully described by $2 \cdot n$ independent parameters, i.e. n coefficients and n exponents; they are called the *macroparameters* of the model.

From the $2 \cdot n$ macroparameters we can try to determine a number of so-called microparameters that represent the transit rates of, and the transfer rates between, the compartments, necessary and sufficient to characterize the model.

The determination of the microparameters is not always unique; they fulfill a useful purpose only if we can give them a physical or physiological interpretation [10].

3.5. DIMENSIONAL ANALYSIS

3.5.1. Allometry

The term *allometry* is derived from the Greek αλλος = other, and μητρον = measure. It means, in general, the relationship between any two non-homogeneous quantities; more particularly, the allometric equation

$$Y = a\,W^b \tag{1}$$

is used to show the relationship between the quantity Y and the body weight of an individual, where a and b are parameters determined from theoretical considerations or experimentally.

To explain the theoretical basis of the allometric equation I shall follow a suggestion of von Bertalanffy [11]; think of a complex system formed by a large number Q of interacting parts; if all those parts are similar, we can make the reasonable hypothesis that the growth of the system is directly proportional to the number of elements present, therefore we can write

$$\mathrm{d}Q/\mathrm{d}t = a \cdot Q,$$

where a is the growth rate, positive or negative; by integration we get

$$Q = C \cdot \exp(a \cdot t);$$

this exponential law is known as the law of natural growth, and is found to be valid in many different circumstances, whether the interacting parts are individuals or cells or inanimate objects.

Now suppose that two kinds of parts are present, and that the increase of each element depends on that element only; we have two differential equations,

$$\mathrm{d}Q_1/\mathrm{d}t = a_1 \cdot Q_1,$$

$$dQ_2/dt = a_2 Q_2,$$

with the integrals

$$Q_1 = C_1 \cdot \exp(a_1 \cdot t),$$

$$Q_2 = C_2 \cdot \exp(a_2 \cdot t).$$

We solve both integrals for t,

$$t = \frac{\ln Q_1 - \ln C_1}{a_1} = \frac{\ln Q_2 - \ln C_2}{a_2}$$

and finally, eliminating t,

$$Q_1 = \frac{C_1}{C_2^{a_1/a_2}} Q_2^{a_1/a_2}.$$

This expression coincides with the allometric equation (1) if we make $a_1/a_2 = b$ and $\frac{C_1}{C_2^{a_1/a_2}} = a$.

If the quantities Y and W are measured experimentally on a number of different individuals, the numerical values of the parameters a and b can be determined by regression. In the current literature there are many examples where equation (1) has been fitted very satisfactorily using data from individuals of different species, even when the body weight covers a range of five orders of magnitude. A few comments, though, are necessary at this point.

The allometric equation (1) is usually fitted using the log transformation

$$\log Y = \log a + b \log W; \tag{2}$$

in other words the parameters a and b are determined by minimizing the sum of squares

$$\sum_i (\log Y_i - \log a - b \log W_i)^2. \tag{3}$$

The base of the logarithms of course is arbitrary, but the minimization of expression (3) implies that the error in log Y is normally distributed, i.e that Y has a logarithmic normal distribution; this means that the larger values of Y have a much smaller weight in the determination of the parameters a and b.

Another related problem has to do with the dimensions of the allometric equation. The coefficient a has the dimension of $Y \cdot W^{-b}$, where Y and W have appropriate dimensions and b is a *rational* number. But, then, in equation (2), what are the dimensions of log W and of log Y?

The dimensional problem can be resolved with a change of variables in the allometric equation; if Y_0 and W_0 are the appropriate quantities of a reference species, then from equation (1),

$$Y_0 = a W_0 W^b, \tag{4}$$

and dividing each side of equation (1) by each side of equation (4),

$$\xi = \omega^b, \tag{5}$$

where $\xi = Y/Y_0$ and $\omega = W/W_0$ are dimensionless variables. The exponent b is a pure number.

If the regression is done on equation (5) instead of equation (2), the species with smaller Y will not have undue larger weight.

For a detailed review of the method see Adolph [12] and Günther [13].

3.5.2. Buckingham's Theorem

The main result of Dimensional Analysis [14] is the theorem of Buckingham [15]: If and only if an equation is dimensionally homogeneous, it can be reduced to a relationship among a complete set of dimensionless products.

To understand the value of this theorem, a few definitions are necessary. An equation is *dimensionally homogeneous* if the form of the equation does not depend on the units of measurements. For instance, the equation

$$\int_0^\infty c(t)\mathrm{d}t = \frac{F \cdot D}{Cl}$$

also known as the Stewart-Hamilton principle, is valid for any units of measurement of $c(t)$, t, F, D, and Cl, while the equation relating the period T of oscillation of a pendulum with its length l,

$$T = 0.2006/l$$

is correct only if T is measured in seconds and l is measured in centimeters. Dimensions cannot be assigned to numbers!

A *dimensionless product* is a product of physical quantities that does not depend on the units of measurement. A set of dimensionless products of given variables is said to be "complete" if each product of the set is independent of the others, and every other dimensionless product of the variables is a product of powers of dimensionless products in the set.

As an example, consider the free ligand **X**, the unoccupied receptor **Y**, and the ligand-receptor complex **Z**; the reaction

$$\mathrm{X} + \mathrm{Y} \leftrightarrow \mathrm{Z}$$

is governed by equations

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -k_1 x(t)\big(y_0 - z(t)\big) + k_2 z(t) - k_3 x(t), \qquad x(0) = x_0$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = +k_1 x(t)\big(y_0 - z(t)\big) - k_2 z(t), \qquad z(0) = 0$$

where

t = time,

$x(t)$ = concentration of X at time t,

$z(t)$ = concentration of Z at time t,

x_0 = concentration of X at time 0,

y_0 = concentration of Y at time 0,

k_1 = formation rate constant of Z,

k_2 = dissociation rate constant of Z,

k_3 = elimination rate constant of X.

There are two dependent variables, $x(t)$ and $z(t)$, one independent variable, t, and five parameters, x_0, y_0, k_1, k_2, k_3. Their dimensions are shown in Table I.

Actually in the matrix of Table I, one row is linearly dependent upon the others; we can represent the dimensions of the quantities involved with the simpler matrix of Table II. In this matrix the rows are linearly independent.

I define now the dimensionless variables

$$\rho(\tau) = x(t)/y_0\ ,\ \sigma(\tau) = z(t)/y_0\ ,\ \tau = k_2 t$$

and the dimensionless parameters

$$\alpha = k_1 y_0/k_2,\ \beta = k_3/k_2,\ \gamma = x_0/y_0.$$

The differential equations become

$$\frac{d\rho}{d\tau} = -\alpha\,\rho(\tau)\big(1-\sigma(\tau)\big) + \sigma(\tau) - \beta\,\rho(\tau) \qquad \rho(0) = \gamma$$

$$\frac{d\sigma}{d\tau} = +\alpha\,\rho(\tau)\big(1-\sigma(\tau)\big) - \sigma(\tau) \qquad \sigma(0) = 0$$

These new equations have three variables and three parameters, therefore the degree of freedom of their solutions is reduced by two units.

Call $\boldsymbol{M}$ the matrix formed with the dimensions of the old quantities shown in Table II,

$$\boldsymbol{M} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 \end{pmatrix}$$

and $\boldsymbol{P}$ the matrix showing the transformation from the old quantities to the new dimensionless products,

Table I. Dimensions of the quantities in the differential equations; the numbers are the exponents of Length, Mass, Time

	t	$x(t)$	$z(t)$	x_0	y_0	k_1	k_2	k_3
L	0	−3	−3	−3	−3	3	0	0
M	0	1	1	1	1	−1	0	0
T	1	0	0	0	0	−1	−1	−1

Table II. Dimensions of the quantities in the differential equations; the numbers are the exponents of Concentration, Time

	t	$x(t)$	$z(t)$	x_0	y_0	k_1	k_2	k_3
M/L^3	0	1	1	1	1	−1	0	0
T	1	0	0	0	0	−1	−1	−1

$$\boldsymbol{P} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Observe that

$$\boldsymbol{M \cdot P = 0};$$

this fact is general: any matrix $\boldsymbol{P}$ such that $\boldsymbol{M \cdot P = 0}$ defines a set of dimensionless products. Furthermore the number of columns in $\boldsymbol{P}$ is equal to the number of columns in $\boldsymbol{M}$ minus the number of its rows; any additional column we may add to matrix $\boldsymbol{P}$ defines a dimensionless quantity not independent upon the previous ones.

3.5.3. Dimensionless quantities

In pharmacokinetics a typical dimensionless parameter is the *dilution factor* δ, defined as the ratio between the amount of drug present in the organism and the amount of drug present in the sampling compartment, when the drug is fed to the sampling compartment and a steady state has been reached. It can be expressed as the ratio between the steady-state volume of distribution and the initial volume of distribution.

We can define the dimensionless function, called *unit response function,*

$$\chi(t) = \frac{x(t)}{D},$$

where $x(t)$ is the amount of drug in a compartment after a bolus administration in any other compartment (see chapter 5). If the dose was administered to the sampling compartment we can also write

$$\chi(t) = \frac{c(t)}{c(0)}.$$

Other dimensionless parameters are the *turnover number,* ν_i, (see chapter 14)

$$\nu_i = \frac{\text{Permanence time in } i}{\text{Turnover time of } i},$$

and the *yield*, γ_{ij}, (see chapter 17)

$$\gamma_{ij} = \frac{\text{Residence time from } i \text{ to } j}{\text{Permanence time in } j}.$$

3.6. REFERENCES

1. A. Rescigno and B. M. Bocchialini, Pharmacokinetics: Unfolding of a Concept, in: *New Trends in Pharmacokinetics,* edited by A. Rescigno and A. K. Thakur (Plenum Press, New York, 1991), p. 1-26.
2. J. S. Newton, *Philosophiae Naturalis Principia Mathematica* (Societatis Regiae, London, 1687).
3. A. Rescigno, Fundamental Concepts in Pharmacokinetics, *Pharmacol. Res.* **35**, 363-90 (1997).
4. P. Veng-Pedersen, Theorems and Implications of a Model Independent Elimination/Distribution Function Decomposition of Linear and Some Nonlinear Drug Dispositions. I. Derivations and Theoretical Analysis, *J. Pharmacokin. Biopharm.* **12**, 627-48 (1984).
5. D. Brockmeier, Model-Free Evaluation and Mean-Time Concept in Pharmacokinetics. Methods and Findings, *Exptl. Clin. Pharmacol.* **8**, 593-602 (1986).
6. V. Guardabasso, P. J. Munson and D. Rodbard, A versatile method for simultaneous analysis of families of curves, *FASEB J.* **2**, 209-15 (1988).
7. K. Yamaoka, T. Nakagawa and T. Uno, Statistical Moments in Pharmacokinetics, *J. Pharmacokin. Biopharm.* **6**, 547-58 (1978).
8. A. Rescigno and G. Segre, *La Cinetica dei Farmaci e dei Traccianti Radioattivi* (Boringhieri, Torino, 1961).
9. J. Mordenti and A. Rescigno, Estimation of Permanence Time, Exit Time, Dilution Factor, and Steady-State Volume of Distribution, *Pharm. Res.* **9**, 17-25 (1992).
10. A. Rescigno and J. S. Beck, The Use and Abuse of Models, *J. Pharmacokin. Biopharm.* **15**, 327-40 (1987).
11. L. von Bertalanffy, *General System Theory* (George Braziller; New York, 1968).
12. E. F. Adolph, Quantitative relations in the physiological constitution of mammals, *Science* **109,** 579-85 (1949).
13. B. Gunther, Dimensional Analysis and Theory of Biological Similarity, *Physiol. Rev.* **55**, 659-99 (1975).
14. H. L. Langhaar, A summary of dimensional analysis, *Journal of the Franklin Institute* **242**, 459-63 (1946).
15. E. Buckingham, On physically similar systems: Illustration of the use of dimensional equations, *Phys. Rev.* **4**, 345-76(1914).

4. THE BIRTH OF COMPARTMENTS

4.1. THE RUTHERFORD EQUATIONS

The first compartmental models were used in physics for the description of radioactive decay. After Becquerel [1, 2] discovered radioactivity, Rutherford and Soddy [3] found experimentally that Thorium X decays in time according to an exponential law, i.e., that the number of radioactive atoms decaying per unit time is proportional to the number of radioactive atoms present. If $X(t)$ is the quantity of radioactive substance present at time t, the law of radioactive decay is

$$\frac{dX}{dt} = -K \cdot X(t), \tag{1}$$

whose integral is

$$X(t) = X_0 \cdot e^{-K(t-t_0)},$$

where X_0 is the value of $X(t)$ at time t_0.

Later Rutherford [4] developed the theory of successive radioactive transformations. If A is transformed into B, B is transformed into C, and so forth, X_a , X_b, X_c, ... be the amounts of A, B, C,... present at any given time; and K_a, K_b, K_c, … be the rates of such transformations, as summarized in Fig. 1. In analogy with equation (1), he wrote

$$\begin{aligned}
\frac{dX_a}{dt} &= -K_a X_a, \\
\frac{dX_b}{dt} &= +K_a X_a - K_b X_b, \\
\frac{dX_c}{dt} &= +K_b X_b - K_c X_c, \\
&\cdots\cdots\cdots,
\end{aligned} \tag{2}$$

and by integration, provided all K_i's are different,

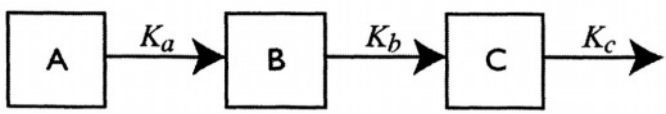

Fig. 1. The radioactive disintegration chain

$$X_a(t) = X_a(t_0) \cdot e^{-K_a(t-t_0)},$$

$$X_b(t) = \frac{K_a}{K_b - K_a} X_a(t_0) \cdot e^{-K_a(t-t_0)} + \left(X_b(t_0) + \frac{K_a}{K_a - K_b} X_a(t_0) \right) \cdot e^{-K_b(t-t_0)},$$

$$X_c(t) = \frac{K_a K_b}{K_b - K_a} X_a(t_0) \cdot e^{-K_a(t-t_0)} +$$
$$+ \left(\frac{K_b}{K_c - K_b} X_b(t_0) + \frac{K_a K_b}{(K_b - K_a)(K_b - K_c)} X_a(t_0) \right) \cdot e^{-K_b(t-t_0)} +$$
$$+ \left(X_c(t_0) + \frac{K_b}{K_b - K_c} X_b(t_0) + \frac{K_a K_b}{(K_c - K_a)(K_c - K_b)} X_a(t_0) \right) \cdot e^{-K_c(t-t_0)}$$

and so forth.

Many experimental observations have shown that this compartmental model is consistent with the behavior of all radioactive substances, thus confirming the hypothesis incorporated into equations (1) and (2), i.e. that radioactive decay is a first order process.

4.2. THE BENKE EQUATIONS

The phenomenon of nitrogen absorption by, and elimination from, the various tissues via the lung and circulation was studied experimentally in dogs and man from the viewpoint of determining cardiac output, determining body composition, and prevention of decompression sickness by Benke et al.[5]. They represented their results (Fig. 2), obtained by measuring accumulated nitrogen elimination from human subjects breathing pure oxygen, with the equation

$$Y(t) = A(1 - e^{-k\,t}), \tag{3}$$

where $Y(t)$ is the amount of nitrogen eliminated up to time t, A is the total amount of nitrogen contained in the body at time $t = 0$, the instant breathing pure oxygen began, and k an appropriate constant.

The above equation becomes easier to interpret if written in differential form. By eliminating the exponential from equation (3) and its derivative,

$$dY/dt = A\,k\,e^{-k\,t},$$

we get

$$dY/dt = k\,[A - Y(t)].$$

Here $A - Y(t)$ is the amount of nitrogen present in the body at time t, therefore k is its fraction eliminated per unit time.

Proceeding in their analysis, Behnke et al. observed that the value of k calculated from equation (3) decreases after the first 25 minutes; their explanation was that during the first part of the experiment the nitrogen is eliminated mostly from the body fluids, while later it is released mostly from fatty tissues. A better model of the experiment is therefore given by equation

$$Y(t) = B\left(1 - e^{-k_1 t}\right) + C\left(1 - e^{-k_2 t}\right), \tag{4}$$

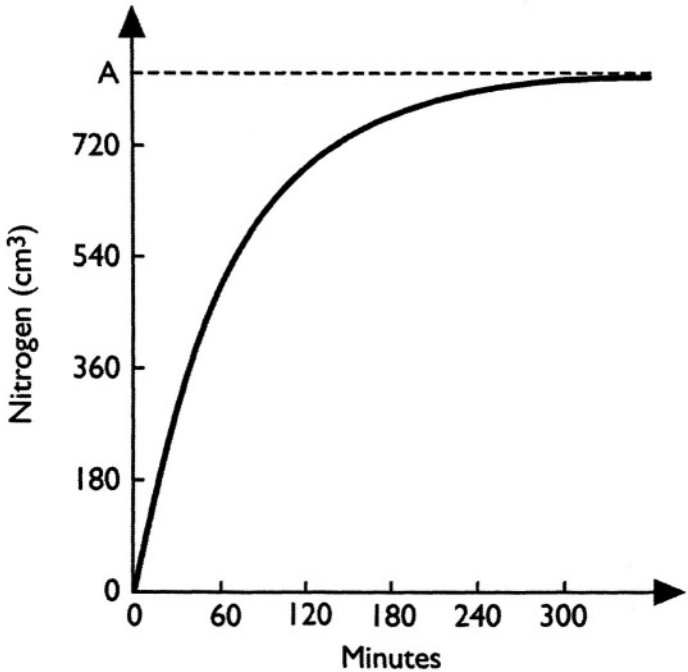

Fig. 2. Elimination of nitrogen from human subjects breathing pure oxygen

where B is the total amount of nitrogen contained in water, and C the total amount contained in fatty tissues, with $A = B + C$.

In modern terminology we would call this a "two compartments in parallel" model (see Fig. 3). Alternatively, we can think of a "two compartments in series" model (see Fig. 4), described by equations

$$\frac{dY}{dt} = k_1 X(t), \qquad Y(0) = 0$$

$$\frac{dX}{dt} = k_2 Z(t) - k_1 X(t), \qquad X(0) = X_0$$

$$\frac{dZ}{dt} = -k_2 Z(t), \qquad Z(0) = Z_0$$

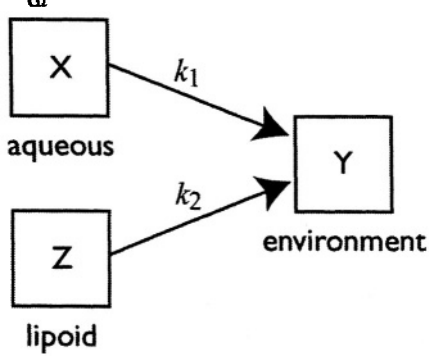

Fig. 3. Parallel model of nitrogen elimination

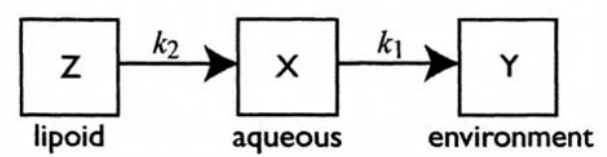

Fig. 4. Series model of nitrogen elimination

where $X(t)$ is the amount of nitrogen present in the body fluids at time t, $Z(t)$ the amount of nitrogen present in fatty tissues at time t, and X_0 and Z_0 are their respective initial quantities, with $X_0 + Z_0 = A$.

The solution of the above equations is the same as given in (4), but with

$$B = X_0 - \frac{k_2}{k_1 - k_2} Z_0, \qquad C = \frac{k_1}{k_1 - k_2} Z_0.$$

As both models give the same good fit, the choice between them should by based on additional experimental evidence.

4.3. THE TEORELL EQUATIONS

In 1937 Teorell [6, 7] published a systematic study of the kinetics of drugs introduced into the mammalian body in various ways. As in the analysis discussed above, the assumption about the transport and the definition of regions or compartments wherein measurements are to be made lead to a set of differential equations with constant coefficients. Beyond that, however, two other interesting considerations appear in this paper.

One is the idea of chemical transformation as a route between compartments where the latter term now has a more general meaning. Theorell's concern was the disappearance of a drug from the blood or tissue. The activity of the drug, being dependent upon its chemical form, could decrease in kinetically identical ways by transport to another spacial region (elimination) or by transformation to another chemical form (inactivation). Thus, compartment is defined here as a state characterized by spacial localization and chemical nature. This is a useful generalization that will be discussed in the next pages.

The other idea is the distinction between what one may call Fick kinetics and what one may call stochastic kinetics. We dealt with stochastic kinetics in the first example of transport of inert gases between compartments associated with pulmonary function. In this case the particles which collectively constitute a variable associated with a compartment each have a constant probability of transport from that compartment to any other. The instantaneous rate of loss thus is proportional to the number of particles (amount of substance) present at that instant. A set of n such compartments then is represented by a set of n equations

$$\frac{dX_i}{dt} = \sum_{j \neq i}^{1 \cdots n} k_{ji} X_j(t) - K_i X_i(t), \quad i = 1, 2, \ldots, n, \tag{5}$$

where $X_i(t)$ is the amount of drug (number of particles) present in compartment i, the constant k_{ji} is the fraction of drug in compartment j that is transferred to compartment i per unit time, and the constant K_i is the total fraction efflux of drug from compartment i per unit time. This is the kinetic form attributed by Teorell to the resorption of a drug from a subcutaneous depot.

On the other hand, Teorell assumed for the transport between blood and tissues what one may call Fick kinetics. This may be expressed by the equation

$$\phi = A(\varphi_j - \varphi_i),$$

where ϕ is the net flux from compartment j to compartment i, φ_j and φ_i are the activity in compartment i and j, respectively, and A is a constant. Here the driving force for transport is activity, a thermodynamic quantity, rather than an amount of substance. Then with the assumption that the activity of a chemical entity is adequately approximated by its concentration and that the rate of change of concentration in a homogeneous constant volume is proportional to the net flux across its boundary, we have the relations

$$\frac{dC_i}{dt} = \sum_{j \neq i}^{1 \cdots n} h_{ij}(C_j - C_i), \quad i = 1, 2, \ldots, n, \tag{6}$$

where h_{ij} is the *permeability constant* for the barrier of constant thickness and area between the compartments i and j. These n equations represent the kinetics of the system of compartments governed by Fick kinetics.

Equations (5) are more general than equations (6). This is to be expected, for the latter set follows from physical conditions that narrow the applicability. We can see the relation between these sets of equations and the respective parameters and variables as follows. Consider equation (6), where the symbols C and h have the meanings given above. Then we define a variable X by

$$C_i = X_i/V_i, \tag{7}$$

where V is simply a parameter which is independent of time. Then equations (6) become

$$\frac{dX_i}{dt} = \sum_{j \neq i}^{1 \cdots n} h_{ij} V_i \left(\frac{X_j}{V_j} - \frac{X_i}{V_i} \right). \tag{8}$$

Now we define the new parameters

$$k_{ji} = h_{ji} \frac{V_j}{V_j}, \quad K_i = \sum_{j \neq i}^{1 \cdots n} h_{ij}; \tag{9}$$

this definition transforms equation (8) into equation (5).

Formally, then, the Fick kinetics is a special case of stochastic kinetics where definitions (7) and (9) hold. Again formally, k_{ji} is the instantaneous time rate of increase of X_i due to X_j expressed as a fraction of X_j. Given the physical interpretation of C and h, one might choose to regard V as a volume, which then leads to the interpretation of X as an amount. Then k_{ji} becomes the *fractional turnover rate,* the fraction per unit time of X_j contributed to X_i. Though equations (6) are very restrictive, the special case of Fick kinetics is an important one, having wide use as a model for biological transport processes.

Returning to Teorell, we look at some of his results. He wrote four differential equations representing

(a) resorption, that is, passage between subcutaneous depot and blood,

(b) elimination, the passage from blood to urine,

(c) tissue take-up, the exchange between blood and tissue,

(d) tissue inactivation.

Processes (a) and (c) were represented by Fick-type equations; thus

$$\text{rate of resorption} = k_1\left(\frac{x}{V_1} - \frac{y}{V_2}\right),$$

$$\text{rate of tissue takeup} = k_3\left(\frac{y}{V_2} - \frac{z}{V_3}\right).$$

Here x is the amount of drug in the depot, y is the amount of drug in the blood, z is the amount of drug in the tissue, V_1 is the volume of the depot, V_2 is the volume of the blood, and V_3 is the volume of the tissue. However, V_1 is very small compared to V_2, in man approximately 50 ml versus 5 liters. Therefore he simplified to

$$\text{rate of resorption} = \frac{k_1}{V_1} \cdot x.$$

Processes (b) and (d) were considered to be monomolecular reactions, that is, reactions of order one, thus

$$\text{rate of elimination} = \frac{k_2}{V_2} \cdot y,$$

$$\text{rate of inactivation} = \frac{k_4}{V_3} \cdot z.$$

Combining these equations Teorell wrote (see Fig. 5)

$$\frac{dx}{dt} = -\frac{k_1}{V_1} \cdot x$$

$$\frac{dy}{dt} = +\frac{k_1}{V_1} \cdot x - \frac{k_2}{V_2} \cdot y - \left(\frac{k_3}{V_2} \cdot y - \frac{k_3}{V_3} \cdot z\right)$$

$$\frac{dz}{dt} = +\frac{k_3}{V_2} \cdot y - \frac{k_3}{V_3} \cdot z - \frac{k_4}{V_3} \cdot z$$

and found that the amounts in blood and tissue as function of time are sums of exponential terms with constant coefficients.

4.4. TRACER KINETICS

The three examples discussed above represent the historical origins of the concept of compartment in physical, physiological and pharmacological problems. The concept of compartment was introduced in the radioactivity problem as a set of particle all with the same probability of transformation, the nitrogen problem as a geometric space defined by certain physical and physico-chemical properties. In Teorell's work the idea was extended to include chemical transformations. In all cases the substance followed was assumed to be chemically identifiable and the mathematics used was presented rather informally.

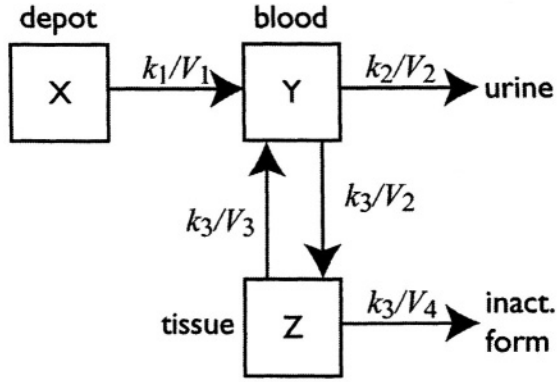

Fig. 5. Teorell model of drug disposition

In a 1938 paper Artom, Sarzana and Segré [8] presented a radioactive tracer study of the formation of phospholipid as affected by dietary fat, in which they gave a more formal analysis than in the two previous examples. They administered inorganic phosphate containing radioactive ^{32}P to rats and measured the radioactivity present in inorganic phosphate of blood, in the lipid of liver and in the skeleton at known times after administration. The physical correlate of compartment, then, is a state determined by the simultaneous existence of a particular location in space and a particular chemical state. For example, the variable representing the amount of ^{32}P in inorganic form in blood is a compartment and is distinct from the variable representing inorganic ^{32}P in the liver and distinct as well from that representing lipid ^{32}P in blood.

As a basis for their analysis Artom et al. specify four assumptions:

(a) that the organism is incapable of distinguishing between ^{32}P and ^{31}P;

(b) that the quantity of P fixed in any form whatever (for example, as lipid P) by a tissue per unit time is proportional to the amount of inorganic P in the blood; and, similarly, that the amount of inorganic P which, in the same time, is returned to the blood from the considered form is proportional to the amount of P present in that form in that tissue;

(c) that the total amount of P in each of the tissues remains constant during the experiment;

(d) that the quantity of P administered is sufficiently small that it does not modify the metabolism of the animal.

They define the following symbols:

1. N_s, N_f, N_ω represent the number of atoms of ^{31}P of the form of interest in blood, liver, and skeleton, respectively.

2. n_s, n_f, n_ω represent the analogous numbers of atoms of ^{32}P.

3. f/N_s represents the probability per unit time of fixation in the form of interest of a given atom of inorganic P by the liver.

4. ω/N_s represents the analogous probability for fixation by bone.

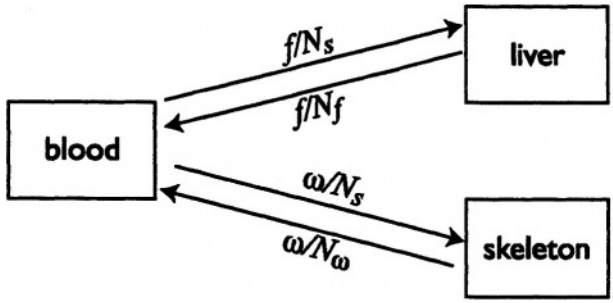

Fig. 6. Distribution of radioactive P in rats according to Artom *et al.*

From these assumptions and definitions and the additional assumption that no other appreciable exchange of P occurs (see Fig. 6), three differential equations follow:

$$\frac{dn_s}{dt} = -(f+\omega)\frac{n_s}{N_s} + \frac{f \cdot n_f}{N_f} + \frac{\omega \cdot n_\omega}{N_\omega},$$

$$\frac{dn_f}{dt} = +\frac{f \cdot n_s}{N_s} - \frac{f \cdot n_f}{N_f},$$

$$\frac{dn_\omega}{dt} = +\frac{\omega \cdot n_s}{N_s} - \frac{\omega \cdot n_\omega}{N_\omega}.$$

These three equations are analogous with equations (5), where, say, $k_{fs} = f/N_f$, and so forth. The solutions as functions of time, Artom et al. go on to say, are in general sums of three exponential. The constants of the exponents are characteristic of the system, that is, they depend upon f, ω, N_s, N_f, N_ω; the coefficients on the other hand are constants dependent upon these parameters and the initial conditions of the experiment.

It is of interest to note here that the parameters f, ω play a two-way role in this case as does the permeability parameter h_{ij} in the case of Fick kinetics in the previous example. The reason is quite different, however. In this case f, ω are number of atoms per unit time transported between compartments. Hence the number of atoms per unit time transported from blood inorganic P to liver lipid P, say, is f. The probability per unit time of transport for a single atom, then, is f/N_s and the number per unit time of radioactive atoms transported is fn_s/N_s. That the same parameter f appears in the term for transport from liver lipid P to blood inorganic P is required by the assumption (c) quoted above. It should be clear that the probabilities per unit time of transport between liver lipid and blood phosphate (f/N_f, f/N_s) are not necessarily equal in the two directions, but what is equal is the number of atoms exchanged in the two directions per unit time (f). Furthermore, if there were a path for transport from liver to bone not including blood inorganic P, then this steady-state assumption would not imply the single parameter f for both directions.

4.5. REFERENCES

1. H. Becquerel, Sur les radiations émises par phosphorescence, *Comptes rendus de l'Académie des Sciences (Paris)* **122**, 420-1 (1896).
2. H. Becquerel, Sur les radiations invisibles émises par les corps phosphorescents, *Comptes rendus de l'Académie des Sciences (Paris)* **122**, 501-3 (1896).

3. E. Rutherford and B. A. Soddy, The cause and nature of radioactivity, *Philosophical Magazine* **4**, 370 (1902).
4. E. Rutherford, The succession of changes in radioactive bodies, *Royal Society of London: Philosophical Transactions* **204**, 169 (1904).
5. A. R. Behnke, R. M. Thomson and L. A. Shaw, The rate of elimination of dissolved nitrogen in man in relation to the fat and water content of the body, *Am. J. Physiol.* **114**, 137-46 (1935).
6. T. Teorell, Kinetics of Distribution of Substances Administered to the Body. I. The Extravascular Modes of Administration. *Archives Internationales de Pharmacodynamie et de Thérapie* **57**, 205-25 (1937).
7. T. Teorell, Kinetics of Distribution of Substances Administered to the Body. II. The Intravascular Modes of Administration. *Archives Internationales de Pharmacodynamie et de Thérapie* **57**, 226-40 (1937).
8. C. Artom, G. Sarzana and E. Segré, Influence des grasses alimentaires sur la formation des phospholipides dans les tissues animaux (nouvelles recherches) *Archives Internationales de Physiologie* **47**, 245-76 (1938).

This page intentionally left blank

5. LINEAR DIFFERENTIAL EQUATIONS

5.1. GENERAL ASSUMPTIONS

In all four examples shown in chapter 4, "The Birth of Compartments," the differential equations arrived at were of the general form

$$\frac{dx_i}{dt} = a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{ni}x, \quad i = 1, 2, \ldots, n; \tag{1}$$

the solution of those equations leads in general to a sum of exponential functions of the form

$$x_i(t) = A_{i1}e^{-\alpha_1 t} + A_{i2}e^{-\alpha_2 t} + \cdots + A_{in}e^{-\alpha_n t}, \quad i = 1, 2, \ldots, n;$$

nevertheless there is a real danger of using the exponential law too frequently. The first reason is that the exponential function is easier to manipulate and it describes most linear phenomena, but not all of them; for instance the diffusion is a linear phenomenon, but it is not exponential at all. A second reason, and the most dangerous of all, is that almost any reasonably smooth function can be fitted with a sum of exponential functions, thus hiding its real meaning under a purely phenomenological description.

I shall try here to develop the compartmental equations keeping always in mind their physical (physiological and pharmacological) meaning, starting from some simple cases and proceeding toward a more general solution.

5.2. ONE COMPARTMENT

For a single compartment we have equation

$$\frac{dx}{dt} = -K\,x(t) + r(t), \tag{2}$$

where

$x(t)$ = amount of substance in the compartment,

$r(t)$ = entry rate into the compartment,

K = fractional exit rate from the compartment.

Observe that equation (2) is simply a conservation equation, i.e., it states that the variation dx/dt of the quantity $x(t)$ present is the difference between its rate of entry and its rate of exit. Furthermore a fundamental hypothesis is declared by equation (2), that the

rate of exit of the drug from the compartment is proportional to the amount of drug present; this implies that the process causing this exit is a process of order one.

The integral of equation (2) is

$$x(t) = e^{-Kt}\left(x(0) + \int_0^t e^{K\tau} \cdot r(\tau)\, d\tau\right), \tag{3}$$

where $x(0)$ is the amount of drug present in the compartment at the initial time. Expression (3) is useful only when $r(t)$ has a very simple form. This is true in the first two cases described below.

5.2.1. No recirculation, single bolus administration

If the drug is administered as a single bolus at time $t = 0$ and there is no recirculation, then $r(t) \equiv 0$ and the integral of equation (2) becomes

$$x(t) = x(0) \cdot e^{-Kt}, \tag{4}$$

where $x(0)$ is the amount of drug administered as a bolus. Expression (4) can be transformed logarithmically into

$$\ln x(t) = \ln x(0) - K t,$$

showing that $\log x(t)$ is a linear functions of t.

5.2.2. No recirculation, constant infusion

If the drug is administered by constant infusion and there is no recirculation, then $r(t) \equiv r = \text{constant}$, and the integral of equation (2) becomes

$$x(t) = r\, e^{-Kt} \int_0^t e^{K\tau}\, d\tau = \frac{r}{K}\left(1 - e^{-Kt}\right).$$

Observe that

$$\lim_{t \to \infty} x(t) = \frac{r}{K},$$

therefore

$$\lim_{t \to \infty} x(t) - x(t) = \frac{r}{K} e^{-Kt},$$

and again a logarithmic transformation shows that

$$\ln\left[\lim_{t \to \infty} x(t) - x(t)\right] = \ln\frac{r}{K} - K t$$

is a linear functions of t.

Usually we cannot measure the amount of drug in an organ, but only its concentration, therefore sometimes it may be convenient to transform equation (2) by dividing both sides by V, the volume of the compartment, to get

$$\frac{dc}{dt} = -K\,c(t) + \frac{r(t)}{V}, \tag{5}$$

where $c(t) = x(t)/V$ is the concentration of the drug in the compartment. From a physical point of view there is an important difference between equation (2) and equation (5); in fact the last one cannot be viewed as a conservation equation, because the concentration is not a conserved quantity; besides, $r(t)/V$ has the somewhat unusual dimension $[L^{-3}MT^{-1}]$, not a flow rate. Nevertheless from a mathematical point of view equations (2) and (5) are formally identical, and any solution of equation (2) can become a solution of equation (5) by substituting $c(t)$ to $x(t)$ and $r(t)/V$ to $r(t)$.

If we multiply and divide the first term of the right-hand side of equation (2) by V we get

$$\frac{dx}{dt} = -V\,K \cdot c(t) + r(t); \tag{6}$$

this form of the one-compartment equation may seem awkward, because it mixes amount of drug and concentration in the same equation, but it is interesting from two points of view; first, it shows that the quantity $x(t)$ is conserved, second it shows that the quantity eliminated per unit time is the product of two quantities, one intensive, $c(t)$, the other extensive, VK. Now this last quantity is the *clearance,* and equation (6) leads to some very interesting conclusion, as shown in later chapters.

Let us now perform a thought experiment. We administer the drug to the compartment with an infusion at a constant flow rate $r(t) \equiv r$; after a sufficiently long time a steady state x_{ss} is reached with $r = K\,x_{ss}$; we can write

$$\frac{\text{amount of drug present}}{\text{rate of elimination of drug}} = \frac{x_{ss}}{K \cdot x_{ss}} = \frac{1}{K};$$

but the ratio between the amount present and its rate of elimination is the time elapsed for eliminating an amount of drug equal to the amount present; this time does not depend on r, and it is called the *turnover time* of the compartment.

5.2.3. Single bolus administration, recirculation possible

This is an important case where we cannot use equation (3), but we can get some information using directly equation (2). When recirculation is possible, function $r(t)$ is generally unknown, but we can certainly say that at time $t = 0$ there is no recirculation, i.e., $r(0) = 0$, therefore from equation (2) we get

$$\lim_{t \to 0} \frac{-dx/dt}{x(t)} = K$$

and from a number of experimental values of dx/dt divided by $x(t)$ we can extrapolate the value of K; this parameter is called *turnover rate,* and it is the inverse of the turnover time defined above.

5.3. TWO COMPARTMENTS

For simplicity I consider two compartments with a single bolus administration in compartment one only. The equations are

$$\begin{aligned} \frac{dx_1}{dt} &= -K_1 x_1(t) + k_{21} x_2(t) \\ \frac{dx_2}{dt} &= +k_{12} x_1(t) - K_2 x_2(t) \end{aligned} \tag{7}$$

with the initial conditions

$$x_1(0) = x_0, \; x_2(0) = 0. \tag{8}$$

In those equations K_1 and K_2 are the fractional rate of elimination from compartments 1 and 2, respectively, while k_{12} and k_{21} are the fractional transfer rates from compartment 1 to compartment 2, and vice-versa, respectively. For the conservation of matter it must necessarily be

$$K_1 \geq k_{12}\,, \; K_2 \geq k_{21}.$$

The integral of the above differential equations is

$$\begin{aligned} x_1(t) &= \frac{x_0}{\beta - \alpha}\left[(K_2 - \alpha)e^{-\alpha t} + (\beta - K_2)e^{-\beta t}\right] \\ x_2(t) &= \frac{x_0 k_{12}}{\beta - \alpha}\left[(e^{-\alpha t} - e^{-\beta t}\right] \end{aligned}$$

where $-\alpha$ and $-\beta$ are the roots of the ordinary equation

$$x^2 + (K_1 + K_2)\,x + K_1 K_2 - k_{12} k_{21} = 0, \tag{9}$$

provided $\alpha \neq \beta$, which is always true if both k_{12} and k_{21} are not zero.

In the special case $k_{12} = 0$ the drug cannot reach the second compartment therefore the differential equations become

$$\begin{aligned} \frac{dx_1}{dt} &= -K_1 x_1(t) + k_{21} x_2(t) \\ \frac{dx_2}{dt} &= -K_2 x_2(t) \end{aligned}$$

and their integral is

$$\begin{aligned} x_1(t) &= x_0\, e^{-K_1 t}, \\ x_2(t) &= 0. \end{aligned}$$

In the other special case $k_{21} = 0$ the drug cannot return from compartment 2 to compartment 1, therefore we have the differential equations

$$\frac{dx_1}{dt} = -K_1 x_1$$
$$\frac{dx_2}{dt} = +k_{12} x_1 - K_2 x_2$$

whose integral is

$$x_1(t) = x_0 e^{-K_1 t}$$
$$x_2(t) = \frac{x_0 k_{12}}{K_2 - K_1}\left[e^{-K_1 t} - e^{-K_2 t}\right]$$

provided $K_1 \neq K_2$. In the special sub-case $K_1 = K_2$, the integral is

$$x_1(t) = x_0 e^{-K_1 t},$$
$$x_2(t) = x_0 k_{12}\, t\, e^{-K_1 t}.$$

5.4. SEVERAL COMPARTMENTS

With more than two compartments the relevant differential equations become

$$\begin{aligned}
\frac{dx_1}{dt} &= -K_1 x_1(t) + k_{21} x_2(t) + \cdots + k_{n1} x_n(t) \\
\frac{dx_2}{dt} &= +k_{12} x_1(t) - K_2 x_2(t) + \cdots + k_{n2} x_n(t) \\
&\cdots\cdots \\
\frac{dx_n}{dt} &= +k_{1n} x_1(t) + k_{2n} x_2(t) + \cdots - K_n x_n(t)
\end{aligned} \tag{10}$$

where

$x_i(t)$ = amount of drug in compartment i,

K_i = fractional rate of exit from compartment i,

k_{ij} = fractional rate of transfer from compartment i to compartment j.

For the principle of conservation of matter we must have

$$K_1 \geq k_{12} + k_{13} + \cdots + k_{1n},\ K_2 \geq k_{21} + k_{23} + \cdots + k_{2n},\ \ldots\ldots,$$

and in general

$$K_i \geq \sum_{j \neq i}^{1 \cdots n} k_{ij}. \tag{11}$$

If we define

$$k_{i0} = K_i - \sum_{j \neq i}^{1 \cdots n} k_{ij}, \quad i = 1, 2, \ldots, n,$$

the parameters k_{io} are the fractional transfer rates from a compartment to outside of the system.

To transform the above equations from amount of drug to concentration, we divide each term by the corresponding volume V_i, thus

$$\begin{aligned}
\frac{dc_1}{dt} &= -K_1 c_1(t) + \frac{V_2 k_{21}}{V_1} c_2(t) + \cdots + \frac{V_n k_{n1}}{V_1} c_n(t), \\
\frac{dc_2}{dt} &= +\frac{V_1 k_{12}}{V_2} c_1(t) - K_2 c_2(t) + \cdots + \frac{V_n k_{n2}}{V_2} c_n(t), \\
&\cdots\cdots, \\
\frac{dc_n}{dt} &= +\frac{V_1 k_{1n}}{V_n} c_1(t) + \frac{V_2 k_{2n}}{V_n} c_2(t) + \cdots - K_n c_n(t).
\end{aligned} \tag{12}$$

Observe that equations (12) are formally identical with equations (10), but the meanings of their coefficients are quite different. Even though concentrations are used more often than amounts, we will use equations (10) in preference to equations (12), because their physical meaning is more apparent.

If we differentiate $n-1$ times the first of equations (10) and $n-2$ times each of the others equations, we get a total of n equations from the first one and $n-1$ equations from each of the other $n-1$, for a grand total of n^2-n+1 differential equations, containing x_1 and its first n derivatives, plus the other $n-1$ functions x_i $(i>1)$ with their first $n-1$ derivatives; now we can eliminate all these last functions with their derivatives and we are left with one differential equation containing only the function x_1 with its first n derivatives. In short, from n first order differential equations in n different functions we can obtain one differential equation of order n in one function. The solution of these equations is much simpler if we use the operational calculus.

5.5. OPERATIONAL NOTATION

Using the operational notation, (see appendix B) equations (10) become

$$\begin{aligned}
s\{x_1\} - x_1(0) &= -K_1\{x_1\} + k_{21}\{x_2\} + \cdots + k_{n1}\{x_n\}, \\
s\{x_2\} - x_2(0) &= +k_{12}\{x_1\} - K_2\{x_2\} + \cdots + k_{n2}\{x_n\}, \\
&\cdots\cdots, \\
s\{x_n\} - x_n(0) &= +k_{1n}\{x_1\} + k_{2n}\{x_2\} + \cdots - K_n\{x_n\},
\end{aligned} \tag{13}$$

and reordering all terms,

$$\begin{aligned}
+(s+K_1)\{x_1\} - k_{21}\{x_2\} - \cdots - k_{n1}\{x_n\} &= x_1(0), \\
-k_{12}\{x_1\} + (s+K_2)\{x_2\} - \cdots - k_{n2}\{x_n\} &= x_2(0), \\
&\cdots\cdots, \\
-k_{1n}\{x_1\} - k_{2n}\{x_2\} - \cdots + (s+K_n)\{x_n\} &= x_n(0).
\end{aligned} \tag{14}$$

In general the dose is administered in one compartment only; there is no loss of generality if we suppose this to be compartment one. With this hypothesis the terms at the

right-hand side of equations (14) become $x_1(0), 0, 0, \ldots$ and the solution of those equations is

$$\frac{\{x_i(t)\}}{x_1(0)} = \frac{\Delta_{i;1}(s)}{\Delta(s)}, \quad i = 1, 2, \ldots, n \tag{15}$$

where

$$\Delta(s) = \begin{vmatrix} s+K_1 & -k_{12} & \cdots & -k_{1n} \\ -k_{21} & s+K_2 & \cdots & -k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -k_{n1} & -k_{n2} & \cdots & s+K_n \end{vmatrix}$$

is the determinant formed by the coefficients on the left-hand side of equations (14), and $\Delta_{i;1}(s)$ is the determinant obtained from $\Delta(s)$ by suppressing row i and column 1 [1]. By developing $\Delta(s)$ we get a polynomial in s of degree n,

$$\Delta(s) = s^n + q_1 s^{n-1} + + q_2 s^{n-2} + \ldots + q_n,$$

where q_i is the sum of all products of the constants K taken i by i, minus all products of the constants k forming rings of length i, for instance

$$k_{12}k_{21},\ k_{13}k_{31},\ k_{23}k_{32},\ \ldots$$
$$k_{12}k_{23}k_{31},\ k_{13}k_{32}k_{21},\ \ldots$$
$$k_{12}k_{23}k_{34}k_{41},\ k_{13}k_{32}k_{24}k_{41},\ \ldots$$

and so forth. For inequalities (11) all coefficients q_i are non-negative.

By developing $\Delta_{1;1}(s)$ we get a polynomial in s of degree $n-1$, like $\Delta(s)$ but without K_1 and all k's with the first or the second subscript equal to 1.

For $\Delta_{i;1}(s)$, with $i \neq 1$, we have the polynomial in s

$$\Delta_{i;1}(s) = p_0 s^{n-1} + p_1 s^{n-2} + \ldots + p_{n-1},$$

where

$$p_0 = k_{1i}$$
$$p_1 = \sum_{l \neq 1,i} (k_{1i}K_l + k_{1l}k_{li})$$
$$p_2 = \sum_{l_1, l_2 \neq 1,i} k_{1i}(K_{l_1}K_{l_2} - k_{l_1 l_2}k_{l_2 l_1}) + \sum_{l_1, l_2 \neq 1,i} k_{1l_1}k_{l_1 i}K_{l_2} + \sum_{l_1, l_2 \neq 1,i} k_{1l_1}k_{l_1 l_2}k_{l_2 i}$$
$$\ldots\ldots\ldots\ldots$$

In short, form all possible strings of k's such that the first subscript of the first k is 1, the last subscript of the last k is i, and the second subscript of each of them is equal to the first subscript of the following one; then complete each of those strings with an appropriate number of K's with a subscript different from the subscripts of the k' they are with, but subtract from those K's all rings.

Observe now that, if $-\lambda_1, -\lambda_2, \ldots, -\lambda_n$ are the roots of equation $\Delta(s) = 0$, then

$$q_1 = \sum_i \lambda_i,\ q_2 = \sum_{i,j} \lambda_i \lambda_j,\ q_3 = \sum_{i,j,l} \lambda_i \lambda_j \lambda_l,\ \ldots$$

We can now rewrite equation (15) as

$$\frac{\{x_i(t)\}}{x_1(0)} = \frac{p_0 s^{n-1} + p_1 s^{n-2} + \cdots + p_{n-1}}{s^n + q_1 s^{n-1} + q_2 s^{n-2} + \cdots + q_n} = \sum_{j=1}^{n} \frac{A_{ij}}{s + \lambda_j}, \quad i = 1, 2, \ldots, n; \tag{16}$$

or, in the non-operational notation,

$$\frac{x_i(t)}{x_1(0)} = \sum_{j=1}^{n} A_{ij} e^{-\lambda_j t}. \tag{17}$$

The rules to determine the coefficients A_{ij} from the p's and q's are described in Appendix B, section B.13.

Function (17) is called the *unit response function*; it is equal to the function measuring the amount of drug in compartment i when a unit dose is given to compartment 1 as a bolus at time $t = 0$. It is dimensionless.

Observe that in the corresponding function (16) the brackets $\{\}$ themselves have dimension of time, [T], therefore the fraction $\{x_i(t)\}/x_i(t)$ has dimension [T]. The homogeneity of the above expression requires that the dimension of p_i and q_i be $[T^{-i}]$, that A_{ij} be dimensionless for all values of i and j from 1 to n, and the λ_j have dimension $[T^{-1}]$.

5.6. EIGENVALUES OF A SYSTEM OF COMPARTMENTS

The *eigenvalues* of a system of compartments, also called *characteristic values,* are the roots of equation $\Delta(s) = 0$, or the exponents of the sum of exponential functions (17); we indicated them by $-\lambda_1, -\lambda_2, \ldots, -\lambda_n$ in the previous section.

The eigenvalues are typical *model parameters,* i.e., quantities that define a property of a model.

Consider, for instance, a two-compartment model with reversible reactions, represented by the differential equations (7) with initial conditions (8); the solution of those equations is

$$x_1(t) = \frac{x_0}{\beta - \alpha}\left((K_2 - \alpha)e^{-\alpha t} - (K_2 - \beta)e^{-\beta t}\right)$$

$$x_2(0) = \frac{x_0 \cdot k_{12}}{\beta - \alpha}\left(e^{-\alpha t} - e^{-\beta t}\right)$$

where the eigenvalues

$$-\alpha = -\frac{1}{2}\left(K_1 + K_2 - \sqrt{(K_1 - K_2)^2 + 4k_{12}k_{21}}\right)$$

$$-\beta = -\frac{1}{2}\left(K_1 + K_2 + \sqrt{(K_1 - K_2)^2 + 4k_{12}k_{21}}\right)$$

are the roots of equation (9). Observe that $-\alpha$ and $-\beta$ do not depend upon the initial dose x_0; therefore the eigenvalues are invariant for different doses.

Suppose now that instead of administering an initial dose x_0 to the first compartment, we feed it continuously with a constant rate of infusion r; equations (7) become

$$\frac{dx_1}{dt} = -K_1 x_1(t) + k_{21} x_2(t) + r$$

$$\frac{dx_2}{dt} = +k_{12} x_1(t) - K_2 x_2(t)$$

with initial conditions

$$x_1(0) = x_2(0) = 0;$$

the solution is now

$$x_1(t) = \frac{r}{\beta - \alpha}\left(\frac{K_2}{\alpha} - \frac{K_2}{\beta} - (\frac{K_2}{\alpha} - 1)e^{-\alpha t} + (\frac{K_2}{\beta} - 1)e^{-\beta t}\right)$$

$$x_2(0) = \frac{r \cdot k_{12}}{\beta - \alpha}\left(\frac{1 - e^{-\alpha t}}{\alpha} - \frac{1 - e^{-\beta t}}{\beta}\right)$$

with the same eigenvalues as before. The eigenvalues, therefore, are invariant not only to the dose, but also to the mode of administration. This result of course applies to the compartmental model, and it is not necessarily true for any pharmacokinetic system, unless it has been specifically confirmed experimentally.

5.7. PROPERTIES OF THE EIGENVALUES

Hadamard [2] did show that, when inequalities (11) hold with all $k_{ij} \geq 0$, the real eigenvalues of $\Delta(s)$ are non-positive, and the complex eigenvalues have the real part non-positive. This property of the eigenvalues will be investigated in more detail in Chapter 11, Matrix Equations. For the time being we observe that if one of the eigenvalues is zero, then q_n must be zero. Remembering that q_n is the sum of all products of the constants K taken i by i, minus all products of the constants k forming rings of length n, then inequalities (11) imply that $q_n = 0$ only if

$$K_i = \sum_{j \neq i}^{1 \cdots n} k_{ij}, \quad i = 1, 2, \ldots, n. \tag{18}$$

When inequalities (11) are strict inequalities, expression (17) converges to zero for $t \to \infty$; when identities (18) hold, the expression (17) contains at least one exponential term with a non-negative coefficient, therefore for $t \to \infty$ it does not converge to zero. When this is the case, we say that the compartmental system is *closed.*

5.8. DETERMINATION OF THE EIGENVALUES

Determining the eigenvalues from the experimental data is not always an easy problem. Many times the observation errors propagate in such a way as to invalidate most of the numerical procedures towards this goal. In general, the easiest eigenvalue that can be computed is the smallest one in absolute value.

Suppose that a particular drug in a particular organ is characterized by three eigenvalues; in other words, the function $c(t)$ representing the concentration of that drug in the plasma is a sum of three exponential functions; then

$$\frac{V \cdot c(t)}{D} = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} + A_3 e^{-\lambda_3 t} \tag{19}$$

where $-\lambda_1, -\lambda_2, -\lambda_3$ are the three eigenvalues, V is the volume of the compartment and D the dose administered. Suppose also that

$$\lambda_1 < \lambda_2 < \lambda_3.$$

When t increases, the last two exponential functions decrease faster than the first, so that after a sufficiently long time

$$\frac{V \cdot c(t)}{D} \approx A_1 e^{-\lambda_1 t},$$

and as a consequence,

$$\ln\left(\frac{V \cdot c(t)}{D}\right) \approx -\lambda_1 \cdot t + \ln A_1. \tag{20}$$

The plot of $\ln\left(V \cdot c(t)/D\right)$ versus t approaches a straight line of slope $-\lambda_1$ when t increases. Therefore λ_1 can easily be determined by plotting $V \cdot c(t)/D$ as a function of t on a semilogarithmic scale and extrapolating for $t \to \infty$.

The interval of time necessary for $V \cdot c(t)/D$ to decrease 50%, in the range of t where the approximation of expression (20) is valid, is called η or *terminal half-life* of that drug. Clearly

$$\eta = \frac{\ln 2}{\lambda_1} = \frac{0.693}{\lambda_1}.$$

Apart from the factor ln 2, the terminal half-life is just one of the eigenvalues of the model, therefore a model parameter [3].

The unreliability of the determination of the terminal half-life has been demonstrated experimentally [4] and theoretically [5]. It is important to remember that the slowest identified eigenvalue of $c(t)$ may be equal to the turnover rate of the plasma itself, or of one of its precursors, or of no compartments of the system.

I mentioned earlier that this particular eigenvalue is easy to determine, but this is not always the case. There are at least two cases when this determination is difficult and inaccurate. If A_1 is very small, the approximation in equation (20) is still valid, but only for values of $V \cdot c(t)/D$ correspondingly small, that is for measurements of $V \cdot c(t)/D$ taken for large values of t, when experimental errors are more likely. This difficulty sometimes can be overcome. In fact, λ_1 is invariant, but A_1 is not; if the initial conditions are modified appropriately, for instance, by using a continuous infusion, A_1 may sufficiently increase while λ_1 stays constant.

Another case when the determination of λ_1 is difficult is when $\lambda_1 \approx \lambda_2$; in this case the approximation (20) is still valid, but only for very large values of t. Suppose that

$$\lambda_2 - \lambda_1 = \varepsilon, \quad \lambda_3 > \lambda_2,$$

where ε is small; the coefficients of equation (19), as shown by Rescigno and Beck [6], are

$$A_1 = \frac{1}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)}, \quad A_2 = \frac{1}{(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3)}, \quad A_3 = \frac{1}{(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)},$$

therefore

$$\frac{V \cdot c(t)}{D} = \frac{(\lambda_3 - \lambda_2)e^{-\lambda_1 t} - (\lambda_3 - \lambda_1)e^{-\lambda_2 t} - (\lambda_2 - \lambda_1)e^{-\lambda_3 t}}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)}.$$

For t very large we may use the approximation

$$\begin{aligned}\frac{V \cdot c(t)}{D} &\simeq \frac{(\lambda_3 - \lambda_2)e^{-\lambda_1 t} - (\lambda_3 - \lambda_1)e^{-\lambda_2 t}}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)} \\ &\simeq \frac{e^{-\lambda_1 t}}{(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)} \cdot \frac{\lambda_3 - \lambda_1 - (\lambda_3 - \lambda_1)e^{-\varepsilon t}}{\varepsilon}\end{aligned}$$

The limit of the last fraction on the right-hand side can be determined using L'Hospital rule; in fact

$$\begin{aligned}\lim_{\varepsilon \to 0} \frac{\lambda_3 - \lambda_1 - (\lambda_3 - \lambda_1)e^{-\varepsilon t}}{\varepsilon} &= \lim_{\varepsilon \to 0} \left(t \cdot (\lambda_3 - \lambda_1)e^{-\varepsilon t} \right) \\ &= (\lambda_3 - \lambda_1) \cdot t\end{aligned}$$

therefore, for ε very small,

$$\frac{V \cdot c(t)}{D} \simeq \frac{t \cdot e^{-\lambda_1 t}}{\lambda_3 - \lambda_2},$$

an expression whose logarithm, for $t \to \infty$, does not approach a straight line.

More details on the eigenvalues of a model can be found in the literature [7, 8].

5.9. COMPLEX EIGENVALUES

I have shown in section 5.7 that some roots of equation $\Delta(s) = 0$ may be complex; we shall examine here what are the conditions for this to happen, and what its physical consequences.

We first observe that with only two compartments

$$\Delta(s) = s^2 + (K_1 + K_2)s + K_1K_2 - k_{12}k_{21},$$

therefore the two eigenvalues are

$$\tfrac{1}{2}\left[-(K_1 + K_2) \pm \sqrt{(K_1 - K_2)^2 + 4k_{12}k_{21}} \right];$$

the expression under the square root is the sum of two non-negative quantities and can never be negative, therefore no complex eigenvalues are possible with two compartments.

With more than two compartments, complex eigenvalues are possible; let us consider as an example the case of three compartments. In this case the eigenvalues, if any, are roots of equation

$$s^3 + (K_1 + K_2 + K_3)s^2 + (K_1K_2 + K_1K_3 + K_2K_3 - k_{12}k_{21} - k_{13}k_{31} - k_{23}k_{32})s$$
$$+K_1K_2K_3 - k_{12}k_{21}K_3 - k_{13}K_2k_{31} - K_1k_{23}k_{32} - k_{12}k_{23}k_{31} - k_{13}k_{32}k_{21} = 0;$$

this equation can also be written as

$$s^3 + (K_1 + K_2 + K_3)s^2 + (K_1K_2 + K_1K_3 + K_2K_3)s + K_1K_2K_3 =$$
$$= (k_{12}k_{21} + k_{13}k_{31} + k_{23}k_{32})s + k_{12}k_{21}K_3 + k_{13}K_2k_{31} + K_1k_{23}k_{32} + k_{12}k_{23}k_{31} + k_{13}k_{32}k_{21}$$

and can graphically be represented by the intersection of a cubic and a straight line; the cubic intersects the abscissa at the points $-K_1$, $-K_2$, $-K_3$ and the ordinate at the point $K_1K_2K_3$; the straight line intersects the ordinate at the point

$$r = k_{12}k_{21}K_3 + k_{13}K_2k_{31} + K_1k_{23}k_{32} + k_{12}k_{23}k_{31} + k_{13}k_{32}k_{21}$$

with condition (11) requiring

$$r \leq K_1K_2K_3;$$

the slope of the straight line is

$$\alpha = k_{12}k_{21} + k_{13}k_{31} + k_{23}k_{32},$$

while the slope of the cubic at the point of intersection with the ordinate is

$$\beta = K_1K_2 + K_1K_3 + K_2K_3,$$

and the condition (11) requires $\beta \geq \alpha$.

In general, the straight line intersects the cubic at three points, with abscissae equal to the real eigenvalues, as shown in Fig. 1; if we make α as small as possible and the folding of the cubic as small as possible, there will be just one intersection between the cubic and the straight line, and consequently one real eigenvalue and two complex ones. To this purpose we can make

$$k_{21} = k_{32} = k_{13} = 0 \text{ and } K_1 = K_2 = K_3$$

and the eigenvalues are the roots of equation

$$(s + K_1)^3 = k_{12}\, k_{23}\, k_{31},$$

i.e.,

$$-K_1 + k, \quad -K_1 - \left(\frac{1}{2} + i\frac{\sqrt{3}}{2}\right)k, \quad -K_1 - \left(\frac{1}{2} - i\frac{\sqrt{3}}{2}\right)k,$$

where

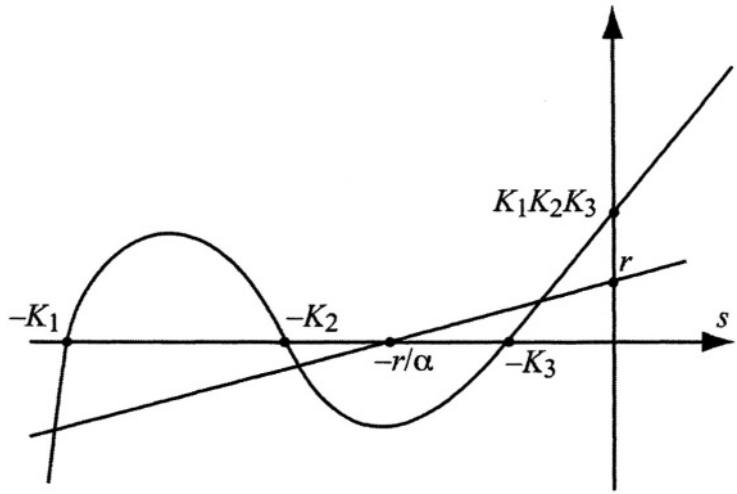

Fig. 1. In a system of three compartments a straight line intersects a cubic where s is equal to the eigenvalues

$$k = \sqrt[3]{k_{12}k_{23}k_{31}}\,.$$

The solution (see appendix B) is

$$\frac{\{x_1\}}{D} = \frac{(s+K_1)^2}{(s+K_1)^3 - k^3} = \frac{1/3}{s+K_1-k} + \frac{2}{3}\cdot\frac{s+K_1+k/2}{(s+K_1+k/2)^2+\frac{3}{4}k^2}$$

thence

$$\frac{x_1(t)}{D} = \frac{1}{3}e^{-(K_1-k)t} + \frac{2}{3}e^{-(K_1+k/2)t}\cdot\cos\left(\frac{\sqrt{3}}{2}kt\right).$$

Some oscillations are present, but they are strongly damped by the exponential terms and hardly noticeable [9].

With four compartments, proceeding in the same way, we find equation

$$(s+K_1)^4 = k_{12}k_{23}k_{34}k_{41}$$

with the eigenvalues

$$-K_1+k,\quad -K_1-k,\quad -K_1+ik,\quad -K_1-ik,$$

where

$$k = \sqrt[4]{k_{12}k_{23}k_{34}k_{41}}\,.$$

The solution is

$$\frac{\{x_1\}}{D} = \frac{(s+K_1)^3}{(s+K_1)^4 - k^4} = \frac{1/4}{s+K-k} + \frac{1}{2}\cdot\frac{s+K_1}{(s+K_1)^2+k^2}$$

thence

$$\frac{x_1(t)}{D} = \frac{1}{4}e^{-(K_1+k)t} + \frac{1}{4}e^{-(K_1-k)t} + \frac{1}{2}e^{-K_1 t} \cdot \cos kt.$$

Again we have some oscillations, but they are strongly damped by the three exponential terms. Even if we make $k = K_1$, i.e., if the four-compartment system is closed, the oscillations are damped by two exponential terms.

Of course, oscillations are possible, and may be noticeable if some non-linear processes are present; but no general theory is available for this case.

5.10. REFERENCES

1. A. Rescigno, A contribution to the theory of tracer methods: Part II, *Biochim. Biophys. Acta* **21,** 111-6 (1956).
2. J. Hadamard, *Léçons sur la Propagation des Hondes et les Equations de l'Hydrodynamigue* (Herman, Paris, 1903).
3. A. Rescigno, Fundamental Concepts in Pharmacokinetics, *Pharmacol. Res.* **35,** 363-90 (1997).
4. N. A. Lassen and P. Sejrsen, Monoexponential extrapolation of tracer clearance curves in kinetic analysis, *Circ. Res.* **29,** 76-87(1971).
5. L. Bass, J. Aisbett and A. J. Bracken, Asymptotic form of tracer clearance curves: theory and applications of improved extrapolations, *J. theoret. Biol.* **111,** 755-85 (1984).
6. A. Rescigno and J. S. Beck, Compartments, in: *Foundations of Mathematical Biology, Volume II,* edited by R. Rosen (Academic Press, New York, 1972), pp. 255-322.
7. A. Rescigno and B. M. Bocchialini, Pharmacokinetics: Unfolding of a Concept, in: *New Trends in Pharmacokinetics,* edited by A. Rescigno and A. K. Thakur (Plenum Press, New York, 1991), pp. 1-26.
8. A. Rescigno and E. Rocca, Terminal Half-Life, *J. Pharmacokin. Biopharm.* **21,** 125-9, 1993.
9. A. Rescigno, R. M. Lambrecht and C. C. Duncan, Mathematical Methods in the Formulation of Pharmacokinetic Models, in: *Tracer Kinetics and Physiologic Modeling,* edited by R. M. Lambrecht and A. Rescigno (Springer-Verlag, Berlin, 1983), pp. 59-119.

6. STOCHASTIC COMPARTMENTS

6.1. STOCHASTIC PROCESS

Consider the number $X(t)$ of particles present in a compartment at time t to be a discrete random variable [1], and call $p_i(t)$ the probability that $X(t) = i$,

$$p_i(t) = \mathrm{P}\{X(t) = i\}. \quad (i = 0, 1, 2, \ldots)$$

The general property of probability requires that

$$\sum_{i=0}^{\infty} p_i(t) = 1, \quad 0 \le i \le \infty.$$

We make the hypothesis that the probability of a particle leaving the compartment in the short interval of time $(t, t + \mathrm{d}t)$ is proportional to $\mathrm{d}t$, but independent on t and on the number and age of the particles present; the probability of this event is thus $\mu \cdot \mathrm{d}t + o(\mathrm{d}t)$, where $o(\mathrm{d}t)$ is an infinitesimal of higher order than $\mathrm{d}t$.

If we call $f(t)\mathrm{d}t + o(\mathrm{d}t)$ the probability that a particle enters the compartment in the interval $(t, t + \mathrm{d}t)$, where $f(t) \ge 0$ is an arbitrary function, we can consider three possible transitions in the compartment during the above interval of time:

$$\mathrm{P}\{X(t + \mathrm{d}t) = i \mid X(t) = i + 1\} = (i + 1) \cdot \mu \cdot \mathrm{d}t + o(\mathrm{d}t)$$

$$\mathrm{P}\{X(t + \mathrm{d}t) = i \mid X(t) = i - 1\} = f(t) \cdot \mathrm{d}t + o(\mathrm{d}t)$$

$$\mathrm{P}\{X(t + \mathrm{d}t) = i \mid X(t) = i\} = 1 - i \cdot \mu \cdot \mathrm{d}t - f(t) \cdot \mathrm{d}t + o(\mathrm{d}t)$$

for any positive integer i.

We also have

$$\mathrm{P}\{X(t + \mathrm{d}t) = 0 \mid X(t) = 1\} = \mu \cdot \mathrm{d}t + o(\mathrm{d}t)$$

$$\mathrm{P}\{X(t + \mathrm{d}t) = 0 \mid X(t) = 0\} = 1 - f(t) \cdot \mathrm{d}t + o(\mathrm{d}t).$$

Clearly, the probability of two simultaneous events (entering and/or leaving the compartment) in the same short interval of time $\mathrm{d}t$, is an infinitesimal of higher order than $\mathrm{d}t$.

Combining the statements above we have

$$p_0(t + \mathrm{d}t) = [1 - f(t)\mathrm{d}t]\, p_0(t) + \mu\, \mathrm{d}t \cdot p_1(t) + o(\mathrm{d}t)$$

$$p_i(t + \mathrm{d}t) = f(t)\mathrm{d}t \cdot p_{i-1}(t) + [1 - i \cdot \mu\, \mathrm{d}t - f(t)\mathrm{d}t]\, p_i(t) + (i + 1)\mu\, \mathrm{d}t \cdot p_{i+1}(t) + o(\mathrm{d}t), \qquad i \ge 1$$

thence, rearranging and taking the limits for $t \to 0$,

$$\frac{dp_0}{dt} = -f(t) \cdot p_0(t) + \mu \cdot p_1(t), \tag{1}$$

$$\frac{dp_i}{dt} = f(t) \cdot p_{i-1}(t) - [i \cdot \mu + f(t)] \cdot p_i(t) + (i+1) \cdot \mu \cdot p_{i+1}(t), \quad i \geq 1. \tag{2}$$

These differential equations must be completed with the initial conditions

$$p_i(0) = 0, \; i \neq n,$$

$$p_n(0) = 1,$$

where n is the number of particles present in the compartment at time $t = 0$.

6.2. GENERATING FUNCTION OF $X(t)$

The generating function $G_x(s, t)$ of the random variable $X(t)$ is defined by [2]

$$G_x(s,t) = \sum_{i=0}^{\infty} p_i(t) \cdot s^i$$

with $|s| < 1$; from the definition we get

$$\frac{\partial G_x}{\partial t} = \sum_{i=0}^{\infty} \frac{dp_i}{dt} \cdot s^i,$$

$$\frac{\partial G_x}{\partial s} = \sum_{i=0}^{\infty} i \cdot p_i(t) \cdot s^{i-1}.$$

Now multiply each of the equations (2) by s^i and add them all together with equation (1):

$$\sum_{i=0}^{\infty} \frac{dp_i}{dt} \cdot s^i = f(t) \cdot \sum_{i=1}^{\infty} p_{i-1}(t) \cdot s^i - [i\mu + f(t)] \sum_{i=0}^{\infty} p_i(t) \cdot s^i + \sum_{i=0}^{\infty} \mu(i+1) p_{i+1}(t)$$

or

$$\frac{\partial G_x}{\partial t} = f(t) \cdot s \cdot G_x(s,t) - \mu \cdot s \cdot \frac{\partial G_x}{\partial s} - f(t) \cdot G_x(s,t) + \mu \cdot \frac{\partial G_x}{\partial s}.$$

Rearranging,

$$-\frac{\partial G_x}{\partial t} + \mu(1-s) \cdot \frac{\partial G_x}{\partial s} = (1-s) \cdot f(t) \cdot G_x(s,t). \tag{3}$$

Proceeding the same way with the initial conditions we get

$$G_x(s,0) = s^n. \tag{4}$$

The partial differential equation (3) with the boundary condition (4) determines the generating function $G_x(s, t)$ uniquely. The solution can be found with Lagrange's method [3], by first solving the auxiliary equations

$$\frac{dt}{-1} = \frac{ds}{\mu(1-s)} = \frac{dG_x}{(1-s)f(t)G_x(s,t)}.$$

From the first auxiliary equation

$$-\mu dt = \frac{ds}{1-s}$$

we get

$$-\mu t + \ln(1-s) = \text{constant},$$

thence,

$$1 - s = C_1 e^{\mu t}. \tag{5}$$

From the second auxiliary equation

$$-dt = \frac{dG_x}{(1-s)f(t)G_x(s,t)}$$

we get

$$\frac{dG_x}{G_x(s,t)} = -(1-s)f(t)dt,$$

thence, using (5),

$$\frac{dG_x}{G_x(s,t)} = -C_1 e^{\mu t} f(t)dt,$$

whose integral is

$$\ln|G_x(s,t)| = -C_1 \int_0^t e^{\mu\tau} f(\tau)d\tau + C_2;$$

using (5) again,

$$\ln|G_x(s,t)| = -(1-s)e^{-\mu t} \int_0^t e^{\mu\tau} f(\tau)d\tau + C_2.$$

The general integral of equation (3) is

$$C_2 = \eta(C_1),$$

where η is an arbitrary function of its argument; therefore

$$\ln|G_x(s,t)| + (1-s)e^{-\mu t} \int_0^t e^{\mu\tau} f(\tau)d\tau = \eta[(1-s)e^{-\mu t}],$$

thence

$$|G_x(s,t)| = \exp\left(\eta[(1-s)e^{-\mu t}]\right) \cdot \exp\left(-(1-s)\int_0^t e^{-\mu(t-\tau)} f(\tau)d\tau\right).$$

The expression on the right-hand side is never negative, therefore we can omit the absolute value sign on the left-hand side; we can also use the notation

$$e^{-\mu t} * f(t) = \int_0^t e^{-\mu(t-\tau)} f(\tau)\, d\tau,$$

where the symbol $*$ means *convolution*; we have thus

$$G_x(s,t) = \exp\left(\eta[(1-s)e^{-\mu t}]\right) \cdot \exp\left(-(1-s)\, e^{-\mu t} * f(t)\right). \tag{6}$$

For $t = 0$ this expression becomes

$$G_x(s,0) = \exp\left(\eta(1-s)\right),$$

and using equation (4),

$$s^n = \exp[\eta(1-s)]$$

or

$$\eta(r) = n \cdot \ln(1-r);$$

now we are able to eliminate the function η from equation (6) and get the final solution

$$G_x(s,t) = \left(1-(1-s)e^{-\mu t}\right)^n \exp\left(-(1-s) \cdot f(t) * e^{-\mu t}\right).$$

The generating function $G_x(s, t)$ can be written as the product of two generating functions,

$$G_y(s,t) = \left(1-(1-s)e^{-\mu t}\right)^n$$

and

$$G_z(s,t) = \exp\left(-(1-s) \cdot f(t) * e^{-\mu t}\right)$$

defining two new random variables $Y(t)$ and $Z(t)$, respectively. From these definitions it follows that $Y(t)$ and $Z(t)$ are stochastically independent and that

$$X(t) = Y(t) + Z(t).$$

The random variable $Y(t)$ represents the number of particles in the compartment if $f(t) \equiv 0$, i.e., if no new particles enter the compartment. Its generating function shows that $Y(t)$ has a binomial distribution corresponding to n particles with probabilities $e^{-\mu t}$ and $1-e^{-\mu t}$ of being inside or outside the compartment, respectively.

The random variable $Z(t)$ represents the number of particles in the compartment if $n = 0$, i.e. if the compartment is empty when $t = 0$. Its generating function shows that $Z(t)$ has a Poisson distribution corresponding to

$$\lambda = f(t) * e^{-\mu t},$$

where λ is the product of the number of particles present and the probability of each particle being present.

6.3. MOMENTS OF THE RANDOM VARIABLE

From the definition of generating function it follows that, for any random variable $X(t)$,

$$\lim_{s\to 1}\frac{\partial G_x}{\partial s} = \sum_{i=1}^{\infty} i\cdot p_i(t) = E[X(t)],$$

where $E[X(t)]$ is the expected value of the random function $X(t)$. Furthermore,

$$\lim_{s\to 1}\frac{\partial^2 G_x}{\partial s^2} = \sum_{i=1}^{\infty} i(i-1)\cdot p_i(t) = E[X^2(t)] - E[X(t)],$$

$$\lim_{s\to 1}\frac{\partial^3 G_x}{\partial s^3} = \sum_{i=1}^{\infty} i(i-1)(i-2)\cdot p_i(t) = E[X^3(t)] - 3\cdot E[X^2(t)] + 2\cdot E[X(t)].$$

We can therefore write,

$$E[X(t)] = \lim_{s\to 1}\left(\frac{\partial G_x}{\partial s}\right), \tag{7}$$

$$E[X^2(t)] = \lim_{s\to 1}\left(\frac{\partial^2 G_x}{\partial s^2} + \frac{\partial G_x}{\partial s}\right), \tag{8}$$

$$E[X^3(t)] = \lim_{s\to 1}\left(\frac{\partial^3 G_x}{\partial s^3} + 3\frac{\partial^2 G_x}{\partial s^2} + \frac{\partial G_x}{\partial s}\right). \tag{9}$$

The Variance of a random variable is defined by

$$\mathrm{Var}[X(t)] = E\left[\left(X(t) - E[X(t)]\right)^2\right],$$

therefore,

$$\mathrm{Var}[X(t)] = E\left[X^2(t) - E^2[X(t)]\right],$$

and using equations (7) and (8),

$$\mathrm{Var}[X(t)] = \lim_{s\to 1}\left[\frac{\partial^2 G_x}{\partial s^2} + \frac{\partial G_x}{\partial s} - \left(\frac{\partial G_x}{\partial s}\right)^2\right].$$

The Third Central Moment is defined by

$$M_3[X(t)] = E\left[\left(X(t) - E[X(t)]\right)^3\right],$$

therefore,

$$M_3[X(t)] = E\left[X^3(t) + 2 \cdot E^2[X(t) - 3 \cdot E[X(t)] \cdot E[X^2(t)]\right];$$

proceeding as before and using equation (9),

$$M_3[X(t)] = \lim_{s \to 1}\left[\frac{\partial^3 G_x}{\partial s^3} + 3 \cdot \frac{\partial^2 G_x}{\partial s^2}\left(1 - \frac{\partial G_x}{\partial s}\right) + 2 \cdot \left(\frac{\partial G_x}{\partial s}\right)^3 - 3 \cdot \left(\frac{\partial G_x}{\partial s}\right)^2 + \frac{\partial G_x}{\partial s}\right].$$

The moments of the random variable $X(t)$ are equal to the sum of the corresponding moments of $Y(t)$ and $Z(t)$. We shall therefore compute separately the moments of these two random variables.

For $Y(t)$ we obtain immediately,

$$\lim_{s \to 1} \frac{\partial G_y}{\partial s} = n \cdot e^{-\mu t},$$

$$\lim_{s \to 1} \frac{\partial^2 G_y}{\partial s^2} = n \cdot (n-1) \cdot e^{-2\mu t},$$

$$\lim_{s \to 1} \frac{\partial^3 G_y}{\partial s^3} = n \cdot (n-1) \cdot (n-2) \cdot e^{-3\mu t},$$

therefore,

$$E[Y(t)] = n \cdot e^{-\mu t}, \tag{10}$$

$$\mathrm{Var}[Y(t)] = n \cdot e^{-\mu t} \cdot (1 - e^{-\mu t}), \tag{11}$$

$$M_3[Y(t)] = n \cdot e^{-\mu t} \cdot (1 - e^{-\mu t}) \cdot (1 - 2e^{-\mu t}), \tag{12}$$

as we must expect from a random variable with a binomial distribution.

For $Z(t)$ we obtain immediately,

$$\lim_{s \to 1} \frac{\partial G_z}{\partial s} = f(t) * e^{-\mu t},$$

$$\lim_{s \to 1} \frac{\partial^2 G_z}{\partial s^2} = \left(f(t) * e^{-\mu t}\right)^2,$$

$$\lim_{s \to 1} \frac{\partial^3 G_z}{\partial s^3} = \left(f(t) * e^{-\mu t}\right)^3,$$

therefore,

$$E[Z(t)] = \mathrm{Var}[Z(t)] = M_3[Z(t)] = f(t) * e^{-\mu t} \tag{13}$$

as we must expect from a random variable with a Poisson distribution.

From (10) and (11) we observe that

$$\frac{\mathrm{Var}[Y(t)]}{\mathrm{E}[Y(t)]} = 1 - e^{-\mu t}$$

and from (13) that

$$\frac{\mathrm{Var}[Z(t)]}{\mathrm{E}[Z(t)]} = 1;$$

as a consequence,

$$\mathrm{Var}[X(t)] < \mathrm{E}[X(t)]. \tag{14}$$

Similarly from (11) and (12) we observe that

$$\frac{\mathrm{M}_3[Y(t)]}{\mathrm{Var}[Y(t)]} = 1 - 2e^{-\mu t}$$

and from (13) that

$$\frac{\mathrm{M}_3[Z(t)]}{\mathrm{Var}[Z(t)]} = 1;$$

as a consequence,

$$\left|\mathrm{M}_3[X(t)]\right| \leq \mathrm{Var}[X(t)].$$

6.4. TURNOVER TIME

In the first section of this chapter I defined a compartment as a pool of particles all having the same probability of transition from their present state to another identifiable state and called $\mu \cdot dt + o(dt)$ the probability that a given particle present in a specified compartment leaves from that compartment in the interval of time t, $t + dt$; for the time being we consider μ a constant, i.e., the said probability depends on the length of time considered but not on the absolute time.

Call

$P(t)$ the probability that a given particle is present in a given compartment at time t,

t_0 the actual time of entrance of a particle into that compartment;

then

$P(t) = 0$ for $t < t_0$,

$P(t_0) = 1$,

$P(t + dt) = P(t) \cdot [1 - \mu\, dt - o(dt)]$;

this last equation can be written

$$dP/dt = -\mu \cdot P(t)$$

and by integration

$$P(t) = 0 \quad \text{for } t < t_0$$
$$= e^{-\mu(t-t_0)} \quad \text{for } t \geq t_0.$$

We now define the continuous random variable T equal to the time spent by a particle in a compartment from its entry to its next exit, and call it *turnover time*; its generating function is

$$G_T(s) = \int_0^\infty p_t s^t dt,$$

where $p_t dt$ is the probability that $t < T < t + dt$.

This last probability is the product of $P(t + t_0)$, the probability of being present at time $t + t_0$, times $\mu\, dt$, the probability of leaving in the immediately following interval of time dt; therefore

$$G_T(s) = \int_0^\infty \mu\, e^{-\mu t} s^t dt$$
$$= \frac{\mu\, e^{-\mu t} s^t}{\ln|s| - \mu}\Bigg|_0^\infty$$

thence,

$$G_T(s) = \frac{1}{1 - 1/\mu \cdot \ln|s|}.$$

This is the generating function of a continuous random variable with an exponential distribution.

Operating as in the previous sections we find that

$$E[T] = \frac{1}{\mu},$$

$$Var[T] = \frac{1}{\mu^2},$$

$$M_3[T] = \frac{2}{\mu^3}.$$

From these last expressions we find that

$$Var[T] = \left(E[T]\right)^2,$$

$$M_3[T] = 2\left(E[T]\right)^3.$$

6.5. RELEVANCE OF STOCHASTIC MODELS

From the proceeding sections it must be clear that the stochastic models we have considered are not providing much more information than the corresponding deterministic models, unless the number of particles present is quite small.

Now, remembering that, for any random variable, the standard deviation is given by

$$\sigma = \sqrt{\mathrm{Var}[X(t)]}$$

and the relative standard deviation by

$$\sigma_R = \frac{\sqrt{\mathrm{Var}[X(t)]}}{\mathrm{E}[X(t)]},$$

when a compartment contains, say, 1 μmole of a substance, corresponding to $6 \cdot 10^{17}$ particles, we have

$$\mathrm{E}[X(t)] = 6 \cdot 10^{17}$$

and using inequality (14)

$$\sigma_R < 1.3 \cdot 10^{-9},$$

a very small value indeed.

Nevertheless there are many pharmacokinetic studies where the primary purpose is the study of certain compartments which happen to have a very small number of particles present even if the given dose is comparatively large. An example [4] is given by the current molecular model of carcinogenesis, where a single molecule may be sufficient to cause a tumor with a "one-hit" in the target organ.

6.6. REFERENCES

1. W. Feller, An Introduction to Probability Theory and Its Applications (Wiley & Sons, New York, 1957), chapter IX.
2. Idem, chapter XI.
3. A. K. Thakur, A. Rescigno and D. E. Shafer, On the Stochastic Theory of Compartments: I. A Single Compartment System, *Bull. Math. Biophysics* **34,** 53-63 (1972).
4. A. Rescigno and J. H. Matis, On the Relevance of Stochastic Compartmental Models to Pharmacokinetic Systems, *Bull. Math. Biology* **43,** 245-7 (1981).

This page intentionally left blank

7. DIRECTED GRAPHS

7.1. INTRODUCTION

One important problem in the analysis of compartment systems is the study of the properties related to the structure of the model, i.e., of the properties depending upon the presence or absence of a connection between any two compartments and not upon the values of such connections.

The topological properties of a system of compartments have been studied by Rescigno and Segre [1] using a *directed graph* [2]. A directed graph was called *réseau orienté* by Sainte-Laguë [3] and *graphe* by Berge [4]; I shall use the term *graph* for brevity in this chapter when referring to a directed graph. A graph consists of a set of *nodes*, representing the compartments, together with a set of *arms* connecting the nodes and representing the transfer between compartments.

7.2. CONNECTIVITY MATRIX

To each graph containing n nodes we can associate a square matrix of order n, called the *connectivity matrix*; the element a_{ij} of row i and column j of the connectivity matrix is equal to 1 if there is an arm from node i to node j, is equal to 0 if not. The sum of two connectivity matrices $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{ij})$ of the same order is the matrix

$$\boldsymbol{A} + \boldsymbol{B} = (a_{ij} + b_{ij}),$$

where the elements are added according to the rules of Boolean algebra (see Appendix D); the product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is

$$\boldsymbol{A} \cdot \boldsymbol{B} = \left(\sum_{l=1}^{n} a_{il} b_{lj} \right),$$

where again addition and multiplication of elements follow the rules of Boolean algebra; the power $\boldsymbol{A}^r$ of a connectivity matrix is defined by

$$\boldsymbol{A}^r = \boldsymbol{A} \cdot \boldsymbol{A}^{r-1}; \quad r = 2, 3, \ldots$$

finally the transpose $\boldsymbol{A}^{\mathrm{T}}$ of $\boldsymbol{A}$ is defined by

$$\boldsymbol{A}^{\mathrm{T}} = (a_{ji}).$$

In a connectivity matrix a column of zeros means that the corresponding node is an *initial node,* a row of zeros means that the corresponding node is a *terminal node.* For

convenience we shall consider only graphs with only one initial node; this kind of graph corresponds to systems of compartments where the drug is introduced only at one point; of course the linearity of the system implies that if the drug enters through several compartments, that system can be considered to be the sum of several systems, each with one initial node.

We shall call the initial node, *node* 0. Node 0 does not correspond to a real compartment of the system, but rather to the ideal point from where the drug enters the system [5].

If an element a_{ij} of a matrix $\boldsymbol{A}$ is equal to 1, we say that i is the *precursor* of j, and j is the *successor* of i.

A succession of arms such that the node entered by each of them (except the last one) is the node at which the next arm begins, is called a *path.* If the starting node of the first arm coincides with the ending node of the last arm of a path, that path is called a *cycle.* The length of a path is equal to the number of its arms. A path, including a cycle, is called *simple* if every arm of it appears only once; it is called *elementary* if every node of it is entered only once.

A graph is called *connected* if there is at least a path from its initial node to any other node; in this section we consider only connected graphs. A graph is called *strongly connected,* or a *strong graph,* if there is at least a path from every node, including node 0, to every other node, excluding node 0. In a strong graph there is at least one cycle.

Deleting row i and column i from matrix $\boldsymbol{A}$ results in the minor $\boldsymbol{A}_{i,i}$ of $\boldsymbol{A}$ and corresponds to disconnecting node i from the graph. The minimum number of nodes that must be deleted to transform a connected graph into a non-connected one, is the *connectivity* of the graph. A graph of connectivity p is also called *p-connected.*

A *subgraph* is a connected graph obtained by suppressing some nodes and their connecting arms from a given graph; the subgraph obtained by suppressing the initial node and the arms leaving it, from a given graph, is called its G_0 *subgraph.* A subgraph in which each node occurs in exactly one cycle is called a *linear subgraph.*

Each set of cycles in which each node of the subgraph occurs in exactly one cycle is called a *strong component.* A *Hamiltonian cycle* of a graph or subgraph is an elementary cycle that joins all the nodes of that graph or subgraph. For instance the graph of Fig. 1 has one Hamiltonian cycle (1→2→3→4→5→6→7→8→1), two strong components (1→2→3→4→5→6→7→8→1; 1→2→7→8→1 and 3→4→5→6→3), three elementary cycles (1→2→3→4→5→6→7→8→1; 1→2→7→8→1; 3→4→5→6→3).

The graph of Fig. 2 too is strongly connected, but it is not the linear subgraph of any graph; it does not have Hamiltonian cycles or strong components, but it has two elementary cycles (1→2→5→6→1, 2→5→4→3→2).

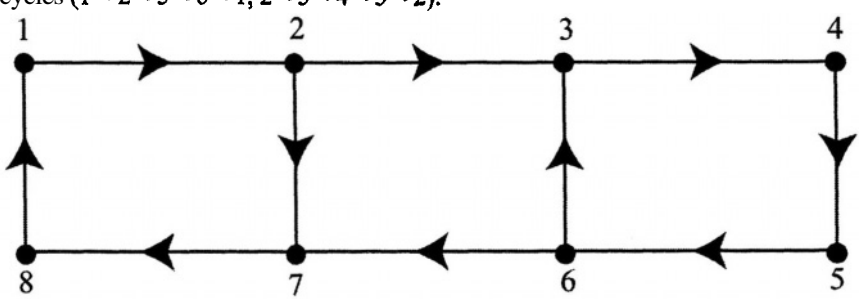

Fig. 1. A connected graph

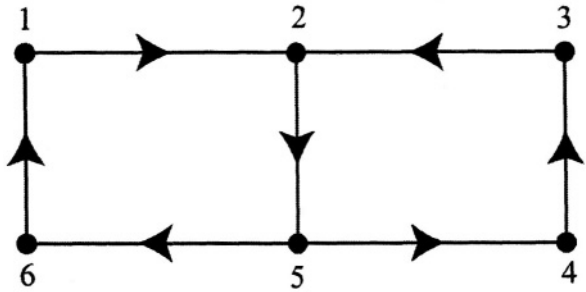

Fig. 2. Another connected graph

A graph is *symmetric* if for any arm connecting a non-initial node to another node, there is an arm in the opposite direction; the connectivity matrix of the G_0 subgraph of a symmetric graph is symmetric, i.e., $A_0 = A_0^T$.

A graph is *asymmetric* if there is no more than one arm between any two nodes; if an asymmetric graph is strongly connected, it admits one Hamiltonian cycle. Define the element-by-element product $A \times B$ of two matrices $A = (a_{ij})$ and $B = (b_{ij})$ by

$$A \times B = (a_{ij} \cdot b_{ij});$$

then for any asymmetric graph A,

$$A \times A^T = 0;$$

if in a graph some of the connections between nodes are symmetric, the set of such connections is given by the non-zero elements of the element-by-element product $A \times A^T$.

Fig. 3. A lineal or catenary graph

A *lineal* or *catenary* graph is a graph with all nodes entered by no more than one arm, as in Fig. 3; it has one initial node, one terminal node, and one path; its connectivity matrix has no more that one non-zero element in each column and in each row.

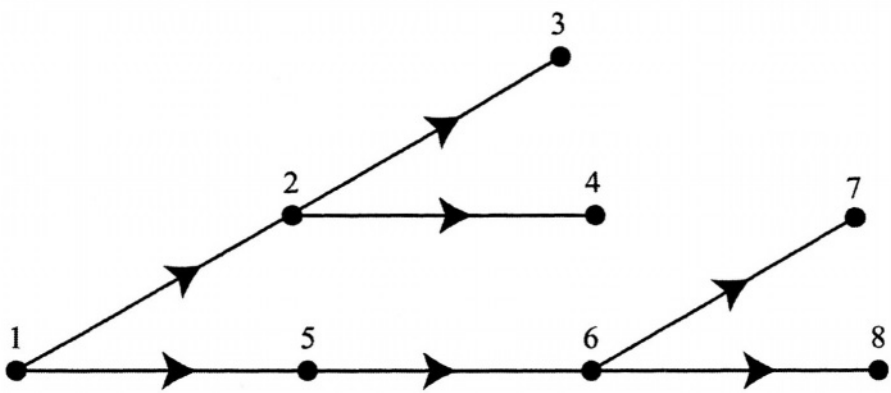

Fig. 4. A tree

A *tree* is a graph in which all nodes, except the initial one, are entered by exactly one arm, and from at least one node more than one arm starts; these last nodes are called *roots*. The connectivity matrix of a tree has no more than one non-zero element in each column and at least one row with more than one; the rows with more than one non-zero element correspond to the roots of the tree. For instance the roots of the graph of Fig 4 are 1, 2, 6; the connectivity matrix is

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

A *mammillary graph* [6] has a *central* node connected with all other nodes, in one or in both directions, while all other nodes are not connected among them. Its connectivity matrix has all elements not on the row or column corresponding to the central compartments equal to zero. For instance the connectivity matrix of the graph of Fig. 5 is

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

The successive powers of the connectivity matrix show the existence of paths in the corresponding graph; in fact the element of row i and column j of matrix A^r is equal to 1 if there is a path of length r from i to j; of course the diagonal elements of A^r show the existence of cycles of length r.

If a graph with n nodes does not contain any cycle, then there is a number $r < n$ such that $A^r = 0$, and matrix A is said to be *nilpotent*. If $A^r = 0$ but $A^{r-1} \neq 0$, then $r - 1$ is the length of the longest path of the graph. Marimont [7] has proved that a matrix is nilpotent

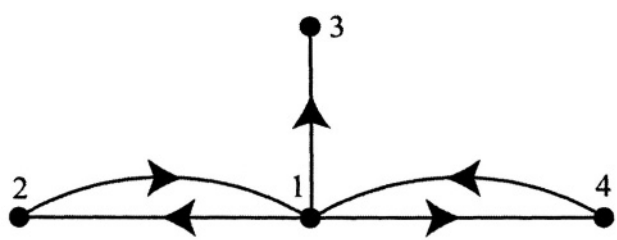

Fig. 5. A mammillary graph

if and only if every principal submatrix has at least one zero row or zero column; as a practical rule for finding whether a matrix is nilpotent, i.e., whether the corresponding graph has no cycles, we can delete successively the rows (or columns) whose elements are all zero and the corresponding columns (or rows); if some non-zero elements are left, the original matrix is not nilpotent.

The sum of successive powers

$$\boldsymbol{R}_k = \boldsymbol{A} + \boldsymbol{A}^2 + \ldots + \boldsymbol{A}^k$$

shows the paths of length up to k. If $\boldsymbol{A}$ is nilpotent and $\boldsymbol{A}^{k+1} = \boldsymbol{0}$, then $\boldsymbol{R}_k$ shows the paths of any length. If $\boldsymbol{A}$ is not nilpotent, and k is the length of the longest simple path in the graph, then adding higher powers of $\boldsymbol{A}$ does not change $\boldsymbol{R}_k$ because it includes all possible connections between nodes. We call *reachability matrix*, $\boldsymbol{R}$, the limit of the above sum for k sufficiently large; then a non-zero element r_{ij} of $\boldsymbol{R}$ shows that there exists a path from node i to node j, i.e., that compartment j can be reached from compartment i. Harary [8] has shown that the element-by-element product of $\boldsymbol{R}$ and its transpose, $\boldsymbol{R}^{\mathrm{T}}$, indicates in row i the nodes belonging to the same cycle as node i.

For instance, from the graph of Fig. 6,

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \boldsymbol{A}^2 = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \boldsymbol{A}^3 = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix},$$

$$\boldsymbol{A}^4 = \boldsymbol{A}^2, \boldsymbol{A}^5 = \boldsymbol{A}^3, \ldots;$$

therefore

$$\boldsymbol{R} = \boldsymbol{A} + \boldsymbol{A}^2 + \boldsymbol{A}^3 = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}, \boldsymbol{R} \times \boldsymbol{R}^T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix};$$

this last matrix shows that nodes 2 and 4 are on one cycle and nodes 3 and 5 on another.

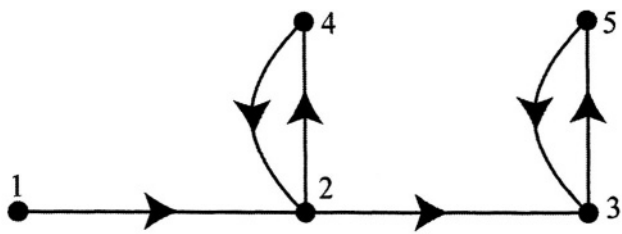

Fig. 6. A graph with two cycles

7.3. PRECURSOR-SUCCESSOR RELATIONSHIP

Another important application of graphs to compartmental analysis is the classification of the precursor-successor relationship [9]. If in a graph there is a path from i to j, then i is a *precursor* of j and j is a *successor* of i; the length of the shortest path from i to j is called the *order* of the precursor. For instance in the graph of Fig. 6, node 1 is the precursor of order one of node 2, of order two of nodes 3 and 4, and of order three of node 5. Precursors of order one can be further classified in different types:

a) *Absolute precursor*: the arm from i to j is the only one leaving i and the only one entering j;

b) *Complete precursor*: there is only one arm leaving i; no cycle may enter j except from i;

c) *Complete precursor with recycling*: there is only one arm leaving i; j belongs to a cycle not including i;

d) *Unique precursor*: there is only one arm entering j;

e) *Total precursor*: there are no paths from i to j of length more than one; if j belongs to a cycle, i belongs to the same cycle;

f) *Total precursor with recycling*: there are no paths from i to j of length more than one; there is a cycle in j not through i;

g) *Partial precursor*: there is a paths from i to j of length more than one; if j belongs to a cycle, no node of the cycle except i has an arm entering j;

h) *Partial precursor with recycling*: there is a paths from i to j of length more than one; there is a cycle in j not through i.

Observe that a precursor i of j is classified "with recycling" only when there is a cycle in j not passing through i. Therefore, in the examples of Fig. 7, i is an absolute precursor of j. The reason for this is that, in the case of Fig. 7, even though j is a precursor of i (of order 2), a knowledge of i is sufficient to explain the behavior of j, while with recycling j is determined by i but also by j itself, and i is not sufficient to explain the behavior of j.

Fig. 8 shows the graphs of the different types of precursors of first order, while Table I shows how they can be classified according to the values of the matrices $\boldsymbol{A}$ and $\boldsymbol{R}$.

If $a_{ij} = 0$ but $a_{ih}a_{hj} = 1$ for some h, then i is a precursor of order two of j; classification of second order precursors according to different types is done as above. More details on the precursor-successor relationship will be given in Chapter 10.

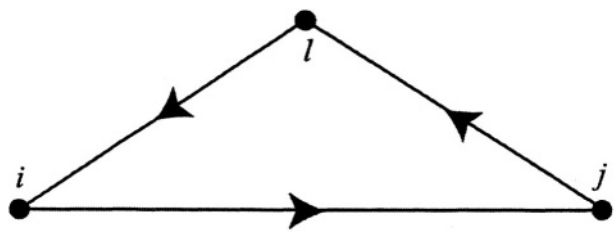

Fig. 7. A graph in which i is an absolute precursor of j of order one

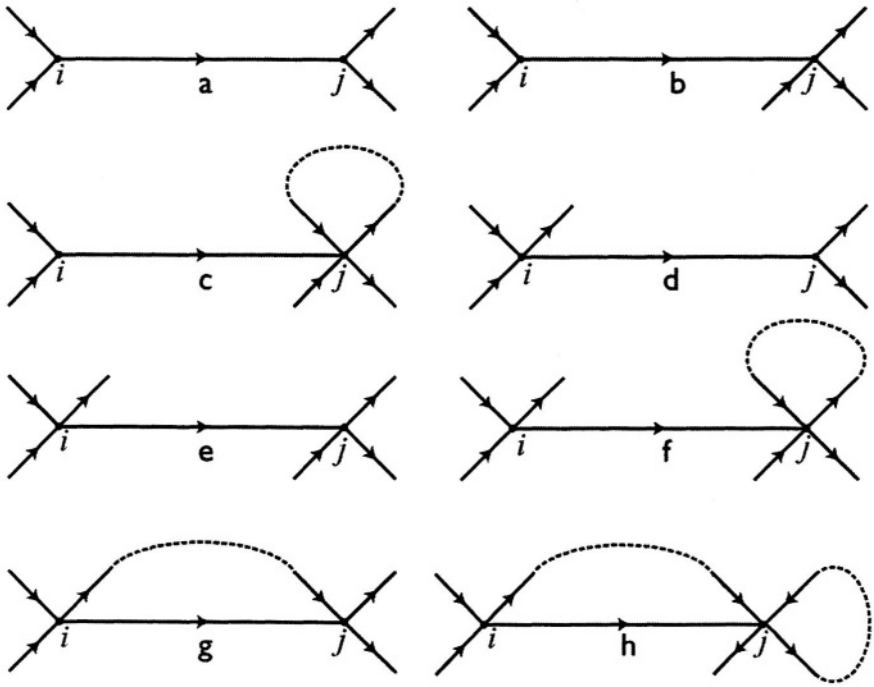

Fig. 8. Different types of precursors of order one

Table I. Precursors of order one

$a_{ij} = 1$ in all cases		$a_{kj} = 0$ for any $k \neq i$	$a_{kj} = 1$ for some $k \neq i$	
			$r_{jk}a_{kj} = 0$ for any $k \neq i$	$r_{jk}a_{kj} = 1$ for some $k \neq i$
$a_{ih} = 0$ for any $h \neq j$		a	b	c
$a_{ih} = 1$ for some $h \neq j$	$a_{ih}r_{hj} = 0$ for any h	d	e	f
	$a_{ih}r_{hj} = 1$ for some h	–	g	h

7.4. REFERENCES

1. A. Rescigno and G. Segre, On Some Topological Properties of the Systems of Compartments, *Bull. Math. Biophysics* **26**, 31-8 (1964).
2. O. Ore, *Theory of Graphs* (American Mathematical Society, Providence, 1962).
3. A. Sainte-Laguë, *Les Reseaux ou Graphes* (Gauthier-Villars, Paris, 1926).
4. C. Berge, *Theorie des Graphes et Ses Applications* (Dunod, Paris, 1958); *The Theory of Graphs and Its Applications* (Methuen & Co., London, 1958).
5. A. Rescigno, Synthesis of a Multicompartmental Biological Model, *Biochim. Biophys. Acta* **37**, 463-8 (1960).

6. C. M. E. Matthews, The Theory of Tracer Experiments with ^{131}I-labeled Plasma Proteins, *Phys. Med. Biol.* **2**, 36-53 (1957).
7. R. B. Marimont, A New Method of Checking the Consistency of Precedence Matrices, *J. Assoc. Computing Machinery* **6**, 164-72 (1959).
8. F. Harary, A Graph Theoretic Method for the Complete Reduction of a Matrix with a View toward Finding its Eigenvalues, *J. Mathem. and Physics* **38**, 104-11 (1959).
9. A. Rescigno and G. Segre, The Precursor-Product Relationship, *J. theoret. Biol.* **1**, 498-513 (1961).

8. LINEAR GRAPHS

8.1. FUNDAMENTAL PROPERTY OF LINEAR GRAPHS

Equations (14) of chapter 5 can be written in a concise form,

$$\{x_i\} = \sum_{j\neq i}^{1\cdots n} \frac{k_{ji}}{s+K_i}\{x_j\} + \frac{1}{s+K_i}x_i(0),\ i = 1, 2, \ldots, n \tag{1}$$

showing how the operator $\{x_i\}$ of compartment i depends on the operators $x_i(0)$ of its initial condition and on the operators $\{x_j\}$ of all other compartments. Equation (1) can be represented graphically with a node for $x_i(0)$, a node for each function $\{x_j\}$ and a node for $\{x_i\}$, plus an arm from each of the former nodes to this last node, these arms equal to the coefficients of the respective terms on the right-hand side of equation (1). Thus node $\{x_i\}$ is equal to the sum of all arms entering it, times their nodes of departure. For instance to equation

$$\{x_1\} = \frac{1}{s+K_1}x_1(0) + \frac{k_{21}}{s+K_1}\{x_2\} + \frac{k_{31}}{s+K_1}\{x_3\} \tag{2}$$

corresponds the graph of Fig. 1 (next page).

Of course an equation like (1) can be written for each compartment of a system, and to each equation corresponds a graph; all those graphs can be combined together, because the arms entering a node of one graph will not change the values of the nodes determined by another equation. If equation (2) is holding with the two additional equations

$$\{x_2\} = \frac{k_{12}}{s+K_2}\{x_1\} + \frac{k_{32}}{s+K_2}\{x_3\} \tag{3}$$

and

$$\{x_3\} = \frac{k_{13}}{s+K_3}\{x_1\} \tag{4}$$

their two corresponding graphs are shown in Fig. 2 and Fig. 3. All those graphs can be combined in a single graph, as shown in Fig. 4.

This last graph includes all the information contained in the given differential equations (2), (3), (4), plus the initial conditions. This kind of graph was introduced in 1953 by Mason [1] who called it *signal-flow graph*; it has been used in compartmental analysis since 1960 [2]. I prefer the name *linear graph* as more indicative of its function, even though this term was used by Kirchhoff [3] in a slightly different context.

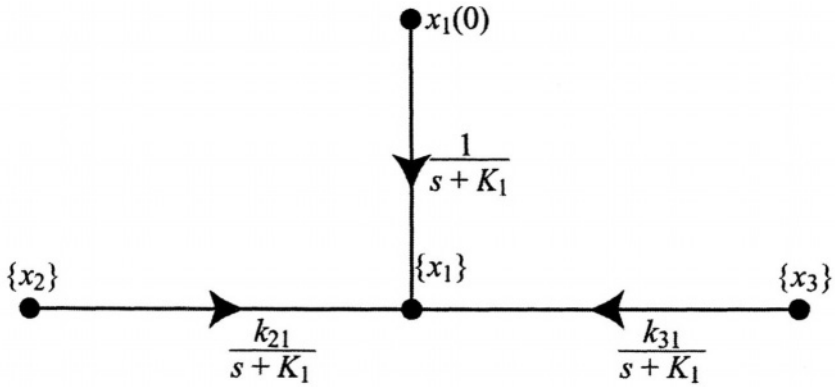

Fig. 1. The graph of equation (2)

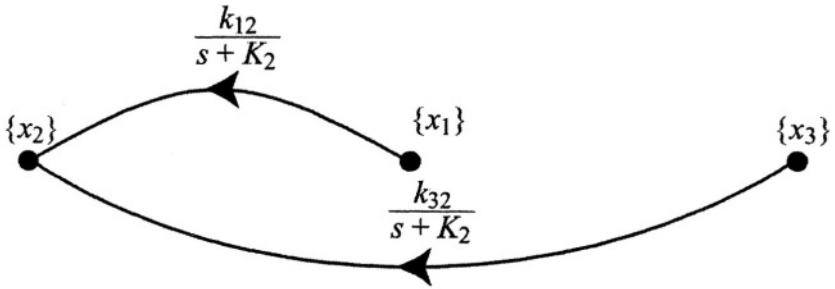

Fig. 2. The graph of equation (3)

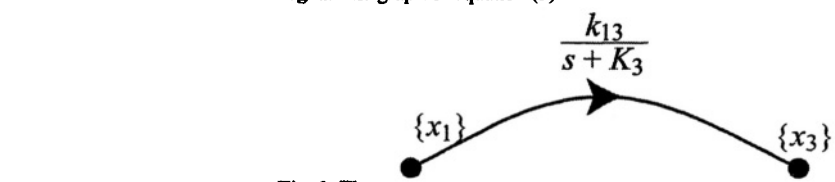

Fig. 3. The graph of equation (4)

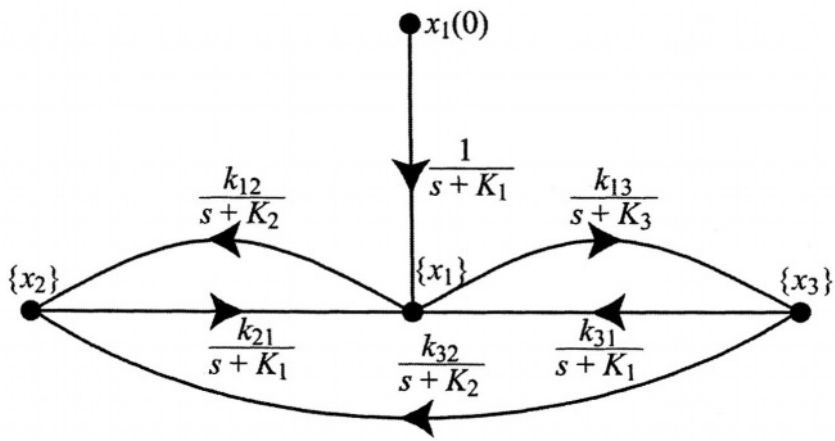

Fig. 4. The graph of equations (2), (3), (4) combined

Almost all definitions given for directed graphs are valid for linear graphs too. In particular a path has a length, as in a directed graph, but also a *value,* equal to the product of its arms.

The fundamental property of a linear graph, as an immediate consequence of its definition, is: "Each node is equal to the sum of the products of the arms entering it times their departing nodes."

8.2. TRANSFORMATIONS OF LINEAR GRAPHS

A number of transformations can be made, including the suppression of some nodes, without changing the fundamental properties of the nodes left. These four transformations were described by Mason:

a) Two tandem arms can be substituted by a single arm equal to their product and the intermediate node suppressed.

b) Two parallel arms can be substituted by a single arm equal to their sum.

c) An arm entering a node can be substituted by arms entering all nodes immediately following it, each new arm being equal to the product of the original arm and the arm connecting the previous to the new node.

d) An arm starting and ending at the same node can be suppressed by dividing all arms entering that node by one minus the value of the suppressed arm.

Repeated application of these four rules leads to a much simpler linear graph that helps interpreting some of the properties of the system of compartments it represents. More details can be found in the literature [4, 5, 6]; here I intend to show only some simple properties of the linear graphs.

Going back to equations (1), observe that the term $x_i(0)$ represents the contribution to compartment i from outside the system of compartments; we can call this term the *input* to compartment i. For simplicity consider the case when only one compartment has an input different from zero; then its linear graph has only one initial node, as defined in chapter 7; without any loss of generality we can suppose this node to be node 1.

If the graph does not contain any cycles, there are only a finite number of paths between its initial node and any other node; repeated application of the first three Mason's rules lead to a graph containing exactly one arm between the source and each other node. For instance the graph of Fig. 5 has one path from $x_1(0)$ to $\{x_1\}$ equal to $\frac{1}{s+K_1}$, one path from $x_1(0)$ to $\{x_2\}$ equal to $\frac{k_{12}}{(s+K_1)(s+K_2)}$, two paths from $x_1(0)$ to $\{x_4\}$ respectively equal to $\frac{k_{12}k_{24}}{(s+K_1)(s+K_2)(s+K_4)}$ and to $\frac{k_{13}k_{34}}{(s+K_1)(s+K_3)(s+K_4)}$, and so forth.

The graph of Fig. 6 is therefore equivalent to the graph of Fig. 5. The values of its arms are given in Table I. In this new graph there is one initial node, the source; all other nodes are terminal nodes, therefore any of them can be suppressed without altering the properties of the rest of the graph. This is important, because when the behavior of only one compartment is of interest, the graph can be reduced to only one arm, between the source and that particular node.

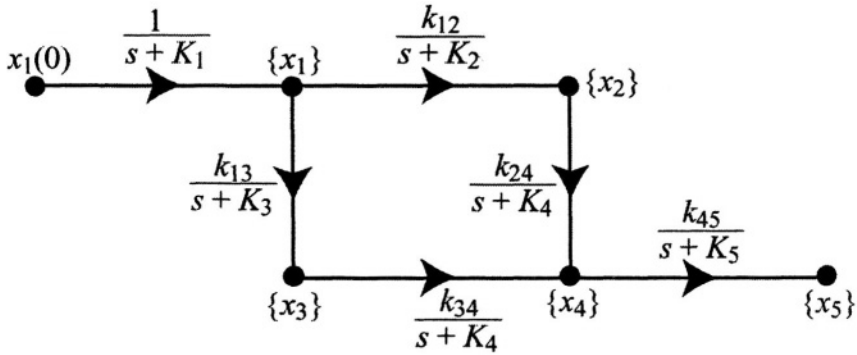

Fig. 5. Example of a linear graph

The problem is slightly more complicated when the original graph contains some cycles, because in this case the number of paths between some nodes is infinite. A first step in the simplification of such graph is to look for *essential nodes*, i.e., nodes that must be removed to interrupt all cycles; their choice is not unique, but in any case it must be such that the number of essential nodes be minimum. Once the essential nodes are chosen, the simplified graph contains:

Table I. Values of the arms of the graph of Fig. 6

From $x_1(0)$ to $\{x_1\}$	$\frac{1}{s+K_1}$
From $x_1(0)$ to $\{x_2\}$	$\frac{k_{12}}{(s+K_1)(s+K_2)}$
From $x_1(0)$ to $\{x_3\}$	$\frac{k_{13}}{(s+K_1)(s+K_3)}$
From $x_1(0)$ to $\{x_4\}$	$\frac{k_{12}k_{24}}{(s+K_1)(s+K_2)(s+K_4)} + \frac{k_{13}k_{34}}{(s+K_1)(s+K_3)(s+K_4)}$
From $x_1(0)$ to $\{x_5\}$	$\frac{k_{12}k_{24}k_{45}}{(s+K_1)(s+K_2)(s+K_4)(s+K_5)} + \frac{k_{13}k_{34}k_{45}}{(s+K_1)(s+K_3)(s+K_4)(s+K_5)}$

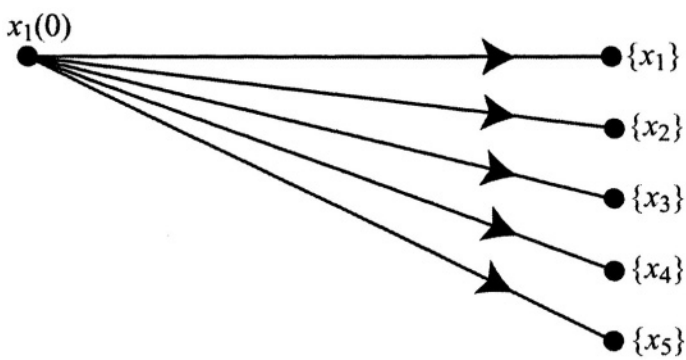

Fig. 6. The simplified graph of Fig. 5

a) an arm from the source to the terminal node of interest,

b) an arm from the source to each essential node,

c) an arm from each essential node to the terminal node,

d) an arm from each essential node to each other essential node,

e) an arm from each essential node to itself.

The value of each of these arms is equal to the sum of the values of the elementary paths between the nodes they connect, excluding all other nodes of the simplified graph. Some of the arms listed above may be missing, and the terminal node itself may be an essential node. Fig. 7 shows a simplified graph with one essential node; Fig. 8 shows a simplified graph where the terminal node is an essential node; Fig. 9 and Fig. 10 show simplified graphs with two essential nodes.

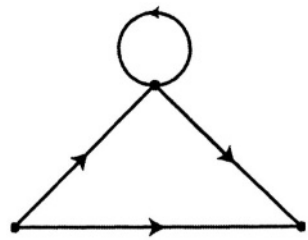

Fig. 7. A simplified graph with one essential node

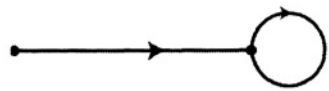

Fig. 8. A simplified graph where the terminal node is an essential node

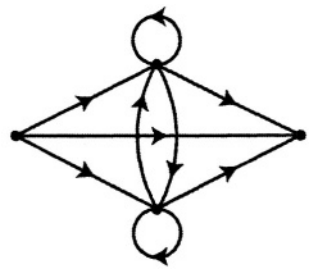

Fig. 9. A simplified graph with two essential nodes, general case

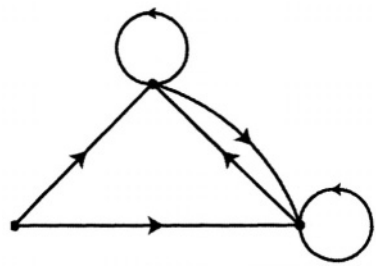

Fig. 10. A simplified graph with two essential nodes, one of them being the terminal node

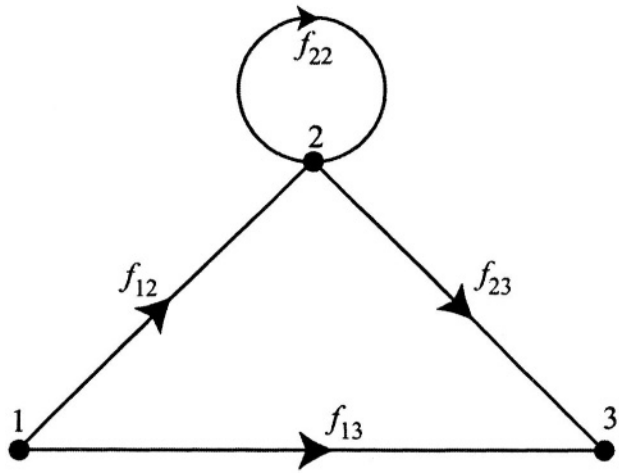

Fig. 11. One essential node

From the simplified graph the closed arms can be eliminated using Mason's forth rule, then the new graph can be further simplified as before, until only one initial and one terminal node are left. The value of the only arm left is called the *transfer function* between those two nodes. The properties of the transfer function will be studied with more details in chapter 10.

For instance the essential graph of Fig. 11 using Mason's fourth rule becomes the graph of Fig. 12 not containing any cycles; then using Mason's first and second rules becomes the graph of Fig. 13 with only one arm.

The rules shown here are conveniently applied when a graph contains very few cycles; if the simplification of the graph involves more than one or two essential nodes, it is more convenient to use the method of the strong components [7] (See Section 8.5).

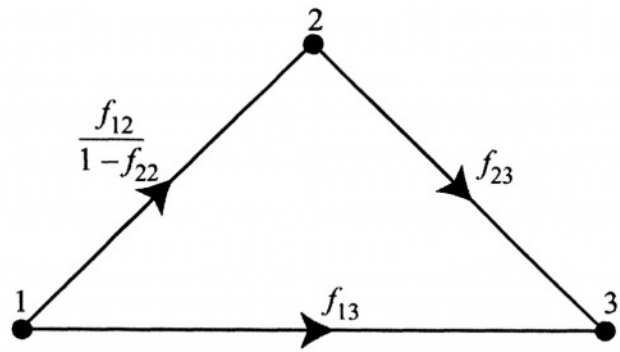

Fig. 12. The closed arm removed

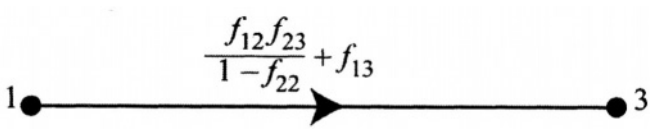

Fig. 13. One arm only

8.3. VARIABLE ADJACENCY MATRIX

The value of an arm from node i to node j in the complete graph of a compartmental system is $k_{ij}/(s + K_j)$; in the simplified graph the arm is an expression formed with the operator s and the transfer rates; let's call in general f_{ij} the value of the arm from node i to node j, at any stage of simplification of the graph. We define now the *variable adjacency matrix* A a square matrix whose elements are the operators f_{ij} of a linear graph, or of the system of compartments it represents. Matrix A^2 is formed by the elements $f_{ij}^{(2)} = \sum_k f_{ik} f_{kj}$ and it represents the values of the paths of length 2 of the associated graph. In general, we can say that matrix A^r represents the values of the path of length r.

8.4. DETERMINANT OF THE VARIABLE ADJACENCY MATRIX

The determinant D (see Appendix C) of the variable adjacency matrix of order n of a G_0 subgraph (i.e., of the subgraph obtained from G by deleting node 0 and the arms leaving it) is called *zero-axial* or *invertebrate* because all elements of its principal diagonal are null. Its development was obtained by Caley [8]. He used the notation

$$|ij| = f_{ij} f_{ji}$$

$$|ijk| = f_{ij} f_{jk} f_{ki}$$

and omitted the second bar when two such symbols were written consecutively in a product, for instance

$$|12|345| = |12| \cdot |345|.$$

We now observe that the symbol $|ijk...|$ represents the product of the values of the arms forming an elementary cycle through the nodes $i, j, k, \ldots$, and that the symbol, say, $|12|3 \ldots |n-1, n|$ represents the product of the values of the arms forming a strong component of G_0. According to Cayley, D is equal to the sum of the products formed with the values of the arms of the strong components of G_0. Each product has the sign + or –, depending on whether the number of cycles with an even number of arms is even or odd. When there are no strong components, D is equal to 0. For instance, with $n = 4$,

$$D = + \sum |12|34| - \sum |1234|,$$

where the two sums are extended to the permutations of the numbers 1, 2, 3, 4, giving different values of the products [9]. As $|12| = |21|$, $|34| = |43|$, with our notation this is equivalent to

$$D = f_{12} f_{21} f_{34} f_{43} + f_{13} f_{31} f_{24} f_{42} + f_{14} f_{41} f_{23} f_{32}$$

$$- f_{12} f_{23} f_{34} f_{41} - f_{12} f_{24} f_{43} f_{31} - f_{13} f_{32} f_{24} f_{41}$$

$$- f_{13} f_{34} f_{42} f_{21} - f_{14} f_{42} f_{23} f_{31} - f_{14} f_{43} f_{32} f_{21}.$$

For $n = 5$

$$D = + \sum |12345| - \sum |123|45|,$$

the first sum being formed by 24 terms and the second by 20 terms [8], For $n = 6$, Muir [10] gave a formula which evidently contains a printing error and should read

$$D = -\sum|12|34|56| + \sum|12|3456| + \sum|123|456| - \sum|123456|;$$

the number of terms in each sum is 15, 90, 40, 120, respectively. The convenience of the Cayley notation is very apparent.

Calling $\bar{D}$ the determinant formed by the elements $-f_{ij}$, the development of $\bar{D}$ has the same terms as the development of D; the terms corresponding to cycles with an even number of arms have the same sign in D as in $\bar{D}$; the terms corresponding to cycles with an odd number of arms have the sign changed. Therefore the sign of the terms of $\bar{D}$ is + or – according to whether there is an even or odd number of cycles in the corresponding strong component.

For instance, with $n = 4$,

$$\bar{D} = +\sum|12|34| - \sum|1234|;$$

with $n = 5$,

$$\bar{D} = -\sum|12345| + \sum|123|45|.$$

8.5. ENUMERATION OF THE STRONG COMPONENTS

The number $\psi(n)$ of terms of an invertebrate determinant of order n was first calculated by Stockwell [11]; he expressed it with the recurrent formula

$$\psi(n) = n \cdot \psi(n-1) + (-1)^n.$$

Balzer [12] showed that

$$\psi(n) = n!\left(\frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right),$$

and later [13] he demonstrated that $\psi(n)$ is the nearest integer to $n!e^{-1}$. Because of what was said previously about D, $\psi(n)$ is the number of strong components in a *complete* graph of order n, i.e., a graph with n nodes and $n(n-1)$ arms. From Stockwell's formula one obtains

$$\psi(n) = (n-1)\left(\psi(n-1) + \psi(n-2)\right),$$

that shows the law of formation of the strong components of a complete graph of order n from the strong components of order $n-1$ and $n-2$.

Given the strong components of a graph of order $n-1$, the strong components of a graph of order n obtained by adding a new node and the corresponding arms are found by

1. Including the new node in the cycles of the $\psi(n-1)$ old strong components after each one of the old nodes, i.e., in $n-1$ different ways, and
2. Connecting the new node in a cycle with each one of the $n-1$ old nodes and coupling each one of these new cycles with the $\psi(n-2)$ strong components containing the other nodes.

The same reasoning is true for a non-complete graph. For instance, given the graph of Fig. 14, the determinant of which is

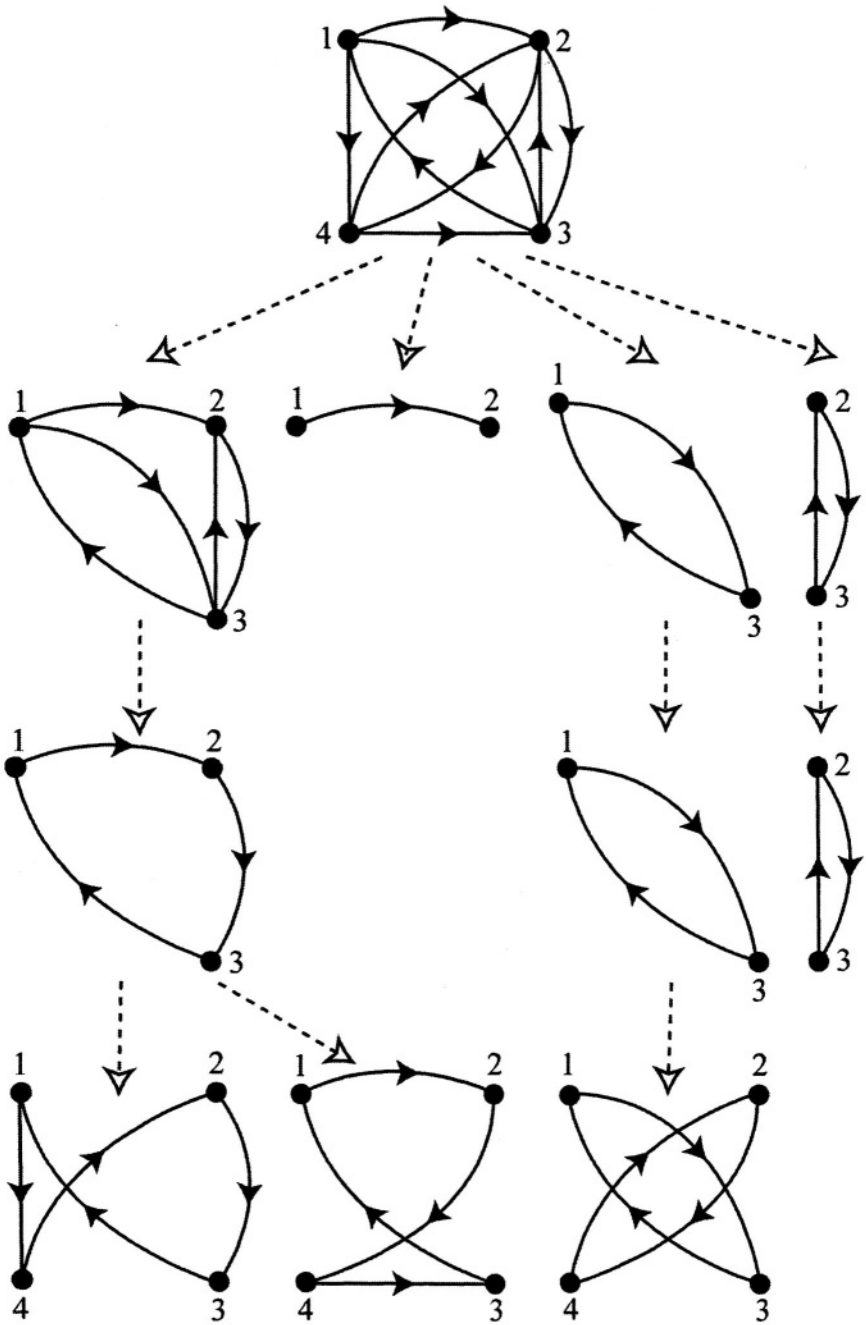

Fig. 14. Construction of the strong components of a graph of order 4.

$$\begin{vmatrix} 0 & f_{12} & f_{13} & f_{14} \\ 0 & 0 & f_{23} & f_{24} \\ f_{31} & f_{32} & 0 & 0 \\ 0 & f_{42} & f_{43} & 0 \end{vmatrix},$$

we first construct the subgraph suppressing node 4, then the three subgraphs suppressing each one of the nodes 3,2, 1. Of each one of them we construct the strong components, if any. In the cycles of the strong components of order 3 we insert node 4 after node 1, after node 2, after node 3, if possible, i.e., if there are the corresponding arms; to the strong components of order 2 we add a cycle formed by node 4 and by the node missing, if possible. As shown in Fig. 14, there are three strong components symbolized by |1423|, |1243|, |13|24|.

8.6. DETERMINANT OF A SYSTEM

The solution of a compartment system corresponds to the solution of a system of first order differential equations with constant coefficients. And any system of first order linear differential equations with constant coefficients can be represented by a linear graph and solved by the method described here.

The system of differential equations is

$$\frac{dx_1}{dt} = -K_1 x_1(t) + k_{21} x_2(t) + \ldots + k_{n1} x_n(t)$$
$$\frac{dx_2}{dt} = +k_{12} x_1(t) - K_2 x_2(t) + \ldots + k_{n2} x_n(t)$$
$$\ldots\ldots$$
$$\frac{dx_n}{dt} = +k_{1n} x_1(t) + k_{2n} x_2(t) + \ldots + K_n x_n(t)$$

with the initial conditions $x_i(0) = k_{oi} x_0$, where $k_{oi} = 1$ or 0, according to whether compartment i is connected or not to the initial compartment. In operational form the system is written

$$+(s + K_1)\{x_1\} - k_{21}\{x_2\} - \ldots - k_{n1}\{x_n\} = k_{01} x_0$$
$$-k_{12}\{x_1\} + (s + K_2)\{x_2\} - \ldots - k_{n2}\{x_n\} = k_{02} x_0$$
$$\ldots\ldots$$
$$-k_{1n}\{x_1\} - k_{2n}\{x_2\} - \ldots + (s + K_n)\{x_n\} = k_{0n} x_0$$

or

$$-f_{01} x_0 + \{x_1\} - f_{21}\{x_2\} - \ldots - f_{n1}\{x_n\} = 0$$
$$-f_{02} x_0 - f_{12}\{x_1\} + \{x_2\} - \ldots - f_{n2}\{x_n\} = 0$$
$$\ldots\ldots$$
$$-f_{0n} x_0 - f_{1n}\{x_1\} - f_{2n}\{x_2\} - \ldots + \{x_n\} = 0$$

and in matrix form

$$\left(I_{n+1} - A^T\right) \cdot \begin{pmatrix} x_0 \\ \{x_i\} \\ \vdots \\ \{x_n\} \end{pmatrix} = \begin{pmatrix} x_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where I_{n+1} is the $(n+1)\cdot(n+1)$ identity matrix, and A^T is the transpose of A.

The solution is obtained by the Cramer's rule,

$$\frac{\{x_i\}}{x_0} = (-1)^i \frac{\det(I_{n-1} - A^T)_{0;i}}{\det(I_{n-1} - A^T)}$$

where in the last fraction the numerator is the minor of the denominator obtained by suppressing its row 0 and its column i.

The row 0 of $\det(I_{n+1} - A^T)$ is 1,0,0,...; furthermore, the value of a determinant does not change if its columns are changed with its rows. Therefore

$$\det(I_{n+1} - A^T) = \det I_n + \overline{D},$$

where the expression on the right hand side is the determinant obtained by changing the sign of all the elements of D and making equal to 1 all elements of its principal diagonal. If we move the first column of $\det(I_{n+1} - A^T)_{0;i}$ after the i^{th} one, and then change the columns with the rows, we obtain the determinant $(-1)^{i-1}\left(\det I_{n-1} + \overline{D}_i\right)$, where $\det I_{n-1} + \overline{D}_i$ is the determinant $\det I_n + \overline{D}$ with its i^{th} row substituted by the elements $-f_{01}, -f_{02}, \ldots, -f_{0n}$. Therefore

$$\frac{\{x_i\}}{x_0} = -\frac{\det I_{n-1} - \overline{D}_i}{\det I_n - \overline{D}}.$$

8.7. DEVELOPMENT OF THE DETERMINANT OF A SYSTEM

If the rows and the columns of $\det I_n + D$ are rearranged such that all nodes from where no paths go to node i are represented after i, and all nodes from where a path goes to node i are represented before i, then all the elements belonging jointly to the rows after the i^{th} and to the columns to the i^{th} are zero. Therefore $\det I_n + \overline{D}$ is equal to its principal minor M corresponding to the node i and to the nodes from where a path goes to node i, times the principal minor N corresponding to the nodes from where no path goes to node i. If from every node a path goes to node i, then $\det I_n + \overline{D} = M$. In the same way it can be shown that

$$\det I_n + \overline{D}_i = M_i \cdot N,$$

where M_i is the determinant M with the i^{th} row substituted by the elements $-f_{01}, -f_{02}, \ldots, -f_{0i}$. Therefore

$$\frac{\{x_i\}}{x_0} = \frac{-M_i}{M}.$$

The determinant

$$M = \begin{vmatrix} 1 & -f_{12} & \cdots & -f_{1n} \\ -f_{21} & 1 & \cdots & -f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -f_{n1} & -f_{n2} & \cdots & 1 \end{vmatrix}$$

can be developed in terms of $\bar{D}$ and its principal minors, following a theorem enunciated by Cayley [14] for skew determinants, but evidently valid for any determinant: M is equal to 1, plus the determinant $\bar{D}$ of the subgraph obtained from the original G_0 suppressing the node from where no paths go to node i, plus all principal minors of this last determinant.

According to Cayley's rule, each of these determinants is equal to the sum of the products of the arms forming a linear subgraph (see Chapter 7, section 7.2) of G, the sum being extended to all possible linear subgraphs, including G_0, if the case, and each product has the sign + or –, according to whether in the corresponding subgraph there is an even or odd number of cycles.

The expansion of

$$-M_i = \begin{vmatrix} 1 & -f_{12} & \cdots & -f_{1i} \\ -f_{21} & 1 & \cdots & -f_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ +f_{01} & +f_{02} & \cdots & +f_{0i} \end{vmatrix}$$

is the same as M, substituting a 0 to the i's appearing as first subscript, multiplying all terms not containing i as first subscript by f_{0i}, and changing the sign of all terms. Therefore $-M_i$ is equal to the sum of the products of the arms forming an elementary path from node 0 to node i, plus the same products times the products of the arms forming a linear subgraph with the nodes untouched by the same path. Each product has the sign + or – according to whether there is an even or an odd number of cycles.

For an example, see Fig. 15. In that graph there is one elementary path from node 0 to node 1, plus two linear subgraphs formed by nodes 2, 3, 4 untouched by that path, therefore we can compute

$$-M_1 = +f_{01} - f_{01}f_{23}f_{32} - f_{01}f_{34}f_{43},$$

two elementary paths from node 0 to node 2, plus a linear subgraph formed by the nodes untouched by the first path, therefore

$$-M_2 = +f_{01}f_{12} - f_{01}f_{12}f_{34}f_{43} + f_{01}f_{14}f_{43}f_{32},$$

two elementary paths from node 0 to node 3, therefore

$$-M_3 = +f_{01}f_{12}f_{23} + f_{01}f_{14}f_{43},$$

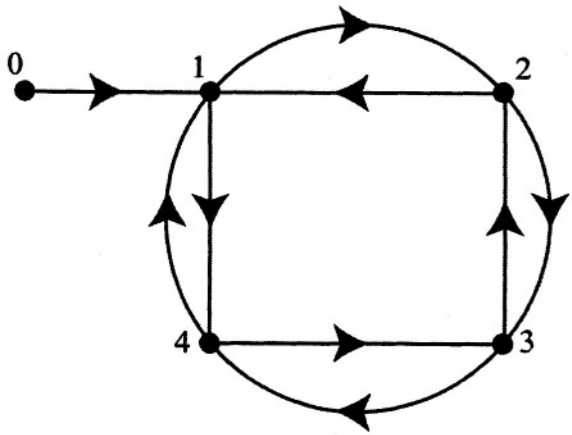

Fig. 15. A graph with 5 nodes

two elementary paths from node 0 to node 4, plus a linear subgraph formed by the nodes untouched by the first path, therefore

$$-M_4 = +f_{01}f_{14} - f_{01}f_{14}f_{23}f_{32} + f_{01}f_{12}f_{23}f_{34},$$

and finally eight linear subgraphs, therefore

$$+M = 1 - f_{12}f_{21} - f_{23}f_{32} - f_{34}f_{43} - f_{14}f_{41} + f_{12}f_{21}f_{34}f_{43} + \\ + f_{14}f_{41}f_{23}f_{32} - f_{12}f_{23}f_{34}f_{41} - f_{14}f_{43}f_{32}f_{21}$$

8.8. TRANSFER FUNCTION

The transfer function $\{g(t)\}$ between any two compartments a and b can be obtained from the ratio of the two functions $\{x_b\}/x_0$ and $\{x_a\}/x_0$, provided a is a precursor of b; therefore

$$\{g(t)\} = \frac{\{x_b\}}{\{x_a\}} = \frac{M_b}{M_a}.$$

For instance from the graph of Fig. 16 we can compute

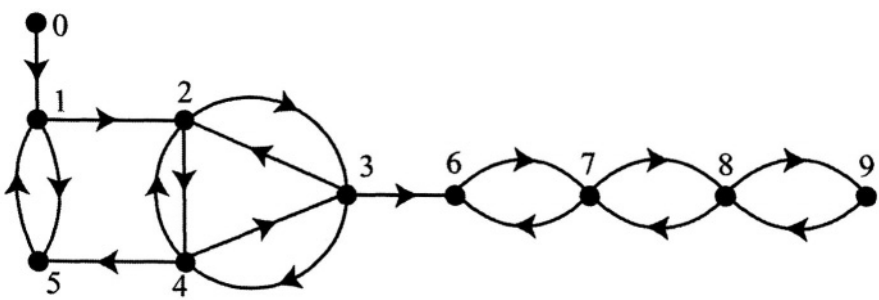

Fig. 16. A graph with 10 nodes

$$+M = 1 - f_{15}f_{51} - f_{23}f_{32} - f_{24}f_{42} - f_{34}f_{43} + f_{15}f_{51}f_{23}f_{32}$$
$$+f_{15}f_{51}f_{24}f_{42} + f_{15}f_{51}f_{34}f_{43} - f_{12}f_{24}f_{45}f_{51}$$
$$+f_{15}f_{51}f_{23}f_{34}f_{42} + f_{15}f_{51}f_{24}f_{43}f_{32} - f_{12}f_{23}f_{34}f_{45}f_{51}$$

$$-M_1 = +f_{01}(1 - f_{23}f_{32} - f_{24}f_{42} - f_{34}f_{43} - f_{23}f_{34}f_{42} - f_{24}f_{43}f_{32})$$

$$-M_2 = +f_{01}f_{12}(1 - f_{34}f_{43})$$

thence the transfer function from compartment 1 to compartment 2 is

$$\frac{M_2}{M_1} = \frac{f_{12}(1 - f_{34}f_{43})}{1 - f_{23}f_{32} - f_{24}f_{42} - f_{34}f_{43} - f_{23}f_{34}f_{42} - f_{24}f_{43}f_{32}}.$$

8.9. REFERENCES

1. S. J. Mason, Feedback Theory - Some Properties of Signal Flow Graphs, *Proc. Inst. Radio Engrs,* **41**, 1144-56(1953).
2. A. Rescigno, Synthesis of a Multicompartmented Biological Model, *Biochim. Biophys. Acta,* **37**, 463-8 (1960).
3. G. Kirchhoff, Ueber die Auflosung der Gleichungen auf welche man bei der Untersichung der linearen Vertheilung galvanischer Strome gefuhrt wird, *Ann. Phys. u. Chem.* (2) **72**, 497-514 (1847).
4. Y. Chow and E. Cassignol, *Linear Signal-flow Graphs and Applications* (John Wiley & Sons, New York, 1962)
5. L. P. A. Robichaud, M. Boisvert and J. Robert, *Signal-flow Graphs and Applications* (Prentice-Hall, Englewood Cliffs, 1962)
6. R. Laue, *Elemente der Graphentheorie und ihre Anwendung in den biologischen Wissenschafien* (Geest & Portig, Leipzig, 1970)
7. A. Rescigno and G. Segre, On Some Metric Properties of the Systems of Compartments, *Bull. Math. Biophysics* **27**, 315-23 (1965).
8. A. Caley, Note on the Theory of Determinants, *Philos. Magazine* (4) **2**, 180-5 (1861).
9. A. Cayley, Note on the Value of Certain Determinants, the Terms of which are the Squared Distances of Points in a Plane or in Space, *Quart. J. Math.* **3**, 275-7 (1860).
10. T. Muir, Question 18033, *Mathematical Questions and Solutions from Educational Times* (2) **29**, 100 (1916).
11. J. N. Stockwell, On the resolution of Symmetrical Equations with Indeterminate Coefficients, *Astron. J.* (Cambridge, Mass.) **6**, 145-9 (1860).
12. R. Balzer, *Theorie und Anwendung der Determinanten,* 3te Auflage (S. Hirzel, Leipzig, 1870), page 29.
13. idem, 4te Auflage (S. Hirzel, Leipzig, 1870), page 38.
14. A. Cayley, Sur les Determinants Gauches, *J. für die reine und angewandte Mathematik* **39**, 93-96 (1849).

9. MOMENTS

9.1. DEFINITION OF MOMENT

Moments have been used in pharmacokinetics for many years, but with different definitions and different names. An early example is given by Rescigno and Segre who used moments to determine the pulmonary transfer function from radiocardiographic data [1] and then generalized the method to include a larger class of problems [2]. Bergner [3] offered a critical analysis of the method and suggested possible applications. Hearon [4] analyzed the distribution of residence times in a compartmental system. Rescigno and Gurpide [5] applied the moments to the study of distribution and metabolism of blood-born compounds. Rescigno [6] defined transit time, residence time, time of entrance and time of exit in terms of moments.

In this chapter I shall try to develop the concept of moment in a simpler way and with a more convenient set of symbols.

Given the function $f(t)$, defined and continuous for any $t \geq 0$, the integral

$$F_i = \int_0^\infty \frac{t^i}{i!} f(t)\mathrm{d}t, \quad i = 0, 1, 2, \ldots \tag{1}$$

if it converges, is called the *moment of order* i, or shortly the ***i-moment,*** of $f(t)$.

Sometimes the moment of a function is defined without the $i!$ at the denominator; I prefer definition (1) because with it all following formulas are considerably simpler.

The moment of order 0 of $\mathrm{d}f/\mathrm{d}t$, if it exists, is given by

$$\int_0^\infty \frac{\mathrm{d}f}{\mathrm{d}t}\mathrm{d}t = -f(0),$$

while for $i \geq 1$ the moment of order i of $\mathrm{d}f/\mathrm{d}t$, if it exists, is given by

$$\int_0^\infty \frac{t^i}{i!}\frac{\mathrm{d}f}{\mathrm{d}t}\mathrm{d}t = \int_{t=0}^{t=\infty} \frac{t^i}{i!}\mathrm{d}f = -\int_0^\infty \frac{t^{i-1}}{(i-1)!} f(t)\mathrm{d}t,$$

i.e. *minus* the moment of order $i - 1$ of $f(t)$.

Consider now the function $\mathrm{d}f^2/\mathrm{d}t^2$; its moment of order 0, if it exists, is given by

$$\int_0^\infty \frac{\mathrm{d}^2 f}{\mathrm{d}t^2}\mathrm{d}t = -\frac{\mathrm{d}f(0)}{\mathrm{d}t},$$

and its moment of order 1, if it exists, by

$$\int_0^\infty t\frac{d^2f}{dt^2}dt = \int_{t=0}^{t=\infty} t\cdot d\frac{df}{dt} = -\int_0^\infty \frac{df}{dt}dt = f(0),$$

while for $i \geq 2$ its moment of order i, if it exists, is given by

$$\int_0^\infty \frac{t^i}{i!}\frac{d^2f}{dt^2}dt = \int_{t=0}^{t=\infty} \frac{t^i}{i!}d\frac{df}{dt} = -\int_0^\infty \frac{t^{i-1}}{(i-1)!}\frac{df}{dt}dt = \int_0^\infty \frac{t^{i-2}}{(i-2)!}f(t)\,dt,$$

i.e. the moment of order $i-2$ of $f(t)$.

By induction we can show that the moment of order i of the function d^jf/dt^j, with $i \geq j$, if it exists, is equal to $(-1)^j$ times the moment of order $i-j$ of function $f(t)$, while for $j > i$ it is equal to $(-1)^{i+1}\dfrac{d^{j-i-1}f(0)}{dt^{j-i-1}}$.

These results can be generalized by defining, for any positive integer i,

$$F_{-i} = (-1)^{i-1}\frac{d^{i-1}f(0)}{dt^{i-1}}, \quad i = 1, 2, \ldots \tag{2}$$

and calling F_{-i} the moment of order $-i$ of $f(t)$; with this notation we can say that the moment of order i of the function d^jf/dt^j, with any integer j, if it exists, is given by

$$\int_0^\infty \frac{t^i}{i!}\frac{d^jf}{dt^j}dt = (-1)^jF_{i-j}, \tag{3}$$

with i, j arbitrary integers.

If a function $f(t)$ has an i-moment F_i, the ratio

$$f_i = \frac{F_i}{F_0} \tag{4}$$

is called *relative moment of order i of $f(t)$*; as seen above, i may be a positive or negative integer.

The i-moment of $f(t)$ has the dimension of $f(t)$ times $[T^{i+1}]$. The corresponding relative moment has the dimension $[T^i]$. The number i may be positive, negative, or null.

9.2. MOMENTS AND CONVOLUTION

Given the convolution equation

$$f(t) = \int_0^t g(\tau)h(t-\tau)d\tau, \tag{5}$$

we can compute the different moments of $f(t)$,

$$\int_0^\infty \frac{t^i}{i!}f(t)dt = \int_0^\infty \frac{t^i}{i!}\int_0^t g(\tau)h(t-\tau)d\tau\,dt;$$

now we change the order of integration,

$$\int_0^\infty \frac{t^i}{i!} f(t)\mathrm{d}t = \int_0^\infty g(\tau)\int_\tau^\infty \frac{t^i}{i!} h(t-\tau)\mathrm{d}t\,\mathrm{d}\tau,$$

then we change the variable of integration of the inner integral,

$$\int_0^\infty \frac{t^i}{i!} f(t)\mathrm{d}t = \int_0^\infty g(\tau)\int_0^\infty \frac{(\tau+\sigma)^i}{i!} h(\sigma)\mathrm{d}\sigma\,\mathrm{d}\tau;$$

we expand the binomial,

$$\int_0^\infty \frac{t^i}{i!} f(t)\mathrm{d}t = \int_0^\infty g(\tau)\int_0^\infty \sum_{j=0}^{i} \frac{1}{i!}\binom{i}{j}\tau^j\sigma^{i-j}h(\sigma)\mathrm{d}\sigma\,\mathrm{d}\tau$$

$$= \int_0^\infty g(\tau)\int_0^\infty \sum_{j=0}^{i} \frac{\tau^j}{j!}\cdot\frac{\sigma^{i-j}}{(i-j)!}h(\sigma)\mathrm{d}\sigma\,\mathrm{d}\tau$$

and finally

$$\int_0^\infty \frac{t^i}{i!} f(t)\mathrm{d}t = \sum_{j=0}^{i}\int_0^\infty \frac{\tau^j}{j!} g(\tau)\mathrm{d}\tau \cdot \int_0^\infty \frac{\sigma^{i-j}}{(i-j)!} h(\sigma)\mathrm{d}\sigma,$$

i.e.,

$$F_i = \sum_{j=0}^{i} G_j H_{i-j} \quad i = 0, \pm1, \pm2, \ldots \tag{6}$$

and

$$f_i = \sum_{j=0}^{i} g_j h_{i-j}. \quad i = \pm1, \pm2, \ldots \tag{7}$$

The continuous convolution of equation (5) has been transformed into the discrete convolutions of equations (6) and (7).

9.3. MOMENT OF ORDER ZERO

The 0-moment of function $f(t)$ its integral from 0 to ∞,

$$F_0 = \int_0^\infty f(t)\mathrm{d}t;$$

it has obviously the dimension of function $f(t)$ times [T].

If $f(t)$ measures the concentration of a drug in a compartment, its 0-moment is called by some authors "Area Under the Curve", abbreviation *AUC*. The *AUC* depends on the dose and on the mode of administration; it is therefore a typical example of an incidental parameter, i.e., of a quantity that defines a property of a pharmacokinetic experiment, not the property of a drug or of a model. I recommend to abandon the use of this symbol because it is often a source of ambiguity.

To understand the meaning of the 0-moment, think of a simple compartment described by the differential equation

$$\frac{dx}{dt} = -K\,x(t) + r(t),$$

where $x(t)$ is the amount of drug in the compartment, K its turnover rate, and $r(t)$ the rate of entry into the compartment from recyrculation. If $x(0) = D$ is the initial dose in the compartment, then, by integration of the above equation,

$$x(\infty) - D = -\int_0^\infty K\,x(t)dt + \int_0^\infty r(t)dt;$$

if the system is open, $x(\infty) = 0$; rearranging we get

$$\int_0^\infty K\,x(t)dt = D + \int_0^\infty r(t)dt;$$

if the turnover rate is constant we can export it from the integral and we get finally,

$$\int_0^\infty x(t)dt = \frac{1}{K}\left(D + \int_0^\infty r(t)dt\right).$$

The term of the left-hand side is the 0-moment of $x(t)$, the term of the right-hand side is the turnover time multiplied by the total amount of drug that went through the compartment.

If we divide the total amount of drug that went through the compartment by the dose D, we get the number of times the drug goes through the compartment; this quantity is called *turnover number*. The product of turnover time by turnover number is the total time spent by the drug in the compartment in all its passages through it; we call it *permanence time*:

$$\frac{\int_0^\infty x(t)dt}{D} = \frac{1}{K}\left(1 + \frac{\int_0^\infty r(t)dt}{D}\right).$$

Both turnover number and permanence time will be discussed in detail in later chapters.

The zero moment has one more important meaning, even when the system is not linear. With the hypothesis that the effect of a drug is proportional to its concentration and time of presence at the site of sampling, the zero moment is a measure of the effect of the drug at the point where it was sampled. But it is important to stress the fact, as observed in section 17.4.1, that this statement is valid even if the processes of absorption and of elimination are not linear, but only if the effect is a linear function of the product time × concentration.

9.4. MOMENT OF ORDER ONE

The 1-moment of function $f(t)$ is the integral

$$F_1 = \int_0^\infty t \cdot f(t)dt;$$

it has obviously the dimension of function $f(t)$ times $[T^2]$.

The relative moment of order one is

$$f_1 = \frac{\int_0^\infty t \cdot f(t)\mathrm{d}t}{\int_0^\infty f(t)\mathrm{d}t}$$

and it has the dimension [T].

If $f(t)$ measures the concentration of a drug in a compartment, its 1-moment is called by some authors "Area Under the Moment Curve", abbreviation *AUMC*. The *AUMC*, as the *AUC*, depends on the dose and on the mode of administration, but also on the time of administration; it is therefore another incidental parameter. I strongly recommend to abandon the use of this symbol, even more ambiguous of *AUC*.

To understand the meaning of the relative moment of order one, observe that, if $x(t)$ is the amount of drug in a compartment and K its turnover rate, then

$K \cdot x(t)$ = rate of exit from the compartment,

$K \cdot x(t)\,\mathrm{d}t$ = amount leaving the compartment in the interval $t, t + \mathrm{d}t$,

t = time of this event,

$$\frac{\int_0^\infty t \cdot K\,x(t)\mathrm{d}t}{\int_0^\infty K\,x(t)\mathrm{d}t} = \frac{\int_0^\infty t \cdot x(t)\mathrm{d}t}{\int_0^\infty x(t)\mathrm{d}t} = \text{average time of exit from the compartment.}$$

Nothing changes if we substitute "concentration" to "amount of substance", because the factor "volume" will appear on both numerator and denominator and will cancel out.

The relative moment of order one of a compartment is called *exit time*, symbol Ω, and will be discussed in more detail in chapter 16. But I will show in section 16.4.1 that if K is not constant, i.e., if we are not dealing with a perfect compartment, the parameter K cannot be exported from the above integrals, therefore in that case the first relative moment is not equal to the exit time.

9.5. MOMENT OF ORDER TWO

The 2-moment of function $f(t)$ is the integral

$$F_2 = \int_0^\infty \frac{t^2}{2} \cdot f(t)\mathrm{d}t;$$

it has obviously the dimension of function $f(t)$ times $[T^3]$.

The relative moment of order two is

$$f_2 = \frac{\int_0^\infty \frac{t^2}{2} \cdot f(t)\mathrm{d}t}{\int_0^\infty f(t)\mathrm{d}t}$$

and it has the dimension $[T^2]$.

Consider the difference

$$2f_2 - f_1^2 = \frac{\int_0^\infty t^2 \cdot f(t)\mathrm{d}t}{\int_0^\infty f(t)\mathrm{d}t} - \left(\frac{\int_0^\infty t \cdot f(t)\mathrm{d}t}{\int_0^\infty f(t)\mathrm{d}t}\right)^2;$$

if $f(t)$ measures the concentration of a drug in a perfect compartment we can multiply each of the four integrand functions on the right-hand side by K; then reasoning as in the previous section we find that the fraction

$$\frac{\int_0^\infty t^2 \cdot Kf(t)dt}{\int_0^\infty Kf(t)dt}$$

is the expected value of the square of the exit time from the compartment, while

$$\left(\frac{\int_0^\infty t \cdot f(t)dt}{\int_0^\infty f(t)dt}\right)^2$$

is the square of the expected value of the exit time from the compartment, therefore the difference

$$2f_2 - f_1^2$$

is the *variance of the exit time* from the compartment.

Some applications of the variance of the exit time can be found in the literature [7, 8].

9.6. MOMENTS OF NEGATIVE ORDER

The –1-moment of function $f(t)$ is its initial value,

$$F_{-1} = f(0);$$

it has obviously the dimension of function $f(t)$.

The $-i$-moment is given by

$$F_{-i} = (-1)^{i-1} \frac{d^{i-1} f(0)}{dt^{i-1}},$$

where i is any positive integer; it has the dimension of function $f(t)$ times $[T^{-(i-1)}]$.

9.7. MOMENTS AND OPERATORS

The different moments of a function $f(t)$ can easily be computed using the operational notation. For instance we know from section B.5 that

$$\left\{\int_0^t f(\tau)d\tau\right\} = \frac{\{f\}}{s}$$

and from section B.15 that

$$\lim_{t\to\infty} f(t) = \lim_{s\to 0} s\{f\};$$

combining the two above expressions we get

$$\lim_{s\to 0}\{f\} = \left\{\int_0^\infty f(t)\mathrm{d}t\right\},$$

therefore

$$F_0 = \lim_{s\to 0}\{f\}.$$

We also know from section B. 10 that

$$\{-t\cdot f(t)\} = \frac{\mathrm{d}\{f\}}{\mathrm{d}s},$$

and combining this with the previous results,

$$\left\{-\int_0^\infty t\cdot f(t)\mathrm{d}t\right\} = \lim_{s\to 0}\frac{\mathrm{d}\{f\}}{\mathrm{d}s},$$

therefore

$$F_1 = -\lim_{s\to 0}\frac{\mathrm{d}\{f\}}{\mathrm{d}s}.$$

Proceeding the same way we obtain, in general,

$$\left\{(-1)^i t^i f(t)\right\} = \frac{\mathrm{d}^i\{f\}}{\mathrm{d}s^i},$$

therefore

$$F_i = \lim_{s\to 0}\frac{(-1)^i}{i!}\frac{\mathrm{d}^i\{f\}}{\mathrm{d}s^i}$$

for any positive integer i.

For the moments of negative order we can use the formula

$$\lim_{t\to 0} f(t) = \lim_{s\to\infty} s\{f\}$$

from section B. 14 to obtain

$$F_{-1} = \lim_{s\to\infty} s\{f\}$$

and identity

$$\{f'\} = s\{f\} - f(0)$$

from section B.5 to obtain

$$F_{-2} = -\lim_{s\to\infty} s\left[s\{f\} - F_{-1}\right].$$

Proceeding the same way we can write

$$F_{-i} = (-1)^{i-1}\lim_{s\to\infty} s\left[s\{f^{(i-2)}\} - F_{i-1}\right]$$

for any positive integer i.

9.8. EXAMPLES

The successive moments of an exponential function

$$f(t) = a \cdot e^{-\alpha t}, \quad \{f\} = \frac{a}{s+\alpha}$$

are

$$F_0 = a/\alpha, \; F_1 = a/\alpha^2, \; F_2 = a/\alpha^3, \; \ldots$$
$$F_{-1} = a, \; F_{-2} = a\alpha, \; F_{-3} = a\alpha^2, \; \ldots$$

The successive moments of a gate function, as defined in section B.7, are

$$F_0 = \mu - \lambda, \; F_1 = \frac{\mu^2 - \lambda^2}{2}, \; F_2 = \frac{\mu^3 - \lambda^3}{3!}, \; \ldots$$
$$F_{-1} = 0, \; F_{-2} = 0, \; F_{-3} = 0, \; \ldots$$

Consider the function $F_{\lambda,\varepsilon}(t)$ defined by

$$\begin{aligned} F_{\lambda,\varepsilon}(t) &= 0 && 0 \le t < \lambda \\ &= 1/\varepsilon && \lambda \le t \le \lambda + \varepsilon \\ &= 0 && \lambda + \varepsilon < t < \infty; \end{aligned}$$

as shown in section B.12, if ε is small, we can write

$$\{F_{\lambda,\varepsilon}\} \approx e^{-\lambda s}.$$

Its successive moments, for ε very small, are approximated by

$$F_0 = 1, \; F_1 = \lambda, \; F_2 = \lambda^2, \; F_3 = \lambda^3, \; \ldots$$
$$F_{-1} = 0, \; F_{-2} = 0, \; F_{-3} = 0, \; F_{-4} = 0, \; \ldots$$

9.9. MOMENTS AND COMPARTMENTAL EQUATIONS

We have seen in section 5.4 that if $x(t)$ is the measure of a drug in a particular compartment of a system of n compartments, then $x(t)$ is the integral of a differential equation of the type

$$\sum_{i=0}^{n} a_i \frac{\mathrm{d}^{n-i}x}{\mathrm{d}t^{n-i}} = 0. \tag{8}$$

We can compute the moments of order j of all terms of this equation as shown in section 9.1; then,

$$\sum_{i=0}^{n} (-1)^i a_i X_{i+j-n} = 0. \quad j = 0, \pm 1, \pm 2, \ldots \tag{9}$$

Observe that each homogeneous equation (9) represents a linear relationship among $n+1$ moments X_i, therefore only n of them are linearly independent.

The persymmetric matrix (see section D.12)

$$\begin{pmatrix} M_i & M_{i+1} & M_{i+2} & \cdots \\ M_{i+1} & M_{i+2} & M_{i+3} & \cdots \\ M_{i+2} & M_{i+3} & M_{i+4} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

with i any integer (positive, negative, or null), is of rank n, because only n of its elements are linearly independent. As a consequence, if we construct the successive persymmetric matrices

$$\left(M_i\right), \begin{pmatrix} M_i & M_{i+1} \\ M_{i+1} & M_{i+2} \end{pmatrix}, \begin{pmatrix} M_i & M_{i+1} & M_{i+2} \\ M_{i+1} & M_{i+2} & M_{i+3} \\ M_{i+2} & M_{i+3} & M_{i+4} \end{pmatrix}, \ \ldots \qquad (10)$$

only the first n of them are non-singular.

In practice, if a number of moments of a linear system have been measured, the first of the matrices (10) that is singular shows the order of the system it was taken from. The vector that annihilates it is formed by the coefficients a_0, a_1, a_2, ..., of differential equation (8).

For instance, if the system is formed by only two compartments, the matrices

$$\begin{pmatrix} M_{-2} & M_{-1} & M_0 \\ M_{-1} & M_0 & M_1 \\ M_0 & M_1 & M_2 \end{pmatrix} = \begin{pmatrix} -\dfrac{dx(0)}{dt} & x(0) & \int_0^\infty x(t)dt \\ x(0) & \int_0^\infty x(t)dt & \int_0^\infty t\,x(t)dt \\ \int_0^\infty x(t)dt & \int_0^\infty t\,x(t)dt & \int_0^\infty \dfrac{t^2}{2} x(t)dt \end{pmatrix},$$

$$\begin{pmatrix} M_{-1} & M_0 & M_1 \\ M_0 & M_1 & M_2 \\ M_1 & M_2 & M_3 \end{pmatrix} = \begin{pmatrix} x(0) & \int_0^\infty x(t)dt & \int_0^\infty t\,x(t)dt \\ \int_0^\infty x(t)dt & \int_0^\infty t\,x(t)dt & \int_0^\infty \dfrac{t^2}{2} x(t)dt \\ \int_0^\infty t\,x(t)dt & \int_0^\infty \dfrac{t^2}{2} x(t)dt & \int_0^\infty \dfrac{t^3}{3!} x(t)dt \end{pmatrix},$$

$$\begin{pmatrix} M_0 & M_1 & M_2 \\ M_1 & M_2 & M_3 \\ M_2 & M_3 & M_4 \end{pmatrix} = \begin{pmatrix} \int_0^\infty x(t)dt & \int_0^\infty t\,x(t)dt & \int_0^\infty \dfrac{t^2}{2} x(t)dt \\ \int_0^\infty t\,x(t)dt & \int_0^\infty \dfrac{t^2}{2} x(t)dt & \int_0^\infty \dfrac{t^3}{3!} x(t)dt \\ \int_0^\infty \dfrac{t^2}{2} x(t)dt & \int_0^\infty \dfrac{t^3}{3!} x(t)dt & \int_0^\infty \dfrac{t^4}{4!} x(t)dt \end{pmatrix}$$

contain the same information; we are free to choose the one with the elements that were measured with the highest accuracy.

9.10. EXAMPLE

Suppose we have observed the following moments:

$$M_{-3} = 10,\ M_{-2} = 6,\ M_{-1} = 4,\ M_0 = 3,\ M_1 = 2.5,\ M_2 = 2.25;$$

we can construct the matrices

$$\begin{pmatrix} 4 & 3 \\ 3 & 2.5 \end{pmatrix} \text{ and } \begin{pmatrix} 6 & 4 & 3 \\ 4 & 3 & 2.5 \\ 3 & 2.5 & 2.25 \end{pmatrix};$$

the first is non-singular and the second singular, therefore the order of the system of compartments they were taken from is 3, and we must find the vector satisfying equation

$$\begin{pmatrix} 6 & 4 & 3 \\ 4 & 3 & 2.5 \\ 3 & 2.5 & 2.25 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ -a_1 \\ a_2 \end{pmatrix} = \mathbf{0}.$$

The solution is

$$\begin{pmatrix} a_0 \\ -a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix},$$

and the observed system has the characteristic equation

$$\frac{dx^2}{dt^2} + 3\frac{dx}{dt} + 2x = 0$$

or, in operational notation,

$$s^2\{f\} - sf(0) - f'(0) + 3\left[s\{f\} - f(0)\right] + 2\{f\} = 0;$$

reordering we get

$$\{f\} = \frac{f(0)\cdot s + f'(0) + 3f(0)}{s^2 + 3s + 2}.$$

We know that $f(0)$ is equal to the –1-moment and $-f'(0)$ to the –2-moment, therefore

$$\{f\} = \frac{4s - 6 + 3\cdot 4}{s^2 + 3s + 2} = \frac{4s + 6}{s^2 + 3s + 2} = \frac{2}{s+1} + \frac{2}{s+2},$$

$$f(t) = 2e^{-t} + 2e^{-2t}.$$

Alternatively, after having found that the order of the system is 3, and therefore it can be represented by a sum of two exponentials, we can write

$$f(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t},$$

and for its moments,

$$F_1 = \frac{A_1}{\lambda_1^2} + \frac{A_2}{\lambda_2^2} = 2.5,$$

$$F_0 = \frac{A_1}{\lambda_1} + \frac{A_2}{\lambda_2} = 3,$$

$$F_{-1} = A_1 + A_2 = 4,$$

$$F_{-2} = A_1\lambda_1 + A_2\lambda_2 = 6;$$

the four equations above determine the parameters $A_1, A_2, \lambda_1, \lambda_{12}$ in a unique way.

9.11. REFERENCES

1. A. Rescigno and G. Segre, Calcolo della funzione di trasferimento polmonare dai dati radiocardiografici, *Minerva Nucleare* **5**, 296-8 (1961).
2. A. Rescigno and G. Segre, *La Cinetica dei Farmaci e dei traccianti radioattivi* (Boringhieri, Torino, 1961).
3. P.-E. E. Bergner, The Significance of Certain Tracer Kinetical Methods, Especially with Respect to the Tracer Dynamic Definition of Metabolic Turnover, *Acta Radiologica Supplement* **210**, 1-59 (1962).
4. J. Z. Hearon, Residence Times in Compartmental Systems and the Moments of a Certain Distribution, *Mathematical Biosciences* **15**, 69-77 (1972).
5. A. Rescigno and E. Gurpide, Estimation of Average Times of Residence, Recycle, and Interconversion of Blood-Borne Compounds Using Tracer Methods, *J. clin. Endocrinol. Metab.* **36**, 263-76 (1973).
6. A. Rescigno, On Transfer Times in Tracer Experiments, *J. theoret. Biol.* **39**, 9-27 (1973).
7. A. Rescigno and L. D. Michels, On Dispersion in Tracer Experiments, *J. theoret. Biol.* **41**, 451 -60, 1973.
8. A. Rescigno and L. D. Michels, Compartment Modeling from Tracer Experiments, *Bull. Math. Biol.* **35**, 245-57, 1973.

This page intentionally left blank

10. INTEGRAL EQUATIONS

10.1. TRANSFER FUNCTION

Consider a particle in a living system and suppose that that particle can be recognized in two different states of the system, where by state I mean a particular location or a particular chemical form, or both. If one state is the precursor of the other (not necessarily the immediate precursor), then we can study the relationship among event **A** (the particle is in the precursor state), event **B** (transition from precursor to successor state), and event **C** (presence of the particle in the successor state).

For any t and τ such that $0 \leq \tau \leq t$, call $A(\tau)$ the probability of **A** at time τ, and $C(t)$ the probability of **C** at time t. Suppose now that **B** depends only on the interval of time separating **A** and **C**, so that we can call now $B(t-\tau)\mathrm{d}\tau$ the conditional probability that a particle is in **C** at time t if it left **A** in the interval $\tau, \tau + \mathrm{d}\tau$.

The product

$$A(\tau) \cdot B(t-\tau)\,\mathrm{d}\tau$$

therefore is the absolute probability that a particle leaves **A** in the interval of time between τ and $\tau + \mathrm{d}\tau$ and is still in **C** at time t.

By integration of the above product we obtain the probability of a particle being in state **C** at time t irrespective of when it left **A**, i.e.

$$\int_0^t A(\tau) \cdot B(t-\tau)\,\mathrm{d}\tau = C(t). \tag{1}$$

The integral on the left-hand side is called *convolution integral*; it represents the relationship among the variables of a *linear, invariant system.* See Appendix A for a discussion on the different aspects of convolution.

If we think of an experiment where a very large number of identical particles is used, then the number of particles present in the precursor and in the successor states are good estimates of functions $A(t)$ and $C(t)$ respectively. Function $B(t)$ represents the probability that a particle that was in **A** at time 0 will still be in **C** at time t; therefore in a hypothetical experiment where all identical particles left the precursor near time zero, the number of particles found in the successor will be given by $B(t)$.

In more general terms, if i and j are two compartments of a system, and if the drug was added to compartment i (or to another compartment from where it could not reach compartment j before going through compartment i), there might exist a function $g_{ij}(t,\tau)$ of two variables such that

$$x_j(t) = \int_0^t g_{ij}(t,\tau)x_j(\tau)d\tau\,, \qquad (2)$$

where $x_i(t)$ and $x_j(t)$ are the amount or the concentration of drug in compartment i and j respectively. This is a Volterra equation of the second type [1]; if any two of the three functions $x_i(t)$, $x_j(t)$, $g_{ij}(t, \tau)$ are known, the third can be computed.

Function $g_{ij}(t, \tau)$, if it exists, is called the *transfer function* from compartment i to compartment j. The use of transfer functions in pharmacokinetic problems began in 1960 [2].

The integral in equation (2), called *convolution,* is a linear operation. This means that two of its solutions can be added to give a new solution. The consequence is that if in one experiment the amount of drug in compartments i and j were $x_i^*(t)$ and $x_j^*(t)$, respectively, and in another experiment they were $x_i^{**}(t)$ and $x_j^{**}(t)$, respectively, then in an experiment where the amount of drug in i is $x_i^*(t) + x_i^{**}(t)$, the amount in j will be $x_j^*(t) + x_j^{**}(t)$.

The existence of a transfer function can be assumed as the definition of *linearity* of a system.

When the transfer function is a function of one variable, equation (2) takes the simpler form of equation (1), i.e.,

$$x_j(t) = \int_0^t g_{ij}(t-\tau)x_i(\tau)d\tau \qquad (3)$$

and the linear system is said to be *state-determined* or *time-invariant.* This means that a solution does not change if the time origin is changed. In fact suppose that $x_i(t)$, $x_j(t)$ are a solution of equation (3), and consider the new function

$$\begin{aligned} x_i^*(t) &= 0, && 0 \le t < t_0 \\ &= x(t-t_0), && t \ge t_0. \end{aligned}$$

For this new function,

$$x_j^*(t) = \int_0^t g_{ij}(t-\tau)x_i^*(\tau)d\tau\,,$$

but

$$\begin{aligned} \int_0^t g_{ij}(t-\tau)x_i^*(\tau)d\tau &= 0, && 0 \le t < t_0 \\ &= \int_{t_0}^t g_{ij}(t-\tau)x_i(\tau-t_0)d\tau, && t \ge t_0 \end{aligned}$$

furthermore

$$\int_{t_0}^t g_{ij}(t-\tau)x_i(\tau-t_0)d\tau = \int_0^{t-t_0} g_{ij}(t-t_0-\sigma)x_i(\sigma)d\sigma$$

therefore

$$\begin{aligned} x_j'^*(t) &= 0, && 0 \le t < t_0 \\ &= x_j(t-t_0), && t \ge t_0 \end{aligned} \qquad \begin{aligned} x_j'^*(t) &= 0, && 0 \le t < t_0 \\ &= x_j(t-t_0), && t \ge t_0 \end{aligned}$$

i.e., $x_i(t)$ and $x_j(t)$ are shifted along the time axis by the same quantity.

10.2. OPERATIONAL FORM OF THE TRANSFER FUNCTION

Appendix B shows that the operational calculus transforms the convolution integral into a product of functions; we shall therefore use the operational notation to transform equation (3) into the form

$$\{g_{ij}\} = \frac{\{x_j\}}{\{x_i\}}. \tag{4}$$

The above ratio is the transfer function from compartment i to compartment j; by an obvious extension, the ratio $\{x_j\}/x_i(0)$ is called the transfer function *from the source* of compartment i to compartment j [3,4].

It is worth repeating that all these definitions are valid only if the drug were added to compartment i, or to another compartment from where it could reach compartment j only through compartment i.

As an example, consider three compartments 1, 2, 3 connected in series (see Fig. 1); the amount of drug in each of them is $x_1(t)$, $x_2(t)$, $x_3(t)$; these three functions are related by the differential equations

$$\frac{dx_2}{dt} = -K_2 x_2(t) + k_{12} x_1(t) \tag{5}$$

$$\frac{dx_3}{dt} = -K_3 x_3(t) + k_{23} x_2(t) \tag{6}$$

with initial conditions

$$x_2(0) = x_3(0) = 0.$$

The integrals of those equations, if $K_2 \neq K_3$, are

$$x_2(t) = \int_0^t k_{12} e^{-K_2\tau} \cdot x_1(t-\tau) \cdot d\tau$$

$$x_3(t) = \int_0^t \frac{k_{12}k_{23}}{K_3 - K_2}\left(e^{-K_2\tau} - e^{-K_3\tau}\right) \cdot x_1(t-\tau) \cdot d\tau$$

therefore the transfer functions are

$$g_{12}(t) = k_{12} e^{-K_2 t}, \tag{7}$$

$$g_{13}(t) = \frac{k_{12}k_{23}}{K_3 - K_2}\left(e^{-K_2 t} - e^{-K_3 t}\right). \tag{8}$$

In operational notation the differential equations (5) and (6) are

$$s\{x_2\} = -K_2\{x_2\} + k_{12}\{x_1\}$$

$$s\{x_3\} = -K_3\{x_3\} + k_{23}\{x_2\}$$

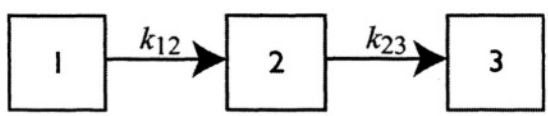

Fig. 1. The three compartments of equations (5) and (6)

thence

$$\{g_{12}\} = \frac{\{x_2\}}{\{x_1\}} = \frac{k_{12}}{s + K_2},$$

$$\{g_{13}\} = \frac{\{x_3\}}{\{x_1\}} = \frac{k_{12}k_{23}}{(s + K_2)(s + K_3)}.$$

As a second example, consider the three compartments as above, with a reversible transfer between the second and third one (see Fig. 2); equation (5) becomes

$$\frac{dx_2}{dt} = -K_2 x_2(t) + k_{12} x_1(t) + k_{32} x_3(t) \tag{9}$$

while equation (6) does not change formally, though the coefficient K_3 in this case includes the transfer rate from compartment 3 to compartment 2.

Equations (6) and (9) can be integrated with the standard methods of analysis, giving

$$x_2(t) = \int_0^t \frac{k_{12}}{\beta - \alpha}\left[(K_3 - \alpha)e^{-\alpha\tau} - (K_3 - \beta)e^{-\beta\tau}\right] \cdot x_1(t - \tau)d\tau$$

$$x_3(t) = \int_0^t \frac{k_{12}k_{23}}{\beta - \alpha}\left[e^{-\alpha\tau} - e^{-\beta\tau}\right] \cdot x_1(t - \tau)d\tau$$

where $-\alpha$ and $-\beta$ are the roots of the characteristic equation

$$x^2 + (K_2 + K_3)x + K_2K_3 - k_{23}k_{32} = 0, \tag{10}$$

that is,

$$\alpha = \tfrac{1}{2}\left(K_2 + K_3 - \sqrt{(K_2 - K_3)^2 + 4k_{23}k_{32}}\right),$$

$$\beta = \tfrac{1}{2}\left(K_2 + K_3 + \sqrt{(K_2 - K_3)^2 + 4k_{23}k_{32}}\right).$$

The transfer functions are therefore

$$g_{12}(t) = \frac{k_{12}}{\beta - \alpha}\left[(K_3 - \alpha)e^{-\alpha t} - (K_3 - \beta)e^{-\beta t}\right],$$

$$g_{13}(t) = \frac{k_{12}k_{23}}{\beta - \alpha}\left[e^{-\alpha t} - e^{-\beta t}\right].$$

In operational notation the differential equations (9) and (6) are

$$s\{x_2\} = -K_2\{x_2\} + k_{12}\{x_1\} + k_{32}\{x_3\}$$

$$s\{x_3\} = -K_3\{x_3\} + k_{23}\{x_2\}$$

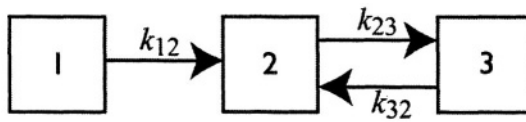

Fig. 2. The three compartments of equations (9) and (6)

thence

$$-k_{12}\{x_1\} + (s + K_2)\{x_2\} - k_{32}\{x_3\} = 0$$
$$-k_{23}\{x_2\} + (s + K_3)\{x_3\} = 0$$

and by eliminating first $\{x_2\}$, then $\{x_3\}$,

$$\{x_3\} = \frac{k_{12}k_{23}}{(s + K_2)(s + K_3) - k_{23}k_{32}}\{x_1\},$$

$$\{x_2\} = \frac{k_{12}(s + K_3)}{(s + K_2)(s + K_3) - k_{23}k_{32}}\{x_1\}.$$

Observe that α and β are always real and different one from the other; in fact the discriminant

$$(K_2 - K_3)^2 + 4k_{23}k_{32}$$

of the characteristic equation (10) is always positive because of the relationships

$$K_2 \geq k_{23} > 0, \quad K_3 \geq k_{32} > 0.$$

An alternative approach to transfer functions was shown in section 8.8. More properties of the transfer functions will be investigated in later chapters.

10.3. INITIAL VALUES OF THE TRANSFER FUNCTION

The numerical computation of the transfer function between two compartments where a drug has been measured is always a difficult operation, and in all cases requires a very detailed knowledge of the two given functions. Nevertheless a partial determination of the transfer function is possible from a limited number of sampling points of the functions $x_i(t)$ and $x_j(t)$. In fact by sampling $x_i(t)$ and $x_j(t)$ at different times we can plot a number of values of $x_j/(t \cdot x_i)$ and extrapolate those values for $t \to 0$; if $x_i(0) \neq 0$, using L'Hospital rule [5],

$$\lim_{t \to 0} \frac{x_j(t)}{t \cdot x_i(t)} = \lim_{t \to 0} \frac{\int_0^t g_{ij}(t - \tau)x_i(\tau)\,d\tau}{t \cdot x_i(t)}$$

$$= \lim_{t \to 0} \frac{\int_0^t g_{ij}'(t - \tau)x_i(\tau)\,d\tau + g_{ij}(0)x_i(t)}{x_i(t) + t \cdot x_i'(t)} = g_{ij}(0).$$

If $x_i(0) = 0$, but $x_i'(0) \neq 0$, we apply L'Hospital rule once more and get

$$\lim_{t \to 0} \frac{x_j(t)}{t \cdot x_i(t)} = \lim_{t \to 0} \frac{\int_0^t g_{ij}''(t - \tau)x_i(\tau)\,d\tau + g_{ij}'(0)x_i(t) + g_{ij}(0)x_i'(t)}{2x_i'(t) + t \cdot x_i''(t)}$$

$$= \frac{g_{ij}(0)}{2}.$$

By induction, if m is the lowest order of derivative of $x_i(t)$ different from zero for $t = 0$, then

$$\lim_{t\to 0}\frac{x_j(t)}{t\cdot x_i(t)}=\frac{g_{ij}(0)}{m+1}.$$

Suppose now it was found that $g_{ij}(0)=0$; we can plot $x_i/(t^2\cdot x_j)$ versus t and extrapolate those values for $t\to 0$; again if $x_i(0)\neq 0$, using L'Hospital rule,

$$\begin{aligned}\lim_{t\to 0}\frac{x_j(t)}{t^2\cdot x_i(t)}&=\lim_{t\to 0}\frac{\int_0^t g_{ij}(t-\tau)x_i(\tau)\mathrm{d}\tau}{t^2\cdot x_i(t)}\\&=\lim_{t\to 0}\frac{\int_0^t g_{ij}'(t-\tau)x_i(\tau)\mathrm{d}\tau+g_{ij}(0)x_i(t)}{2t\,x_i(t)+t^2\cdot x_i'(t)}\\&=\lim_{t\to 0}\frac{\int_0^t g_{ij}''(t-\tau)x_i(\tau)\mathrm{d}\tau+g_{ij}'(0)x_i(t)+g_{ij}(0)x_i'(t)}{2x_i(t)+4t\,x_i'(t)+t^2x_i''(t)}\\&=\frac{g_{ij}'(0)}{2}.\end{aligned}$$

If $g_{ij}(0)=0$ and $x_i(0)=0$, but $x_i'(0)\neq 0$, we can apply L'Hospital rule once more,

$$\begin{aligned}\lim_{t\to 0}\frac{x_j(t)}{t^2\cdot x_i(t)}&=\lim_{t\to 0}\frac{\int_0^t g_{ij}'''(t-\tau)x_i(\tau)\mathrm{d}\tau+g_{ij}''(0)x_i(t)+g_{ij}'(0)x_i'(t)+g_{ij}(0)x_i''(t)}{6x_i'(t)+6t\,x_i''(t)+t^2x_i'''(t)}\\&=\frac{g_{ij}'(0)}{6}.\end{aligned}$$

Proceeding in the same way we can prove that, if

$$\begin{gathered}x_i(0)=x_i'(0)=\cdots=x_i^{(m-1)}(0)=0,\quad x_i^{(m)}(0)\neq 0,\\g_{ij}(0)=g_{ij}'(0)=\cdots=g_{ij}^{(n-1)}(0)=0,\end{gathered}\tag{11}$$

then

$$\lim_{t\to 0}\frac{x_j(t)}{t^{n+1}x_i(t)}=\frac{m!}{(m+n+1)!}g_{ij}^{(n)}(0).$$

Conditions (11) are not very restrictive, for if they do not hold for a specific transfer function $g_{ij}(t)$, we can define the function

$$g_{ij}^{*}=g_{ij}(t)-g_{ij}(0)-t\,g_{ij}'(0)-\frac{t^2}{2!}g_{ij}''(0)-\cdots-\frac{t^{n-1}}{(n-1)!}g_{ij}^{(n-1)}(0);$$

conditions (11) hold for function $g_{ij}^{*}(t)$, and since

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n}g_{ij}^{*}(0)=\frac{\mathrm{d}^n}{\mathrm{d}t^n}g_{ij}(0),$$

we can write

$$g_{ij}^{*(n)}(0) = \frac{(m+n+1)!}{m!} \lim_{t\to 0} \frac{x_j^*(t)}{t^{n+1} x_i(t)}$$

with

$$x_j^*(t) = x_j(t) - g_{ij}(0)\int_0^t x_j(t)dt - g_{ij}'(0)\iint x_j(t)dt - g_{ij}''(0)\iiint x_j(t)dt - \cdots$$

and compute in succession $g_{ij}(0)$, $g_{ij}'(0)$, $g_{ij}''(0)$, ..., $g_{ij}^{(n-1)}(0)$, $g_{ij}^{(n)}(0)$.

The physical meaning of the initial value of $g(t)$ and of its successive derivatives may be explained with the following considerations.

From the definition of function $g_{ij}(t)$ we know that

$g_{ij}(t-\tau)d\tau$ = conditional probability that a particle that left compartment i in the interval $\tau, \tau+d\tau$, is in compartment j at time $t \geq \tau$,

consequently

$g_{ij}(0)d\tau$ = conditional probability that a particle that left compartment i in the interval $\tau, \tau+d\tau$, is in compartment j at the same time τ,

thence $g_{ij}(0)$ is the fractional rate at which the particles in compartment i enter compartment j, and $g_{ij}(0) = 0$ means that no particles are transferred from compartment i to compartment j instantly.

If $g_{ij}(0) = 0$, from differentiating equation (3) we get

$$x_j'(t) = \int_0^t g_{ij}'(t-\tau)x_i(\tau)d\tau,$$

thence, when $g_{ij}(0) = 0$, the meaning of $g_{ij}'(0)$ is the derivative of the fractional rate at which the particles in compartment i enter compartment j.

Similarly, if $g_{ij}(0) = g_{ij}'(0) = 0$, with a second differentiation we get

$$x_j''(t) = \int_0^t g_{ij}''(t-\tau)x_i(\tau)d\tau,$$

thence in this case the meaning of $g_{ij}''(0)$ is the second derivative of the fractional rate at which the particles in compartment i enter compartment j.

For example consider the transfer function (7); its initial value is $k_{12} \neq 0$. That means that the particles leaving compartment 1 enter compartment 2 without any intermediary step with fractional rate k_{12}.

For the transfer function (8) the initial value is zero; in fact there is an intermediate compartment between 1 and 3 and the initial transfer rate from compartment 1 to compartment 3 is zero. The derivative of that transfer function is

$$g_{13}'(t) = \frac{k_{12}k_{23}}{K_3 - K_2}\left(-K_2 e^{-K_2 t} + K_3 e^{-K_3 t}\right)$$

whose initial value is $k_{12} \cdot k_{23}$; in fact from equation (6) we get

$$\frac{d^2x_3}{dt^2} = -K_3\frac{dx_3}{dt} + k_{23}\frac{dx_2}{dt}$$

and by substituting (5) and (6) into this one,

$$\frac{d^2x_3}{dt^2} = -K_3\left(-K_3x_3 + k_{23}x_2\right) + k_{23}\left(-K_2x_2 + k_{12}x_1\right)$$
$$= +k_{12}k_{23}x_1 - k_{23}(K_2 + K_3)x_2 + K_3^2x_3$$

showing that the fractional transfer rate from 1 to 3 increases with rate $k_{12}k_{23}$.

10.4. INTEGRAL OF THE TRANSFER FUNCTION

We go back now to equation (3), valid for a linear state-determined system; by integration of both sides from 0 to ∞ we get

$$\int_0^\infty x_j(t)dt = \int_0^\infty \int_0^t g_{ij}(t-\tau)x_i(\tau)d\tau\, dt,$$

and with a change of the order of integration,

$$\int_0^\infty x_j(t)dt = \int_0^\infty x_i(\tau)dt \cdot \int_0^\infty g_{ij}(t)dt.$$

This last identity can be written in the form

$$\int_0^\infty g_{ij}(t)dt = \frac{\int_0^\infty x_j(t)dt}{\int_0^\infty x_i(\tau)dt}, \tag{12}$$

analogous to (4).

In the above fraction, the denominator is the number of particles present in compartment i times the interval of time they spend there, the numerator is the same for compartment j, therefore the integral on the left-hand side is equal to the fraction of particles transferred from i to j, times the ratio of the turnover time of the two compartments.

For example, for the transfer function (7) we get

$$\int_0^\infty g_{12}(t)dt = \frac{k_{12}}{K_2} = \frac{k_{12}}{K_1} \cdot \frac{K_1}{K_2},$$

where k_{12}/K_1 is the fraction transferred from compartment 1 to compartment 2, and K_1/K_2 is the ratio of the turnover times of the two compartments.

For the transfer function (8) we get

$$\int_0^\infty g_{13}(t)dt = \frac{k_{12}k_{23}}{K_3 - K_2}\left(\frac{1}{K_2} - \frac{1}{K_3}\right) = \frac{k_{12}}{K_1} \cdot \frac{k_{23}}{K_2} \cdot \frac{K_1}{K_3},$$

where $k_{12}/K_1 \cdot k_{23}/K_2$ is the fraction transferred from compartment 1 to compartment 3, and K_1/K_3 is the ratio of the turnover times of the two compartments.

With the operational notation the computation of the integral of the transfer function is much simpler. From

$$\{g_{12}\} = \frac{k_{12}}{s + K_2}$$

using the corollary 1 (section B.15) we get

$$\int_0^\infty g_{12}(t)dt = \lim_{s\to 0} \frac{k_{12}}{s+K_2} = \frac{k_{12}}{K_2};$$

from

$$\{g_{13}\} = \frac{k_{12}k_{23}}{(s+K_2)(s+K_3)}$$

we get

$$\int_0^\infty g_{13}(t)dt = \lim_{s\to 0} \frac{k_{12}k_{23}}{(s+K_2)(s+K_3)} = \frac{k_{12}k_{23}}{K_2K_3}.$$

10.5. MOMENTS OF THE TRANSFER FUNCTION

In the two previous sections I have shown the meaning of the initial value of the transfer function and of its derivative, and of the integral of the transfer function from 0 to ∞. In this section I will show how those results can be unified and generalized using the method of moments, as outlined in chapter 9.

Observe that using equation (6) of section 9.2 we can transform the convolution integral equation (1) into the discrete convolution

$$\sum_{j=0}^{i} A_j B_{i-j} = C_i, \tag{13}$$

where i is an arbitrary integer, positive, negative, or null.

In particular, for $i = -2, -1, 0, +1$, from equation (13) we get,

$$\int_0^\infty x_1(t)\mathrm{d}t \cdot \frac{\mathrm{d}g_{12}(0)}{\mathrm{d}t} + x_1(0) \cdot g_{12}(0) + \frac{\mathrm{d}x_1(0)}{\mathrm{d}t} \cdot \int_0^\infty g_{12}(t)\mathrm{d}t = \frac{\mathrm{d}x_2(0)}{\mathrm{d}t}, \tag{14}$$

$$\int_0^\infty x_1(t)\mathrm{d}t \cdot g_{12}(0) + x_1(0) \cdot \int_0^\infty g_{12}(t)\mathrm{d}t = x_2(0), \tag{15}$$

$$\int_0^\infty x_1(t)\mathrm{d}t \cdot \int_0^\infty g_{12}(t)\mathrm{d}t = \int_0^\infty x_2(t)\mathrm{d}t, \tag{16}$$

$$\int_0^\infty x_1(t)\mathrm{d}t \cdot \int_0^\infty t \cdot g_{12}(t)\mathrm{d}t + \int_0^\infty t \cdot x_1(t)\mathrm{d}t \cdot \int_0^\infty g_{12}(t)\mathrm{d}t = \int_0^\infty t \cdot x_2(t)\mathrm{d}t. \tag{17}$$

Expression (16) is identical to identity (12) shown in the previous section; all other expressions can be solved sequentially to compute the moments of a transfer function from the moments of two compartments.

For instance, if $\int_0^\infty g_{12}(t)\mathrm{d}t$ has been computed from (16), we can use (15) to compute $g_{12}(0)$, then (14) to compute $\mathrm{d}g_{12}(0)/\mathrm{d}t$, and so forth.

I will show the use of expression (17) in the next section.

A special case is presented by the transfer function from the source of a compartment to that same compartment (see section 10.2); that transfer function is defined as the ratio

$\{x\}/x(0)$, therefore its moments are identical to the corresponding moments of $x(t)$ divided by its initial value $x(0)$; we can write

$$G_i = \frac{X_i}{x(0)}$$

for any i, positive, negative or null. Substituting this identity into equation (13) we conclude that all moments of the source are 0 except the 0-moment equal to the initial value $x(0)$.

The only function whose moments are $F_0 = 1$, $F_l = 0$ for $l \neq 0$, is the unit impulse function $F_{\lambda,\varepsilon}(t)$ with $\lambda = 0$ defined in section B.12, and whose moments where computed in section 9.8. We conclude that the transfer function from the source of a compartment to that same compartment is equal to the Dirac delta function multiplied by the initial value of the compartment.

10.6. FIRST RELATIVE MOMENT

In expression (17) we divide all terms by the 0-moments of $x_1(t)$ and $g_{12}(t)$,

$$\frac{\int_0^\infty t \cdot g_{12}(t)\mathrm{d}t}{\int_0^\infty g_{12}(t)\mathrm{d}t} + \frac{\int_0^\infty t \cdot x_1(t)\mathrm{d}t}{\int_0^\infty x_1(t)\mathrm{d}t} = \frac{\int_0^\infty t \cdot x_2(t)\mathrm{d}t}{\int_0^\infty x_1(t)\mathrm{d}t \cdot \int_0^\infty g_{12}(t)\mathrm{d}t},$$

we substitute expression (16) into the denominator of the right-hand side, and rearrange,

$$\frac{\int_0^\infty t \cdot g_{12}(t)\mathrm{d}t}{\int_0^\infty g_{12}(t)\mathrm{d}t} = \frac{\int_0^\infty t \cdot x_2(t)\mathrm{d}t}{\int_0^\infty x_2(t)\mathrm{d}t} - \frac{\int_0^\infty t \cdot x_1(t)\mathrm{d}t}{\int_0^\infty x_1(t)\mathrm{d}t}.$$

The two fractions on the right-hand side are equal to the times of exit from compartments 2 and 1, respectively. Their difference, called *transfer time* from compartment i to compartment j, is the first relative moment of the transfer function from i to j; it is equal to the average time spent by the particles from the time of exit from i to the time of exit from j.

As an example consider the compartments described by equations (5) and (6). The first relative moment of $\{g_{12}\}$ is

$$\frac{\int_0^\infty t \cdot k_{12}e^{-K_2 t}\mathrm{d}t}{\int_0^\infty k_{12}e^{-K_2 t}\mathrm{d}t} = \frac{k_{12}/K_2^2}{k_{12}/K_2} = \frac{1}{K_2};$$

in fact compartment 1 is the immediate precursor of compartment 2, and the transfer time from 1 to 2 is just the time spent in the second compartment.

The first relative moment of $\{g_{13}\}$ is

$$\frac{\int_0^\infty t \cdot \frac{k_{12}k_{23}}{K_3 - K_2}\left(e^{-K_2 t} - e^{-K_3 t}\right)\mathrm{d}t}{\int_0^\infty \frac{k_{12}k_{23}}{K_3 - K_2}\left(e^{-K_2 t} - e^{-K_3 t}\right)\mathrm{d}t} = \frac{1/K_2^2 - 1/K_3^2}{1/K_2 - 1/K_3} = \frac{1}{K_2} + \frac{1}{K_3};$$

in fact after leaving compartment 1 the particles have to spend time first in compartment 2 and then in compartment 3.

Consider now the compartments described by equations (6) and (9); the first relative moment of $\{g_{12}\}$ is

$$\frac{\int_0^\infty t \cdot \frac{k_{12}}{\beta-\alpha}\left[(K_3-\alpha)e^{-\alpha t}-(K_3-\beta)e^{-\beta t}\right]dt}{\int_0^\infty \frac{k_{12}}{\beta-\alpha}\left[(K_3-\alpha)e^{-\alpha t}-(K_3-\beta)e^{-\beta t}\right]dt} =$$

$$= \frac{(K_3-\alpha)/\alpha^2-(K_3-\beta)/\beta^2}{(K_3-\alpha)/\alpha-(K_3-\beta)/\beta} = \frac{1}{\alpha}+\frac{1}{\beta}-\frac{1}{K_3},$$

and the first relative moment of $\{g_{13}\}$ is

$$\frac{\int_0^\infty t \cdot \frac{k_{12}k_{23}}{\beta-\alpha}(e^{-\alpha t}-e^{-\beta t})dt}{\int_0^\infty \frac{k_{12}k_{23}}{\beta-\alpha}(e^{-\alpha t}-e^{-\beta t})dt} = \frac{1/\alpha^2-1/\beta^2}{1/\alpha-1/\beta} = \frac{1}{\alpha}+\frac{1}{\beta}.$$

We shall start analyzing this second result; the transfer time from compartment 1 to compartment 3 is $1/\alpha + 1/\beta$; we can write

$$\frac{1}{\alpha}+\frac{1}{\beta} = \frac{\alpha+\beta}{\alpha\cdot\beta};$$

then, observing the characteristic equation (10) and remembering that $-\alpha$ and $-\beta$ are its roots,

$$\frac{1}{\alpha}+\frac{1}{\beta} = \frac{K_2+K_3}{K_2K_3-k_{23}k_{32}};$$

we can divide numerator and denominator of the right-hand side fraction by K_2K_3,

$$\frac{1}{\alpha}+\frac{1}{\beta} = \left(\frac{1}{K_2}+\frac{1}{K_3}\right)\frac{1}{1-\frac{k_{23}}{K_2}\frac{k_{32}}{K_3}}.$$

Observe that k_{ij}/K_j is the fraction of substance leaving compartment i that is transferred to compartment j; therefore the product $\frac{k_{23}}{K_2}\cdot\frac{k_{32}}{K_3}$ is the fraction of substance recirculated through compartments 2 and 3. We shall call this product r_{23}; this product is certainly smaller that one, therefore we can write

$$\frac{1}{1-\frac{k_{23}}{K_2}\frac{k_{32}}{K_3}} = 1+r_{23}+r_{23}^2+r_{23}^3+\cdots.$$

The transfer time from compartment 1 to compartment 3 can now be written,

$$\left(\frac{1}{K_2}+\frac{1}{K_3}\right)\left(1+r_{23}+r_{23}^2+r_{23}^3+\cdots\right);$$

this expression shows that, after leaving compartment 1, the substance spends the time $1/K_2$ in the second compartment and the time $1/K_3$ in the third one, plus the same time multiplied by r_{23} for the fraction recirculated once, plus the same time multiplied by r_{23}^2 for the fraction recirculated twice, and so forth.

For the transfer time from compartment 1 to compartment 2 we found above the value

$$\frac{1}{\alpha}+\frac{1}{\beta}-\frac{1}{K_3};$$

the term $1/K_3$ is subtracted from the time spent in both compartments 2 and 3 because the time spent by the substance in its first passage through compartment 3 does not accrue toward the time of exit from compartment 2. A more detailed analysis of the transfer time will be in Chapter 16.

10.7. OPERATIONAL FORM OF THE TRANSFER TIME

From section 9.7 we know that the first moment of $f(t)$ is

$$\left\{\int_0^\infty t\cdot f(t)\mathrm{d}t\right\}=-\lim_{s\to 0}\frac{\mathrm{d}\{f\}}{\mathrm{d}s}$$

and its 0-moment is

$$\left\{\int_0^\infty f(t)\mathrm{d}t\right\}=\lim_{s\to 0}\frac{\mathrm{d}\{f\}}{\mathrm{d}s},$$

therefore the first relative moment of $f(t)$ is

$$f_1=\frac{\int_0^\infty t\cdot f(t)\mathrm{d}t}{\int_0^\infty f(t)\mathrm{d}t}=-\lim_{s\to 0}\frac{\mathrm{d}\ln\{f\}}{\mathrm{d}s}.$$

It follows that the first relative moment of the transfer function $g_{12}(t)$ is

$$\begin{aligned}\frac{\int_0^\infty t\cdot g_{12}(t)\mathrm{d}t}{\int_0^\infty g_{12}(t)\mathrm{d}t}&=-\lim_{s\to 0}\frac{\mathrm{d}\ln\{x_2\}}{\mathrm{d}s}+\lim_{s\to 0}\frac{\mathrm{d}\ln\{x_1\}}{\mathrm{d}s}\\&=\lim_{s\to 0}\frac{\mathrm{d}}{\mathrm{d}s}\ln\frac{\{x_1\}}{\{x_2\}}.\end{aligned}\qquad(18)$$

10.8. EXAMPLES

10.8.1. From Model to Transfer Function

Consider the model described by the differential equations

$$\frac{dx_1}{dt} = -K_1 x_1(t) + k_{21} x_2(t)$$

$$\frac{dx_2}{dt} = +k_{12} x_1(t) - K_2 x_2(t)$$

$$\frac{dx_3}{dt} = +k_{23} x_2(t) - K_3 x_3(t)$$

and the initial conditions

$$x_1(0) = x_0, \quad x_2(0) = x_3(0) = 0.$$

As shown in section 8.1 to this model corresponds the linear graph of Fig. 3.

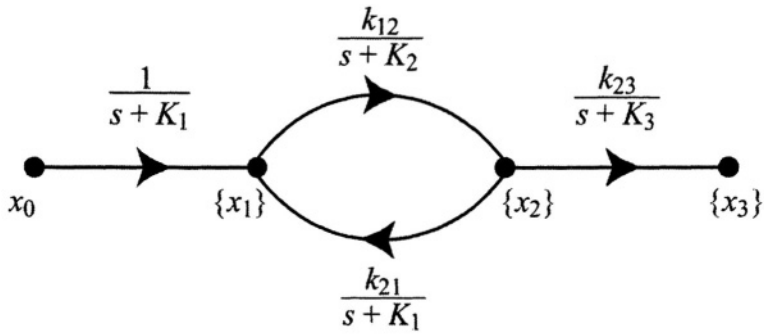

Fig. 3. Linear graph of the model in section 10.8.1

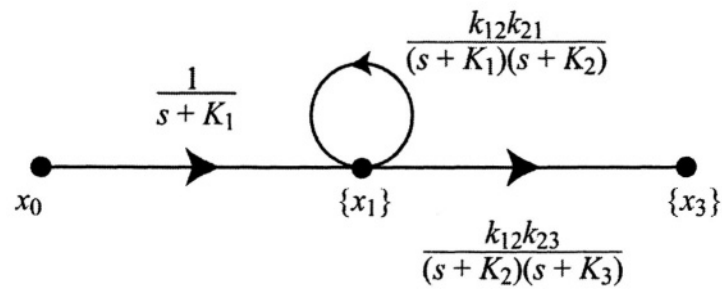

Fig. 4. Essential graph of the model of Fig. 3

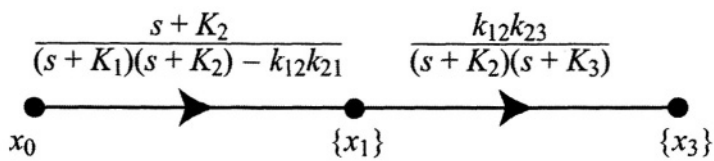

Fig. 5. The closed arm removed from the graph of Fig. 4

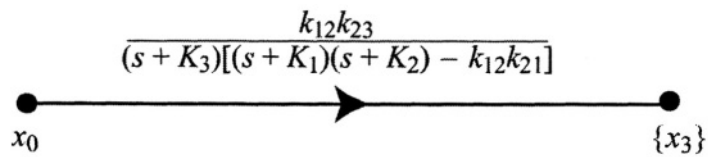

Fig. 6. A linear graph with just one arm

With the rules illustrated in section 8.2 the graph of Fig. 3 can be transformed into the essential graph of Fig. 4 containing the initial node x_0, the essential node $\{x_1\}$, and the terminal node $\{x_3\}$. In Fig. 5 the closed arm has been removed, and finally in Fig. 6 it is transformed into a graph with just one arm, equal to the transfer function from x_0 to $\{x_3\}$.

Operating in the same way we can find the transfer functions from x_0 to $\{x_1\}$ and to $\{x_2\}$; the results are

$$\{f\} = \frac{\{x_1\}}{x_0} = \frac{s + K_2}{(s + K_1)(s + K_2) - k_{12}k_{21}},$$

$$\{g\} = \frac{\{x_2\}}{x_0} = \frac{k_{12}}{(s + K_1)(s + K_2) - k_{12}k_{21}},$$

$$\{h\} = \frac{\{x_3\}}{x_0} = \frac{k_{12}k_{23}}{(s + K_3)\left[(s + K_1)(s + K_2) - k_{12}k_{21}\right]}.$$

Of the transfer function $\{f\}$ we can compute different moments as shown in section 9.7; they are

$$F_0 = \lim_{s \to 0}\{f\} = \frac{K_2}{K_1K_2 - k_{12}k_{21}} = \frac{1}{K_1} \cdot \frac{1}{1 - \dfrac{k_{12}}{K_1}\dfrac{k_{21}}{K_2}},$$

$$F_1 = -\lim_{s \to 0}\frac{d\{f\}}{ds} = \frac{K_2^2 + k_{12}k_{21}}{\left(K_1K_2 - k_{12}k_{21}\right)^2},$$

$$f_1 = \frac{F_1}{F_0} = \frac{K_2 + k_{12}k_{21}/K_2}{K_1K_2 - k_{12}k_{21}} = \frac{1}{K_1} \cdot \frac{1}{1 - \dfrac{k_{12}}{K_1} \cdot \dfrac{k_{21}}{K_2}} + \frac{1}{K_2} \cdot \frac{\dfrac{k_{12}}{K_1} \cdot \dfrac{k_{21}}{K_2}}{1 - \dfrac{k_{12}}{K_1} \cdot \dfrac{k_{21}}{K_2}}.$$

The moments of negative order can be computed with the formulas of section 9.7 or directly from the differential equations; they are

$$F_{-1} = \lim_{s \to \infty} s\{f\} = 1$$

$$F_{-2} = -\lim_{s \to \infty} s[s\{f\} - 1] = K_1$$

$$F_{-3} = \lim_{s \to \infty} s[s\{f'\} - K_1] = K_1^2 + k_{12}k_{21}$$

Proceeding the same way with the transfer function $\{g\}$ we find

$$G_0 = \lim_{s \to 0}\{g\} = \frac{k_{12}}{K_1K_2 - k_{12}k_{21}} = \frac{k_{12}}{K_1} \cdot \frac{1}{K_2} \cdot \frac{1}{1 - \dfrac{k_{12}}{K_1}\dfrac{k_{21}}{K_2}},$$

$$G_1 = -\lim_{s\to 0}\frac{d\{g\}}{ds} = \frac{k_{12}(K_1+K_2)}{\left(K_1K_2 - k_{12}k_{21}\right)^2},$$

$$g_1 = \frac{G_1}{G_0} = \frac{K_1+K_2}{K_1K_2 - k_{12}k_{21}} = \frac{1}{K_1}\cdot\frac{1}{1-\frac{k_{12}}{K_1}\cdot\frac{k_{21}}{K_2}} + \frac{1}{K_2}\cdot\frac{1}{1-\frac{k_{12}}{K_1}\cdot\frac{k_{21}}{K_2}}.$$

The moments of negative order can be computed with the formulas of section 9.7 or directly from the differential equations; they are

$$G_{-1} = \lim_{s\to\infty} s\{g\} = 0,$$

$$G_{-2} = -\lim_{s\to\infty} s[s\{0\}] = k_{12}.$$

$$G_{-3} = \lim_{s\to\infty} s[s\{g'\} - k_{12}] = k_{12}(K_1+K_2).$$

For the transfer function $\{h\}$ we find

$$H_0 = \lim_{s\to 0}\{h\} = \frac{k_{12}k_{23}}{K_3(K_1K_2 - k_{12}k_{21})} = \frac{k_{12}}{K_1}\cdot\frac{k_{23}}{K_2}\cdot\frac{1}{K_3}\cdot\frac{1}{1-\frac{k_{12}}{K_1}\frac{k_{21}}{K_2}},$$

$$H_1 = -\lim_{s\to 0}\frac{d\{h\}}{ds} = \frac{k_{12}k_{23}(K_1K_2 + K_1K_3 + K_2K_3 - k_{12}k_{21})}{K_3^2\left(K_1K_2 - k_{12}k_{21}\right)^2},$$

$$h_1 = \frac{H_1}{H_0} = \frac{K_1K_2 + K_1K_3 + K_2K_3 - k_{12}k_{21}}{K_3(K_1K_2 - k_{12}k_{21})} = \left(\frac{1}{K_1}+\frac{1}{K_2}\right)\frac{1}{1-\frac{k_{12}}{K_1}\frac{k_{21}}{K_2}} + \frac{1}{K_3}.$$

The moments of negative order can be computed with the formulas of section 9.7 or directly from the differential equations; they are

$$H_{-1} = \lim_{s\to\infty} s\{h\} = 0,$$

$$H_{-2} = -\lim_{s\to\infty} s[s\{h\}] = 0,$$

$$H_{-3} = \lim_{s\to\infty} s[s\{h'\}] = k_{12}k_{23}.$$

The 0-moment of the transfer function $f(t)$ is equal to the permanence time of the first compartment, i.e., the turnover time $1/K_1$ of that compartment times its turnover number, as shown in section 9.3. The first relative moment of the transfer function $f(t)$ is the exit time from the first compartment, as shown in section 9.4; it is equal to the sum of two terms, the time spent in compartment 1 (equal to its permanence time), plus the time spent in compartment 2 before the exit from compartment 1; this last time is the permanence time of compartment 2 times the fraction

$$\frac{k_{12}}{K_1}\cdot\frac{k_{21}}{K_2}$$

that is actually recyrculated between the two compartments.

The 0-moment of the transfer function $g(t)$ is equal to the *residence time* of the second compartment, i.e., its permanence time

$$\frac{1}{K_2}\cdot\frac{1}{1-\dfrac{k_{12}}{K_1}\cdot\dfrac{k_{21}}{K_2}}$$

times the fraction k_{12}/K_1 that reaches compartment 2 from compartment 1 (see chapter 16). The first relative moment of the transferfunction $g(t)$ is the exit time from the second compartment, as shown in section 9.4; it is equal to the sum of two terms, the time spent in compartment 1 plus the time spent in compartment 2, both equal to their respective permanence times.

The 0-moment of the transfer function $h(t)$ is equal to the residence time of the third compartment, i.e., to its permanence time times the fraction

$$\frac{k_{12}}{K_1}\cdot\frac{k_{23}}{K_2}$$

that reaches compartment 3 from compartment 1. The first relative moment of the transfer function $h(t)$ is the exit time from the third compartment, as shown in section 9.4; it is equal to the sum of three terms, the time spent in compartment 1 plus the time spent in compartment 2 (both equal to their permanence times), plus the turnover time of compartment 3, where there is no recyrculation.

10.8.2. From Transfer Function to Model

If in an actual experiment the moments f_1, F_0, F_{-1}, F_{-2}, F_{-3}, have been measured, what conclusions can be drawn about the model?

There are 4 parameters, K_1, K_2, k_{12}, k_{21}, to be determined by the four equations

$$\frac{\dfrac{1}{K_2}+\dfrac{1}{K_1}\cdot\dfrac{k_{12}}{K_1}\dfrac{k_{21}}{K_2}}{1-\dfrac{k_{12}}{K_1}\dfrac{k_{21}}{K_2}}=f_1,$$

$$\frac{1}{K_1}\cdot\frac{1}{1-\dfrac{k_{12}}{K_1}\dfrac{k_{21}}{K_2}}=F_0,$$

$$K_1=F_{-2},$$

$$K_1^2+k_{12}k_{21}=F_{-3}.$$

The fourth equation is redundant; from the third equation we determine K_1; from the first and the second, the product $k_{12}k_{21}$ and K_2:

$$K_1 = F_{-2},$$

$$K_2 = \frac{F_0F_{-2} - 1}{f_1 - F_0},$$

$$k_{12}k_{21} = \frac{(F_0F_{-2} - 1)^2}{F_0(f_1 - F_0)}.$$

With the data available we cannot determine the separate values of k_{12} and k_{21}, but we can determine a range for them; in fact we know that

$$0 \le k_{12} \le K_1, \quad 0 \le k_{21} \le K_2$$

and that to the maximum value or k_{12} corresponds the minimum value of k_{21} and vice versa; therefore

$$k_{12} \le \frac{(F_0F_{-2} - 1)^2}{F_0(f_1 - F_0)} \cdot \frac{f_1 - F_0}{F_0F_{-2} - 1} = F_{-2} - \frac{1}{F_0},$$

$$k_{21} \le \frac{(F_0F_{-2} - 1)^2}{F_0(f_1 - F_0)} \cdot \frac{1}{F_{-2}} = \frac{(F_0F_{-2} - 1)^2}{F_0F_{-2}(f_1 - F_0)}.$$

If, in addition to the moments of $f(t)$, the 0-moment of $g(t)$ were determined, we could write

$$\frac{G_0}{F_0} = \frac{k_{12}}{K_2}$$

and compute the exact value of k_{12}. Alternatively, this last value could be obtained from the –2-moment G_{-2}.

Observe that no information on k_{23} and on K_3 can be obtained from the transfer functions $f(t)$ and $g(t)$ and from their moments. We can determine the parameter K_3 from

$$h_1 - g_1 = \frac{1}{K_3}$$

and the parameter k_{23} from

$$\frac{H_0}{G_0} = \frac{k_{23}}{K_3}.$$

10.9. REFERENCES

1. V. Volterra, *Theory of functionals and of integral and integro-differential equations* (Blaskie, London and Glasgow, 1930).
2. A. Rescigno, Synthesis of a Multicompartmented Biological Model, *Biochim. Biophys. Acta* **37**, 463-8 (1960).
3. A. Rescigno and G, Segre, Analysis of Multicompartmented Biological Systems. *J. Theoret. Biol.* **3**, 149-63 (1962).
4. A. Rescigno, Flow Diagrams of Multi-Compartment Systems. *Ann. N. Y. Acad. Sci.* **108,** 204-16 (1963).
5. J. S. Beck and A. Rescigno, Determination of Precursor Order and Particular Weighting Functions from Kinetic Data, *J. theoret. Biol.* **6**, 1-12 (1964).

This page intentionally left blank

11. MATRIX EQUATIONS

11.1. DEFINITIONS

Equations (10) from chapter 5 can be written in matrix form,

$$\left(\frac{dx_1}{dt} \quad \frac{dx_2}{dt} \quad \cdots \quad \frac{dx_n}{dt}\right) = -\left(x_1(t) \quad x_2(t) \quad \cdots \quad x_n(t)\right) \cdot \begin{pmatrix} K_1 & -k_{12} & \cdots & -k_{1n} \\ -k_{21} & K_2 & \cdots & -k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -k_{n1} & -k_{n2} & \cdots & K_n \end{pmatrix}$$

or with a more synthetic notation,

$$\frac{d\boldsymbol{X}}{dt} = -\boldsymbol{X}(t) \cdot \boldsymbol{K}. \tag{1}$$

Observe that for the transfer rate *from* i *to* j I use the symbol k_{ij}, while some authors use the symbol k_{ji}; the reason for my choice is fourfold:

i) The physical meaning of k_{ij} becomes more evident if, while reading the subscripts, one thinks of the transfer of the drug from one compartment to another as happening in the same direction as the letters are read in English and most other European languages;

ii) If the drug is transferred through a succession of compartments, the product of the transfer constants involved, written as a string with the second subscript of a constant equal to the first subscript of the following constant, has a particularly useful physical and mathematical meaning [1];

iii) This definition is consistent with the notation used in the theory of Markov processes, as shown by Thakur, Rescigno and Shafer [2, 3]

iv) This definition conforms with the standards adopted by the Journal of Pharmacokinetics and Biopharmaceutics [4] and by the European Journal of Clinical Pharmacology [5].

On the other hand, the alternative notation was used because, when in equations (1) the matrix $\boldsymbol{K}$ is post-multiplied by the column vector $\boldsymbol{X}$, then the element of $\boldsymbol{K}$ in the i^{th} row and j^{th} column is the transfer rate *to* i *from* j. In my opinion there is no reason to prefer post-multiplying $\boldsymbol{K}$ by a column vector $\boldsymbol{X}$ rather than pre-multiplying it by a row vector.

Physical realizability of the system requires that

$$\begin{aligned} k_{ij} &\geq 0 && \text{for any } i, j \\ K_i &\geq \sum_{i \neq j} k_{ij} && \text{for any } i \end{aligned} \tag{2}$$

as seen in chapter 5.

Most of the following definitions are due to Hearon. [6]

A matrix (a_{ij}) is called *diagonal dominant* if either

$$(a_{ii}) \geq \sum_{j \neq i} (a_{ij}) \quad i = 1, 2, \ldots, n$$

or

$$(a_{ii}) \geq \sum_{j \neq i} (a_{ji}) \quad i = 1, 2, \ldots, n;$$

obviously conditions (2) make matrix $\boldsymbol{K}$ diagonal dominant. The determinant of such a matrix has been called *unisignant* by Muir, [7] and it has been shown to be non-negative.

Matrix $\boldsymbol{K}$ is said to be *decomposable* if, with a number of permutations of its rows and corresponding columns, it can be put in the quasi-diagonal form

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{K}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K}_2 \end{pmatrix},$$

where $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are two square matrices, and 0 are null matrices of appropriate size. When this is the case, if $n \times n$ is the size of $\boldsymbol{K}$ and $m \times m$ the size of $\boldsymbol{K}_1$, then the n compartments can be partitioned in two different systems, one of m and the other of $n - m$ compartments, completely independent of each other. A system whose matrix is not decomposable is said to be *connected.* Unless otherwise explicitly stated, I shall only consider non-decomposable matrices.

Matrix $\boldsymbol{K}$ is said to be *reducible* if, with an appropriate permutation of its rows and corresponding columns, it can be put in the form

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{K}_1 & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{K}_2 \end{pmatrix}, \tag{3}$$

where $\boldsymbol{K}_1$ is $m \times m$, $\boldsymbol{K}_2$ is $[n - m] \times [n - m]$, and $\mathbf{B}$ is $m \times [n - m]$, but different from $\boldsymbol{0}$. When this is the case, the n compartments can be relabeled in such a way that the first m of them are independent from the remaining $n - m$. A system whose matrix is not reducible is said to be *strongly connected.*

Observe that if a matrix is decomposable, its corresponding linear graph is not connected. To a connected matrix corresponds a connected linear graph. To a strongly connected matrix corresponds a strong graph.

11.2. PROPERTIES OF MATRIX *K*

11.2.1. Theorem 1

If $\boldsymbol{K}$ is irreducible, then it is singular if and only if the second of conditions (2) is a strict equality, i.e., if

$$K_i = \sum_{i \neq j}^{1 \cdots n} k_{ij} \quad \text{for any } i; \tag{4}$$

when this is the case, the system is closed, i.e., the drug can move from one compartment to another in many different ways, but it never leaves the system.

The first part of this theorem is obvious; in fact, if conditions (4) hold, then the sum of the elements of each row is zero, and the matrix is singular.

To prove the second part of the theorem, i.e., to show that matrix $\boldsymbol{K}$ is singular only if conditions (4) hold, consider a matrix $\boldsymbol{K}$ of size n, singular by hypothesis; this means that there is a linear relationship among the elements of its columns, i.e., we can find a set of numbers $a_1, a_2, \ldots, a_n$, not all zero, such that

$$\begin{aligned} &+a_1K_1 - a_2k_{21} - a_3k_{31} - \cdots - a_nk_{n1} = 0, \\ &-a_1k_{12} + a_2K_2 - a_3k_{32} - \cdots - a_nk_{n2} = 0, \\ &\cdots \\ &-a_1k_{1n} - a_2k_{2n} - a_3k_{3n} - \cdots + a_nK_n = 0. \end{aligned} \tag{5}$$

By adding all the above expressions we get

$$a_1\left(K_1 - \sum_{i \neq 1} k_{1i}\right) + a_2\left(K_2 - \sum_{i \neq 2} k_{2i}\right) + \cdots + a_n\left(K_n - \sum_{i \neq n} k_{ni}\right) = 0; \tag{6}$$

all terms in parenthesis are non-negative by hypothesis, therefore the above sum is zero only if some of the a's are negative; now we write expression (6) in the form

$$a_l\left(K_l - \sum_{i \neq l} k_{li}\right) + \cdots = -a_m\left(K_m - \sum_{i \neq m} k_{mi}\right) - \cdots, \tag{7}$$

where on the left-hand side we put all terms with a positive a coefficient, and on the right-hand side those with a negative a coefficient; the terms with an a coefficient equal to zero of course can be ignored. Now from expressions (5) choose only those containing a $\boldsymbol{K}$ appearing on the left-hand side of expression (7); add them together and reorder the result as in expression (7). On the right-hand side the terms in parenthesis are all negative because all $\boldsymbol{K}$'s have disappeared, while on the left-hand side all terms in parenthesis are still positive; because a negative sum cannot be equal to a positive sum, we conclude that all terms in parenthesis are zero, q.e.d.

11.2.2. Theorem 2

If $\boldsymbol{K}$ is irreducible and singular, then 0 is a simple eigenvalue of $\boldsymbol{K}$; in other words, if n is the order of the irreducible matrix $\boldsymbol{K}$, its rank cannot be less than $n - 1$.

If $\boldsymbol{K}$ is singular, for theorem 1 we can write expression (4) for each of its rows. If its rank is less than $n - 1$, any of the principal minors of $\boldsymbol{K}$ is singular; consider for instance the principal minor obtained by suppressing row 1 and column 1 from $\boldsymbol{K}$; for all of its rows we can write the corresponding expression (4), i.e.,

$$K_i = \sum_{i \neq j}^{2 \cdots n} k_{ij}, \quad i = 2, 3, \ldots, n; \tag{8}$$

subtracting (8) from (4) we get

$$k_{i1} = 0, \quad i = 2, 3, \ldots, n;$$

this means that $\boldsymbol{K}$ is reducible, against the hypothesis.

11.2.3. Theorem 3

If $\boldsymbol{K}$ is reducible and singular, it can be written in form

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{K}_1 & \boldsymbol{B}_{12} & \boldsymbol{B}_{13} & \cdots & \boldsymbol{B}_{1m} \\ \boldsymbol{0} & \boldsymbol{K}_2 & \boldsymbol{B}_{23} & \cdots & \boldsymbol{B}_{2m} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{K}_3 & \cdots & \boldsymbol{B}_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{K}_m \end{pmatrix},$$

where $\boldsymbol{K}_1, \boldsymbol{K}_2, \ldots, \boldsymbol{K}_m$ are all irreducible matrices, and only matrix $\boldsymbol{K}_m$ is singular.

In fact if $\boldsymbol{K}$ is reducible, we can put it in the form (3); if either $\boldsymbol{K}_1$ or $\boldsymbol{K}_2$, or both, are further reducible, we can transform them in the same way and proceed until $\boldsymbol{K}$ has the above form, where $\boldsymbol{K}_1, \boldsymbol{K}_2, \ldots, \boldsymbol{K}_m$ are all irreducible matrices. If any of these last matrices, say $\boldsymbol{K}_i$, is singular, then 0 is a simple eigenvalue of it, and the corresponding subsystem is closed; this implies that all matrices on the same row as $\boldsymbol{K}_i$ are zero, and the system is decomposable. It follows that if $\boldsymbol{K}$ is not decomposable, only matrix $\boldsymbol{K}_m$ can be singular.

11.3. INTEGRATION OF THE MATRIX EQUATION

11.3.1. All eigenvalues of *K* are real and separate

To integrate equation (1) we need to specify its initial conditions, i.e. the vector $\boldsymbol{X}(0)$; we can make equation (1) even more general if we allow the administration of the drug after the initial time, namely if we include a vector

$$\boldsymbol{R}(t) = \begin{pmatrix} r_1(t) & r_2(t) & \cdots & r_n(t) \end{pmatrix},$$

where $r_i(t)$ represents the feeding function into compartment i; equation (1) thus becomes

$$\frac{d\boldsymbol{X}}{dt} = -\boldsymbol{X}(t) \cdot \boldsymbol{K} + \boldsymbol{R}(t); \tag{9}$$

its integral is

$$\boldsymbol{X}(t) = \left(\boldsymbol{X}(0) + \int_0^t \boldsymbol{R}(\tau) e^{\tau \boldsymbol{K}} \mathrm{d}\tau\right) \cdot e^{-t\boldsymbol{K}}, \tag{10}$$

where by definition (see Appendix D)

$$e^{t\boldsymbol{K}} = \boldsymbol{I} + t\boldsymbol{K} + \frac{t^2}{2!}\boldsymbol{K}^2 + \frac{t^3}{3!}\boldsymbol{K}^3 + \cdots,$$

$\boldsymbol{I}$ being the $n \times n$ identity matrix. The properties of the exponential matrix $e^{t\boldsymbol{K}}$ depend on the eigenvalues of $\boldsymbol{K}$, therefore we shall spend the next few pages on the analysis of these eigenvalues.

If $\boldsymbol{\Lambda}$ is the diagonal matrix formed by the eigenvalues of $\boldsymbol{K}$, and $\boldsymbol{P}$ is the matrix of the eigenvectors of $\boldsymbol{K}$, then

$$\boldsymbol{K} = \boldsymbol{P} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{P}^{-1},$$

therefore expression (10) becomes

$$\boldsymbol{X}(t) = \left(\boldsymbol{X}(0) + \int_0^t \boldsymbol{R}(\tau) \boldsymbol{P} e^{\tau \boldsymbol{\Lambda}} \boldsymbol{P}^{-1} \mathrm{d}\tau\right) \cdot \boldsymbol{P} e^{-t\boldsymbol{\Lambda}} \boldsymbol{P}^{-1}.$$

This expression can be solved only when $\boldsymbol{R}(t)$ has some simple form. For instance if $\boldsymbol{R}(t) = 0$,

$$\boldsymbol{X}(t) = \boldsymbol{X}(0) \cdot \boldsymbol{P} e^{-t\boldsymbol{\Lambda}} \boldsymbol{P}^{-1}.$$

If the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ in $\boldsymbol{\Lambda}$ are all real and different, then

$$e^{-t\boldsymbol{\Lambda}} = \begin{pmatrix} e^{-\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{-\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{-\lambda_n t} \end{pmatrix},$$

therefore

$$\boldsymbol{X}(t) = \left(\sum_{i=1}^{n} a_{1i} e^{-\lambda_i t} \quad \sum_{i=1}^{n} a_{2i} e^{-\lambda_i t} \quad \cdots \quad \sum_{i=1}^{n} a_{ni} e^{-\lambda_i t}\right),$$

where the coefficients a_{ij} are functions of the transfer constants and of the initial conditions.

11.3.2. Some eigenvalues of K are multiple

If the eigenvalues are not all different, we can transform $\boldsymbol{K}$ into a quasi-diagonal matrix, i.e., into a matrix of the form

$$\boldsymbol{P}\cdot\boldsymbol{K}\cdot\boldsymbol{P}^{-1}=\begin{pmatrix}\lambda_1 & 1 & 0 & \cdots & 0\\ 0 & \lambda_1 & 0 & \cdots & 0\\ 0 & 0 & \lambda_3 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & \lambda_n\end{pmatrix},$$

where λ_1 is a double eigenvalue, λ_3 is a simple one, and so forth (see section D.8).

Observe now that to the Jordan submatrix

$$\begin{pmatrix}\lambda_1 & 1\\ 0 & \lambda_1\end{pmatrix}$$

corresponds the differential equation (in matrix form)

$$\left(\frac{dx_1}{dt}\quad\frac{dx_2}{dt}\right)=-\left(x_1(t)\quad x_2(t)\right)\cdot\begin{pmatrix}\lambda_1 & 1\\ 0 & \lambda_1\end{pmatrix}$$

or, in scalar form, the two equations

$$\frac{dx_1}{dt}=-\lambda_1 x_1(t),$$
$$\frac{dx_2}{dt}=-x_1(t)-\lambda_1 x_2(t).$$

The first of these two equations can be integrated into

$$x_1(t)=x_1(0)e^{-\lambda_1 t};$$

substituting the value of $x_1(t)$ into the second one we get

$$\frac{dx_2}{dt}+\lambda_1 x_2(t)=-x_1(0)e^{-\lambda_1 t}$$

whose integral is

$$x_2(t)=[x_2(0)-x_1(0)\cdot t]\cdot e^{-\lambda_1 t}.$$

If we put these results back in matrix form we can write

$$\left(x_1(t)\quad x_2(t)\right)=\left(x_1(0)\quad x_2(0)\right)\cdot\begin{pmatrix}e^{-\lambda_1 t} & -te^{-\lambda_1 t}\\ 0 & e^{-\lambda_1 t}\end{pmatrix}.$$

For the general case with eigenvalues of multiplicity three or more, see Rescigno et al. [8].

11.3.3. Some eigenvalues of *K* are complex

If some of the eigenvalues are complex, we can transform $\boldsymbol{K}$ into a quasi-diagonal matrix, i.e., into a matrix of the form

$$\boldsymbol{P}\cdot\boldsymbol{K}\cdot\boldsymbol{P}^{-1}=\begin{pmatrix}\lambda_1 & \omega & 0 & \cdots & 0\\ -\omega & \lambda_1 & 0 & \cdots & 0\\ 0 & 0 & \lambda_3 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & \lambda_n\end{pmatrix},$$

where λ_1 is the real part of a complex eigenvalue and ω its imaginary part, while λ_3, ..., λ_n, are the real eigenvalues (see section D.9).

Observe now that to the submatrix

$$\begin{pmatrix}\lambda_1 & \omega\\ -\omega & \lambda_1\end{pmatrix}$$

corresponds the differential equation (in matrix form)

$$\left(\frac{dx_1}{dt}\quad\frac{dx_2}{dt}\right)=-\begin{pmatrix}x_1(t) & x_2(t)\end{pmatrix}\cdot\begin{pmatrix}\lambda_1 & \omega\\ -\omega & \lambda_1\end{pmatrix}$$

or, in scalar form, the two differential equations

$$\frac{dx_1}{dt}=-\lambda_1 x_1(t)+\omega x_2(t),$$
$$\frac{dx_2}{dt}=-\omega x_1(t)-\lambda_1 x_2(t).$$

These two equations can be integrated into

$$x_1(t)=e^{-\lambda_1 t}\left(x_1(0)\cdot\cos\omega t+x_2(0)\sin\omega t\right),$$
$$x_2(t)=e^{-\lambda_1 t}\left(x_2(0)\cdot\cos\omega t-x_1(0)\sin\omega t\right).$$

If we put these results back in matrix form we can write

$$\begin{pmatrix}x_1(t) & x_2(t)\end{pmatrix}=\begin{pmatrix}x_1(0) & x_2(0)\end{pmatrix}\cdot\begin{pmatrix}e^{-\lambda_1 t}\cos\omega t & -e^{-\lambda_1 t}\sin\omega t\\ e^{-\lambda_1 t}\sin\omega t & e^{-\lambda_1 t}\cos\omega t\end{pmatrix}.$$

I leave to the interested reader the solution of the case of multiple eigenvalues.

11.4. MATRIX *K* AND ITS POWERS

From equation (1) by successive differentiations we get

$$\frac{d^p\boldsymbol{X}}{dt^p}=-\frac{d^{p-1}\boldsymbol{X}}{dt^{p-1}}\cdot\boldsymbol{K},\quad p=1,2,\ldots$$

thence, by induction,

$$\frac{d^p \boldsymbol{X}}{dt^p} = (-1)^p \boldsymbol{X}(t) \cdot \boldsymbol{K}^p$$

and

$$\frac{d^p \boldsymbol{X}(0)}{dt^p} = (-1)^p \boldsymbol{X}(0) \cdot \boldsymbol{K}^p \tag{11}$$

for any non-negative integer p.

To understand the meaning of identity (11) think of an experiment where the drug is given only to compartment i as a bolus; in that case vector $\boldsymbol{X}(0)$ has all its elements equal to 0 except the one in position i equal to $x_i(0)$. The product on the right-hand side of (11), except for the sign, is equal to $x_i(0)$ times row i of matrix $\boldsymbol{K}^p$.

Call $k_{ij}^{(p)}$ the element of row i and column j of matrix $\boldsymbol{K}^p$; call $x_{ij}(t)$ the amount of drug present in compartment j that was given to compartment i at time $t = 0$, and $x_{ij}^{(p)}(t)$ its p^{th} time derivative. From identity (11) we get

$$k_{ij}^{(p)} = (-1)^p \frac{x_{ij}^{(p)}(0)}{x_i(0)}. \tag{12}$$

In the special case $p = 1$ we have

$$k_{ij} = -\frac{dx_{ij}(0)/dt}{x_i(0)},$$

another formulation of the transfer rate as defined in section 5.4. I will describe in chapter 16 with more details the general properties of the elements $k_{ij}^{(p)}$ of $\boldsymbol{K}^p$.

11.5. INVERSION OF MATRIX *K*

We return now to equation (1); if the system is open, $\boldsymbol{K}$ is not singular, therefore it has an inverse (see section D.6); we call it $\boldsymbol{T}$. Multiply both members of equation (1) to the right by $\boldsymbol{T}$:

$$\frac{d\boldsymbol{X}}{dt} \cdot \boldsymbol{T} = -\boldsymbol{X}(t); \tag{13}$$

integrate from 0 to t:

$$\int_0^t \frac{d\boldsymbol{X}}{dt} d\tau \cdot \boldsymbol{T} = -\int_0^t \boldsymbol{X}(\tau) d\tau$$

and

$$\left(\boldsymbol{X}(t) - \boldsymbol{X}(0)\right) \cdot \boldsymbol{T} = -\int_0^t \boldsymbol{X}(\tau) d\tau .$$

The system is open by hypothesis, therefore $\boldsymbol{X}(\infty) = \boldsymbol{0}$ and

$$\boldsymbol{X}(0)\cdot\boldsymbol{T}=\int_0^\infty \boldsymbol{X}(\tau)\mathrm{d}\tau\,. \tag{14}$$

All elements of $\boldsymbol{T}$ are non-negative. To prove it, observe that, as it was shown in section 11.1, $\boldsymbol{K}$ is diagonal dominant, therefore its determinant is non-negative; but $\boldsymbol{K}$ is also non-singular, therefore its determinant is strictly positive. The elements of $\boldsymbol{T}$ are (see section D.4),

$$t_{ij}=\frac{K_{ji}}{\det(\boldsymbol{K})}, \tag{15}$$

where K_{ji} is the cofactor of the element $-k_{ji}$ of $\det(\boldsymbol{K})$. This cofactor, when $i=j$, is non-negative because all principal minors of $\boldsymbol{K}$ are also diagonal dominant. If $i\neq j$, we can use expression (1) of section D.2 to transform it into a sum of products of elements of $\boldsymbol{K}$ times cofactors of $\boldsymbol{K}$ of order $n-2$, one of them a principal cofactor, therefore non-negative. We can continue this transformation until the numerator of expression (15) contains only principal cofactors of $\boldsymbol{K}$ of all orders from $n-1$ to 1. Each of those cofactors is multiplied by a string of non-diagonal elements of $\boldsymbol{K}$, all non-positive. But if one of those terms contains an even numbers of elements of $\boldsymbol{K}$, the sign of the corresponding cofactor is positive, if it contains an odd numbers of elements of $\boldsymbol{K}$, the sign of the corresponding cofactor is negative; because all non-diagonal elements of $\boldsymbol{K}$ are non-positive, all terms from expression (15) are non-negative, q.e.d.

We can now say that all elements of $\boldsymbol{T}$ are strictly positive if and only if $\boldsymbol{K}$ is irreducible. In fact if $\boldsymbol{K}$ is reducible, there is at least a value of j and a value of i such that the cofactor K_{ji} is zero, and the corresponding element t_{ij} of $\boldsymbol{T}$ is zero. On the other hand suppose that an element of $\boldsymbol{T}$ is zero; if it is an element of the principal diagonal, say t_{ii}, then the cofactor K_{ii} is zero and its corresponding submatrix is singular; the sum of the elements of each of its rows is zero, and consequently the elements of the same rows and the missing column in this submatrix are zero, and $\boldsymbol{K}$ is reducible. If the null element of $\boldsymbol{T}$ is not a diagonal one, say t_{ij}, then the cofactor K_{ji} is zero; we can transform it into a sum of products of elements of $\boldsymbol{K}$ times principal cofactors of $\boldsymbol{K}$ of all orders from $n-1$ to 1, as done in the previous paragraph; all those terms are non-negative; their sum is zero, therefore each of those terms is zero, and matrix $\boldsymbol{K}$ is reducible, q.e.d.

11.6. MATRIX *T* AND ITS POWERS

Back to equation (13), by successive integrations we get

$$\int_0^t \frac{t^p}{p!}\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t}\mathrm{d}t\cdot\boldsymbol{T}=-\int_0^t \frac{t^p}{p!}\boldsymbol{X}(t)\mathrm{d}t,\quad p=1,2,\ldots$$

and, with an integration by parts,

$$\left(\frac{t^p}{p!}\boldsymbol{X}(t)-\int_0^t \frac{t^{p-1}}{(p-1)!}\boldsymbol{X}(t)\mathrm{d}t\right)\cdot\boldsymbol{T}=-\int_0^t \frac{t^p}{p!}\boldsymbol{X}(t)\mathrm{d}t;\quad p=1,2,\ldots$$

with the hypothesis that the system is open,

$$\int_0^\infty \frac{t^{p-1}}{(p-1)!} X(t)\mathrm{d}t \cdot T = \int_0^\infty \frac{t^p}{p!} X(t)\mathrm{d}t$$

and by induction,

$$X(0) \cdot T^p = \int_0^\infty \frac{t^{p-1}}{(p-1)!} X(t)\mathrm{d}t \tag{16}$$

for any positive integer p.

Using the same thought experiment as in section 11.4, we can say that

$$x_i(0) \cdot \left(t_{i1}^{(p)} \quad t_{i2}^{(p)} \quad \cdots \quad t_{in}^{(p)} \right) =$$

$$= \left(\int_0^\infty \frac{t^{p-1}}{(p-1)!} x_{i1}(t)\mathrm{d}t \quad \int_0^\infty \frac{t^{p-1}}{(p-1)!} x_{i2}(t)\mathrm{d}t \quad \cdots \quad \int_0^\infty \frac{t^{p-1}}{(p-1)!} x_{in}(t)\mathrm{d}t \right)$$

where $t_{ij}^{(p)}$ is the element of row i and column j of matrix T^p, and $x_{ij}(t)$ is the amount of substance that, having entered compartment i at time 0, is present in compartment j at time t; it follows

$$t_{ij}^{(p)} = \frac{\int_0^\infty \frac{t^{p-1}}{(p-1)!} x_{ij}(t)\mathrm{d}t}{x_i(0)}. \tag{17}$$

In the special case $p = 1$ we have

$$t_{ij} = \frac{\int_0^\infty x_{ij}(t)\mathrm{d}t}{x_i(0)} = \frac{1}{K_j} \int_0^\infty \frac{K_j x_{ij}(t)}{x_i(0)} \mathrm{d}t,$$

where in the last integral $x_{ij}(t)/x_i(0)$ is the fraction of substance that entered i at time 0, present in j at time τ, while $K_j x_{ij}(t)/x_i(0)$ is its rate of exit from j, and the integrand at the right-hand side is the fraction that leaves j in the interval $t, t + dt$; therefore the integral

$$\int_0^\infty \frac{K_j x_{ij}(t)}{x_i(0)} \mathrm{d}t$$

is the total fraction that leaves from j at any time. This last number may be larger than one if the same particles exit from j more than once due to recirculation. The factor $1/K_j$, the *turnover time* in j, is the average time spent by the particles in each passage through j; therefore t_{ij} is the average interval of time spent in compartment j by particles introduced into compartment i.

If $i = j$, this time is called *permanence time.* If there is no recirculation, the permanence time coincides with the turnover time; in general the permanence time divided by the transit time is equal to the average number of passages through a compartment, or *turnover number.*

If $i \neq j$, this time is called *residence time.* Observe that the ratio t_{ij}/t_{jj} is equal to the

fraction of particles administered to compartment i that reach compartment j. We call this ratio *yield* from i to j:

$$\gamma_{ij} = t_{ij} / t_{jj}.$$

All these parameters will be examined in mote details in chapter 16.

11.7. RECONSTRUCTION OF MATRICES *K* AND *T*

If all compartments of a system can be controlled and all observed, then from n experiments and n measurements for each experiment, all elements t_{ij} of $\boldsymbol{T}$ can be computed, hence by inversion $\boldsymbol{K}$ is obtained, which describes completely the compartment system. More often than not only very few compartments are controllable and few observable, so that only very few elements of $\boldsymbol{T}$, and the corresponding elements of its powers, can be computed. The problem is the reconstruction of $\boldsymbol{K}$ from these known elements.

Call

$$p(s) = s^n + c_1 s^{n-1} + c_2 s^{n-2} + \cdots + c_n$$

the characteristic polynomial of $\boldsymbol{K}$; as shown in D. 11, $p(s)$ annihilates $\boldsymbol{K}$, that is,

$$\boldsymbol{K}^n + c_1 \boldsymbol{K}^{n-1} + c_2 \boldsymbol{K}^{n-2} + \cdots + c_n \boldsymbol{I} = \boldsymbol{0};$$

we multiply each term of this identity by $\boldsymbol{T}^n$, $\boldsymbol{T}^{n-1}$, $\boldsymbol{T}^{n-2}$, ...

$$\begin{aligned}
&\boldsymbol{I} + c_1 \boldsymbol{T} + c_2 \boldsymbol{T}^2 + c_3 \boldsymbol{T}^3 + \cdots + c_n \boldsymbol{T}^n = \boldsymbol{0} \\
&\boldsymbol{K} + c_1 \boldsymbol{I} + c_2 \boldsymbol{T} + c_3 \boldsymbol{T}^2 + \cdots + c_n \boldsymbol{T}^{n-1} = \boldsymbol{0} \\
&\boldsymbol{K}^2 + c_1 \boldsymbol{K} + c_2 \boldsymbol{I} + c_3 \boldsymbol{T} + \cdots + c_n \boldsymbol{T}^{n-2} = \boldsymbol{0} \\
&\cdots\cdots\cdots
\end{aligned} \tag{18}$$

If only compartment i is controllable and only compartment j is observable (i and j may be equal or different), function $x_{ij}(t)$ can be measured for that particular set of values of i, j, and we can rewrite identities (18) for row i and column j only:

$$\begin{aligned}
&\delta_{ij} + c_1 t_{ij} + c_2 t_{ij}^{(2)} + \cdots + c_n t_{ij}^{(n)} = 0 \\
&-k_{ij} + c_1 \delta_{ij} + c_2 t_{ij} + \cdots + c_n t_{ij}^{(n-1)} = 0 \\
&k_{ij}^{(2)} - c_1 k_{ij} + c_2 \delta_{ij} + \cdots + c_n t_{ij}^{(n-2)} = 0 \\
&\cdots\cdots\cdots
\end{aligned} \tag{19}$$

where $k_{ij}^{(p)}$ is computed with identity (12) in section 11.4, $t_{ij}^{(p)}$ is computed with identity (17) in section 11.6, and δ_{ij} is the Kronecker delta, i.e.

$$\begin{aligned}
\delta_{ij} &= 1 \quad i = j \\
&= 0 \quad i \neq j.
\end{aligned}$$

Any set of n equations chosen from equations (19) may be used to determine the coefficients $c_1, c_2, \ldots, c_n$, of the characteristic polynomial $p(s)$ of $\boldsymbol{K}$.

With the coefficients of the characteristic polynomial of $\boldsymbol{K}$ we can write n ordinary equations that determine the n^2 elements of $\boldsymbol{K}$ with $n^2 - n$ degrees of freedom.

11.8. ENDOGENOUS PRODUCTION

If the vector $\boldsymbol{R}(t)$ in equation (9) is constant but not identically zero, after a sufficiently long time a steady state $\boldsymbol{X}_{ss}$ is reached for the vector $\boldsymbol{X}(t)$; then,

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{0};$$

therefore at steady state,

$$\boldsymbol{X}_{ss} \cdot \boldsymbol{K} = \boldsymbol{R}.$$

From this last identity, multiplying on the right by $\boldsymbol{T}$ we get

$$\boldsymbol{X}_{ss} = \boldsymbol{R} \cdot \boldsymbol{T}, \tag{17}$$

where

$$\boldsymbol{T} = \boldsymbol{K}^{-1}$$

is the matrix of the *permanence times* and *residence times* of the compartmental system as defined in section 11.6. This is the case when the substance under observation is endogenously produced by the system. If this endogenous production stays constant, then by administration of a dose $\boldsymbol{A}$ at time $t = 0$ equation (10) becomes

$$\boldsymbol{X}(t) = \left(\boldsymbol{X}_{ss} + \boldsymbol{A} + \boldsymbol{R} \int_0^t e^{\tau \boldsymbol{K}} d\tau \right) \cdot e^{-t\boldsymbol{K}},$$

thence, by integration,

$$\boldsymbol{X}(t) = \left(\boldsymbol{X}_{ss} + \boldsymbol{A} \right) \cdot e^{-t\boldsymbol{K}} + \boldsymbol{R} \cdot \boldsymbol{K}^{-1} \left(\boldsymbol{I} - e^{-t\boldsymbol{K}} \right),$$

and finally, using identity (17),

$$\boldsymbol{X}(t) - \boldsymbol{X}_{ss} = \boldsymbol{A} \cdot e^{-t\boldsymbol{K}}.$$

This identity shows that by simply subtracting from the measured values of the state variables the corresponding base values, one gets the same result as though there were no endogenous production.

This is not true when the rate of endogenous production changes with the concentration of the substance under observation. For instance, suppose that $\boldsymbol{R}(t)$ is proportional to the deviation of $\boldsymbol{X}(t)$ from a fixed value $\boldsymbol{X}_{ref}$, provided that all elements of $\boldsymbol{R}(t)$ are always positive definite; we can write

$$\boldsymbol{R}(t) = \left(\boldsymbol{X}_{ref} - \boldsymbol{X}(t) \right) \cdot \boldsymbol{B},$$

where

$$\boldsymbol{B} = \begin{pmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_n \end{pmatrix}$$

is a diagonal matrix of constant coefficients $b_1, b_2, \ldots, b_n$. Equation (9) now becomes

$$\frac{d\boldsymbol{X}}{dt} = -\boldsymbol{X}(t) \cdot \boldsymbol{K} + \left(\boldsymbol{X}_{ref} - \boldsymbol{X}(t)\right) \cdot \boldsymbol{B}, \tag{18}$$

and the steady-state value $\boldsymbol{X}_{ss}$ of the state variables is given by

$$\boldsymbol{X}_{ss} \cdot (\boldsymbol{K} + \boldsymbol{B}) = \boldsymbol{X}_{ref} \cdot \boldsymbol{B}. \tag{19}$$

Using identity (19), equation (18) can be written in the form

$$\frac{d\boldsymbol{X}}{dt} = -\left(\boldsymbol{X}(t) - \boldsymbol{X}_{ss}\right) \cdot (\boldsymbol{K} + \boldsymbol{B}),$$

and its integral is

$$\boldsymbol{X}(t) - \boldsymbol{X}_{ss} = \boldsymbol{A} \cdot \exp\left[-t(\boldsymbol{K} + \boldsymbol{B})\right].$$

Again subtraction of the base values from the measured values of the state variables leads to an exponential function, but in this case matrix $\boldsymbol{K}$ has been substituted by matrix $\boldsymbol{K} + \boldsymbol{B}$, i.e. to the turnover rate of each compartment has been added the coefficient b_i measuring the rate at which the endogenous production in that compartment is controlled. In other words, by subtracting the base values, the compartments whose concentration is regulated by endogenous production will appear to have turnover rates larger than their real value.

11.9. REFERENCES

1. A. Rescigno, R. M. Lambrecht and C. C. Duncan, Stochastic Modeling and Physiologic Processes with Radiotracers and Positron Emission Tomography. In *Applications of Physics to Medicine and Biology,* edited by G. Alberi, Z. Bajzer and P. Baxa (World Scientific Publ. Co., Singapore, 1983), pp. 303-18.
2. A. K. Thakur, A. Rescigno and D. E. Schafer, On the Stochastic Theory of Compartments: I. A Single Compartment System, *Bull. Math. Biophysics* **34,** 53-63 (1972).
3. A. K. Thakur, A. Rescigno and D. E. Schafer, On the Stochastic Theory of Compartments: II. Multi-Compartment Systems, *Bull. Math. Biophysics* **35,** 263-71 (1973).
4. M. Rowland and G. Tucker, Symbols in Pharmacokinetics, *J. Pharmacokin. Biopharm.* **8,** 497-507 (1980).
5. J. K. Aronson, H. J. Dengler, L. Dettli and F. Follath, Standardization of Symbols in Clinical Pharmacology, *Eur. J. Clin. Pharmacol.* **35,** 1-7 (1988).
6. J. Z. Hearon, Theorems on Linear Systems, *Annals N. Y. Acad. Sci.* **108,** 36-68 (1963).
7. T. Muir, *A Treatise on the Theory of Determinants* (Longmans, Green and Co., London, 1933).
8. A. Rescigno, R. M. Lambrecht and C. C. Duncan, Mathematical Methods in the Formulation of Pharmacokinetic Models. In *Tracer Kinetics and Physiologic Modeling,* edited by R. M. Lambrecht and A. Rescigno (Springer-Verlag, Berlin, 1983), pages 59-119.

This page intentionally left blank

12. NON-LINEAR MODELS

12.1. LINEAR VERSUS NON LINEAR

All models so far considered were linear. There are two important reasons for this; the first is that many pharmacokinetic systems are linear, or exhibit a linear behavior under proper conditions, or at least can yield considerable information under linear analysis using tracer techniques; the second is that linear models have an important role in understanding the qualitative behavior of non-linear systems and in the definition of most pharmacokinetic parameters. Hearon [1] adds an additional reason: "Linear theory is pleasant; the concepts are relatively simple but far reaching, and, in principle at least, the formalism is tractable".

The truth is that there is no general treatment of non-linear models; each particular model must be solved separately. All I can do in this chapter is to show a few examples of selected non-linear models with general indications on their possible applications.

12.2. REACTIONS OF ORDER ZERO

The simplest non-linear model is the zero-order elimination from a compartment. An elimination of this type was observed in the disappearance of ethyl alcohol [2, 3] as well as other volatile substances (ether, acetone) from the blood, and in the elimination of hippuric acid in the rabbit upon administration of benzoic acid [4].

A reaction of this type is to be expected whenever the systems responsible for detoxication, or for elimination, are saturated by the given substance; in these cases the amount of substance that undergoes detoxication or elimination per unit time is constant and not proportional to its concentration.

The equation of first order elimination is

$$\frac{dx}{dt} = -K + r(t), \tag{1}$$

where

$x(t)$ = amount of substance in the compartment,

$r(t)$ = entry rate into the compartment,

K = exit rate from the compartment, dimension of $x(t)$ times $[T^{-1}]$.

Compare this equation with equation (2) of chapter 5.

The integral of equation (1) is

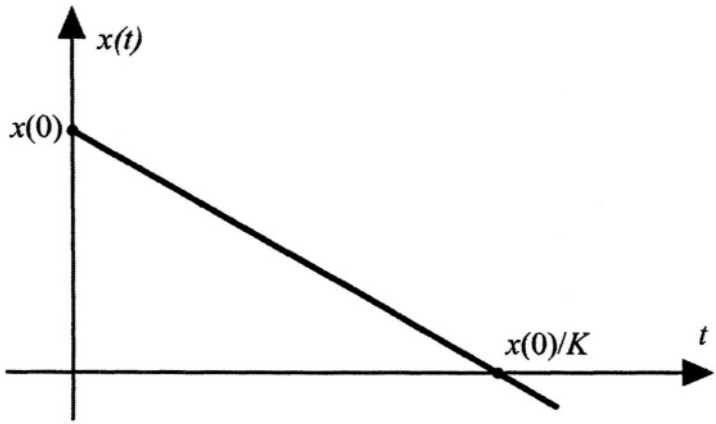

Fig.1. Solution of equation (1) when $r(t) = 0$

$$x(t) = x(0) + \int_0^t r(t)dt - K \cdot t; \tag{2}$$

so far this model is linear because equation (1) is linear and its integral (2) has the linear property described in section 10.1. The problem is, this model is not physically realizable because, as Fig. 1 shows, the variable $x(t)$ becomes negative after a certain value of t.

To make the model physically realizable we must modify equation (1); the simplest realizable model is

$$\begin{aligned} \frac{dx}{dt} &= -K + r(t) \qquad x(t) > 0 \\ &= 0 \qquad\qquad\quad x(t) \le 0 \end{aligned} \tag{3}$$

and its integral, when $r(t) \equiv 0$, is

$$\begin{aligned} x(t) &= x(0) - K \cdot t \quad t < x(0)/K \\ &= 0. \qquad\quad t \ge x(0)/K \end{aligned}$$

as shown in fig. 2.

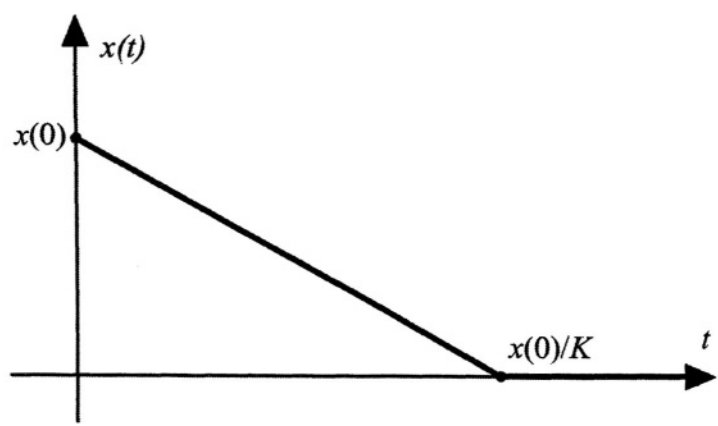

Fig.2. Solution of equation (3) when $r(t) = 0$

We can now compute the moments of this function. The 0-moment is the area of the triangle with base $x(0)/K$ and height $x(0)$, i.e.,

$$X_0 = \frac{[x(0)]^2}{2K};$$

the higher moments are given by

$$X_i = \int_0^{x(0)/K} \frac{t^i}{i!}[x(0) - K \cdot t]dt$$
$$= \frac{(i+1)}{K^{i+1}} \cdot \frac{[x(0)]^{i+2}}{(i+2)!}.$$

For the moments of negative order we use formula (2) of section 9.1 and get

$$X_{-1} = x(0), \quad X_{-2} = K, \quad X_{-i} = 0 \text{ for } i > 2. \tag{4}$$

I will show in later chapters that all those moments have a different meaning from the moments of the linear compartment systems and can be misleading if not used in the appropriate context.

It is impossible to give a general explicit solution of equation (3) in the general case with $r(t)$ not identically zero, because there is no way to know a priori when function $x(t)$ reaches a zero value and the elimination process stops.

12.3. REACTIONS OF MIXED ORDER ZERO AND ONE

A variation of the previous model allows elimination of order zero when the concentration is above a certain critical value c_0, and elimination of order one when the concentration is below that critical value.

The differential equation of this model, again with $r(t) \equiv 0$, is

$$\begin{aligned} \frac{dx}{dt} &= -K_0 && x(t) > c_0V \\ &= -K_1 \cdot x(t) && x(t) \le c_0V \end{aligned} \tag{5}$$

where K_0 is the zero-order rate constant (dimension of $x(t)$ times $[T^{-1}]$), and K_1 is the one-order rate constant (dimension $[T^{-1}]$). For reasons of continuity it must be

$$K_0 = K_1c_0V.$$

If the initial value $x(0)$ is larger than c_0V, the solution is

$$\begin{aligned} x(t) &= x(0) - K_0t && t < t_0 \\ &= c_0V \cdot e^{-K_1(t-t_0)} && t \ge t_0 \end{aligned} \tag{6}$$

where

$$t_0 = \frac{x(0) - c_0V}{K_0}$$

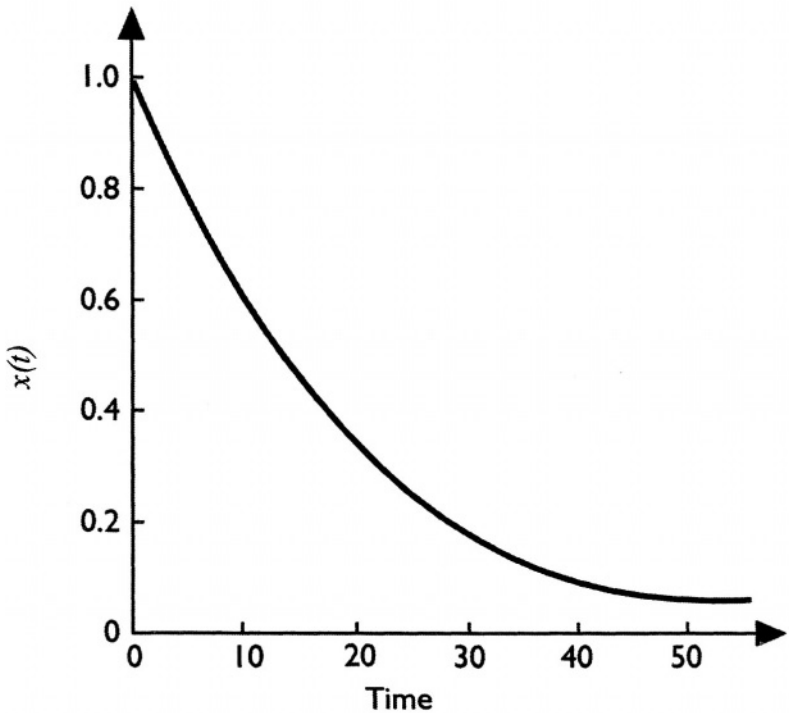

Fig.3. Elimination of mixed order zero and one

is the time when the critical concentration c_0 is reached (see fig. 3).

The moments of order zero and up of function (6) are computed from the integrals

$$X_i = \int_0^{t_0} \frac{t^i}{i!}\left[x(0) - K_0 t\right]dt + \int_{t_0}^{\infty} \frac{t^i}{i!} c_0 V \cdot e^{-K_1(t-t_0)}dt$$

$$= \frac{t_0^{i+1}}{(i+1)!}x(0) - \frac{t_0^{i+2}}{(i+2)!}(i+1)K_0 + \frac{c_0 V}{K_1^{i+1}}$$

then substituting the value of t_0 and of K_1,

$$X_0 = \frac{\left[x(0) - c_0 V\right]\left[x(0) + c_0 V\right]}{2K_0} + \frac{c_0 V}{K_1},$$

$$X_1 = \frac{\left[x(0) - c_0 V\right]^2\left[x(0) + 2c_0 V\right]}{6K_0^2} + \frac{c_0 V}{K_1^2},$$

and so forth. For the moment of negative order we have the same result as in (4).

12.4. REACTIONS OF ORDER TWO

Elimination from a compartment with a reaction of second order is described by the differential equation

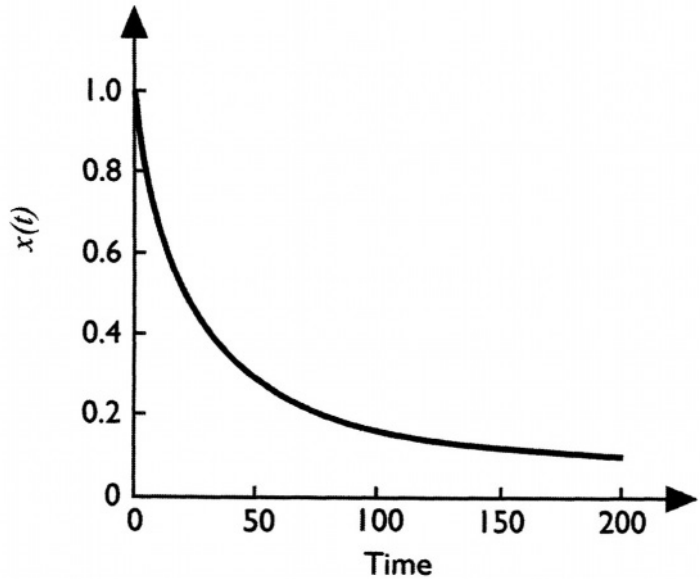

Fig.4. Example of elimination of order two

$$\frac{dx}{dt} = -K\left[x(t)\right]^2 + r(t),$$

where the rate constant K has the dimension of $1/x(t)$ times $[T^{-1}]$.

The integration of this differential equation, when $r(t) \equiv 0$, leads to

$$x(t) = \frac{x(0)}{1 + x(0) \cdot K \cdot t}. \tag{7}$$

See fig. 4 for a typical case of elimination of order two.

We can compute the moments of negative order directly from the differential equation; they are

$$X_{-1} = x(0), \quad X_{-2} = K\left[x(0)\right]^2, \quad X_{-3} = 2K^2\left[x(0)\right]^3, \quad X_{-4} = 6K^3\left[x(0)\right]^4, \quad \ldots$$

For the 0-moment we must find the integral of expression (7), i.e.,

$$\int_0^t \frac{x(0)}{1 + x(0) \cdot K \cdot t}\,dt = \frac{1}{K}\ln\left|1 + x(0) \cdot K \cdot t\right|;$$

unfortunately the expression on the right-hand side does not converge for $t \to \infty$, even though the integrand function does; therefore the 0-moment, and *a fortiori* all higher moments, do not exist [5]. This is an important case when the so-called *AUC* computed with the formula

$$AUC = \int_0^{t_1} c(t)dt + \int_{t_1}^{\infty} c(t_1)e^{-K(t-t_1)}dt,$$

where t_1 is the time the last sample was measured and K is the estimated slope of $c(t)$ at time t_1, may lead to wrong conclusions, as will be shown in the next chapter.

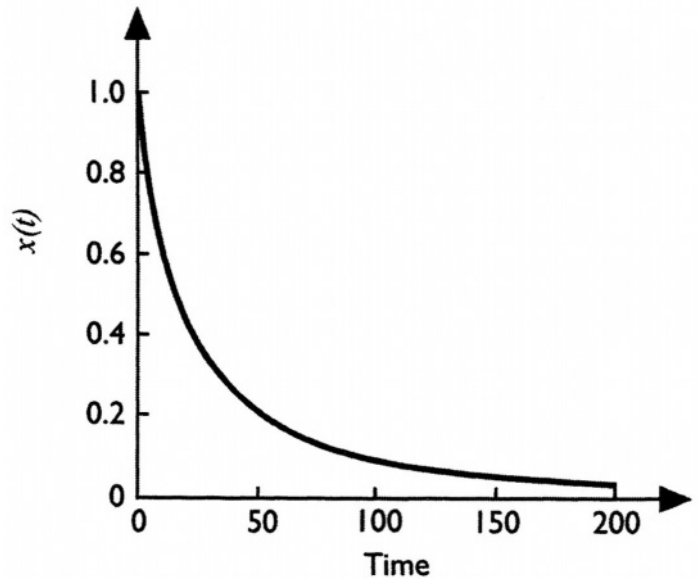

Fig.5. Example of elimination of order one and two

12.5. REACTIONS OF MIXED ORDER ONE AND TWO

We consider now the simultaneous elimination of a drug by a process of order one and a process of order two. This model is described by the differential equation

$$\frac{dx}{dt} = -K_1 x(t) - K_2\left[x(t)\right]^2,$$

where K_1 is the rate constant of the first-order process (dimension $[T^{-1}]$), and K_2 the rate constant of the second-order process (dimension of $1/x(t)$ times $[T^{-1}]$) (see fig. 5).

The moments of negative order are

$$X_{-1} = x(0),\quad X_{-2} = x(0)\left[K_1 + K_2 x(0)\right],\quad X_{-3} = x(0)\left[K_1 + K_2 x(0)\right]\left[K_1 + 2K_2 x(0)\right],\quad \ldots$$

The solution of the differential equation is

$$x(t) = \frac{1}{\left[\frac{1}{x(0)} + \frac{K_2}{K_1}\right]e^{K_1 t} - \frac{K_2}{K_1}}$$

and the integral of this function is

$$\int_0^t \frac{dt}{\left[\frac{1}{x(0)} + \frac{K_2}{K_1}\right]e^{K_1 t} - \frac{K_2}{K_1}} = \frac{1}{K_2}\ln\left[x(0)\left(\frac{1}{x(0)} + \frac{K_2}{K_1}\right) - \frac{K_2}{K_1}e^{-K_1 t}\right],$$

therefore the 0-moment converges to the value

$$X_0 = \frac{1}{K_2} \ln\left(1 + \frac{K_2}{K_1} x(0)\right).$$

12.6. MICHAELIS-MENTEN ELIMINATION

A model frequently invoked for non-linear processes is the Michaelis-Menten equation

$$\frac{dx}{dt} = \frac{-K_0 x(t)}{K_1 + x(t)},$$

where K_0 (dimension of $x(t)$ times $[T^{-1}]$) and K_1 (dimension of $x(t)$) are appropriate rate constants, whose meaning will be discussed later. The solution of this equation is given by the implicit expression

$$x(t) - x(0) + K_1 \ln\left(\frac{x(t)}{x(0)}\right) + K_0 t = 0. \tag{8}$$

Fig 6 shows a typical graph of equation (8).

Observe that when $x(t) >> K_1$, that is, for the initial part of the curve, the rate of elimination is

$$\frac{dx}{dt} \approx -K_0,$$

therefore we can consider K_0 as a zero-order elimination constant valid when $x(t)$ is large enough. When $x(t) << K_1$, that is, for the final part of the curve, the rate of elimination is

$$\frac{dx}{dt} \approx -\frac{K_0}{K_1} x(t),$$

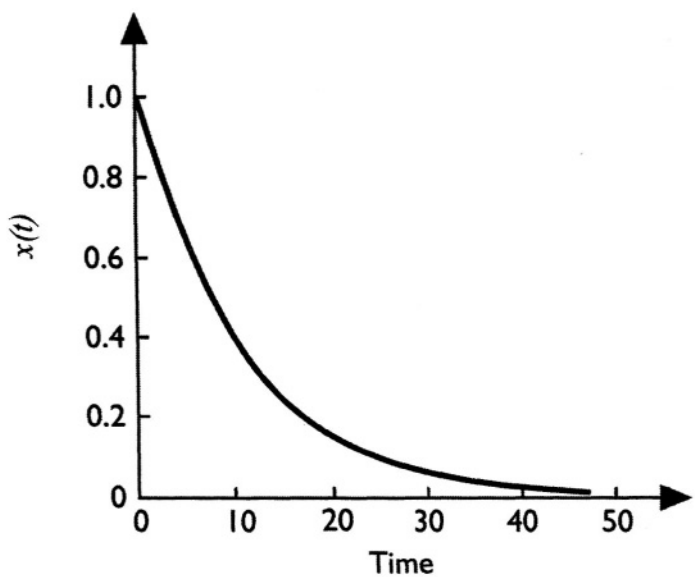

Fig.6. Michaelis-Menten elimination

therefore we can consider K_0/K_1 as a first-order elimination constant valid when $x(t)$ is small enough. The model of section 12.3 is similar to the one of this section, but in the last one the transition from order zero to order one is smoother.

12.7. REFERENCES

1. J. Z. Hearon, Theorems on Linear Systems, *Ann. N. Y. Acad. Sci.* **108**, 36-68 (1963).
2. E. M. P. Widmark, *Die wissenshaftliche Grundlagen und die praktische Verwendbarkeit der gerichtlich-medizinshen Alcoholbestimmung* (Urban & Schwarzenberg, Berlin, 1932).
3. F. Lundquist and H. Wolthers, The Kinetics of Alcohol Elimination in Man, *Acta Pharmacol. Toxicol.* **14**, 265-89 (1958).
4. H. G. Bray, W. V. Thorpe and K. White, Kinetics Studies of the Metabolism of Foreign Organic Compounds: I - The formation of benzoic acid from benzamide, toluene, benzyl alcohol and benzaldehyde and its conjugation with glycine and glucoronic acid in the rabbit, *Biochem. J.* **48**, 88-96 (1951).
5. A. Rescigno and A. Marzo, Area Under the Curve, Bioavailability, and Clearance, *J. Pharmacokin. Biopharm.* **19**, 473-80 (1991).

13. CLEARANCE

13.1. DEFINITION OF CLEARANCE

Clearance is one of the most important pharmacokinetic parameters. It is defined as the volume containing the amount of drug eliminated per unit time by a specified organ; it has the dimension of a flow (i.e., $[L^3\ T^{-1}]$). It is usually represented by the symbol Cl; some authors use the symbol *CL*, but the use of two capital letters for a symbol is contrary to all scientific conventions.

Note: I once heard that the use of the symbol *CL* is necessary because in print the small case letter *l* is easily confused with the capital letter *I*; but there is no danger of confusing Cl with a non-existing symbol *CI* or Ci.

A word of caution is necessary at this point: the drug may be eliminated by an organ reversibly or irreversibly; furthermore, we may distinguish between total elimination from an organ, and transfer from an organ to another specific organ. Thus, we can define four different clearances: Cl_i^{rev}, Cl_i^{irr}, Cl_{ij}^{rev}, Cl_{ij}^{irr}; the subscripts mean *out* of organ i, or *from* organ i to organ j; the superscripts mean *reversibly* or *irreversibly*.

If not indicated differently, by "clearance" I mean "total irreversible clearance from a specified organ", i.e., the volume containing the amount of substance eliminated, or transferred, or inactivated, per unit time, from the specified organ; where by "organ" I mean any anatomical organ or physico-chemical space properly defined.

13.2. PROPERTIES OF CLEARANCE

If the drug is eliminated from an organ with a uniform concentration $c(t)$, then, from the definition of clearance it follows that

$$Cl \cdot c(t) = r(t) \tag{1}$$

is the rate of elimination from the organ at time t, and

$$Cl \cdot c(t) \cdot dt = r(t) \cdot dt$$

is the amount of drug eliminated in the interval of time from t to $t + dt$. By integration over the interval of time t_1, t_2 we get

$$\int_{t_1}^{t_2} Cl \cdot c(t) \cdot dt = \int_{t_1}^{t_2} r(t) \cdot dt\ ;$$

the integral at the right-hand side is the amount of drug eliminated in that interval of time. If we call $Q(t_1, t_2)$ that amount, then

$$\int_{t_1}^{t_2} Cl \cdot c(t) \cdot dt = Q(t_1, t_2). \tag{2}$$

If the clearance is constant in time, we can export Cl from the integral and write

$$Cl = \frac{Q(t_1, t_2)}{\int_{t_1}^{t_2} c(t) \cdot dt}. \tag{3}$$

If the clearance is not constant in time, from (2) we get

$$\frac{\int_{t_1}^{t_2} Cl \cdot c(t) \cdot dt}{\int_{t_1}^{t_2} c(t) \cdot dt} = \frac{Q(t_1, t_2)}{\int_{t_1}^{t_2} c(t) \cdot dt},$$

where the fraction at the left-hand side is the average value of the clearance weighted for the concentration; therefore when the clearance is not constant in time, expression (3) becomes

$$\langle Cl \rangle = \frac{Q(t_1, t_2)}{\int_{t_1}^{t_2} c(t) \cdot dt}, \tag{4}$$

where the brackets $\langle\ \rangle$ mean "average value".

If a drug is eliminated from an organ where its concentration is not uniform, then consider the clearance dCl from the infinitesimal volume dV where the concentration is $c_V(t)$; the subscript means that the concentration depends upon the location through the volume V. The amount of drug cleared from that infinitesimal volume in the interval of time t, $t + dt$ is

$$c_V(t) \cdot dCl \cdot dt = dr(t) \cdot dt;$$

taking the integral of the above expression for the interval of time t_1, t_2 and for the whole volume of the organ, we get the total amount $Q(t_1, t_2)$ of drug eliminated in that interval of time,

$$\int_{t_1}^{t_2} \left(\int_V c_V(t) \cdot dCl \right) \cdot dt = Q(t_1, t_2). \tag{5}$$

Define, for any time t,

$$\frac{\int_V c_V(t) \cdot dCl}{Cl} = \langle c_V(t) \rangle,$$

where $\langle c_V(t) \rangle$ is the average value of the concentration at time t in the volume V computed with a weight proportional to the clearance, and Cl is the total clearance from the volume V; now expression (5) can be written as

$$\int_{t_1}^{t_2} \left(Cl \cdot \langle c_V(t) \rangle \right) \cdot dt = Q(t_1, t_2); \tag{6}$$

if the total clearance is constant in time, we can export it from the above integral and write

$$Cl = \frac{Q(t_1, t_2)}{\int_{t_1}^{t_2} \langle c_V(t) \rangle \cdot dt}, \tag{7}$$

similar to expression (3).

If the total clearance is not constant in time, from (6) we get

$$\frac{\int_{t_1}^{t_2} \langle Cl \cdot \langle c_V(t) \rangle \rangle \cdot dt}{\int_{t_1}^{t_2} \langle c_V(t) \rangle \cdot dt} = \frac{Q(t_1, t_2)}{\int_{t_1}^{t_2} \langle c_V(t) \rangle \cdot dt},$$

where the fraction at the left-hand side is the average value of the clearance weighted for the average value of the concentration; therefore when the clearance is not constant in time, expression (7) becomes

$$\langle Cl \rangle = \frac{Q(t_1, t_2)}{\int_{t_1}^{t_2} \langle c_V(t) \rangle \cdot dt}. \tag{8}$$

Observe that (3) and (7) give the exact value of the clearance, while (4) and (8) give only an approximate value; but only (3) is of practical use, because in general we cannot measure the average concentration $\langle c(t) \rangle$; nevertheless the above expressions are useful to define correctly the other pharmacokinetic parameters.

These results are summarized in Table I.

Table I

Clearance	Concentration	Formula
Constant	Uniform	$Cl = \frac{Q(t_1, t_2)}{\int_{t_1}^{t_2} c(t) \cdot dt}$
Non constant	Uniform	$\langle Cl \rangle = \frac{Q(t_1, t_2)}{\int_{t_1}^{t_2} c(t) \cdot dt}$
Constant	Non uniform	$Cl = \frac{Q(t_1, t_2)}{\int_{t_1}^{t_2} \langle c(t) \rangle \cdot dt}$
Non constant	Non uniform	$\langle Cl \rangle = \frac{Q(t_1, t_2)}{\int_{t_1}^{t_2} \langle c(t) \rangle \cdot dt}$

13.3. DETERMINATION OF CLEARANCE

If the concentration of the drug in an organ is uniform and the clearance is constant, its determination can be done using formula (3), provided the amount of drug eliminated in the chosen interval of time has been measured and its concentration in the clearing organ has been monitored for that same interval of time. Actually in general in formula (3) it is customary to put $t_1 = 0$ and $t_2 = \infty$, so that

$$Cl = \frac{Q(0,\infty)}{\int_0^\infty c(t)\cdot dt}. \tag{9}$$

Two observations are necessary at this point:

1. If the amount of drug eliminated $Q(0, \infty)$ cannot be determined directly, it can be substituted by the product $F.D$, where F, *bioavailability,* is the fraction of the given dose D that reaches the organ investigated. The bioavailability, of course, must be determined by a separate experiment.
2. The integral in the denominator can be determined exactly only if the concentration $c(t)$ can be monitored until no drug is left inside the organ, an unrealistic assumption. One must use the extrapolating formula

$$\int_0^\infty c(t)dt = \int_0^{t^*} c(t)dt + \int_{t^*}^\infty c(t)dt,$$

where the integral $\int_0^{t^*} c(t)dt$ is computed with the values of the concentration monitored up to time t^*, and the integral $\int_{t^*}^\infty c(t)dt$ is computed with values of concentration estimated using a selected model; see for instance Bass et al. [1].
3. Formula (9), like formula (3) from where (9) is derived, is valid only if the clearance is constant, a condition hard to verify. In fact sometimes the integral $\int_0^\infty c(t)dt$ does not converge at all [2]; in that case the resulting error may be very large.

13.4. VERIFICATION OF CONSTANCY OF CLEARANCE

To verify whether clearance depends or not on concentration, think of a pharmacokinetic system at steady state, i.e. a system where a drug is administered in a continuous way to keep its concentration in all organs constant in time, but not necessarily uniform. Call c the concentration at the point of elimination from a given organ and r the rate of elimination therefrom; by definition the ratio

$$Cl = r/c$$

is the *clearance* of the drug from that organ. Clearance may depend upon concentration or not; but in either case all quantities involved are constant in time, therefore we can write

$$Cl = \frac{Q(t_2) - Q(t_1)}{(t_2 - t_1) \cdot c};$$

by experimenting with different values of r and c we can check the dependency of Cl upon c.

13.5. SCALING OF CLEARANCE

Sometimes the clearance is divided by the body weight; for instance the "Notice to applicants for marketing authorizations for medicinal products for human use in the Member States of the European Community" (III/118/87-EN) explicitly requires that the clearance, computed as *Dose/AUC,* be referred, if possible, to 1 kg of body mass. The result is a quantity with the dimension $[L^3\ M^{-1}\ T^{-1}]$, and should not be called *clearance*, but with a different name, for instance *specific clearance* or *normalized clearance.* But aside from the name, what is the rational for dividing clearance by body mass?

As I have shown in Chapter 3, a quantity may be divided by another quantity only if their quotient is an invariant, i.e., if the result of that division is a meaningful quantity whose value does not depend, within the limits of validity of some specific hypotheses, on the particular experimental conditions. But can we say that clearance is proportional to body mass? According to many authors [3], the experimental evidence points to a relationship of the form

$$Cl = a \cdot W^b,$$

where W is the body mass, a is a constant depending upon the particular substance, and b is a number between 0.69 and 0.89.

If the clearance is not proportional to the body mass, what other quantity can be used to scale its value in a biologically sensible way?

From expression (1), dividing and multiplying the left-hand side by the volume V of the organ under consideration,

$$\frac{Cl}{V} \cdot V\, c(t) = r(t),$$

or

$$\frac{Cl}{V} = \frac{r(t)}{V\, c(t)},$$

where the fraction at the right-hand side is the inverse of the turnover time, an important pharmacokinetic parameter, as shown in chapter 14. It appears that dividing clearance by volume of the eliminating organ, in general plasma, is a more sensible way of scaling than dividing by the body mass. Of course the volume of the plasma is not always known with precision, but neither is the clearance; using estimates of well defined quantities is certainly better than using exact values of quantities that do not have any physical or physiological meaning.

13.6. ENDOGENOUS SUBSTANCES

A peculiar problem is presented by endogenous substances when determining the clearance with formula (9) using $Q(0,\infty) = F \cdot D$.

As shown by Marzo, Rescigno and Thakur [4, 5], for endogenous substances we can define F, bioavailability, as "fraction of exogenous substance absorbed" or as "fractional increase of the level of exogenous plus endogenous substances".

In an analogous way, by clearance we may mean "volume containing the amount of exogenous substance eliminated per unit time" or "volume containing the amount of exogenous plus endogenous substance eliminated per unit time"; both definitions are meaningful. Even if the eliminating organ does not distinguish between the exogenous and the endogenous substance, the two clearances will be different when a homeostatic mechanism reduces the endogenous production after administration of the exogenous substance.

13.7. EXAMPLES

Consider the case of zero-order elimination described by equation (3) in section 12.2. If $x(t)$ is the amount of drug in a compartment whose volume is V, the concentration there is $x(0)/V$; using formula (9) above we get

$$Cl = \frac{x(0)}{\int_0^\infty \frac{x(t)}{V}\,dt} = \frac{2KV}{x(0)};$$

but this is only an average value of the clearance, that is changing with the change in concentration. The actual instantaneous value of the clearance at time t is the ratio between its rate of elimination and its concentration at the same time, as shown by equation (1) at the beginning of this chapter, i.e.,

$$Cl(t) = KV/x(t);$$

the clearance starts with initial value

$$Cl(0) = KV/x(0),$$

reaches the value $2KV/x(0)$ when $x(t) = x(0)/2$, then continues to increase with the decrease of the concentration, and suddenly stops when the concentration becomes zero.

Consider now the case of second-order elimination described by equation (7) in section 12.4. The true value of the clearance at time t is the rate of elimination divided by the concentration, i.e.,

$$Cl(t) = \frac{K\left[x(t)\right]^2}{x(t)/V} = KV \cdot x(t).$$

Strictly speaking we cannot use formula (9) above because the integral at the denominator does not converge; but suppose we succumb to the temptation and use that formula stopping the integration at a time t_1 where we think the concentration is small enough. We have

$$\int_0^{t_1} c(t)\mathrm{d}t = \frac{1}{V}\int_0^{t_1} \frac{x(0)}{1 + x(0)\cdot K\cdot t}\mathrm{d}t = \frac{1}{KV}\ln\left[1 + x(0)\cdot K\cdot t_1\right];$$

thence we estimate the value of the clearance as

$$Cl_{est} = \frac{x(0)}{\int_0^{t_1} c(t)\mathrm{d}t} = \frac{K\cdot V\cdot x(0)}{\ln\left[1 + x(0)\cdot K\cdot t_1\right]}.$$

The ratio between the true value of the clearance and this "estimated" value is

$$\frac{Cl(t)}{Cl_{est}} = \frac{x(t)}{x(0)}\ln\left[1 + x(0)\cdot K\cdot t_1\right]$$

and can be very large. Paradoxically, the error increases when t_1 increases, even though "it seems" that with a larger value of t_1 we have a better approximation of $\int_0^{\infty} c(t)\mathrm{d}t$.

I invite the reader to experiment with other examples of non-linear models.

13.8. REFERENCES

1. L. Bass, J. Aisbett and A. J. Bracken, Asymptotic Forms of Tracer Clearance Curves: Theory and Applications of Improved Extrapolations, *J. theoret.Biol.* **111**, 755-85 (1984).
2. A. Rescigno and A. Marzo, Area Under the Curve, Bioavailability, and Clearance, *J. Pharmacokin. Biopharm.*; **19,** 473-82 (1991).
3. K. Krishnan and M. E. Andersen, Interspecies Scaling in Pharmacokinetics. In: *New Trends in Pharmacokinetics,* edited by A. Rescigno and A. K. Thakur (Plenum Press, New York, 1991), pp. 203-26.
4. A. Marzo and A. Rescigno, Pharmacokinetics of Endogenous Substances: Some Problems and Some Solutions. *Eur. J. Drug Metab. Pharmacokin.*; **18**, 77-88 (1993).
5. A. Rescigno, A. K. Thakur and A. Marzo, On Definition and Use of the Term Bioavailability, *Arzneim.-Forsch.* **44**, 1167-9 (1994).

This page intentionally left blank

14. TURNOVER

14.1. TURNOVER TIME

Another important pharmacokinetic parameter is *turnover time.* The turnover time of an organ is defined as the average time spent by the drug in one passage through that organ.

Consider first an organ where the drug has a uniform concentration. From expression (1) of chapter 13, after multiplying both sides by the volume V of the organ, we get

$$\frac{V \cdot c(t)}{r(t)} = \frac{V}{Cl};$$

but the product $V \cdot c(t)$ is the amount of drug present in the organ, therefore the ratio $V \cdot c(t)/r(t)$ is the average time spent by the drug in that organ. We call it *turnover time* [1], symbol T, and write

$$T = \frac{V}{Cl}. \tag{1}$$

Consider now an organ where the drug is at a steady state, i.e. it has a constant concentration, not necessarily uniform. If Q is the amount of drug present in the organ and r its rate of elimination therefrom, the ratio Q/r is the interval of time taken by the organ to eliminate an amount of drug equal to the amount present, or in other words, the average time spent by the drug in one passage through that organ,

$$T = \frac{Q}{r}.$$

In general, in an organ where the concentration is neither constant or uniform, the time T spent by the drug in the infinitesimal volume dV of the organ before being cleared is

$$T = \frac{\mathrm{d}V \cdot c(t)}{c(t) \cdot \mathrm{d}Cl} = \frac{\mathrm{d}V}{\mathrm{d}Cl};$$

by integration over the volume V we get

$$V = \int_V T \cdot \mathrm{d}Cl;$$

but the ratio of volume cleared over clearance is equal to the time spent in the whole organ, therefore the turnover time of the organ is

$$T = \frac{\int_V T \cdot dCl}{\int_V dCl}, \tag{2}$$

identical to definition (1) above; we conclude that this definition of turnover time is valid even if the concentration is not uniform and constant. Therefore the turnover time may be considered a pharmacokinetic parameter [2].

This expression shows the general relationship among turnover time, clearance, and volume of the organ the drug is distributed in.

14.2. DETERMINATION OF TURNOVER TIME

If the physical volume of an organ is known and the clearance can be measured, the turnover time can be calculated using definition (1). In some instances the turnover time can be measured directly. For instance if a drug can be administered to an organ as a bolus at time $t = 0$, and we can monitor its uniform concentration there, the differential equation

$$\frac{dx}{dt} = -\frac{1}{T}x(t) + r(t)$$

holds, where the amount of drug present in the organ at time t is $x(t) = V \cdot c(t)$, the rate at which the drug leaves the organ is $x(t)/T$, and $r(t)$ is the rate of reentry, if any. But $r(0) = 0$, therefore

$$\lim_{t \to 0} \frac{x(t)}{-dx/dt} = T,$$

and dividing numerator and denominator by V,

$$\lim_{t \to 0} \frac{c(t)}{-dc/dt} = T.$$

14.3. TURNOVER TIME AND COMPARTMENTS

The turnover time T of a compartment is defined as the expected interval of time spent by the drug in one passage through it. The general equation of a compartment can be written

$$\frac{dx}{dt} = -K x(t) + r(t), \tag{3}$$

where $x(t)$ is the amount of drug present in the compartment, K the fraction eliminated per unit time, and $r(t)$ the rate of entry due to external feeding or internal recycling, or both.

The turnover time is the inverse of the fraction eliminated per unit time,

$$T = 1/K;$$

in fact at any moment the amount of the drug present in the compartment is $x(t)$, its rate of exit is $Kx(t)$, therefore the ratio

$$\frac{x(t)}{Kx(t)} = \frac{1}{K}$$

is the time spent by the drug, on the average, in that compartment. We can call K the *turnover rate*; its dimension is $[T^{-1}]$.

The turnover time of a pool formed by a number of compartments is still defined as the expected interval of time spent by the drug in one passage through it; its determination though is possible only if the detailed fate of the drug in all components of that pool is known.

For instance consider the two-compartment system described by equations

$$\frac{dx_1}{dt} = -K_1 x_1(t) + k_{21} x_2(t)$$

$$\frac{dx_2}{dt} = +k_{12} x_1(t) - K_2 x_2(t)$$

with initial conditions

$$x_1(0) = D, \quad x_2(0) = 0.$$

We know that the turnover time of compartment 1 is $1/K_1$ and of compartment 2 is $1/K_2$. What is the turnover time of the two compartments pooled together?

The volume of the pool is $V_1 + V_2$; the clearance of the first compartment is V_1K_1 and the clearance of the second one is V_2K_2; therefore using identity (2) the turnover time of the pool is

$$T_{pool} = \frac{V_1 + V_2}{V_1K_1 + V_2K_2}. \tag{4}$$

If the pool includes many compartments, the turnover time T_{pool} of the pool is given by

$$T_{pool} = T_i + \frac{1}{1-r}(T_j - T_i),$$

where T_i is the turnover time of the compartment from where the drug enters the pool, T_j the time employed by the recirculating particles for a complete cycle, and r the fraction of particles recirculated [3,4].

14.4. TURNOVER RATE

The *turnover rate,* symbol K, is the inverse of the turnover time, as seen in the previous section. By definition,

$$K = 1/T. \tag{5}$$

As seen above, the turnover rate of a compartment is equal to its fractional elimination rate; in a pool of compartments like the one described in the previous section, using expression (4) we have

$$K_{pool} = \frac{V_1K_1 + V_2K_2}{V_1 + V_2}.$$

In general, for any organ, we can write, using expression (2),

$$K = \frac{\int_V \mathrm{d}Cl}{\int_V T \cdot \mathrm{d}Cl}.$$

14.5. TURNOVER NUMBER

The *turnover number,* symbol ν, is the number of times a particle goes through the same site; it is a typical dimensionless parameter. If there is no recirculation of course the turnover number is one.

Take for instance equation (3) of a compartment; by integration of both sides from 0 to ∞ we get

$$x(\infty) - x(0) = -\int_0^\infty Kx(t)\mathrm{d}t + \int_0^\infty r(t)\mathrm{d}t;$$

if the system is open, $x(\infty) = 0$, therefore

$$\int_0^\infty Kx(t)\mathrm{d}t = x(0) + \int_0^\infty r(t)\mathrm{d}t.$$

Observe that in this expression the left-hand side shows the total amount of drug leaving the compartment, and the right-hand side the amount of drug entering it at the initial time and later by recirculation. Without recirculation the left-hand side integral is equal to $x(0)$ because the number of particles leaving is identical to the number of particles entering. With recirculation the number of particles leaving divided by the number of particle that were introduced the first time is equal to the number of times the particles passed through the compartment, therefore

$$\nu = \frac{\int_0^\infty Kx(t)\mathrm{d}t}{x(0)}. \tag{6}$$

14.6. TURNOVER AND MOMENTS

If the turnover rate K is constant, it can be exported from the integral in the expression (6) above, and we can write

$$\nu = K\frac{\int_0^\infty x(t)\mathrm{d}t}{x(0)}.$$

Observe that the above integral is the 0-moment of $x(t)$, and $x(0)$ is its -1-moment, therefore we can write

$$\frac{X_0}{X_{-1}} = \frac{\nu}{K} = \nu \cdot T.$$

The turnover number times the turnover time is the time spent by the drug in a particular organ. This important pharmacokinetic parameter called *permanence time* will be discussed in detail in chapter 16.

14.7. REFERENCES.

1. A. Rescigno, Clearance, Turnover Time, and Volume of Distribution, *Pharmacol. Res.* **35**, 189-93 (1997).
2. A. Rescigno, Fundamental Concepts in Pharmacokinetics, *Pharmacol. Res.* **35**, 363-90 (1997).
3. A. Rescigno and E. Gurpide, Estimation of average times of residence, recycle, and interconversion of blood-borne compounds using tracer methods, *J. Clin. Endocrinol. Metab.* **36**, 263-76 (1973).
4. A. Rescigno, H. Bushe, A. B. Brill, M. Rusckowski, T. W. Griffin and D. J. Hnatowich, Pharmacokinetic Modeling of Radiolabeled Antibody Distribution in Man, *Am. J. Physiologic Imaging* **5**, 141-50 (1990).

This page intentionally left blank

15. VOLUME AND DILUTION FACTOR

15.1. INITIAL VOLUME

If the concentration of the drug in an organ is uniform, the volume of that organ is the ratio between the amount of drug present in it and its concentration.

Consider a simple pharmacokinetic experiment. If a drug is injected as a bolus in an organ at time $t = 0$, call D the dose administered and $c(t)$ its concentration there at time t. If we ignore the short interval of time necessary for the drug to distribute uniformly in the organ, then by extrapolation we get an approximate value of the concentration of the drug had the mixing been instantaneous. Call $c^*(0)$ this extrapolated value; the ratio $D/c^*(0)$ is called *initial volume*. See Fig. 1.

Several observations are necessary at this point. The ratio $D/c^*(0)$ is not necessarily the volume of the sampled organ, or even a physical volume, even though it has the dimension of a volume, $[L^3]$. There may be several reasons for this discrepancy; for instance, the drug may be bound to some other organs before being distributed in the sampled organ, or the mixing may never be complete. In any case if, when repeating the ex-

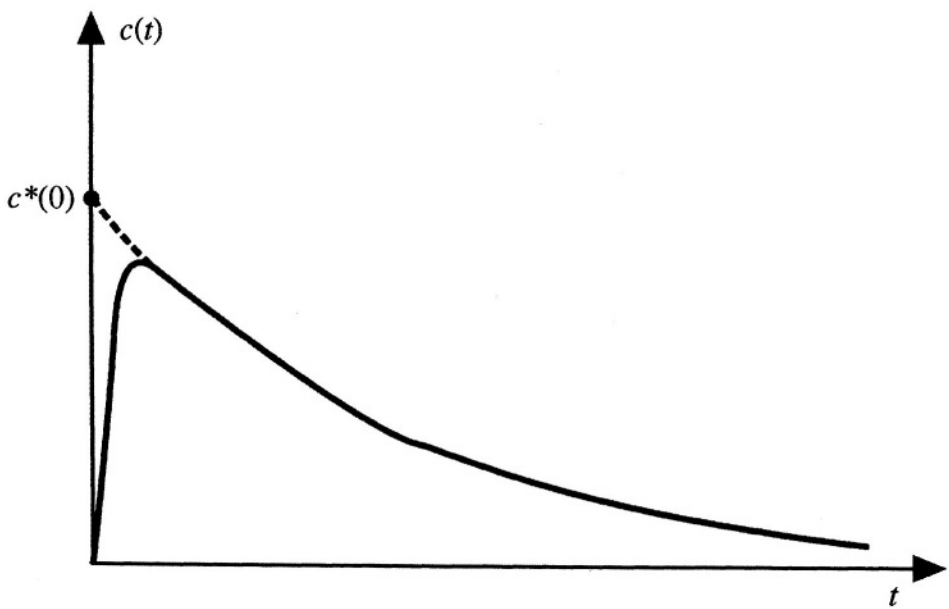

Fig. 1. Plasma concentration after I.V. administration

periment with a different dose D, the concentration $c(t)$ changes in the same proportion, the ratio

$$V_{in} = \frac{D}{c^*(0)} \tag{1}$$

defined above will not change. In this case, and only in this case, we can say that the quantity V_{in} defined by expression (1) is an *invariant,* as explained in chapter 3.

Nevertheless the initial volume, as defined by expression (1), cannot be considered a pharmacokinetic parameter because its value depends on the choice of a particular model, namely upon the hypotheses of rapid and complete mixing, only rarely satisfied, and upon the site of sampling. It is a classical example of a model parameter.

15.2. VOLUME OF DISTRIBUTION.

We have already implicitely given a model-independent definition of the *volume of distribution* V of an organ with expression (1) of chapter 14; it can be rewritten as

$$V = T \cdot Cl. \tag{2}$$

The product (turnover time)×(clearance), called "volume of distribution" by definition, is a pharmacokinetic parameter and it coincides with the initial volume computed with expression (1) if and only if the mixing of the drug in the organ is rapid and complete.

15.3. STEADY-STATE VOLUME

In the first section of this chapter we defined the initial volume in an organ as the ratio between the given dose and the extrapolated value of the concentration of the drug in the sampling organ; we can rewrite that definition as

$$V_{in} = \frac{D}{\lim_{t \to 0} c(t)}, \tag{3}$$

where the limit at the denominator is taken on the modified function $c(t)$.

Consider now the ratio $Q(t)/c(t)$, where $Q(t)$ is the total amount of drug in the body and $c(t)$ the concentration of the sampling compartment, usually the plasma; some authors call it *apparent volume,*

$$V_{app} = \frac{Q(t)}{c(t)}; \tag{4}$$

both numerator and denominator vary with time in a complex way, and we cannot consider V_{app} an invariant quantity. What can be considered an invariant quantity is the limit of V_{app} for $t \to \infty$, if and when such limit exists.

A non-trivial steady state, i.e., a steady state with both $Q(t)$ and $c(t)$ different from zero, may be reached if the system is closed, or if the open system if fed with a constant infusion. In this last case we define the *steady-state volume,* V_{ss}, by

$$V_{ss} = \lim_{t\to\infty} \frac{Q(t)}{c(t)}. \tag{5}$$

In general we can say that

$$V_{in} < V_{app} < V_{ss}$$

as the drug gets distributed from the feeding compartment through other parts of the system.

I will show in the next sections that V_{ss} is an incidental parameter, because its value depends on the site of infusion.

15.4. EXAMPLES

15.4.1. Two compartments fed with a bolus

In the general case of two compartments fed with a single bolus in the first compartment, as shown in section 5.3, we can write

$$\lim_{t\to\infty} \frac{x_2(t)}{x_1(t)} = \lim_{t\to\infty} \frac{k_{12}(e^{-\alpha t} - e^{-\beta t})}{(K_2 - \alpha)e^{-\alpha t} + (\beta - K_2)e^{-\beta t}};$$

without loss of generality we may suppose $\beta > \alpha$; divide numerator and denominator of the right-hand side fraction by $e^{-\alpha t}$, then take the limit,

$$\lim_{t\to\infty} \frac{x_2(t)}{x_1(t)} = \lim_{t\to\infty} \frac{k_{12}(1 - e^{-(\beta-\alpha)t})}{(K_2 - \alpha) + (\beta - K_2)e^{-(\beta-\alpha)t}}$$

$$= \frac{k_{12}}{K_2 - \alpha}.$$

But

$$\alpha = \tfrac{1}{2}\left[K_1 + K_2 - \sqrt{(K_1 - K_2)^2 + 4k_{12}k_{21}}\right],$$

therefore

$$\lim_{t\to\infty} \frac{x_2(t)}{x_1(t)} = \frac{2k_{12}}{K_2 - K_1 + \sqrt{(K_1 - K_2)^2 + 4k_{12}k_{21}}}.$$

This result is valid in general when a steady state is reached, but it is of practical interest only if the steady state is not a trivial one, i.e., if the two compartments are not both empty. With a single bolus administration we have a non-trivial steady state if and only if the system is closed, i.e., when

$$k_{12} = K_1, \quad k_{21} = K_2,$$

consequently,

$$\lim_{t\to\infty}\frac{x_2(t)}{x_1(t)}=\frac{K_1}{K_2}.$$

Now from definition (5) we get

$$V_{ss}=\lim_{t\to\infty}\frac{x_1(t)+x_2(t)}{x_1(t)}V_1=\left(1+\frac{K_1}{K_2}\right)V_1,$$

where V_1 is the volume of the sampling compartment.

15.4.2. Two compartments with continuous infusion

More interesting is the case when there is a constant infusion. If the drug is fed into the sampling compartment, then at steady state we have

$$\begin{aligned}-K_1x_1+k_{21}x_2&=-r\\ +k_{12}x_1-K_2x_2&=0\end{aligned}$$

therefore

$$\begin{aligned}x_1&=\frac{K_2r}{K_1K_2-k_{12}k_{21}}\\ x_2&=\frac{k_{12}r}{K_1K_2-k_{12}k_{21}}\end{aligned}$$

consequently

$$V_{ss}=\left(1+\frac{k_{12}}{K_2}\right)V_1.$$

If the drug is fed into the second compartment, then at steady state we have

$$\begin{aligned}-K_1x_1+k_{21}x_2&=0\\ +k_{12}x_1-K_2x_2&=-r\end{aligned}$$

therefore

$$\begin{aligned}x_1&=\frac{k_{21}r}{K_1K_2-k_{12}k_{21}}\\ x_2&=\frac{K_1r}{K_1K_2-k_{12}k_{21}}\end{aligned}$$

consequently

$$V_{ss}=\left(1+\frac{k_{21}}{K_1}\right)V_1.$$

15.5. DEFINITION OF DILUTION FACTOR

We call *dilution factor,* δ, the ratio of the total amount of drug in the body and its amount in the sampling compartment at steady state [1]; the dilution factor is a dimensionless parameter.

If a steady state is reached with a constant infusion, δ is defined by

$$\delta = \lim_{t \to \infty} \frac{Q(t)}{x_1(t)}, \tag{6}$$

where $Q(t)$ and $x_1(t)$ are the amount of drug in the whole body and in the sampling compartment, respectively, at time t.

By multiplying both sides of definition (6) by V we obtain

$$\delta \cdot V = \lim_{t \to \infty} \frac{Q(t)}{x(t)/V} = \lim_{t \to \infty} \frac{Q(t)}{c(t)},$$

therefore

$$\delta \cdot V = V_{ss}.$$

The dilution factor is an important model parameter that links the total amount of drug in the body with other parameters easily measured in the sampling site; it must be known when planning a chronic therapeutic regimen. In the next sections I will show how it can be determined in a number of cases.

15.6. DILUTION FACTOR AND COMPARTMENTS

15.6.1. Two open compartments

From the results of section 15.4.2 we know that with the model

$$\frac{dx_1}{dt} = -K_1 x_1 + k_{21} x_{21}$$

$$\frac{dx_2}{dt} = +k_{12} x_1 - K_2 x_2$$

where $x_1(t)$ is the sampling compartment, the dilution factor is given by

$$\delta = 1 + \frac{k_{12}}{K_2} \tag{7}$$

if feeding the first compartment, and by

$$\delta = 1 + \frac{k_{21}}{K_1} \tag{8}$$

if feeding the second compartment.

The practical problem is to determine the dilution factor from an experiment done with a bolus administration in the sampling compartment.

Suppose that in a one-bolus experiment we have measured the concentration $c_1(t)$ of a compartment and have found it to be the sum of two exponential functions,

$$c_1(t) = a\,e^{-\alpha t} + b\,e^{-\beta t};$$

in operational notation,

$$\{c_1\} = \frac{a}{s+\alpha} + \frac{b}{s+\beta} = \frac{(a+b)s + a\beta + b\alpha}{s^2 + (\alpha+\beta)s + \alpha\beta};$$

we know from section 5.5 that

$$\begin{aligned} \alpha + \beta &= K_1 + K_2, \\ \alpha \cdot \beta &= K_1 K_2 - k_{12}k_{21}, \\ \frac{a\beta + b\alpha}{a+b} &= K_2. \end{aligned}$$

We have three equations in the four unknown parameters K_1, K_2, k_{12}, k_{21}; we can compute, sequentially,

$$\begin{aligned} K_2 &= \frac{a\beta + b\alpha}{a+b}, \\ K_1 &= \alpha + \beta - K_2, \\ k_{12}k_{21} &= K_1 K_2 - \alpha\beta, \end{aligned}$$

with the explicit solution

$$\begin{aligned} K_1 &= \frac{a\alpha + b\beta}{a+b}, \\ K_2 &= \frac{a\beta + b\alpha}{a+b}, \\ k_{12}k_{21} &= \frac{ab(\alpha-\beta)^2}{(a+b)^2}. \end{aligned}$$

We cannot compute the exact values of the separate transfer rates k_{12} and k_{21}, but we can determine a range for them; in fact knowing that

$$\begin{aligned} 0 \le k_{12} \le K_1, \\ 0 \le k_{21} \le K_2, \end{aligned}$$

the smallest value of k_{12} corresponds to the largest value of k_{21}, and the smallest value of k_{21} to the largest value of k_{12}, therefore

$$\frac{ab(\alpha-\beta)^2}{(a+b)^2} \Big/ \frac{a\beta + b\alpha}{a+b} \le k_{12} \le K_1,$$

$$\frac{ab(\alpha-\beta)^2}{(a+b)^2} \Big/ \frac{a\alpha + b\beta}{a+b} \le k_{21} \le K_2,$$

thence

$$\frac{ab(\alpha-\beta)^2}{(a+b)(a\beta+b\alpha)} \le k_{12} \le \frac{a\alpha+b\beta}{a+b},$$

$$\frac{ab(\alpha-\beta)^2}{(a+b)(a\alpha+b\beta)} \le k_{21} \le \frac{a\beta+b\alpha}{a+b}.$$

If the drug was fed into the first compartment, from (7) we get

$$1+\frac{ab(\alpha-\beta)^2}{(a+b)(a\beta+b\alpha)}\cdot\frac{a+b}{a\beta+b\alpha} \le \delta \le 1+\frac{a\alpha+b\beta}{a+b}\cdot\frac{a+b}{a\beta+b\alpha},$$

thence

$$\frac{(a+b)(a\beta^2+b\alpha^2)}{(a\beta+b\alpha)^2} \le \delta \le \frac{(a+b)(\alpha+\beta)}{a\beta+b\alpha}. \tag{9}$$

If the drug was fed into the second compartment, from (8) we get

$$1+\frac{ab(\alpha-\beta)^2}{(a+b)(a\alpha+b\beta)}\cdot\frac{a+b}{a\alpha+b\beta} \le \delta \le 1+\frac{a\beta+b\alpha}{a+b}\cdot\frac{a+b}{a\alpha+b\beta},$$

thence

$$\frac{(a+b)(a\alpha^2+b\beta^2)}{(a\alpha+b\beta)^2} \le \delta \le \frac{(a+b)(\alpha+\beta)}{a\alpha+b\beta}. \tag{10}$$

15.6.2. Many Compartments

Suppose that the drug is distributed among n compartments, but only one can be sampled; suppose also that the concentration $c(t)$ in that same compartment after a bolus administration at time $t = 0$ can be approximated reasonably well by a sum of exponential functions, then

$$c(t) = A_1e^{-\lambda_1 t} + A_2e^{-\lambda_2 t} + \cdots + A_ne^{-\lambda_n t}$$

and in operational form

$$\{c\} = \frac{A_1}{s+\lambda_1} + \frac{A_2}{s+\lambda_2} + \cdots + \frac{A_n}{s+\lambda_n},$$

or

$$\{c\} = \frac{p_0s^{n-1} + p_1s^{n-2} + \cdots + p_{n-1}}{s^n + q_1s^{n-1} + q_2s^{n-2} + \cdots + q_n}.$$

The dilution factor is given by

$$\delta = \lim_{t\to\infty} \frac{x_1(t) + x_2(t) + \cdots + x_n(t)}{x_1(t)};$$

at steady state the amount of drug in each compartment is proportional to the time spent in it, therefore,

$$\delta = \frac{\text{time spent by the drug in the body}}{\text{time spent in the central compartment}}.$$

The time spent in the sampling compartment, as seen in section 14.6, is given by the 0-moment of $c(t)$ divided by its –1-moment; using B.14 and B.15, the denominator of the above fraction is

$$\frac{\int_0^\infty c(t)\mathrm{d}t}{c(0)} = \frac{p_{n-1}}{p_0 q_n}.$$

We cannot determine the exact value of the numerator, but we can determine a lower and an upper bound for it. We start by observing that the drug spends a minimum time in the body when it is eliminated exclusively from the sampling compartment; the time of elimination from it, as seen in section 9.4, is given by the 1-moment divided by the 0-moment of $c(t)$, i.e.,

$$\frac{\int_0^\infty t \cdot c(t)\mathrm{d}t}{\int_0^\infty c(t)\mathrm{d}t} = \frac{q_{n-1}}{q_n} - \frac{p_{n-2}}{p_{n-1}}.$$

Observe that the term q_{n-1}/q_n is the sum of the products of the turnover rates taken $n-1$ by $n-1$, divided by the product of all turnover rates, therefore

$$\frac{q_{n-1}}{q_n} = \frac{1}{K_1} + \frac{1}{K_2} + \cdots + \frac{1}{K_n},$$

i.e., the time spent in the whole system if there were no recirculation. Incidentally, the term p_{n-2}/p_{n-1}, difference between the time of exit from the sampling compartment and the time spent in all compartments in one passage, is called the *short circuit term* [2]; it represents the time spent in all passages through the system after leaving the sampling compartment.

We can now write

$$\left(\frac{q_{n-1}}{q_n} - \frac{p_{n-2}}{p_{n-1}}\right)\frac{p_0 q_n}{p_{n-1}} \le \delta \le \frac{q_{n-1}}{q_n}\frac{p_0 q_n}{p_{n-1}}$$

or

$$\frac{p_0}{p_{n-1}}\left(q_{n-1} - \frac{p_{n-2} q_n}{p_{n-1}}\right) \le \delta \le \frac{p_0}{p_{n-1}} \cdot q_{n-1}. \qquad (11)$$

15.7. DILUTION FACTOR AND MOMENTS

Inequalities (9) and (10) can be transformed to show how to compute the dilution factor from the moments. When the feeding is in the sampled compartment, inequalities (9) can be written, after dividing numerator and denominator of the left-hand side fraction by $\alpha^2\beta^2$ and numerator and denominator of the right-hand side fraction by $\alpha\beta$, as shown in section 9.8,

$$\frac{X_{-1}X_1}{(X_0)^2} \le \delta \le \frac{X_{-1}\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)}{X_0}. \tag{12}$$

When the feeding is in the second compartment, inequalities (10) can be written,

$$\frac{X_{-1}X_{-3}}{(X_{-2})^2} \le \delta \le \frac{X_{-1}(\alpha+\beta)}{X_{-2}}. \tag{13}$$

It is important to remember that we are considering here a system of two compartments, therefore, as shown in section 9.9, the persymmetric matrices

$$\begin{pmatrix} X_{-3} & X_{-2} & X_{-1} \\ X_{-2} & X_{-1} & X_0 \\ X_{-1} & X_0 & X_1 \end{pmatrix}, \begin{pmatrix} X_{-2} & X_{-1} & X_0 \\ X_{-1} & X_0 & X_1 \\ X_0 & X_1 & X_2 \end{pmatrix}, \begin{pmatrix} X_{-1} & X_0 & X_1 \\ X_0 & X_1 & X_2 \\ X_1 & X_2 & X_3 \end{pmatrix}, \ \ldots,$$

are all singular; this means that the moment X_{-3} in (13) can be expressed in terms of other moments using equation

$$\det\begin{pmatrix} X_{-3} & X_{-2} & X_{-1} \\ X_{-2} & X_{-1} & X_0 \\ X_{-1} & X_0 & X_1 \end{pmatrix} = 0,$$

and that the moment X_{-1} in (12) can be expressed in terms of other moments using equation

$$\det\begin{pmatrix} X_{-1} & X_0 & X_1 \\ X_0 & X_1 & X_2 \\ X_1 & X_2 & X_3 \end{pmatrix} = 0.$$

Inequalities (11), remembering that

$$\sum_{i=1}^{n}\frac{1}{K_i} = \sum_{i=1}^{n}\frac{1}{\lambda_i},$$

can be written

$$\frac{X_{-1}X_1}{(X_0)^2} \le \delta \le \sum_{i=1}^{n}\frac{1}{\lambda_i}\cdot\frac{X_{-1}}{X_0}, \tag{14}$$

of which (12) is just a special case with $n = 2$.

15.8. NON-COMPARTMENTAL SYSTEMS

We have seen that it is impossible to determine the exact value of the dilution factor if we don't have a detailed knowledge of the compartmentalization of the system under observation. We have also seen that a lower bound for the dilution factor is given, in general, by the ratio of exit time over permanence time, i.e.,

$$\delta \geq \frac{\Omega_1}{T_1} = \frac{c(0) \cdot \int_0^\infty t \cdot c(t)\mathrm{d}t}{\left(\int_0^\infty c(t)\mathrm{d}t\right)^2}.$$

This estimate is correct if the drug is eliminated only from the sampling compartment; this is not true in most cases. A better estimate is obtained if we use the apparent exit time Ω^*_{system}, when it is available, because we know that it is closer to the correct exit time from the system, Ω_{system}, than the exit time from the sampling site, Ω_1. Therefore we can write

$$\delta \approx \frac{\Omega^*_{\text{system}}}{T_1} = \frac{c(0) \cdot \int_0^\infty t \cdot x(t)\mathrm{d}t}{\int_0^\infty c(t)\mathrm{d}t \cdot \int_0^\infty x(t)\mathrm{d}t},$$

where $x(t)$ is the total amount of drug present in the system at time t.

Unfortunately, though, we don't know whether this estimate is smaller or larger than the true value.

15.9. EXAMPLES

Christine Matthews [3] studied the metabolic rate and mass of proteins using radioactive iodine as a tracer in human, rabbit and rat. In a typical experiment she found

$$\frac{x(t)}{x(0)} = 0.33e^{0.080t} + 0.53e^{1.00t} + 0.14e^{14.4t}$$

with t measured in hours. In operational form,

$$\frac{\{x\}}{x(0)} = \frac{0.33}{s+0.080} + \frac{0.53}{s+1.00} + \frac{0.14}{s+14.4} = \frac{s^2 + 12.96s + 5.374}{s^3 + 15.48s^2 + 15.63s + 1.152}.$$

Here the plasma is both the feeding and the sampling compartment; using expression (11) we get

$$\frac{1}{5.374}\left(15.63 - \frac{12.96 \cdot 1.152}{5.374}\right) \leq \delta \leq \frac{1}{5.374} \cdot 15.63$$

$$2.39 \leq \delta \leq 2.91.$$

We can also compute the moments of order –1,0, and 1 as shown in section 9.7,

$$X_{-1} = 0.33 + 0.53 + 0.14 = 1.00$$

$$X_0 = \frac{0.33}{0.080} + \frac{0.53}{1.00} + \frac{0.14}{14.4} = 4.66$$

$$X_1 = \frac{0.33}{0.080^2} + \frac{0.53}{1.00^2} + \frac{0.14}{14.4^2} = 52.09$$

and the sum of the inverse of the eigenvalues,

$$\sum_{i=1}^{3} \frac{1}{\lambda_i} = \frac{1}{0.080} + \frac{1}{1.00} + \frac{1}{14.4} = 13.57,$$

then use expression (14) to get the dilution factor, thus

$$\frac{1 \cdot 52.09}{4.66^2} \leq \delta \leq 13.57 \cdot \frac{1}{4.66},$$

yielding the same result.

15.10. REFERENCES

1. J. Mordenti and A. Rescigno, Estimation of Permanence Time, Exit Time, Dilution Factor, and Steady-State Volume of Distribution, *Pharm. Res.* **9**, 17-25 (1992).
2. A. Rescigno and G. Segre, *Drug and Tracer Kinetics* (Blaisdell, Waltham, Mass., 1966), page 162.
3. C. M. E. Matthews, The Theory of Tracer Experiments with ^{131}I-Labelled Plasma Proteins, *Physics Med. Biol.* **2**, 36-53 (1957).

This page intentionally left blank

16. TIME PARAMETERS

16.1. PERMANENCE TIME

16.1.1. Definition of Permanence Time

The permanence time, symbol P, is the expected time spent by a drug administered to an organ in all its passages through that organ. By definition it is equal to the product of the turnover time and the turnover number:

$$P = T \cdot \nu.$$

The pharmacokinetic parameters turnover time and turnover number were defined in sections 14.1 and 14.5; as a consequence the permanence time is another pharmacokinetic parameter, as anticipated in section 3.4.1.

16.1.2. Determination of Permanence Time

If $x(0)$ is the total number of particles introduced into an organ at time $t = 0$ and $x(t)$ the number of particles present there at time t, we have

$x(t)/x(0)$ = fraction of particles present in the organ at time t,

$\frac{x(t)}{x(0)}dt$ = expected time spent in the organ by all particles in the interval $t,t+dt$,

$\int_0^\infty \frac{x(t)}{x(0)}dt$ = average time spent by all particles in the organ.

This last integral, if it converges, is just the permanence time. We can therefore write

$$P = X_0/X_{-1}.$$

Observe that the above derivation is valid even if the drug is not uniformely distributed inside the organ it was measured in; but it is necessary that the integral of $x(t)$ converges. In this sense, even though the permanence time is defined as a pharmacokinetic parameter, its determination depends upon the choice of model.

16.1.3. Permanence Time and Compartments

I have shown in section 11.6 that the diagonal elements of matrix $\boldsymbol{T} = \boldsymbol{K}^{-1}$ are equal to the permanence times of all compartments of a system. If matrix $\boldsymbol{K}$ of a system is known, by inverting it we can know the permanence times of all compartments of the system.

Consider a hypothetical experiment where a drug is given as a bolus in a unit dose in the first compartment; suppose the amount of drug in the first compartment is

$$x_1(t) = 0.33e^{-4t} + 0.67e^{-t}, \tag{1}$$

where t is measured in minutes; we have immediately

$$X_0 = \frac{0.33}{4} + \frac{0.67}{1} = 0.75 \text{ min}$$
$$X_{-1} = 0.33 + 0.67 = 1$$

therefore the permanence time in that compartment is 0.75 min.

With the data available we can determine more parameters; first we write equation (1) in operational form,

$$\{x_1\} = \frac{0.33}{s+4} + \frac{0.67}{s+1} = \frac{s+3}{s^2+5s+4};$$

using the results of section 15.6.1 we can write

$$K_1 + K_2 = 5$$
$$K_1K_2 - k_{12}k_{21} = 4$$
$$K_2 = 3$$

thence

$$K_1 = 2, \quad K_2 = 3,$$
$$k_{12} = \kappa, \quad k_{21} = 2/\kappa,$$

where κ is an unspecified parameter. The matrix $\boldsymbol{K}$ is

$$\boldsymbol{K} = \begin{pmatrix} 2 & -\kappa \\ -2/\kappa & 3 \end{pmatrix}$$

and shows that the turnover rate of the first compartment is 2 min^{-1} while the turnover rate of the second compartment is 3 min^{-1}. By inversion we get

$$\boldsymbol{T} = \begin{pmatrix} 0.75 & \kappa/4 \\ 0.50/\kappa & 0.50 \end{pmatrix},$$

therefore the permanence time of the second compartment is 0.50 min.

16.2. RESIDENCE TIME

The *residence time,* symbol R, is the expected time a drug administered to an organ spends in all its passages through another organ. We can use two subscripts to indicate the fed organ and the sampled organ.

If $x(0)$ is the total number of particles introduced into an organ at time $t = 0$ and $y(t)$ the number of particles present in a second organ at time t, we have

$y(t)/x(0)$ = fraction of particles introduced in the first organ that are present in

the second one at time t,

$\frac{y(t)}{x(0)}dt$ = expected time spent in the second organ in the interval $t, t + dt$, by the particles introduced in the first one,

$\int_0^\infty \frac{y(t)}{x(0)}dt$ = average time spent by all particles in the organ.

This last integral, if it converges, is just the residence time. We can therefore write

$$R_{xy} = Y_0/X_{-1}.$$

As shown in section 11.6, the element of row i and column j of matrix $\boldsymbol{T}$ is equal to the residence times from compartment i to compartment j.

Matrix $\boldsymbol{T}$ in the previous section shows that $R_{12} = \kappa/4$ and $R_{21} = 0.50/\kappa$ are the residence times of the system above.

In the previous section I said that κ is an unspecified parameter, but we can put some bounds on it; we know that

$$0 \le \kappa \le K_1,$$
$$0 \le 2/\kappa \le K_2,$$

therefore

$$2/3 \le \kappa \le 2$$

and

$$0.167 \le R_{12} \le 0.50,$$
$$0.25 \le R_{21} \le 0.75.$$

16.3. YIELD

The fraction of drug fed into an organ that eventually reaches a second organ is called the *yield,* symbol γ. Here again two subscripts can be used to indicate the fed organ and the sampled organ. Evidently,

$$\gamma_{ij} \cdot P_j = R_{ij}.$$

From the example of the previous section we can compute

$$\frac{0.167}{0.50} \le \gamma_{12} \le \frac{0.50}{0.50},$$
$$\frac{0.25}{0.75} \le \gamma_{21} \le \frac{0.75}{0.75},$$

thence

$$0.33 \le \gamma_{12} \le 1,$$
$$0.33 \le \gamma_{21} \le 1.$$

16.4. EXIT TIME

The *exit time* from an organ, symbol Ω, is defined as the interval of time spent by a substance from the time of entry into the system to the time of exit from that organ. While turnover time and permanence time are periods of time spent by a drug in an organ in one single or in many passages through it, the exit time includes the time spent outside that organ before entering it and between different passages through it. In general we can write

$$\text{Turnover time} \leq \text{Permanence time} \leq \text{Exit time}.$$

We have

$$\text{Turnover time} = \text{Permanence time}$$

when the turnover number is one, i.e., when the drug is not recirculated; we have

$$\text{Permanence time} = \text{Exit time}$$

when the drug is fed to the sampled organ and is not recirculated.

The exit time depends upon the site, the mode, and the time, of administration; it is therefore an incidental parameter.

16.4.1. Exit Time from a Compartment

We have seen in section 9.4 that, if $x(t)$ is the amount of drug in a compartment, the 1-moment divided by the 0-moment of $x(t)$ is the exit time from that compartment. Nothing changes, of course, if we use a concentration function instead of an amount-of-drug function, because both the first and the zero moments will then be divided by the same volume.

It is important to remember that the expression

$$\Omega = \frac{\int_0^\infty t \cdot x(t)\mathrm{d}t}{\int_0^\infty x(t)\mathrm{d}t}$$

is not the definition of exit time, but the property of a linear, homogeneous compartment [1,2] and a consequence of the fact that

$$\frac{\int_0^\infty t \cdot K \cdot x(t)\mathrm{d}t}{\int_0^\infty K \cdot x(t)\mathrm{d}t} = \frac{\int_0^\infty t \cdot x(t)\mathrm{d}t}{\int_0^\infty x(t)\mathrm{d}t}$$

only if the turnove rate K is constant and can therefore be exported from the integrals.

In the literature the first relative moment of the drug concentration in the plasma is often called "mean residence time in the systemic circulation" [3]; this is correct only if the drug is distributed uniformely. In general, without any hypotheses on the distribution of the drug in the organ it is measured in, the first relative moment should more correctly be called "age of the drug" [4, 5, 6].

Consider for instance the case of two compartments pooled together. Call $x_i(t)$ and $x_j(t)$ the amount of drug in each of the two compartments, with only their sum

$$x(t) = x_i(t) + x_j(t)$$

accessible to measurement.

We can compute

$$\Omega^*_{i,j} = \frac{\int_0^\infty t \cdot x(t)\,dt}{\int_0^\infty x(t)\,dt} = \frac{\int_0^\infty t\left(x_i(t) + x_j(t)\right)dt}{\int_0^\infty \left(x_i(t) + x_j(t)\right)dt}$$

and call it *apparent* exit time from the pool [7]; I will show that the apparent exit time is different from the real exit time, except in some very special situations.

In fact, following the procedure of section 9.5, the real exit time from the pool should be

$$\Omega_{i,j} = \frac{\int_0^\infty t\left(k_{io}x_i(t) + k_{jo}x_j(t)\right)dt}{\int_0^\infty \left(k_{io}x_i(t) + k_{jo}x_j(t)\right)dt}$$

where k_{io} and k_{jo} are the respective fractional rates of exit of the drug from the two compartments out of the pool. From the above fraction those two constants cannot be eliminated, but we can write

$$\Omega_i = \frac{\int_0^\infty t \cdot k_{io}x_i(t)\,dt}{\int_0^\infty k_{io}x_i(t)\,dt} = \frac{\int_0^\infty t \cdot x_i(t)\,dt}{\int_0^\infty x_i(t)\,dt},$$

$$\Omega_j = \frac{\int_0^\infty t \cdot k_{jo}x_j(t)\,dt}{\int_0^\infty k_{jo}x_j(t)\,dt} = \frac{\int_0^\infty t \cdot x_j(t)\,dt}{\int_0^\infty x_j(t)\,dt}.$$

The difference

$$\Omega_{i,j} - \Omega_i = \frac{\int_0^\infty t\left(k_{io}x_i(t) + k_{jo}x_j(t)\right)dt}{\int_0^\infty \left(k_{io}x_i(t) + k_{jo}x_j(t)\right)dt} - \frac{\int_0^\infty t \cdot k_{io}x_i(t)\,dt}{\int_0^\infty k_{io}x_i(t)\,dt}$$

can be written as

$$\frac{\int_0^\infty t\left(k_{io}x_i(t) + k_{jo}x_j(t)\right)dt \cdot \int_0^\infty k_{io}x_i(t)\,dt - \int_0^\infty \left(k_{io}x_i(t) + k_{jo}x_j(t)\right)dt \cdot \int_0^\infty t \cdot k_{io}x_i(t)\,dt}{\int_0^\infty \left(k_{io}x_i(t) + k_{jo}x_j(t)\right)dt \cdot \int_0^\infty k_{io}x_i(t)\,dt}$$

and after some simplifications,

$$\Omega_{i,j} - \Omega_i = \frac{\int_0^\infty t \cdot k_{jo}x_j(t)dt \cdot \int_0^\infty k_{io}x_i(t)dt - \int_0^\infty k_{jo}x_j(t)dt \cdot \int_0^\infty t \cdot k_{io}x_i(t)dt}{\left(\int_0^\infty k_{io}x_i(t)dt + \int_0^\infty k_{jo}x_j(t)dt\right)\int_0^\infty k_{io}x_i(t)dt} =$$

$$= (\Omega_j - \Omega_i)\frac{\int_0^\infty k_{jo}x_j(t)dt}{\left(\int_0^\infty k_{io}x_i(t)dt + \int_0^\infty k_{jo}x_j(t)dt\right)}.$$

This identity shows that

$$\Omega_{i,j} > \Omega_i \text{ if } \Omega_j > \Omega_i;$$

in the same way we can show that

$$\Omega_{i,j} > \Omega_j \text{ if } \Omega_i > \Omega_j,$$

therefore

$$\mathrm{Min}\left(\Omega_i, \Omega_j\right) \le \Omega_{i,j} \le \mathrm{Max}\left(\Omega_i, \Omega_j\right),$$

where $\mathrm{Min}(\Omega_i, \Omega_j)$ and $\mathrm{Max}(\Omega_i, \Omega_j)$ mean the smaller and the larger of the quantities in parenthesis, respectively.

Now we observe the difference

$$\Omega^*_{i,j} - \Omega_{i,j} = \frac{\int_0^\infty t\left(x_i(t) + x_j(t)\right)\mathrm{d}t}{\int_0^\infty \left(x_i(t) + x_j(t)\right)\mathrm{d}t} - \frac{\int_0^\infty t\left(k_{io}x_i(t) + k_{jo}x_j(t)\right)\mathrm{d}t}{\int_0^\infty \left(k_{io}x_i(t) + k_{jo}x_j(t)\right)\mathrm{d}t}$$

and after some obvious simplifications we find that it has the sign of the product

$$(k_{io} - k_{jo}) \cdot (\Omega_j - \Omega_i).$$

It follows that the apparent exit time and the true exit time coincide when the two compartments have the same exit time or the same fractional rate of exit. When this is not the case, the true exit time will be smaller than the apparent one if the compartment with the larger exit time has the smaller fractional rate of exit, and vice-versa.

In conclusion, without any hypothesis on the compartmentalization of a system, we can only say that the first relative moment is an approximation of the exit time, and the approximation depends on the disuniformity of the concentration of the drug inside the system.

16.4.2. Exit Time from the Organism

An important problem is the determination of the time a drug spends in the whole organism when only the central compartment, say the plasma, can be sampled [7].

Suppose after a bolus administration at time $t = 0$ of a unit dose of a drug we have found that the amount of drug in the plasma can be approximated reasonably well by a sum of exponential functions,

$$x(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} + \cdots + A_n e^{-\lambda_n t};$$

in operational form

$$\{x\} = \frac{A_1}{s + \lambda_1} + \frac{A_2}{s + \lambda_2} + \cdots + \frac{A_n}{s + \lambda_n} = \frac{p_0 s^{n-1} + p_1 s^{n-2} + \cdots + p_{n-1}}{s^n + q_1 s^{n-1} + q_2 s^{n-2} + \cdots + q_n}.$$

As shown in section 9.7, the 0-moment of $x(t)$ is

$$X_0 = \lim_{s\to 0}\{x\} = \frac{p_{n-1}}{q_n}$$

and the 1-moment is

$$X_1 = -\lim_{s\to 0}\frac{d\{x\}}{ds} = \frac{p_{n-1}q_{n-1} - p_{n-2}q_n}{q_n^2},$$

therefore the first relative moment is

$$\frac{X_1}{X_0} = \frac{q_{n-1}}{q_n} - \frac{p_{n-2}}{p_{n-1}}.$$

This is the exit time from the plasma, Ω_{plasma}, if the drug is homogeneously distributed in it. It is also the exit time from the organism if the drug leaves the organism only from the plasma, and not from any peripheral compartments. If the drug spends some time in other compartments before leaving the system, the exit time from the organism, $\Omega_{organism}$, will be larger, therefore

$$\frac{q_{n-1}}{q_n} - \frac{p_{n-2}}{p_{n-1}} \le \Omega_{organism}.$$

Now that we have a lower bound for $\Omega_{organism}$, we shall look for a higher bound. As we saw in section 15.6.2, the term p_{n-2}/p_{n-1}, difference between the time of exit from the sampling compartment and the time spent in all compartments in one passage, is called the *short circuit term* [1, 2]; it represents the time spent in all passages through the system after leaving the sampling compartment. When the short-circuit term is zero, the time spent by the drug in the organism will be maximum, therefore q_{n-1}/q_n is the upper bound of the exit time from the organism:

$$\frac{q_{n-1}}{q_n} - \frac{p_{n-2}}{p_{n-1}} \le \Omega_{organism} \le \frac{q_{n-1}}{q_n}.$$

As shown before,

$$\frac{q_{n-1}}{q_n} = \frac{1}{K_1} + \frac{1}{K_2} + \cdots + \frac{1}{K_n}.$$

16.4.3. Example

We can use the data of section 15.9 for the computation of the exit time from an organism.

From the function

$$\frac{\{x\}}{x(0)} = \frac{s^2 + 12.96s + 5.374}{s^3 + 15.48s^2 + 15.63s + 1.152}$$

we can compute

$$\frac{15.63}{1.152} - \frac{12.96}{5.374} \le \Omega_{organism} \le \frac{15.63}{1.152}$$

and

$$11.16 \le \Omega_{organism} \le 13.57.$$

16.5. TRANSFER TIME

The difference between the exit time from j and the exit time from i, provided that no drug enters j without first passing through i (i.e., i is a unique precursor of j of any order) is called the *transfer time* from compartment i to compartment j, symbol is T_{ij}:

$$T_{ij} = \Omega_j - \Omega_i.$$

We can say that the transfer time is the time spent by a drug from the exit from compartment i to the exit from compartment j.

Even though the definition of transfer time depends on exit time, an incidental parameter, the transfer time is a model parameter, because it is an invariant quantity; in fact, given a specific compartmental model, the difference between exit times in the above definition does not depend upon the site, the mode, and the time of administration, but only on the structure of the model itself.

For instance consider the model described by the graph of Fig. 3 in section 10.8.1; from the data in that section we can write

$$\{x_1\} = \frac{s + K_2}{(s + K_1)(s + K_2) - k_{12}k_{21}} x_0$$

and

$$\{x_3\} = \frac{k_{12}k_{23}}{(s + K_3)\left[(s + K_1)(s + K_2) - k_{12}k_{21}\right]} x_0;$$

as shown in sections 9.4 and 9.7 the exit time from compartment 1 is

$$\Omega_1 = \lim_{s \to 0} \frac{d\{x_1\}/ds}{\{x_1\}} = \frac{\frac{1}{K_1} + \frac{1}{K_2}}{1 - \frac{k_{12}}{K_1} \cdot \frac{k_{21}}{K_2}} - \frac{1}{K_2}$$

and the exit time from compartment 3 is

$$\Omega_3 = \lim_{s \to 0} \frac{d\{x_3\}/ds}{\{x_3\}} = \frac{\frac{1}{K_1} + \frac{1}{K_2}}{1 - \frac{k_{12}}{K_1} \cdot \frac{k_{21}}{K_2}} + \frac{1}{K_3};$$

therefore the transfer time T_{13} is

$$T_{13} = \Omega_3 - \Omega_1 = \frac{1}{K_2} + \frac{1}{K_3}.$$

Observe that the transfer time does not depend on the dose x_0.

After leaving compartment 1 the drug spends the time $1/K_2$ in compartment 2 and the time $1/K_3$ in compartment 3 before leaving this last one.

16.6. TRANSFER TIME AND TRANSFER FUNCTION

There is obviously a very close relationship between transfer time and transfer function. In section 10.2 the transfer function from compartment i to compartment j was defined by

$$\{g_{ij}\} = \frac{\{x_j\}}{\{x_i\}}.$$

Take the logarithms of both sides of that definition,

$$\ln\{g_{ij}\} = \ln\{x_j\} - \ln\{x_i\},$$

then the derivatives with respect to s,

$$\frac{d}{ds}\ln\{g_{ij}\} = \frac{\dfrac{d\{x_j\}}{ds}}{\{x_j\}} - \frac{\dfrac{d\{x_i\}}{ds}}{\{x_i\}},$$

and finally the limits for $s \to 0$,

$$\lim_{s\to 0}\frac{d}{ds}\ln\{g_{ij}\} = \lim_{s\to 0}\frac{\dfrac{d\{x_j\}}{ds}}{\{x_j\}} - \lim_{s\to 0}\frac{\dfrac{d\{x_i\}}{ds}}{\{x_i\}}.$$

Remembering the results of section 9.7 we can now write

$$-\lim_{s\to 0}\frac{\dfrac{d\{x_j\}}{ds}}{\{x_j\}} + \lim_{s\to 0}\frac{\dfrac{d\{x_i\}}{ds}}{\{x_i\}} = \Omega_j - \Omega_i = T_{ij},$$

therefore

$$T_{ij} = -\lim_{s\to 0}\frac{d}{ds}\ln\{g_{ij}\}.$$

Going back to the example of the previous section, we could compute, sequentially,

$$\{g_{13}\} = \frac{\{x_3\}}{\{x_1\}} = \frac{k_{12}k_{23}}{(s+K_2)(s+K_3)},$$

$$\ln\{g_{13}\} = \ln k_{12}k_{23} - \ln(s+K_2) - \ln(s+K_3),$$

$$\frac{d}{ds}\ln\{g_{13}\} = -\frac{1}{s+K_2} - \frac{1}{s+K_3},$$

and finally

$$-\lim_{s\to 0}\frac{d}{ds}\ln\{g_{13}\} = \frac{1}{K_2} + \frac{1}{K_3},$$

the same result as before, but in a much simpler way.

16.7. REFERENCES

1. A. Rescigno and G. Segre. *La Cinetica dei Farmaci e dei Traccianti Radioattivi* (Boringhieri, Torino, 1961).
2. A. Rescigno and G. Segre. *Drug and Tracer Kinetics* (Blaisdell, Waltham, Massachusetts, 1966).
3. K. Yamaoka, T. Nakagawa and T. Uno, Moment analysis for disposition kinetics of several cephalosporin antibiotics in rats, *J. Pharm. Pharmacol.* **35**, 19-22 (1983).
4. A. Rescigno, A. K. Thakur, A. B. Brill, and G. Mariani. Tracer Kinetics: A Proposal for Unified Symbols and Nomenclature, *Phys. Med. Biol.* **35**, 449-465 (1990).
5. A. Rescigno and B. M. Bocchialini, Pharmacokinetics: Unfolding of a Concept, in: *New Trends in Pharmacokinetics,* edited by A. Rescigno and A. K. Thakur (Plenum Press, New York, 1991), pp. 1-25.
6. A. Rescigno, Fundamental Concepts in Pharmacokinetics, *Pharmacol. Res.* **35**, 363-90 (1997).
7. J. Mordenti and A. Rescigno, Estimation of Permanence Time, Exit Time, Dilution Factor, and Steady-State Volume of Distribution, *Pharm. Res.* **9**, 17-25 (1992).

17. BIOAVAILABILITY AND BIOEQUIVALENCE

17.1. DEFINITION OF BIOAVAILABILITY

In 1933 Walker and Nelson [1] observed that vitamin B given to rats as yeast is absorbed differently if the yeast is fresh or dried. This observation marked the beginning of a number of studies that eventually formed the branch of pharmacokinetics called "Bioavailability".

Unfortunately the term bioavailability has not a unique meaning.

Allen et al. [2] define "bioavailability" as the "relative amount of drug which enters the systemic circulation from an administered dosage form and the rate at which the drug appears in systemic circulation". The symbol proposed for the "fraction" of administered dose systemically available is ***f***, but no symbol is proposed for the "rate" of this absorption.

A paper [3] published by a panel under the patronage of the Fédération Internationale de Pharmacie discusses a number of different definitions of "bioavailability", concluding that the definition given by the U.S. Food and Drug Administration (FDA) is sufficiently precise, but might be slightly improved. The original version of the FDA definition is [4]:

> Bioavailability means the rate and extent to which the active drug ingredient or therapeutic moiety is absorbed from a drug product and becomes available at the site of drug action.

The revised definition proposed by the above mentioned panel is:

> Bioavailability means the rate and extent to which the active drug ingredient or therapeutic moiety from a drug product becomes available at the site of drug action or in a biological medium believed to reflect accessibility to a site of action.

The above examples could be multiplied ad libitum. While nobody in the current pharmacokinetic literature contests the official definition of bioavailability, which includes both the fraction absorbed and the rate of this absorption, the common use of the term "bioavailability" refers only to the "fraction absorbed", excluding any reference to the "rate of absorption", which is considered a separate parameter.

Rowland and Tucker [5] avoided this problem by proposing the use of the term *availability*, symbol *F,* for the fraction of drug absorbed, independently of the rate of this process.

17.2. DEFINITION OF BIOEQUIVALENCE

Strictly connected to the concept of bioavailability is the concept of bioequivalence. Here again the official definition accepted by the FDA [4] is,

> "Bioequivalent drug products means pharmaceutical equivalents or pharmaceutical alternatives whose rate and extent of absorption do not show a significant difference when administered at the same molar dose of the therapeutic moiety under similar experimental conditions, either single dose or multiple dose. Bioequivalent means ... drug products whose rate and extent of absorption do not show a significant difference when administered at the same dose under similar conditions."

More recently, the Center for Drug Evaluation and Research, FDA, (CDER), provided a similar definition [6]:

> Bioequivalence is defined ... as "the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study."

Similarly, the U. S. Center for Veterinary Medicine [7],

> Two products are considered to be bioequivalent when they are equally bioavailable; that is, equal in the rate and extent to which the active ingredient(s) or therapeutic ingredient(s) is (are) absorbed and become(s) available at the site(s) of drug action.

The European Agency for the Evaluation of Medicinal Products [8] also provided a comparable definition of product bioequivalence:

> Two medicinal products are bioequivalent if they are pharmaceutically equivalent or pharmaceutical alternatives and if their bioavailabilities after administration in the same molar dose are similar to such degree that their effects, with respect to both efficacy and safety, will be essentially the same. . . . A bioequivalence study is basically a comparative bioavailability study designed to establish equivalence between test and reference products.

Also [9],

> Bioequivalence exists between veterinary medicinal products or between routes of administration if, under identical and appropriate experimental conditions, the bioavailability of the active substance only differs within acceptable limits. Limits must be qualified, a priori, according to the aim of the tests. Bioavailability of a veterinary medicinal product is defined by the rate and extent to which the active substance reaches the systemic circulation and becomes available to the site(s) of action. Rate and extent are typically measured by C_{max} (peak concentration) and AUC (Area under the Curve), respectively.

It is important to note that in all these cases, the definition of bioequivalence is strictly connected with the definition of bioavailability, as given in the previous section. These definitions are based on the tacit, but universally accepted, assumption that [10]

> Two formulations that do not differ very much in the *rate* at and the *extent* to

which they make the active ingredient available in the circulating blood will not differ much in their therapeutic efficacy.

Because of this connection, I shall treat the problems of bioavailability and of bioequivalence together.

17.3. DETERMINATION OF BIOAVAILABILITY

From a scientific point of view, as opposed to the legalistic point of view, what is really important to know is "how" a drug given to an experimental animal or to a volunteer or to a patient, becomes available to the target organ, usually the systemic circulation. This is the first step of the pharmacokinetic investigation. This "how" is nothing more and nothing less than the curve $c(t)$, i.e., *concentration versus time* of the drug in the target organ. This $c(t)$ curve contains all needed information regarding the rate of absorption and the extent of absorption; in short, the $c(t)$ curve tells us how much drug is available and when it is available.

Unfortunately, bioavailability and bioequivalence were born with an original sin. The original sin is in the definition of bioavailability that mentions "rate" and "extent" of absorption as two separate entities; as a consequence, a great effort was spent in trying to determine a value for the rate of absorption and a value for the extent of absorption. In the following sections I shall discuss these two problems, then offer my opinion on their best solution.

17.3.1. Extent of absorption

For the extent of absorption the generally accepted measure is the so-called *AUC,* i.e., the 0-moment of the curve $c(t)$ depicting the systemic drug concentration versus time after a bolus administration.

If $f(t)$ is the rate of drug appearance into the systemic circulation, and $g(t)$ is the rate of its elimination from the central (blood) compartment, provided that the elimination process is linear and invariant (see section 10.1), the concentration $c(t)$ of the drug in the plasma is given by the convolution integral

$$c(t) = \int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau, \tag{1}$$

The corresponding *AUC* is given by equation:

$$\int_0^\infty c(t)\mathrm{d}t = \int_0^\infty \left[\int_0^t f(\tau)\, g(t-\tau)\,\mathrm{d}\tau\right]\mathrm{d}t$$

$$= \int_0^\infty f(t)\,\mathrm{d}t \cdot \int_0^\infty g(t)\,\mathrm{d}t.$$

Observe that $f(t)\mathrm{d}t$ is the amount of drug entering the plasma in the interval of time from t to $t + \mathrm{d}t$, therefore $\int_0^\infty f(t)\mathrm{d}t$ is the total amount of drug absorbed; if D is the administered dose and F is the fraction absorbed, we can write

$$F \cdot D = \int_0^\infty f(t)\mathrm{d}t, \tag{2}$$

therefore

$$AUC = F \cdot D \cdot \int_0^\infty g(t)\mathrm{d}t; \tag{3}$$

in other words, *AUC* is proportional to the fraction of drug absorbed, but only if the elimination process is linear and invariant, and the integral $\int_0^\infty g(t)dt$ is constant.

If the drug is eliminated from the plasma by a non-linear process (a not so rare occurrence) the convolution integral (1) no longer holds and we cannot reach the conclusion of identity (3). In this regard, it is even possible, as shown in section 12.4, that the integral in (2) does not converge at all, and the so-called *AUC* cannot be expressed with a finite number [11,12].

Furthermore there is no a priori guarantee that the elimination rate may not be modified by an excipient; it this happens, the integral $\int_0^\infty g(t)dt$ is not constant from one preparation to another of the same drug product, and equation (3) contains one unknown term too many [13, 14].

The inadequacy of the common measures of extent of absorption is well known [15, 16], but it has not yet been taken care of in the current legislation, as I will show in section 17.4.1.

17.3.2. Rate of absorption

The problem of determining the rate of absorption is even more complex. For one thing, except in some very special cases, there is not a **rate** of absorption, but an absorption **profile**. The absorption of a drug may involve the disaggregation of a formulation, the dissolution of the active molecule, the crossing of membranes along the G.I. tract (a long and diversified tract from the oral cavity to the rectum) [17, 18, 19]; given the vast heterogeneity of the gastrointestinal tract, there is clearly a very low probability that the absorption profile of a compound will be best described as either a simple mono-exponential or even bi-exponential function.

For this reason, we must conclude that in the vast majority of cases, there will be no single parameter that can reliably measure the rate of absorption. The maximal drug concentration, C_{max}, has been widely used as an indicator of the rate of absorption, but it depends more on the fraction absorbed than on the rate of absorption. Similarly, T_{max}, the time of occurrence of the maximal concentration, depends in a complex way on both absorption and elimination rates, and is very ill determined when the plasma drug concentration does not exhibit clearly defined peaks [20]. A typical example of a drug product with poorly defined peaks is zeranol [21].

This obvious fact was recently "rediscovered" by Macheras and Argyrakis [22] who proposed of abandoning the use of C_{max} and T_{max} "in bioequivalence studies for heterogeneous drugs"; I will show in section 17.4.2 that this distinction between heterogeneous and non heterogeneous drugs is superfluous.

17.4. DETERMINATION OF BIOEQUIVALENCE

17.4.1. The present position of the FDA

The Center for Drug Evaluation and Research (CDER), FDA, is aware of the shortcomings of parameters like *AUC*, C_{max} and T_{max} for the evaluation of the equivalence of two medicinal products [23]:

> Both direct (e.g., rate constant, rate profile) and indirect (e.g., *Cmax, Tmax,* mean absorption time, mean residence time, *Cmax* normalized to *AUC)*

> pharmacokinetic measures are limited in their ability to assess rate of absorption. This guidance, therefore, recommends a change in focus from these direct or indirect measures of absorption rate to measures of systemic exposure.

Surprisingly, although the CDER introduces three new parameters, namely *Early Exposure,* measured by *AUC* from 0 to T_{max}, *Peak Exposure,* measured by C_{max}, and *Total Exposure,* measured by *AUC* from 0 to ∞, these three "new" parameters are new only in name, and are not much different from the old parameters *AUC,* C_{max}, T_{max}, whose shortcomings are well recognized.

The simple fact is that equality of *AUC* (either from 0 to T_{max} or from 0 to ∞) and of C_{max} are necessary but not sufficient conditions for establishing bioequivalence [24], and this fact does not disappear by changing the names of the parameters. We can easily think of two plasma concentration curves with very different shapes but the same values of C_{max} and the same values of the respective integrals from 0 to T_{max} and from 0 to ∞ [25]; those two preparations can be considered bioequivalent, (i.e., they have the same biological efficacy) only if one makes the additional hypothesis that the effect of a drug is a linear function of its concentration and of the time that concentration is present. This is a hypothesis implicitly made by Tozer et al. [26] when they introduced the term "exposure", and by the CDER when it proposed to use it. One cannot help thinking of the words of Doctor Pangloss (Voltaire, Candide, Chapter One):

> "Il est démontré, disait-il, que les choses ne peuvent être autrement: car tout étant fait pour une fin, tout est nécessairement pour la meilleure fin."

Unfortunately, in the world we live in, this hypothesis may be true for some drugs but there is no scientific basis to accept it as a universal truth. Upon some reflection, we see that it is no more valid to attempt to assess exposure on the basis of a small number of parameters, than it is to use a single metric for assessing rate. However, for the determination of product bioequivalence, the primary issue is whether "exposure itself", not just total exposure, or peak exposure, or early exposure, is the same for two preparations.

17.4.2. The proposed solution

In 1992 I observed [27] that "to determine that a new formulation is bioequivalent to an old one, it is certainly not necessary to estimate its rate and its extent of absorption, [but] it should be sufficient to show that the two formulations have plasma concentration functions sufficiently similar". As a corollary, I suggested that rather than considering similarity of a number of selected parameters, it would be preferable to use a metric as a measure of the dissimilarity between two plasma concentration curves. When such a metric is sufficiently small (as defined for that drug entity), the "rates" and "extent" of absorption (as well as the "rates of elimination" and any other parameter you care to define) of those two formulations would be declared to be sufficiently close (i.e., bioequivalent). This simple fact was implicitly recognized in one of the Guidelines [28], but was not brought to its full logical consequences:

> Bioequivalence testing aims to demonstrate that two medicinal products produce plasma concentrations similar enough to conclude that the systemic effects of the two products, in respect to efficacy (and possibly safety), are the same.

In very simple terms, for two formulations to be bioequivalent, their time-concentration profiles must be "almost superimposable", that is, the two curves must have the same shape. The attractive feature of this approach is that it applies equally well to homogeneous and to heterogeneous drugs [22], to drug with linear and with non-linear elimination processes, to drugs with well defined C_{max} and with a flat $c(t)$ profile, and so forth.

As a matter of fact, the technique of using a "distance between curves" as a method for deciding if two curves are similar enough is well known in mathematics. Given two functions $f(t)$ and $g(t)$, we can define the function $[f(t) - g(t)]^2$; this function is zero when $f(t)$ and $g(t)$ are identical, and is positive in all other cases; it is also symmetric, i.e., it does not change if the two functions $f(t)$ and $g(t)$ are switched. Its integral from 0 to ∞ is called the "distance of two functions in a Hilbert space" [29] and its properties are analogous to the properties of the "Euclidean distance between points".

Statisticians too have been interested in the problem of deciding whether two curves come from the same distribution. The Kolmogorov-Smirnov "D" statistic [30, 31] has been available to the applied statistician for years as well as the Chi-Square goodness-of-fit procedure along with the Mann Whitney "U" statistic.

Recently Bartoszynski et al. [32] have described a procedure for the comparison of the shape of dissolution curves to determine if they came from the same population of curves; their method can easily be extended to the comparison of time-concentration curves.

17.5. REFERENCES

1. R. Walker and E. M. Nelson, Fresh and Dried Yeast as Source of Vitamin C, *Am. J. Physiol.* **103**, 25-9 (1933).
2. L. Allen, K. Kimura, J. Mackichan and W. A. Ritschel, *Manual of Symbols, Equations & Definitions in Pharmacokinetics* (Committee for Pharmacokinetic Nomenclature of the American College of Clinical Pharmacology, Philadelphia, 1982).
3. L. P. Balant, L. Z. Benet, H. Blume, G. Bozler, D. D. Breimer, M. Eichelbaum, U. Gunder-Remy, J. L. Hirtz, E. Mutschler, K. K. Midha, A. G. Rauws, W. A. Ritschel, L. N. Sansom, J. P. Skelly and K. O. Vollmer, Is there a need for a more precise definition of bioavailability?, *Eur. J. Clin. Pharmacol.* **40**, 123-6 (1991).
4. Bioavailability and Bioequivalence Requirements, *Federal Register,* Vol. **42**, no. 5, pp. 1624-53; January 7, 1977.
5. M. Rowland and G. Tucker, Symbols in Pharmacokinetics, *J. Pharmacokin. Biopharm.* **8**, 497-507 (1980).
6. CDER, *Guidance for Industry, Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations* (U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Rockville, Maryland, October 2000).
7. CVM, *Docket No. 94D-0401, Bioequivalence Guidance* (U. S. Center for Veterinary Medicine, Food and Drug Administration, Rockville, Maryland, October 10,2000).
8. Committee for Proprietary Medicinal Products, *Note for Guidance on the Investigation of Bioavailability and Bioequivalence* CPMP/EWP/QWP/1401/98 (The European Agency for the Evaluation of Medicinal Products, Evaluation of Medicines for Human Use, London, 26 July 2001).
9. Committee for Veterinary Medicinal Products, *Guidelines for the Conduct of Bioequivalence Studies for Veterinary Medicinal Products, EMEA/CVMP/016/00-corr-FINAL.* (The European Agency for the Evaluation of Medicinal Products, Veterinary Medicines and Information Technology, London, 11 July 2001).
10. C. M. Metzler, Bioavailability - A Problem in Equivalence, *Biometrics* **30**, 309-17 (1974)
11. A. Rescigno and A. Marzo, Area Under the Curve, Bioavailability, and Clearance, *J. Pharmacokin. Biopharm,* **19**, 473-82(1991).
12. A. Rescigno, Area Under the Curve and Bioavailability, *Pharmacol. Res.* **42**, 539-40 (2000).
13. A. Marzo and A. Rescigno, Pharmacokinetics of endogenous substances: Some problems and some solutions, *Eur. J. Drug Metab. Pharmacokin.* **18**, 77-88 (1993).

14. A. Rescigno, A. K. Thakur and A. Marzo, On Definition and Use of the Term Bioavailability, *Arzneim.-Forsch.* **44**, 1167-9 (1994).
15. F. Y. Bois, T. N. Tozer, W. W. Hauck, M. L. Chen, R. Patnail and R. L. Williams, Bioequivalence: performance of several measures of extent of absorption, *Pharm. Res.* **11**, 715-22 (1994).
16. R. Schall and H. G. Luus, Comparison of absorption rates in bioequivalence studies of immediate release drug fomulations, *Internat. J. Clin. Pharmacol. Therapy Toxicol.* **30**, 153-9 (1992).
17. A. Rescigno, On Absorption Rate and Fraction Absorbed, *J. Pharmacokin. Biopharm.* **22**, 255-7 (1994).
18. A. Rescigno, Rate of Absorption, *Pharmacol. Res.* **35**, 5-6 (1997).
19. M. Farolfi, J. D. Powers and A. Rescigno, On the determination of bioequivalence, *Pharmacol. Res.* **37**, 93-5 (1998).
20. F. Y. Bois, T. N. Tozer, W. W. Hauck, M. L. Chen, R. Patnail and R. L. Williams, Bioequivalence: performance of several measures of rate of absorption, *Pharm. Res.* **11**, 966-74 (1994).
21. A. E. Pusateri and D. C. Kenison, Measurement of zeranol in plasma from three blood vessles in steers implanted with zeranol, *J. Animal Sci.* **71**, 415-9 (1993).
22. P. Macheras and P. Argyrakis, Gastrointestinal drug absorption: is it time to consider heterogeneity as well as homogeneity?, *Pharm. Res.* **14**, 842-7 (1997).
23. Reference [6], pages 8-9.
24. A. Rescigno and J. D. Powers, AUC and Cmax are not sufficient to prove bioequivalence, *Pharmacol. Res.* **37**, 93-5 (1998).
25. A. Rescigno, J. D. Powers and E. E. Herderick, Bioequivalent or Nonbioequivalent?, *Pharmacol. Res.* **43**, 543-6(2001).
26. T. N. Tozer, F. Y. Bois, W. W. Hauck, M. L. Chen and R. L. Williams, Absorption Rate Vs. Exposure: Which Is More Useful for Bioequivalence Testing?, *Pharm. Res.* **13**, 453-6 (1996).
27. A. Rescigno, Bioequivalence, *Pharm. Res.* **9**, 925-8 (1992).
28. Reference [9], page 2.
29. W. Schmeidler and W. Dreetz, Functional Analysis, in: *Fundamentals of Mathematics, Volume III,* edited by H. Behnke, F. Bachmann, K. Fladt and W. Süss (The MIT Press, Cambridge, Mass., 1974), pp. 391-445.
30. A. Kolmogoroff, Sulla determinazione empirica di una legge di distribuzione, *Giornale dell'Istituto Italiano degli Attuari,* **4**, 83-91 (1933).
31. N. V. Smirnov, On the Estimation of the Discrepancy between Empirical Curves of Distribution for Two Independent Samples (in Russian), *Bulletin ofthe Moscow University*, **2**, 3-16 (1939).
32. R. Bartoszynski, J. D. Powers, E. E. Herderick and J. A. Pultz, Statistical Comparison of Dissolution Curves, *Pharmacol. Res.* **43**, 369-87 (2001).

This page intentionally left blank

18. CONCLUSION

Coming full circle, I return now to the definition of pharmacokinetics; have I helped the reader better to understand how to study absorption, distribution, metabolism, and elimination of drugs? I have introduced a number of definitions and properties of pharmacokinetic parameters; are they of any use in understanding the real world we live in? What is the real meaning of the equations I have written?

In literature —and "scientific literature" is "literature"—the authors must frequently ask the readers for a *willing suspension of disbelief* [1] when they want a semblance of truth sufficient to generate shadows of imagination. Each model used in pharmacokinetics, or in any other experimental science, is a product of our imagination and must have a semblance of truth.

How close is a model to the real world? Each model contains different levels of reality. Let's go back for a moment to equation

$$\frac{dx}{dt} = -K \cdot x(t) + r(t)$$

describing a single compartment (see section 5.2). On the right-hand side there are a negative and a positive term, rate of exit from, and rate of entry into, the compartment; what I had in mind when I wrote that equation was the axiom of conservation of mass. This is the first level of reality, that is, a reality that I accept unconditionally. Somebody could object that another term is necessary because some energy can be transformed into matter, or matter can be transformed into energy, but unless some very strong evidence for that appears, I accept the conservation of mass as an absolute truth. But another level of reality is present in the same equation; that other level of reality is represented by the fact that for the rate of exit I wrote $K \cdot x(t)$, i.e., I supposed that rate to be proportional to the amount of drug present. That is not an absolute level of reality, it is just a hypothesis: the hypothesis is that I am dealing with a perfect compartment. The axiom is one level of reality, the hypothesis is another level of reality, which is contingent to the result of my experiment; the experiment may show that it was a good hypothesis, or it may show that it was a bad hypothesis. I accept it as a contingency.

I can find many other levels of reality. When I solve equations, I get identities; those identities are not just formal expressions, but are instruments to be used to check against the results of the experiment. I move to the laboratory or the clinic and I check the experimental results and they should fit those identities — that is a third level of reality. I don't expect them to match exactly 100%; some random errors are possible. Those possible errors are not written explicitly into the differential equation, but they constitute a

fourth level of reality; the errors should be random, i.e. they should have a certain distribution; whatever distribution I expect is a fourth level of reality [2].

Science, like comedy, often demands that we look at familiar things in unfamiliar ways. What may seem unfamiliar to some readers of this book is its intense use of mathematics; what possible excuses may I adduce? I gave a partial answer to this question in the preface, but for the real reason I cannot do any better than quote what Feynman said in one of his lectures [3]:

> But the real reason is that the subject is enjoyable, and although we human cut nature up in different ways, and we have different courses in different departments, such compartmentalization is really artificial, and we should take our intellectual pleasures where we find them.

REFERENCES

1. S. T. Coleridge, *Biographia Literaria* (Fenner, London, 1817), chapter XIV.
2. A. Rescigno, Pharmacokinetics, Science or Fiction?, P*harmacol. Res.* **33**, 227-33 (1996).
3. R. P. Feynman, R. B. Leighton and M. Sands, *The Feynman Lectures in Physics, Vol. I* (Addison-Wesley, Reading, Mass., 1989), page 22.1.

A. CONVOLUTION

A.1. DISCRETE CONVOLUTION

A.1.1. Memoryless System

A memoryless system is a system where the effects are simultaneous with their causes. In such a system the causes and effects are related in a one-to-one way that can be expressed by a simple mapping:

Cause	a_0	a_1	a_2	a_3	...
Effect	c_0	c_1	c_2	c_3	...

If this system is also linear, we can write

$$c_i = a_i b + d, \quad i = 0, 1, 2, \ldots \tag{1}$$

where b and d are constant parameters characteristic of the system.

Equation (1) can be further simplified if the effects are measured as difference from their rest values, i.e. if we assume that

$$a_i = 0 \text{ implies } c_i = 0;$$

in this case,

$$c_i = a_i b. \quad i = 0, 1, 2, \ldots \tag{2}$$

From now on, we consider only linear systems of the type described by equation (2), i.e. such that to a null cause corresponds a null effect.

A.1.2. Memory System

A memory system is a system whose state depends upon present and past causes. If such system is linear, a cause a_0 at time $t = 0$ may generate the sequence of effects,

At time	$t = 0$	$t = 1$	$t = 2$	$t = 3$	...
the effect is	$a_0 \cdot b_0$	$a_0 \cdot b_1$	$a_0 \cdot b_2$	$a_0 \cdot b_3$	...

If the above system is *invariant,* i.e. at a later time it behaves exactly the same way as at the previous time, then a cause a_1 at time $t = 1$ will generate the sequence of effects,

At time	$t = 1$	$t = 2$	$t = 3$	$t = 4$	...
the effect is	$a_1 \cdot b_0$	$a_1 \cdot b_1$	$a_1 \cdot b_2$	$a_1 \cdot b_3$	...

Similarly a cause a_2 at time $t = 2$ will generate the sequence of effects,

At time	$t=2$	$t=3$	$t=4$	$t=5$	...
the effect is	$a_2 \cdot b_0$	$a_2 \cdot b_1$	$a_2 \cdot b_2$	$a_2 \cdot b_3$	...

The linearity of the system implies that all those effects are additive, therefore a sequence of causes

At time	$t=0$	$t=1$	$t=2$	$t=3$	...
the cause	a_0	a_1	a_2	a_3	...

will generate the effects

At time	$t=0$	$t=1$	$t=2$	$t=3$	...
the effect is	$a_0 \cdot b_0$	$a_0 \cdot b_1 + a_1 \cdot b_0$	$a_0 \cdot b_2 + a_1 \cdot b_1 + a_2 \cdot b_0$	$a_0 \cdot b_3 + a_1 \cdot b_2 + a_2 \cdot b_1 + a_3 \cdot b_0$	...

This result can be summarized by the following statement:

> If a memory system is linear and invariant, and a unit cause at time $t = 0$ generates the sequence of effects
>
> $$b_0, b_1, b_2, b_3, \ldots$$
>
> then the sequence of causes
>
> $$a_0, a_1, a_2, a_3, \ldots$$
>
> generates the sequence of effects
>
> $$a_0 \cdot b_0, a_0 \cdot b_1 + a_1 \cdot b_0, a_0 \cdot b_2 + a_1 \cdot b_1 + a_2 \cdot b_0, , a_0 \cdot b_3 + a_1 \cdot b_2 + a_2 \cdot b_1 + a_3 \cdot b_0, \ldots$$

The above sequence is called the *convolution* of the sequences $a_0, a_1, a_2, a_3, \ldots$ and $b_0, b_1, b_2, b_3, \ldots$

In short we can write

$$\{a_i\} * \{b_i\} = \{c_i\}, \tag{3}$$

where

$$c_i = \sum_{j=0}^{i} a_j \cdot b_{i-j}, \quad i = 0, 1, 2, \ldots \tag{4}$$

The sequence

$$b_0, b_1, b_2, b_3, \ldots$$

is called the *unit response* of the system.

A.1.3. Numerical Convolution

Given the sequences $a_0, a_1, a_2, a_3, \ldots$ and $b_0, b_1, b_2, b_3, \ldots$, their convolution can be computed numerically exactly as in the algebraic multiplication:

a_0	a_1	a_2	...	...	...	...
b_0	b_1	b_2	...	...	...	...
$a_0 \cdot b_0$	$a_1 \cdot b_0$	$a_2 \cdot b_0$	...	...	...	...
	$a_0 \cdot b_1$	$a_1 \cdot b_1$	$a_2 \cdot b_1$	...	...	...
		$a_0 \cdot b_2$	$a_1 \cdot b_2$	$a_2 \cdot b_2$	...	...
			...	...	...	...
c_0	c_1	c_2	c_3	c_4	...	...

A.1.4. Numerical Deconvolution

The deconvolution consists in determining the sequence $\{b_i\}$ given the sequences $\{a_i\}$ and $\{c_i\}$, or determining the sequence $\{a_i\}$ given the sequences $\{b_i\}$ and $\{c_i\}$. The deconvolution is computed as in the algebraic division:

				a_0	a_1	a_2	a_3	
b_0	b_1	b_2	...	c_0	c_1	c_2	c_3	...
				$-a_0b_0$	$-a_0b_1$	$-a_0b_2$	$-a_0b_3$	...
				0	a_1b_0	$c_2-a_0b_2$	$c_3-a_0b_3$	...
					$-a_1b_0$	$-a_1b_1$	$-a_1b_2$	...
					0	a_2b_0	$c_3-a_0b_3-a_1b_2$	...
						...	...	...

A.1.5. Example

Let X and Y be non-negative independent integral-valued random variables with probability distribution $\mathbf{P}\{X = i\} = a_i$ and $\mathbf{P}\{Y = i\} = b_i$. The event $(X = j, Y = k)$ has probability $a_j\, b_k$. The sum $S = X + Y$ is a new random variable, and the event $S = r$ is the sum of the mutually exclusive events

$$(X = 0, Y = r), (X = 1, Y = r - 1), (X = 2, Y = r - 2), \ldots, (X = r, Y = 0).$$

Therefore the distribution $c_r = \mathbf{P}\{S = r\}$ is given by

$$c_r = a_0\, b_r + a_1\, b_{r-1} + a_2\, b_{r-2} + \ldots + a_r\, b_0.$$

A.1.6. Summary

If the unit response of the system is given by the sequence $b_0, b_1, b_2, \ldots$ then the sequence $a_0, a_1, a_2, \ldots$ has the effect given by:

At time	$t=0$	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	...
a_0 causes	$a_0 \cdot b_0$	$a_0 \cdot b_1$	$a_0 \cdot b_2$	$a_0 \cdot b_3$	$a_0 \cdot b_4$	$a_0 \cdot b_5$	...
a_1 causes		$a_1 \cdot b_0$	$a_1 \cdot b_1$	$a_1 \cdot b_2$	$a_1 \cdot b_3$	$a_1 \cdot b_4$	...
a_2 causes			$a_2 \cdot b_0$	$a_2 \cdot b_1$	$a_2 \cdot b_2$	$a_2 \cdot b_3$	...
a_3 causes				$a_3 \cdot b_0$	$a_3 \cdot b_1$	$a_3 \cdot b_2$	...
a_4 causes					$a_4 \cdot b_0$	$a_4 \cdot b_1$	...
a_5 causes						$a_5 \cdot b_0$	...
The total is	c_0	c_1	c_2	c_3	c_4	c_5	...

where

$$c_0 = a_0 \cdot b_0$$
$$c_1 = a_0 \cdot b_1 + a_1 \cdot b_0$$
$$c_2 = a_0 \cdot b_2 + a_1 \cdot b_1 + a_2 \cdot b_0$$
$$c_3 = a_0 \cdot b_3 + a_1 \cdot b_2 + a_2 \cdot b_1 + a_3 \cdot b_0$$
$$c_4 = a_0 \cdot b_4 + a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1 + a_4 \cdot b_0$$
$$c_5 = a_0 \cdot b_5 + a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1 + a_5 \cdot b_0$$

and so forth.

A.2. CONTINUOUS CONVOLUTION

A.2.1. Linear Invariant Systems

We have seen in chapter 10 that we can consider a particle in a living system and suppose that that particle can be recognized in two different states of the system, where by state we mean a particular location or a particular chemical form, or both. If one state is the precursor of the other (not necessarily the immediate precursor), then we can study the relationship among events **A** (the particle is in the precursor state), event **B** (transiti on from precursor to successor state), and event **C** (presence of the particle in the successor state).

If we suppose now that **B** depends only on the interval of time separating **A** and **C**, we can call $A(\tau)$ the probability of **A** at time τ, $B(t-\tau)d\tau$ the conditional probability that a particle is in **C** at time t if it left **A** in the interval τ, $\tau + d\tau$, and $C(t)$ the probability of **C** at time t. As shown in section 10.1, with the above hypotheses the relationship among the functions $A(t)$, $B(t)$, $C(t)$ is

$$\int_0^t A(\tau) \cdot B(t-\tau)\,d\tau = C(t); \tag{5}$$

this is the well known *convolution integral* representing the relationship among the variables of a *linear, invariant* system.

By linear system we mean that two different solutions of equation (5) can be added to give a new solution; by invariant system we mean that a solution does not change if the time origin is changed. These two properties, i.e. linearity and invariance, of equation (5), are often called the *principle of superposition.* For a proof see section 10.1.

If we think of an experiment where a very large number of identical particles is used, then the number of particles present in the precursor and in the successor states are good estimates of functions $A(t)$ and $C(t)$ respectively. Function $B(t)$ represents the probability that a particle that left **A** at time 0 will still be in **C** at time t; therefore in a hypothetical

experiment where all identical particles left the precursor near time zero, the number of particles found in the successor will be given by $B(t)$.

A.2.2. Properties of the convolution

For short, we can write

$$\int_0^t A(\tau)B(t-\tau)\mathrm{d}\tau = A(t) * B(t).$$

It is easy to prove that the convolution integral is a *linear operation,* i.e. that if

$$A_1(t) * B(t) = C_1(t) \text{ and } A_2(t) * B(t) = C_2(t),$$

then

$$\left[A_1(t) + A_2(t)\right] * B(t) = C_1(t) + C_2(t).$$

It is also *commutative,*

$$A(t) * B(t) = B(t) * A(t),$$

associative,

$$\left[A(t) * B(t)\right] * C(t) = A(t) * \left[B(t) * C(t)\right],$$

and *distributive with respect to addition,*

$$A(t) * \left[B(t) + C(t)\right] = A(t) * B(t) + A(t) * C(t).$$

This page intentionally left blank

B. OPERATIONAL CALCULUS

B.1. CONVOLUTION ALGEBRA

Manipulation of the convolution integral is considerably simplified by the use of the operational calculus as developed by Mikusinski [1]. An alternative approach to operational calculus was due to Erdelyi [2], and it was extended to two variables by Rescigno [3].

If $\boldsymbol{C}$ is the class of all real-valued, defined and continuous functions of the variable $t \geq 0$, we shall use the symbol $\{f\}$ to represent the function f of class $\boldsymbol{C}$, while with $f(t)$ we represent the value of that function for a particular value t of the independent variable.

We define the operations of addition and multiplication of functions by

$$\{f\} + \{g\} = \{f(t) + g(t)\}, \tag{1}$$

$$\{f\} \cdot \{g\} = \left\{\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau\right\}, \tag{2}$$

It is easy to verify that the addition is closed, i.e., the sum of two functions of class $\boldsymbol{C}$ is always a function of class $\boldsymbol{C}$; it is commutative, i.e.

$$\{f\} + \{g\} = \{g\} + \{f\},$$

and it is associative, i.e.

$$(\{f\} + \{g\}) + \{h\} = \{f\} + (\{g\} + \{h\}).$$

For any $\{f\}$ there exists an additive identity, i.e.

$$\{f\} + \{0\} = \{f\},$$

and an additive inverse, i.e.

$$\{f\} + \{-f\} = \{0\}.$$

The multiplication of functions too is closed, commutative and associative; in fact the convolution of two functions of class $\boldsymbol{C}$ is a function of class $\boldsymbol{C}$; the commutativity of the multiplication is proved by writing

$$\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau = \int_t^0 f(t-\sigma)g(\sigma)(-\mathrm{d}\sigma) = \int_0^t g(\tau)f(t-\tau)\mathrm{d}\tau;$$

to prove the associativity put

$$\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau = \varphi(t), \quad \int_0^t g(\tau)h(t-\tau)\mathrm{d}\tau = \psi(t);$$

then write

$$\begin{aligned}\int_0^t \varphi(\tau)h(t-\tau)\mathrm{d}\tau &= \int_0^t \left(\int_0^\tau f(\sigma)g(\tau-\sigma)\mathrm{d}\sigma\right)h(t-\tau)\mathrm{d}\tau \\ &= \int_0^t f(\sigma)\left(\int_\sigma^t g(\tau-\sigma)h(t-\tau)\mathrm{d}\tau\right)\mathrm{d}\sigma \\ &= \int_0^t f(\sigma)\left(\int_0^{t-\sigma} g(\rho)h(t-\sigma-\rho)\mathrm{d}\rho\right)\mathrm{d}\sigma \\ &= \int_0^t f(\tau)\psi(t-\tau)\mathrm{d}\tau, \quad \text{q.e.d.}\end{aligned}$$

Finally it is easy to prove that multiplication is distributive with respect to addition, i.e.

$$\{f\}\cdot(\{g\} + \{h\}) = \{f\}\cdot\{g\} + \{f\}\cdot\{h\}.$$

The functions of class $\boldsymbol{C}$ with the two operations (1) and (2) therefore constitute a *commutative ring*.

It is also useful to define the *power* of a function, thus

$$\{f\}^2 = \{f\}\cdot\{f\},$$
$$\{f\}^{n+1} = \{f\}\cdot\{f\}^n. \quad n = 2, 3, \ldots$$

B.2. CONSTANT FUNCTION {1}

From the above identities we get

$$\{1\}^2 = \{t\},$$
$$\{1\}^3 = \{t^2/2\},$$

and in general

$$\{1\}^n = \left\{\frac{t^{n-1}}{(n-1)!}\right\} \tag{3}$$

for any positive integer n.

This last identity can be generalized to non-integer powers of n using the definition of the Euler's gamma integral,

$$\Gamma(\lambda) = \int_0^\infty t^{\lambda-1}e^{-t}\mathrm{d}t \tag{4}$$

which has the following important properties:

$$\Gamma(\lambda+1) = \lambda\cdot\Gamma(\lambda), \tag{5}$$

$$\Gamma(n) = (n-1)! \quad n \text{ a positive integer}, \tag{6}$$

$$\frac{\Gamma(\lambda)\cdot\Gamma(\mu)}{\Gamma(\lambda+\mu)} = \int_0^1 t^{\lambda-1}(1-t)^{\mu-1}\,dt, \tag{7}$$

$$\Gamma(1/2) = \sqrt{\pi}. \tag{8}$$

Property (5) can be proved with an integration by parts,

$$\Gamma(\lambda+1) = \int_0^\infty t^\lambda e^{-t}dt = \lambda\int_0^\infty t^{\lambda-1}e^{-t}dt = \lambda\cdot\Gamma(\lambda).$$

Property (6) is proved by induction after showing that

$$\Gamma(1) = \int_0^\infty e^{-t}dt = 1.$$

To prove property (7) we first show that

$$\Gamma(\lambda)\cdot\Gamma(\mu) = \iint_D x^{\lambda-1}y^{\mu-1}e^{-x}dxdy,$$

with the integral extended to the domain D: $x > 0$, $y > 0$. Now with the substitutions $x = t\cdot u$, $y = (1-t)\cdot u$ we get

$$\Gamma(\lambda)\cdot\Gamma(\mu) = \iint_{D'} t^{\lambda-1}(1-t)^{\mu-1}u^{\lambda+\mu-1}e^{-u}dtdu,$$

with the integral extended to the domain D': $0 < t < 1$, $u > 0$. It follows

$$\begin{aligned}\Gamma(\lambda)\cdot\Gamma(\mu) &= \int_0^1 t^{\lambda-1}(1-t)^{\mu-1}dt\cdot\int_0^\infty u^{\lambda+\mu-1}e^{-u}du\\ &= \int_0^1 t^{\lambda-1}(1-t)^{\mu-1}dt\cdot\Gamma(\lambda+\mu+1),\ \text{q.e.d.}\end{aligned}$$

Using property (7) we can write

$$\Gamma(1/2)\cdot\Gamma(1/2) = \frac{\Gamma(1/2)\cdot\Gamma(1/2)}{\Gamma(1/2+1/2)} = \int_0^1 \frac{dt}{\sqrt{t(1-t)}} = \pi,$$

thence property (8) follows immediately.

Now we can generalize the result of identity (3) by writing

$$\{1\}^\lambda = \left\{\frac{t^{\lambda-1}}{\Gamma(\lambda)}\right\}$$

for all positive values of λ. This equality is to be regarded as a definition of the operator $\{1\}^\lambda$. This definition preserves the fundamental property of power, i. e.,

$$\{1\}^\lambda\cdot\{1\}^\mu = \{1\}^{\lambda+\mu},\quad \lambda, \mu > 0.$$

B.3. QUOTIENT OF FUNCTIONS

According to the theorem of Titchmarsh [4], the product $\{f\}\cdot\{g\}$ is identically zero if at least one of the two factors $\{f\}$, $\{g\}$, is identically zero.

Now we can define the quotient of two functions

$$\frac{\{f\}}{\{g\}} = \{h\}$$

if $\{g\} \neq \{0\}$ and if a function $\{h\}$ exists such that

$$\{f\} = \{g\}\cdot\{h\}.$$

If the quotient of two functions exists, it is unique; in fact suppose that we can find a second function $\{h_1\}$ such that

$$\{f\} = \{g\}\cdot\{h_1\};$$

subtracting the last two identities one from the other,

$$\{g\}\cdot\{h\} - \{g\}\cdot\{h_1\} = \{0\};$$

for the distributive property,

$$\{g\}\cdot\big(\{h\} - \{h_1\}\big) = \{0\};$$

but by hypothesis $\{g\} \neq \{0\}$, therefore from the theorem of Titchmarsh it follows that $\{h\}$ and $\{h_1\}$ are identical.

Not for all pairs of functions does the quotient exists. To make this operation always possible, with the only restriction that the divisor be different from zero, we define the *operator,* represented by two functions separated by a bar, with the following defining properties:

$$\frac{\{f\}}{\{g\}} = \frac{\{\varphi\}}{\{\psi\}} \quad \text{if and only if} \quad \{f\}\cdot\{\psi\} = \{g\}\cdot\{\varphi\}, \tag{9}$$

$$\frac{\{f\}}{\{g\}} + \frac{\{\varphi\}}{\{\psi\}} = \frac{\{f\}\cdot\{\psi\} + \{g\}\cdot\{\varphi\}}{\{g\}\cdot\{\psi\}}, \tag{10}$$

$$\frac{\{f\}}{\{g\}} \cdot \frac{\{\varphi\}}{\{\psi\}} = \frac{\{f\}\cdot\{\varphi\}}{\{g\}\cdot\{\psi\}}. \tag{11}$$

Definition (9) shows that the same operator can be written in different forms; for instance $\{2t\}/\{1\}$ and $\{t^2\}/\{t\}$ are the same operator because $\{2t\}\cdot\{t\} = \{1\}\cdot\{t^2\}$.

Definitions (10) and (11) are formally identical with the addition and multiplication of rational numbers; it is easy to verify that they are commutative, associative and distributive, as the addition and multiplication of functions of class $\boldsymbol{C}$.

Furthermore given any two operators $\{f\}/\{g\}$ and $\{\varphi\}/\{\psi\}$, with $\{\varphi\} \neq \{0\}$, we can always find their quotient; in fact

$$\frac{\{f\}/\{g\}}{\{\varphi\}/\{\psi\}} = \frac{\{f\}\cdot\{\psi\}}{\{g\}\cdot\{\varphi\}},$$

as can be proved by using definition (9).

The operators thus defined with the three operations (9), (10), (11) constitute a *field.* We can operate on them exactly as on ordinary fractions.

For example from the definition of equality,

$$\frac{\{f\}\cdot\{\varphi\}}{\{g\}\cdot\{\varphi\}}=\frac{\{f\}}{\{g\}},$$

i.e. numerator and denominator of an operator can be divided by the same function, as in the reduction of ordinary fractions.

From the definition of quotient,

$$\{f\}=\frac{\{f\}\cdot\{g\}}{\{g\}},$$

where the left hand side is a function; the right hand side too is a function, but it can also be looked at as an operator. It follows that any function can be looked at as an operator, but not the other way around; in other words the ring of functions is *embedded* in the field of operators.

B.4. NUMERICAL OPERATORS

Observe that

$$\frac{\{a\cdot f\}}{\{f\}}=\frac{\{a\cdot g\}}{\{g\}},$$

where a is a constant and $\{f\}$, $\{g\}$ are arbitrary functions; this operator, being independent of the function $\{f\}$, can be represented simply by the letter a, i.e. we put by definition

$$\frac{\{a\cdot f\}}{\{f\}}=a,$$

and call it the *numerical operator.*

It is easy to verify that

$$a\cdot\{f\}=\{a\cdot f\};$$

in fact, by definition,

$$a\cdot\{f\}=\frac{\{a\cdot\varphi\}}{\{\varphi\}}\cdot\{f\}=\frac{\{a\cdot\varphi\}\cdot\{f\}}{\{\varphi\}}=\frac{\left\{\int_0^t a\varphi(\tau)\cdot f(t-\tau)\mathrm{d}\tau\right\}}{\{\varphi\}}$$
$$=\frac{\left\{\int_0^t \varphi(\tau)\cdot af(t-\tau)\mathrm{d}\tau\right\}}{\{\varphi\}}=\frac{\{\varphi\}\cdot\{af\}}{\{\varphi\}}=\{af\}.$$

An important property of the numerical operator 0 is

$$0=\{0\};$$

in fact

$$0=\frac{\{0\}}{\{f\}}=\frac{\{0\}\cdot\{f\}}{\{f\}}=\{0\};$$

this is the only numerical operator that can be written with or without brackets.

Sum, product, and quotient of numerical operators have exactly the same formal properties as sum, product, and quotient of real numbers. In fact, the sum of two numerical operators is

$$\frac{\{af\}}{\{f\}} + \frac{\{bg\}}{\{g\}} = \frac{\left\{a\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau\right\} + \left\{b\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau\right\}}{\left\{\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau\right\}} = a+b;$$

their product is

$$\frac{\{af\}}{\{f\}} \cdot \frac{\{bg\}}{\{g\}} = \frac{\left\{ab\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau\right\}}{\left\{\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau\right\}} = a \cdot b;$$

their quotient is

$$\frac{\{af\}}{\{f\}} \Big/ \frac{\{bg\}}{\{g\}} = \frac{\left\{a\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau\right\}}{\left\{b\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau\right\}} = \frac{a}{b}.$$

B.5. DIFFERENTIAL OPERATOR

The *inverse* of the operator $\{f\}/\{g\}$, with $\{f\} \neq \{0\}$, is the operator $\{g\}/\{f\}$. Obviously the product of an operator by its inverse is the numerical operator 1.

Because of identity

$$\{1\} \cdot \{f\} = \left\{\int_0^t f(\tau)\mathrm{d}\tau\right\},$$

we can call the constant function $\{1\}$, *integral operator.* Call s its inverse; then

$$\{1\} \cdot s = 1.$$

Now consider a function $f(t)$ having a derivative $f'(t)$; we can write

$$\int_0^t f'(\tau)\mathrm{d}\tau = f(t) - f(0)$$

or

$$\{1\} \cdot \{f'\} = \{f\} - \{f(0)\};$$

divide both sides by $\{1\}$,

$$\{f'\} = s\{f\} - f(0). \tag{12}$$

The derivative of a function can thus be expressed in terms of the original function, its initial value, and the operator s. For this reason s can be called *differential operator.*

If $\{f\}$ does not have a derivative, then the expression

$$s \cdot \{f\} - f(0)$$

is not equal to any function, but it is a well defined operator with a number of properties formally similar to the properties of a derivative; we can conventionally call it "derivative of $\{f\}$".

By induction we can prove that, if $\{f\}$ has higher derivatives, then

$$\left\{f^{(n)}\right\} = s^n\{f\} - s^{n-1}f(0) - s^{n-2}f'(0) - s^{n-3}f''(0) - \cdots - f^{(n-1)}(0).$$

B.6. FUNCTIONAL CORRELATES

From the previous definition,

$$\{1\} = \frac{1}{s}; \tag{13}$$

thence, multiplying both sides repeatedly by $\{1\}$,

$$\{t\} = \frac{1}{s^2}, \tag{14}$$

$$\{t^2\} = \frac{2}{s^3},$$

and in general,

$$\{t^n\} = \frac{n!}{s^{n+1}}. \tag{15}$$

From the identities

$$\frac{de^{\alpha t}}{dt} = \alpha \cdot e^{\alpha t},\ e^0 = 1,$$

using (12) we get

$$\left\{\alpha \cdot e^{\alpha t}\right\} = s \cdot \left\{e^{\alpha t}\right\} - 1,$$

thence

$$\left\{e^{\alpha t}\right\} = \frac{1}{s-\alpha}. \tag{16}$$

From the identities

$$\frac{d\sin\beta t}{dt} = \beta \cdot \cos\beta t, \quad \frac{d\cos\beta t}{dt} = -\beta \cdot \sin\beta t$$

$$\sin 0 = 0, \quad \cos 0 = 1,$$

we get

$$\{\sin\beta t\} = \frac{\beta}{s^2+\beta^2}, \tag{17}$$

$$\{\cos \beta t\} = \frac{s}{s^2 + \beta^2}. \tag{18}$$

Note that all expressions thus found represent the fact that the functions listed in them are solutions of differential equations with given initial conditions. Thus from the differential equation

$$a \cdot f''(t) + b \cdot f'(t) + c \cdot f(t) = g(t)$$

with initial conditions

$$f(0) = p, f'(0) = q,$$

we get

$$a\left(s^2\{f\} - s \cdot p - q\right) + b\left(s\{f\} - p\right) + c\{f\} = \{g\},$$

thence

$$\{f\} = \frac{a(ps + q) + bp + \{g\}}{as^2 + bs + c}.$$

For instance identity (17) corresponds to the differential equation

$$f''(t) + \beta^2 f(t) = 0$$

with initial conditions

$$f(0) = 0, \quad f'(0) = \beta;$$

similarly identity (13) simply means that $\{1\}$ is the particular integral of the differential equation

$$f'(t) = 0$$

with initial condition

$$f(0) = 1;$$

identity (14) shows that $\{t\}$ is the particular integral of the differential equation

$$f''(t) = 0$$

with initial conditions

$$f(0) = 0, f'(0) = 1.$$

B.7. DISCONTINUOUS FUNCTIONS

So far we have considered operators defined in terms of functions of class $\mathcal{C}$. Now consider the class $\mathcal{K}$ of real functions $f(t)$ defined for any $t \geq 0$, such that

$$\int_0^t f(\tau)\, d\tau$$

is of class $\mathcal{C}$, plus those having a finite number of discontinuities in any finite interval of t and such that

$$\int_0^t |f(\tau)|\,d\tau$$

is finite in the whole domain of t.

Thus if $\{f\}$ is of class $\mathcal{K}$ we can write

$$\{f\} = \frac{\left\{\int_0^t f(\tau)\,d\tau\right\}}{\{1\}},$$

i.e. any function of class $\mathcal{K}$ can be written as a ratio of two functions of class $\mathcal{C}$; in other words the functions of class $\mathcal{K}$ are embedded in the field of operators defined by functions of class $\mathcal{C}$.

As an example of a function of class $\mathcal{K}$ consider the *jump function* (see Fig. 1)

$$H_\lambda(t) = 0 \quad \text{for } 0 \le t < \lambda$$
$$= 1 \quad \text{for } t \ge \lambda.$$

Define the *translation operator*

$$h_\lambda = s \cdot \{H_\lambda\};$$

for any function $\{f\}$ we have

$$h_\lambda\{f\} = s \cdot \{H_\lambda\}\{f\}$$
$$= s \cdot \left\{\int_0^t H_\lambda(t-\tau) f(\tau)\,d\tau\right\};$$

for the definition of H_λ this integral vanishes for $\tau > t - \lambda$, therefore

$$h_\lambda\{f\} = s \cdot \left\{\int_0^{t-\lambda} f(\tau)\,d\tau\right\} \quad \text{for } t \ge \lambda$$
$$= 0 \qquad\qquad\qquad\quad \text{for } t < \lambda,$$

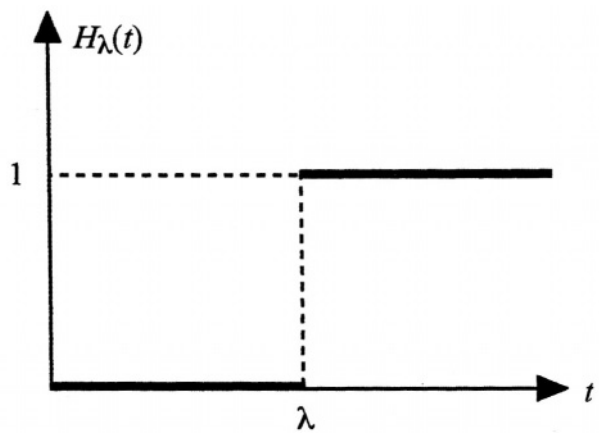

Fig. 1. The jump function

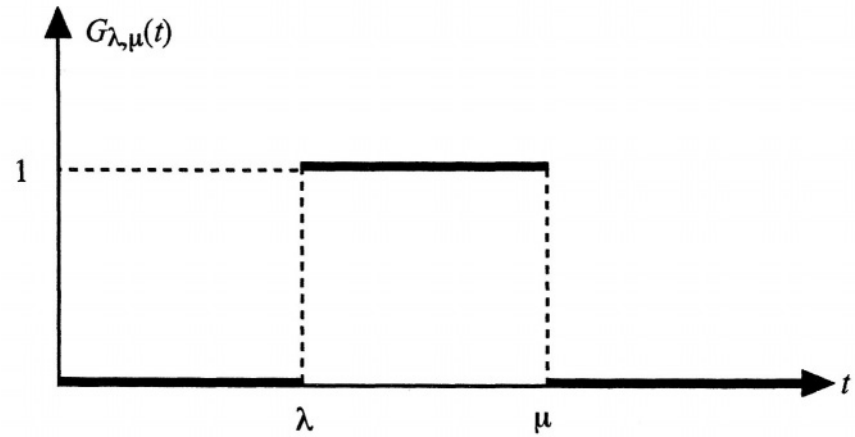

Fig. 2. The gate function

and finally,

$$h_\lambda\{f\} = 0 \qquad \text{for } t < \lambda$$
$$= \{f(t-\lambda)\} \text{ for } t \geq \lambda.$$

This result shows that the operator h_λ shifts a function $f(t)$ by a quantity λ along the t axis and makes it zero for $t < \lambda$, thence its name.

The jump function can thus be written in operational form,

$$\{H_\lambda(t)\} = \frac{h_\lambda}{s}.$$

Many other functions can be written in operational form using the translation operator. For instance the *gate Junction* (see Fig. 2)

$$G_{\lambda,\mu}(t) = 0 \quad \text{for } 0 \leq t < \lambda \text{ or } t > \mu$$
$$= 1 \quad \text{for } \lambda \leq t \leq \mu$$

can be written

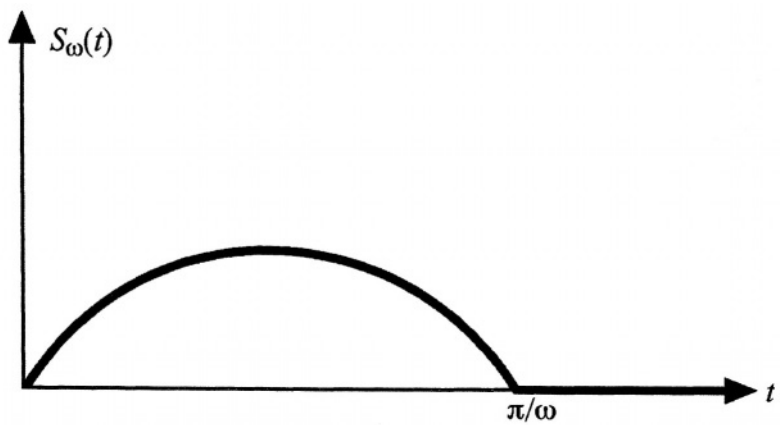

Fig. 3. The single sinusoidal pulse

$$\{G_{\lambda,\mu}\} = \frac{h_\lambda - h_\mu}{s}.$$

A *single sinusoidal pulse* (see Fig. 3)

$$S_\omega(t) = \sin \omega t \text{ for } 0 \le t \le \pi/\omega$$
$$= 0 \qquad \text{for } t > \pi/\omega$$

can be written

$$\{S_\omega\} = \omega \cdot \frac{1 - h_{\pi/\omega}}{s^2 + \omega^2}.$$

B.8. CONTINUOUS DERIVATIVE OF AN OPERATOR

Consider an operator $f(\lambda)$ depending upon the parameter λ; if we can write

$$f(\lambda) = p \cdot \{f_1(\lambda, t)\}, \tag{19}$$

where p is an operator not depending upon λ, and $f_1(\lambda, t)$ is a function such that $\partial f_1/\partial\lambda$ exists and is continuous for $t \ge 0$ and for a certain domain of λ, then we can write

$$\frac{df}{d\lambda} = p \cdot \left\{\frac{\partial f_1}{\partial \lambda}\right\}, \tag{20}$$

called the *continuous derivative* of the operator $f(\lambda)$.

The continuous derivative thus defined is unique. In fact, suppose that $f(\lambda)$ can be represented by (19) and also by

$$f(\lambda) = q \cdot \{f_2(\lambda, t)\};$$

in this case we have

$$p \cdot \left\{\frac{\partial f_1}{\partial \lambda}\right\} = q \cdot \left\{\frac{\partial f_2}{\partial \lambda}\right\}.$$

Now chose a function $\phi(t)$ such that

$$p = \frac{\{g_1\}}{\{\phi\}}, \quad q = \frac{\{g_2\}}{\{\phi\}},$$

with $g_1(t)$ and $g_2(t)$ functions of class C. It follows,

$$\{g_1\} \cdot \{f_1\} = \{g_2\} \cdot \{f_2\},$$
$$\left\{\int_0^t g_1(\tau) f_1(\lambda, t-\tau)\, d\tau\right\} = \left\{\int_0^t g_2(\tau) f_2(\lambda, t-\tau)\, d\tau\right\},$$
$$\left\{\int_0^t g_1(\tau) \cdot \frac{\partial}{\partial\lambda} f_1(\lambda, t-\tau)\, d\tau\right\} = \left\{\int_0^t g_2(\tau) \cdot \frac{\partial}{\partial\lambda} f_2(\lambda, t-\tau)\, d\tau\right\},$$

$$\{g_1\} \cdot \left\{\frac{\partial f_1}{\partial \lambda}\right\} = \{g_2\} \cdot \left\{\frac{\partial f_2}{\partial \lambda}\right\};$$

finally dividing both sides by $\{\phi\}$ we obtain

$$p \cdot \left\{\frac{\partial f_1}{\partial \lambda}\right\} = q \cdot \left\{\frac{\partial f_2}{\partial \lambda}\right\}, \text{ q.e.d.}$$

Consider now the translation operator h_λ; we have seen that it can be written in the form

$$h_\lambda = s \cdot \{H_\lambda(t)\},$$

but $H_\lambda(t)$ does not have a continuous partial derivative with respect to λ, therefore we cannot use definition (19) in this case. Define the two functions

$$H^*_\lambda(t) = 0 \quad \text{for } 0 \le t < \lambda$$
$$= t - \lambda \quad \text{for } t \ge \lambda$$

and

$$H^{**}_\lambda(t) = 0 \quad \text{for } 0 \le t < \lambda$$
$$= \tfrac{1}{2}(t-\lambda)^2 \quad \text{for } t \ge \lambda$$

and observe that

$$\{H_\lambda(t)\} = s \cdot \{H^*_\lambda(t)\} = s^2 \cdot \{H^{**}_\lambda(t)\},$$

$$\left\{\frac{\partial H^{**}_\lambda(t)}{\partial \lambda}\right\} = -\{H^*_\lambda(t)\}.$$

The translation operator can alternatively be written in the form

$$h_\lambda = s^3 \cdot \{H^{**}_\lambda(t)\},$$

and now definition (20) can be used; thence

$$\frac{dh_\lambda}{d\lambda} = -s^3 \cdot \{H^*_\lambda(t)\} = -s^2\{H_\lambda(t)\},$$

$$\frac{dh_\lambda}{d\lambda} = -s \cdot h_\lambda.$$

Other properties of the translation operator are

$$h_0 = 1, \quad h_\lambda \cdot h_\mu = h_{\lambda+\mu}.$$

The only ordinary function with these three properties is the exponential function; therefore we can write

$$h_\lambda = e^{-\lambda s}.$$

B.9. INTEGRAL OF AN OPERATOR

Consider an operator $f(\lambda)$ depending upon the parameter λ; if we can write

$$f(\lambda) = p \cdot \{f_1(\lambda, t)\},$$

where p is an operator not depending on λ, and $f_1(\lambda, t)$ is a function such that $\int_\alpha^\beta f_1(\lambda, t)\,dt$ exists for $t \geq 0$ and for a certain domain of λ, then we can write

$$\int_\alpha^\beta f(\lambda)\,d\lambda = p \cdot \left\{\int_\alpha^\beta f_1(\lambda, t)\,d\lambda\right\},$$

called the *integral* of the operator $f(\lambda)$.

The integral so defined is unique. In fact, suppose that $f(\lambda)$ can be represented in two different ways, for instance

$$p \cdot \{f_1(\lambda, t)\} = q \cdot \{f_2(\lambda, t)\};$$

we can choose a function $\phi(t)$ such that

$$p = \frac{\{g_1\}}{\{\phi\}}, \quad q = \frac{\{g_2\}}{\{\phi\}},$$

with $g_1(t)$ and $g_2(t)$ functions of class C. It follows

$$\{g_1\} \cdot \{f_1\} = \{g_2\} \cdot \{f_2\},$$

$$\left\{\int_0^t g_1(\tau)\, f_1(\lambda, t-\tau)\,d\tau\right\} = \left\{\int_0^t g_2(\tau)\, f_2(\lambda, t-\tau)\,d\tau\right\};$$

integrating both sides with respect to λ and changing the order of integration,

$$\left\{\int_0^t g_1(\tau)\left[\int_\alpha^\beta f_1(\lambda, t-\tau)\,d\lambda\right]d\tau\right\} = \left\{\int_0^t g_2(\tau)\left[\int_\alpha^\beta f_2(\lambda, t-\tau)\,d\lambda\right]d\tau\right\},$$

$$\{g_1\}\left\{\int_\alpha^\beta f_1(\lambda, t)\,d\lambda\right\} = \{g_2\}\left\{\int_\alpha^\beta f_2(\lambda, t)\,d\lambda\right\};$$

finally dividing both sides by $\{\phi\}$ we obtain

$$p \cdot \left\{\int_\alpha^\beta f_1(\lambda, t)\,d\lambda\right\} = q \cdot \left\{\int_\alpha^\beta f_2(\lambda, t)\,d\lambda\right\}, \quad \text{q.e.d.}$$

For any numerical operator $g(\lambda)$ we can write

$$e^{-\lambda s} \cdot g(\lambda) = s \cdot \{H_\lambda\} \cdot g(\lambda)$$

or

$$e^{-\lambda s} \cdot g(\lambda) = s \cdot \{g_1(\lambda, t)\},$$

where

$$g_1(\lambda, t) = 0 \qquad \text{for } 0 \le t < \lambda$$
$$= g(\lambda) \quad \text{for } t \ge \lambda.$$

According to the definition of integral of an operator,

$$\int_0^{\lambda_1} e^{-\lambda s} \cdot g(\lambda)\,d\lambda = s \cdot \left\{ \int_0^{\lambda_1} g_1(\lambda, t)\,d\lambda \right\};$$

but

$$\int_0^{\lambda_1} g_1(\lambda, t)\,d\lambda = \int_0^{t} g(\lambda)\,d\lambda \quad \text{for } 0 \le t < \lambda_1$$
$$= \int_0^{\lambda_1} g(\lambda)\,d\lambda \quad \text{for } t \ge \lambda_1,$$

therefore

$$\int_0^{\lambda_1} e^{-\lambda s} \cdot g(\lambda)\,d\lambda = s \cdot \left\{ \int_0^{t} g(\lambda)\,d\lambda \right\} = \{g\} \quad \text{for } 0 \le t < \lambda_1$$
$$= s \cdot \left\{ \int_0^{\lambda_1} g(\lambda)\,d\lambda \right\} \qquad \text{for } t \ge \lambda_1.$$

For $\lambda_2 > \lambda_1$,

$$\int_{\lambda_1}^{\lambda_2} e^{-\lambda s} \cdot g(\lambda)\,d\lambda = \{g\} \quad \text{for } \lambda_1 \le t < \lambda_2$$
$$= 0 \qquad \text{everywhere else.}$$

With the notation

$$\int_0^{\infty} \quad \text{for} \lim_{\substack{t_1 \to 0 \\ t_2 \to \infty}} \int_{t_1}^{t_2}$$

we can write

$$\int_0^{\infty} e^{-\lambda s} \cdot g(\lambda)\,d\lambda = \{g\}$$

formally identical with the definition of Laplace transform; but it must be clear that in the Laplace transform s is a complex variable, while here s is the differential operator.

B.10. ALGEBRAIC DERIVATIVE

Define the operation D on a function by

$$D\{f(t)\} = \{-t \cdot f(t)\};$$

D *is not* an operator because the ratio $\dfrac{\{t \cdot f\}}{\{f\}}$ *does depend* upon $f(t)$. This operation has the properties

$$D\big(\{f\} + \{g\}\big) = D\{f\} + D\{g\},$$

$$D\big(\{f\} \cdot \{g\}\big) = D\{f\} \cdot \{g\} + \{f\} \cdot D\{g\},$$

$$\{f\} = \frac{\{g\}}{\{h\}} \Rightarrow \mathrm{D}\{f\} = \frac{\mathrm{D}\{g\}\cdot\{h\} - \{g\}\cdot\mathrm{D}\{h\}}{\{h\}^2};$$

the first property is obvious; to prove the second write

$$\begin{aligned}\mathrm{D}\big(\{f\}\cdot\{g\}\big) &= \mathrm{D}\left\{\int_0^t f(\tau)\,g(t-\tau)\,\mathrm{d}\tau\right\} = \left\{-t\int_0^t f(\tau)\,g(t-\tau)\,\mathrm{d}\tau\right\}\\ &= \left\{\int_0^t(-\tau)\,f(\tau)\,g(t-\tau)\,\mathrm{d}\tau\right\} + \left\{\int_0^t f(\tau)\cdot(-t+\tau)g(t-\tau)\,\mathrm{d}\tau\right\}\\ &= \{-t\cdot f(t)\}\cdot\{g(t)\} + \{f(t)\}\cdot\{-t\cdot g(t)\}\\ &= \mathrm{D}\{f\}\cdot\{g\} + \{f\}\cdot\mathrm{D}\{g\},\ \text{q.e.d.}\end{aligned}$$

To prove the third property write

$$\mathrm{D}\{f\}\cdot\{h\} + \{f\}\cdot\mathrm{D}\{h\} = \mathrm{D}\{g\}$$

$$\begin{aligned}\mathrm{D}\{f\} &= \frac{\mathrm{D}\{g\}}{\{h\}} - \frac{\{f\}\cdot\mathrm{D}\{h\}}{\{h\}}\\ &= \frac{\mathrm{D}\{g\} - \mathrm{D}\{h\}\cdot\{g\}/\{h\}}{\{h\}},\ \text{q.e.d.}\end{aligned}$$

The operation D on an operator $p = \{f\}/\{g\}$ is defined by

$$\mathrm{D}p = \frac{\mathrm{D}\{f\}\cdot\{g\} - \{f\}\cdot\mathrm{D}\{g\}}{\{g\}^2};$$

it has the properties

$$\begin{aligned}\mathrm{D}(p+q) &= \mathrm{D}p + \mathrm{D}q,\\ \mathrm{D}(p\cdot q) &= \mathrm{D}p\cdot q + q\cdot\mathrm{D}p,\\ \mathrm{D}\left(\frac{p}{q}\right) &= \frac{\mathrm{D}p\cdot q - q\cdot\mathrm{D}p}{q^2},\end{aligned} \qquad (21)$$

analogous to the properties of the operation D on a function, and easy to verify.

For a numerical operator,

$$\mathrm{D}a = 0;$$

in fact

$$\mathrm{D}a = \mathrm{D}\left(\frac{\{a\}}{\{1\}}\right) = \frac{\mathrm{D}\{a\}\cdot\{1\} - \{a\}\cdot\mathrm{D}\{1\}}{\{1\}^2}$$

thence

$$\mathrm{D}a = \frac{\{-at\}\cdot\{1\} - \{a\}\cdot\{-t\}}{\{1\}^2} = 0,\ \text{q.e.d.}$$

For the differential operator,

$$Ds = 1;$$

in fact

$$Ds = D\left(\frac{1}{\{1\}}\right) = \frac{D1 \cdot \{1\} - 1 \cdot D\{1\}}{\{1\}^2}$$

$$= \frac{-\{-t\}}{\{1\}^2} = 1, \quad \text{q.e.d.}$$

Using repeatedly property (21),

$$Ds^2 = 2 \cdot s, \quad D\,s^3 = 3 \cdot s^2, \ldots\ldots, Ds^n = n \cdot s^{n-1}.$$

For the translation operator,

$$De^{-\lambda s} = D\left(s \cdot \{H_\lambda\}\right) = Ds \cdot \{H_\lambda\} + s \cdot D\{H_\lambda\} = \{H_\lambda\} + s \cdot \{-t \cdot H_\lambda\}$$

$$= s \cdot \left\{\int_0^t H_\lambda(\tau)\,d\tau - t \cdot H_\lambda\right\} = s \cdot \{-\lambda \cdot H_\lambda\},$$

$$De^{-\lambda s} = -\lambda \cdot e^{-\lambda s}.$$

If R(s) is a rational expression in s and in the translation operator $e^{-\lambda s}$, for the properties just shown the operator DR(s) can be computed by formal differentiation of R(s) with respect to s, as though s were a variable. This fact can be expressed by the formula

$$DR(s) = \frac{dR(s)}{ds}. \tag{22}$$

For instance from identity (17) we get

$$\{t \cdot \sin \beta t\} = \frac{2\beta s}{(s^2 + \beta^2)^2};$$

from (18) we get

$$\{t \cdot \cos \beta t\} = \frac{s^2 - \beta^2}{(s^2 + \beta^2)^2}.$$

Many other functional correlates can be found by applying the operation D several times.

B.11. OPERATION T_α

The operation T_α on a function is defined by the formula

$$T_\alpha\{f(t)\} = \{e^{-\alpha t} f(t)\};$$

this operation has the properties

$$T_\alpha\left(\{f\} + \{g\}\right) = T_\alpha\{f\} + T_\alpha\{g\},$$

$$T_\alpha\big(\{f\}\cdot\{g\}\big) = T_\alpha\{f\}\cdot T_\alpha\{g\},$$

$$\{f\} = \frac{\{g\}}{\{h\}} \Rightarrow T_\alpha\{f\} = \frac{T_\alpha\{g\}}{T_\alpha\{h\}}.$$

The first property is obvious; to prove the second property write

$$\begin{aligned} T_\alpha\big(\{f\}\cdot\{g\}\big) &= T_\alpha\left\{\int_0^t f(\tau)\,g(t-\tau)\,d\tau\right\} = \left\{e^{-\alpha t}\int_0^t f(\tau)\,g(t-\tau)\,d\tau\right\} \\ &= \left\{\int_0^t e^{-\alpha\tau} f(\tau)\cdot e^{-\alpha(t-\tau)} g(t-\tau)\,d\tau\right\} = \left\{e^{-\alpha t} f(t)\right\}\cdot\left\{e^{-\alpha t} g(t)\right\} \\ &= T_\alpha\{f\}\cdot T_\alpha\{g\}, \quad \text{q.e.d.} \end{aligned}$$

The third property is an immediate consequence of the second.

The operation T_α on an operator $p = \{f\}/\{g\}$ is defined by

$$T_\alpha p = \frac{T_\alpha\{f\}}{T_\alpha\{g\}};$$

it has the properties

$$\begin{aligned} T_\alpha(p+q) &= T_\alpha p + T_\alpha q, \\ T_\alpha(p\cdot q) &= T_\alpha p\cdot T_\alpha q, \\ p = \frac{q}{r} &\Rightarrow T_\alpha p = \frac{T_\alpha q}{T_\alpha r}, \end{aligned} \tag{23}$$

analogous to the properties of the operation T_α on a function, and easy to verify.

For a numerical operator,

$$T_\alpha a = a;$$

in fact

$$T_\alpha a = T_\alpha\left(\frac{\{a\}}{\{1\}}\right) = \frac{T_\alpha\{a\}}{T_\alpha\{1\}} = \frac{\{a\}}{\{1\}} = a, \quad \text{q.e.d.}$$

For the differential operator,

$$T_\alpha s = s + \alpha;$$

in fact

$$T_\alpha s = \frac{T_\alpha 1}{T_\alpha\{1\}} = \frac{1}{\{e^{-\alpha t}\}},$$

and from identity (16) the above proposition follows.

Using property (23) repeatedly,

$$T_\alpha s^2 = (s+\alpha)^2,$$
$$T_\alpha s^3 = (s+\alpha)^3,$$

and in general,

$$T_\alpha s^n = (s+\alpha)^n.$$

For the translation operator,

$$\begin{aligned} T_\alpha e^{-\lambda s} &= T_\alpha\left(s \cdot \{H_\lambda\}\right) = T_\alpha s \cdot T_\alpha \{H_\lambda\} \\ &= (s+\alpha)\cdot\{e^{-\alpha t}H_\lambda\} = (s+\alpha)\cdot e^{-\lambda s}\{e^{-\alpha\lambda}\cdot e^{-\alpha t}\} \\ &= e^{-\lambda(s+\alpha)}. \end{aligned}$$

If R(s) is a rational expression in s and in the translation operator $e^{-\lambda s}$, then $T_\alpha R(s)$ is equal to the same expression with s substituted by $s+\alpha$. This fact can be expressed by the formula

$$T_\alpha R(s) = R(s+\alpha). \tag{24}$$

For instance from identities (17) and (18) we get

$$\{e^{-\alpha t}\cdot\sin\beta t\} = \frac{\beta}{(s+\alpha)^2+\beta^2}, \tag{25}$$

$$\{e^{-\alpha t}\cdot\cos\beta t\} = \frac{s+\alpha}{(s+\alpha)^2+\beta^2}. \tag{26}$$

Many other functional correlates can be found by applying the operation T_α several times.

B.12. FUNCTIONS APPROXIMATING SOME OPERATORS

Define

$$\begin{aligned} F_{\lambda,\varepsilon}(t) &= 0 && 0 \le t < \lambda \\ &= 1/\varepsilon && \lambda \le t \le \lambda+\varepsilon \\ &= 0 && \lambda+\varepsilon < t < \infty, \end{aligned}$$

thence

$$\begin{aligned} \int_0^t F_{\lambda,\varepsilon}(\tau)\,d\tau &= 0 && 0 \le t < \lambda \\ &= \frac{t-\lambda}{\varepsilon} && \lambda \le t \le \lambda+\varepsilon \\ &= 1 && \lambda+\varepsilon < t < \infty, \end{aligned}$$

If ε is small we can write

$$\int_0^t F_{\lambda,\varepsilon}(\tau)\,d\tau \approx H_\lambda(t),$$

which is exactly correct for $t \leq \lambda$ and for $t \geq \lambda + \varepsilon$; therefore

$$\{1\} \cdot \{F_{\lambda,\varepsilon}\} \sim \{H_\lambda\},$$

$$e^{-\lambda s} \sim \{F_{\lambda,\varepsilon}\} \quad \text{for } \varepsilon \sim 0.$$

This function is also called the *Dirac delta function* or the *unit impulse function.* Define now the *trapezoidal function* (see Fig. 4)

$$\begin{aligned} F_{\lambda,\varepsilon,\eta}(t) &= 0 && 0 \leq t < \lambda \\ &= \frac{t-\lambda}{\varepsilon\eta} && \lambda \leq t < \lambda + \eta \\ &= 1/\varepsilon && \lambda + \eta \leq t < \lambda + \varepsilon \\ &= -\frac{t-\lambda-\varepsilon-\eta}{\varepsilon\eta} && \lambda + \varepsilon \leq t < \lambda + \varepsilon + \eta \\ &= 0 && \lambda + \varepsilon + \eta \leq t < \infty; \end{aligned}$$

thence

$$\begin{aligned} \int_0^t F_{\lambda,\varepsilon,\eta}(\tau)\,d\tau &= 0 && 0 \leq t < \lambda \\ &= \frac{(t-\lambda)^2}{2\varepsilon\eta} && \lambda \leq t < \lambda + \eta \\ &= \frac{\eta}{2\varepsilon} + \frac{t-\lambda-\eta}{\varepsilon} && \lambda + \eta \leq t < \lambda + \varepsilon \\ &= 1 - \frac{\eta}{2\varepsilon} - \frac{(t-\lambda-\varepsilon)(t-\lambda-\varepsilon-2\eta)}{2\varepsilon\eta} && \lambda + \varepsilon \leq t < \lambda + \varepsilon + \eta \\ &= 1 && \lambda + \varepsilon + \eta \leq t < \infty, \end{aligned}$$

and if ε and η are small, i.e. ignoring the interval $\lambda, \lambda + \varepsilon + \eta$,

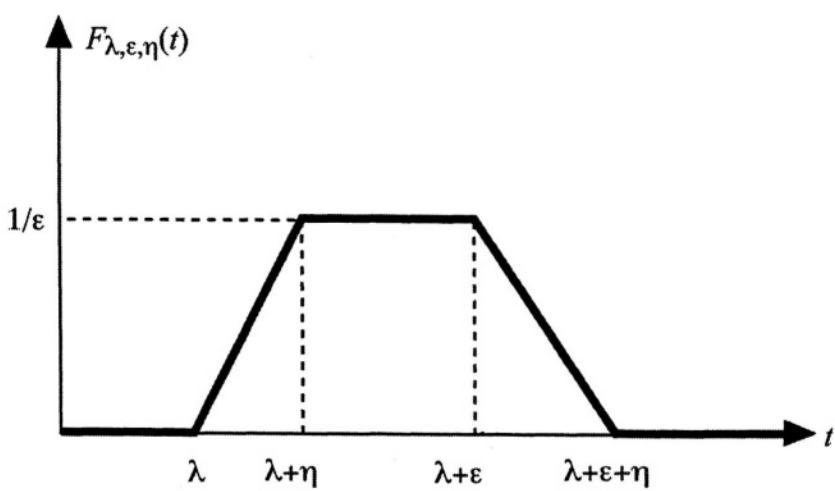

Fig. 4. The trapezoidal function

$$\{1\} \cdot \{F_{\lambda,\varepsilon,\eta}\} \sim \{H_\lambda\},$$
$$e^{-\lambda s} \sim \{F_{\lambda,\varepsilon,\eta}\} \quad \text{for } \varepsilon, \lambda \sim 0.$$

Furthermore

$$\begin{aligned}
\frac{dF_{\lambda,\varepsilon,\eta}}{dt} &= 0 && 0 \le t < \lambda \\
&= \frac{1}{\varepsilon\eta} && \lambda \le t < \lambda + \eta \\
&= 0 && \lambda + \eta \le t < \lambda + \varepsilon \\
&= -\frac{1}{\varepsilon\eta} && \lambda + \varepsilon \le t < \lambda + \varepsilon + \eta \\
&= 0 && \lambda + \varepsilon + \eta \le t < \infty, \\
& F_{\lambda,\varepsilon,\eta}(0) = 0,
\end{aligned}$$

thence

$$s \cdot e^{-\lambda s} \sim \left\{\frac{dF_{\lambda\varepsilon,\eta}}{dt}\right\} \quad \text{for } \varepsilon, \eta \sim 0.$$

In a similar way we can approximate the operators $s^2 \cdot e^{-\lambda s}$, $s^3 \cdot e^{-\lambda s}$, ...; then if λ too is very small, we get the functions approximating the operators $1, s, s^2, \ldots$

B.13. ELEMENTARY TRANSFORMATIONS

Given the operator

$$\frac{p(s)}{q(s)} = \frac{p_m s^m + p_{m-1} s^{m-1} + \cdots + p_1 s + p_0}{q_n s^n + q_{n-1} s^{n-1} + \cdots + q_1 s + q_0},$$

with $q^n \neq 0$, if $m \geq n$ we can write

$$\frac{p(s)}{q(s)} = a_{m-n} s^{m-n} + a_{m-n-1} s^{m-n-1} + \cdots + a_0 + \frac{p^*(s)}{q(s)},$$

where $p^*(s)$ is a polynomial in s of degree less than the degree of $q(s)$.

Suppose now that $m < n$; the operator $p(s)/q(s)$ can be written as a sum of operators of the forms

$$\frac{A}{(s-\alpha)^\mu}, \quad \frac{Bs + C}{\left((s-\beta)^2 + \gamma\right)^\nu},$$

where $A, B, C, \alpha, \beta, \gamma$ are numerical operators, and μ, ν are integers. We first prove a Lemma and a Theorem.

B.13.1. Lemma.

if

$$a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0 = b_n s^n + b_{n-1} s^{n-1} + \cdots + b_1 s + b_0 \tag{27}$$

then

$$a_0 = b_0, \quad a_1 = b_1, \quad \ldots, \quad a_n = b_n \tag{28}$$

and vice versa.

In fact, multiplying both sides of (27) by $\{1\}^{n+1}$,

$$a_n \{1\} + a_{n-1} \{1\}^2 + \cdots + a_0 \{1\}^{n+1} = b_n \{1\} + b_{n-1} \{1\}^2 + \cdots + b_0 \{1\}^{n+1},$$

thence

$$\left\{ a_n + a_{n-1} t + \cdots + a_0 \frac{t^n}{n!} \right\} = \left\{ b_n + b_{n-1} t + \cdots + b_0 \frac{t^n}{n!} \right\}, \tag{29}$$

which implies the identities (28). On the other hand, if the identities (28) hold, equation (29) is valid for any value of t, and multiplying both sides by s^{n+1} we obtain identity (27).

B.13.2. Theorem

If

$$\frac{a_m s^m + a_{m-1} s^{m-1} + \ldots + a_1 s + a_0}{b_n s^n + b_{n-1} s^{n-1} + \ldots + b_1 s + b_0} = \frac{c_p s^p + c_{p-1} s^{p-1} + \ldots + c_1 s + c_0}{d_q s^q + d_{q-1} s^{q-1} + \ldots + d_1 s + d_0},$$

then for any number x (real or complex) such that

$$b_n x^n + b_{n-1} x^{n-1} + \cdots + b_1 x + b_0 \neq 0,$$

$$d_q x^q + d_{q-1} x^{q-1} + \cdots + d_1 x + d_0 \neq 0,$$

it is

$$\frac{a_m x^m + a_{m-1} x^{m-1} + \ldots + a_1 x + a_0}{b_n x^n + b_{n-1} x^{n-1} + \ldots + b_1 x + b_0} = \frac{c_p x^p + c_{p-1} x^{p-1} + \ldots + c_1 x + c_0}{d_q x^q + d_{q-1} x^{q-1} + \ldots + d_1 x + d_0}, \tag{30}$$

and vice versa.

In fact this hypothesis is true if and only if the identity

$$(a_m s^m + a_{m-1} s^{m-1} + \ldots + a_1 s + a_0) \cdot (d_q s^q + d_{q-1} s^{q-1} + \ldots + d_1 s + d_0) =$$
$$= (b_n s^n + b_{n-1} s^{n-1} + \ldots + b_1 s + b_0) \cdot (c_p s^p + c_{p-1} s^{p-1} + \ldots + c_1 s + c_0)$$

is true; for the Lemma this implies, and is implied by, the identities

$$a_0 \cdot d_0 = b_0 \cdot c_0$$

$$a_1 \cdot d_0 + a_0 \cdot d_1 = b_1 \cdot c_0 + b_0 \cdot c_1$$

$$a_2 \cdot d_0 + a_1 \cdot d_1 + a_0 \cdot d_2 = b_2 \cdot c_0 + b_1 \cdot c_1 + b_0 \cdot c_2$$

and so forth; they are the necessary and sufficient conditions for (30) being an identity, q.e.d.

We now return to the rational operator

$$\frac{p(s)}{q(s)} = \frac{p_m s^m + p_{m-1} s^{m-1} + \ldots + p_1 s + p_0}{q_n s^n + q_{n-1} s^{n-1} + \ldots + q_1 s + q_0},$$

with $q_n \neq 0$ and $m < n$. If the polynomial

$$q(x) = q_n x^n + \ldots + q_1 x + q_0$$

contains a real zero α with multiplicity μ, i.e. it contains μ times the factor $(x - \alpha)$, we can write

$$\frac{p(x)}{q(x)} = \frac{A}{(x-\alpha)^\mu} + \frac{p^*(x)}{(x-\alpha)^{\mu-1} q^*(x)},$$

where

$$q(x) = (x - \alpha)^\mu \cdot q^*(x),$$

$$A = \frac{p(\alpha)}{q^*(\alpha)};$$

furthermore the new rational function

$$\frac{p^*(x)}{(x-\alpha)^{\mu-1} q^*(x)}$$

contains in the denominator the factor $(x - \alpha)$ only $\mu - 1$ times, and the same decomposition can be repeated until the polynomial $q(x)$ does not contain any more real zeros. For the theorem just proved we can find the operators $A/(x - \alpha)^\mu$ making the same computations on the operator $p(s)/q(s)$ as though s were a variable.

If $\beta + i\gamma$ and $\beta - i\gamma$ are complex conjugate zeros of multiplicity ν of the polynomial $q(x)$, then we can write

$$\frac{p(x)}{q(x)} = \frac{Bx + C}{\left((x-\beta)^2 + \gamma^2\right)^\nu} + \frac{p^*(x)}{\left((x-\beta)^2 + \gamma^2\right)^{\nu-1} \cdot q^*(x)}$$

where

$$q(s) = \left((x-\beta)^2 + \gamma^2\right)^\nu \cdot q^*(x),$$

$$B = \frac{1}{\gamma} \cdot \Im\left(\frac{p(\beta + i\gamma)}{q^*(\beta + i\gamma)}\right)$$

$$C = \Re\left(\frac{p(\beta + i\gamma)}{q^*(\beta + i\gamma)}\right) - \frac{\beta}{\gamma} \cdot \Im\left(\frac{p(\beta + i\gamma)}{q^*(\beta + i\gamma)}\right),$$

and the same considerations as for the real zeros hold.

B.14. INITIAL VALUE THEOREM

For any function $\{f\}$ of class $\mathcal{K}$, if $f(t)$ has a limit for $t \to 0$, then

$$\lim_{s\to\infty} s\{f\} = \lim_{t\to 0} f(t). \tag{31}$$

In fact, using the results of B.9,

$$\int_0^\infty e^{-\lambda s} f'(\lambda)\mathrm{d}\lambda = s\{f\} - f(0) \tag{32}$$

and

$$\lim_{s\to\infty} \int_0^\infty e^{-\lambda s} f'(\lambda)\mathrm{d}\lambda = \lim_{s\to\infty}\left[s\{f\} - f(0)\right];$$

since s and λ are independent, we can change the order of the limit process, therefore

$$\lim_{s\to\infty}\left[s\{f\} - f(0)\right] = 0,$$

q.e.d.

B.15. FINAL VALUE THEOREM

If $f(t)$ is a function of class $\mathcal{K}$ and it has a limit for $t \to \infty$, then

$$\lim_{s\to 0} s\{f\} = \lim_{t\to\infty} f(t). \tag{33}$$

In fact, from (32),

$$\lim_{s\to 0} \int_0^\infty e^{-\lambda s} f'(\lambda)\mathrm{d}\lambda = \lim_{s\to 0}\left[s\{f\} - f(0)\right];$$

since s and λ are independent, we can change the order of the limit process, therefore

$$\int_0^\infty f'(\lambda)\mathrm{d}\lambda = \lim_{s\to 0}\left[s\{f\} - f(0)\right]; \tag{34}$$

but

$$\int_0^\infty f'(\lambda)\mathrm{d}\lambda = \lim_{t\to\infty} f(t) - \lim_{t\to 0} f(t) \tag{35}$$

and combining (34) with (35) we get (33), q.e.d.

B.15.1. Corollary 1.

If $f(t)$ is a function of class $\mathcal{K}$, then

$$\lim_{s\to 0}\{f\} = \int_0^\infty f(t)\mathrm{d}t\,. \tag{36}$$

In fact if in (33) we substitute $\int_0^\infty f(t)\mathrm{d}t$ to $f(t)$ and use (12), we get (36), q.e.d.

B.15.2. Corollary 2.

If $f(t)$ is a function of class $\mathcal{K}$, then

$$\lim_{s\to 0}\frac{\mathrm{d}}{\mathrm{d}s}\ln\{f\} = -\frac{\int_0^\infty t\cdot f(t)\mathrm{d}t}{\int_0^\infty f(t)\mathrm{d}t}. \tag{37}$$

In fact

$$\frac{\mathrm{d}}{\mathrm{d}s}\ln\{f\} = \frac{\mathrm{d}\{f\}}{\mathrm{d}s}\Big/\{f\},$$

and using the properties of the algebraic derivative,

$$\frac{\mathrm{d}}{\mathrm{d}s}\ln\{f\} = \{-t\cdot f\}/\{f\};$$

now using formula (36) we obtain formula (37).

B.16. REFERENCES

1. J. *Mikusinski, Operational Calculus* (Pergamon, London, 1959).
2. A. Erdelyi, *Operational Calculus and Generalized Functions* (Holt, Rinehart and Winston, New York, 1962).
3. A. Rescigno, The Two-Variable Operational Calculus in the Construction of Compartmental Ecological Models. In *Compartmental Analysis of Ecosystem Models*, edited by J. H. Matis, B. C. Patten and G. C. White (International Co-operative Publishing House, Maryland, 1979) pp. 335-58.
4. E. C. Titchmarsh, The zeros of certain integral functions, *Proceedings of the London Mathematical Society* **25**, 283-302 (1926).

C. BOOLEAN ALGEBRA

There are two elements in the Boolean algebra, 0 and 1, and two operations, addition and multiplication.

The two operations are defined by the following identities:

Addition

$$0 + 0 = 0,$$

$$0 + 1 = 1,$$

$$1 + 0 = 1,$$

$$1 + 1 = 1.$$

Multiplication

$$0 \cdot 0 = 0,$$

$$0 \cdot 1 = 0,$$

$$1 \cdot 0 = 0,$$

$$1 \cdot 1 = 1.$$

Properties of the two operations:

Both addition and multiplication are commutative,

$$a + b = b + a,$$

$$a \cdot b = b \cdot a,$$

and associative,

$$a + (b + c) = (a + b) + c,$$

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c,$$

and multiplication is distributive with respect to addition:

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

This page intentionally left blank

D. MATRIX ALGEBRA

D.1. MATRICES

A rectangular array of homogeneous quantities is called a *matrix;* symbol

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

where there are m rows and n columns.

Two matrices are equal if their elements are equal one by one.

If two matrices have the same number of rows and of columns, their sum is defined by

$$A + B = (a_{ij} + b_{ij}) = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix};$$

it is easy to verify that the sum of matrices has the properties of commutativity and associativity, i.e.,

$$A + B = B + A$$

$$A + (B + C) = (A + B) + C.$$

Given two matrices such that the number of columns of the first is equal to the number of rows of the second, their product is given by

$$A \cdot B = \begin{pmatrix} \sum_{p=1}^{m} a_{1p} b_{p1} & \sum_{p=1}^{m} a_{1p} b_{p2} & \cdots & \sum_{p=1}^{m} a_{1p} b_{pn} \\ \sum_{p=1}^{m} a_{2p} b_{p1} & \sum_{p=1}^{m} a_{2p} b_{p2} & \cdots & \sum_{p=1}^{m} a_{2p} b_{pn} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{p=1}^{m} a_{lp} b_{p1} & \sum_{p=1}^{m} a_{lp} b_{p2} & \cdots & \sum_{p=1}^{m} a_{lp} b_{pn} \end{pmatrix},$$

where A has dimension $l \times m$, B has dimension $m \times n$, and their product had dimension $l \times n$.

Observe that the product of matrices is not always commutative; for instance

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ 2 & -2 \end{pmatrix}$$

but

$$\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The product of matrices, though, is distributive over addition.

Given a number λ and a $m \times n$ matrix A with elements a_{ij}, the product $\lambda \cdot A$ is the $m \times n$ matrix with elements $\lambda \cdot a_{ij}$; obviously the product of a number by a matrix is commutative.

The identity matrix of order n is

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix};$$

it is easy to verify that, for any matrix A,

$$I \cdot A = A$$

$$A \cdot I = A$$

provided A and I have the appropriate size.

The null matrix of order n is

$$0 = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix};$$

it is easy to verify that, for any matrix A,

$$0 \cdot A = 0$$

$$A \cdot 0 = 0$$

provided A and 0 have the appropriate size. We have already seen above that $A \cdot B = 0$ does not necessarily imply that $A = 0$ or $B = 0$.

D.2. DETERMINANT OF A MATRIX.

Given a square matrix, i.e. a matrix of dimension $n \times n$, choose n elements, one from each row and one from each column, and form their product; to this product give a + or a

– sign according to whether the number of inversions in its columns is even or odd; the sum of all such products is called the *determinant* of the matrix; symbol

$$\det(A) = |a_{ij}| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Examples:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21},$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$

$$-a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33}.$$

It is easy to verify that if two rows are equal, the determinant is zero.

The determinant obtained from $\det(A)$ by deleting from it the same number of rows and columns is called a *minor* of $\det(A)$; if the rows and columns deleted had the same ordinal position, the resulting determinant is called a *principal minor.* We call *cofactor* of the element a_{ij} of $\det(A)$, symbol A_{ij}, the quantity

$$A_{ij} = (-1)^{i+j} \det(A)_{ij},$$

where the symbol $\det(A)_{ij}$ means the minor of $\det(A)$ obtained by deleting row i and column j.

A simple algorithm to compute the value of a determinant is

$$\det(A) = \sum_{j=1}^{n} a_{ij} A_{ij}, \tag{1}$$

where i is any integer from 1 to n. To prove this proposition observe that the cofactor A_{ij} above contains one element from each row and each column of A except row i and column j, with a sign + or – according to the number of inversions of rows and columns; therefore the sum includes all possible products of the elements of A with their proper sign.

Another important formula that will be useful in the computation of determinants is

$$\sum_{j=1}^{n} a_{ij} A_{lj} = 0, \quad i \neq l. \tag{2}$$

To prove it observe that this sum is equal to a determinant where row i is equal to row l.

D.3. RANK OF A MATRIX

The *rank* of a matrix is the largest order of a non-zero determinant contained in the matrix. For instance, the matrix

$$\begin{pmatrix} 1 & 5 & 3 \\ -3 & -2 & -4 \\ 2 & -3 & 1 \end{pmatrix}$$

has rank two because its determinant is zero, but the determinant

$$\begin{vmatrix} 1 & 5 \\ -3 & -2 \end{vmatrix}$$

formed by the first two rows and two columns is different from zero; matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 8 & 12 \end{pmatrix}$$

has rank one because all three determinants of order two it contains, i.e.

$$\begin{vmatrix} 1 & 2 \\ 4 & 8 \end{vmatrix}, \begin{vmatrix} 1 & 3 \\ 4 & 12 \end{vmatrix}, \begin{vmatrix} 2 & 3 \\ 8 & 12 \end{vmatrix}$$

are equal to zero.

A square matrix whose determinant is zero is called *singular,* an $n \times n$ non-singular matrix has rank n by definition.

D.4. INVERSE OF A MATRIX

To any non-singular matrix A corresponds an inverse A^{-1} such that

$$A \cdot A^{-1} = A^{-1} \cdot A = I.$$

For instance,

$$\begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & -1 & 2 & -3 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & -1 & 2 & -3 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & -1 \end{pmatrix} = I,$$

$$\begin{pmatrix} 2 & 2 & 3 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -4 & -3 \\ 1 & -5 & -3 \\ -1 & 6 & 4 \end{pmatrix} = I,$$

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \boldsymbol{I}.$$

The inverse of the product of two matrices is equal to the product of their inverses in the inverted order, i.e.,

$$[\boldsymbol{A}\cdot\boldsymbol{B}]^{-1} = \boldsymbol{B}^{-1}\cdot\boldsymbol{A}^{-1};$$

in fact

$$[\boldsymbol{A}\cdot\boldsymbol{B}]\cdot[\boldsymbol{B}^{-1}\cdot\boldsymbol{A}^{-1}] = \boldsymbol{A}\cdot[\boldsymbol{B}\cdot\boldsymbol{B}^{-1}]\cdot\boldsymbol{A}^{-1} = \boldsymbol{A}\cdot\boldsymbol{I}\cdot\boldsymbol{A} = \boldsymbol{A}\cdot\boldsymbol{A}^{-1} = \boldsymbol{I}.$$

Given a matrix $\boldsymbol{A} = (a_{ij})$, the element b_{ij} of its inverse $\boldsymbol{B}$ is given by

$$b_{ij} = \frac{A_{ji}}{\det(\boldsymbol{A})};$$

in fact the element of row i and column j of the product $\boldsymbol{A}\cdot\boldsymbol{B}$ is

$$\sum_{l=1}^{n} a_{il}b_{lj} = \sum_{l=1}^{n} \frac{a_{il}A_{jl}}{\det(\boldsymbol{A})},$$

and the sum on the right-hand side is equal to 1 when $i = j$, due to identity (1), and is equal to 0 when $i \neq j$, due to identity (2).

D.5. SIMILAR MATRICES

If $\boldsymbol{A}$ is a square matrix and

$$\boldsymbol{Q}\cdot\boldsymbol{A}\cdot\boldsymbol{Q}^{-1} = \boldsymbol{B},$$

then $\boldsymbol{A}$ and $\boldsymbol{B}$ are called *similar*.

Think about $\boldsymbol{A}$ as an operation transforming vector $\boldsymbol{X}$ into vector $\boldsymbol{Y}$:

$$\boldsymbol{A}\cdot\boldsymbol{X} = \boldsymbol{Y};$$

multiply both sides of this equation on the left by the non-singular matrix $\boldsymbol{Q}$ (if it has an inverse it must be non-singular):

$$\boldsymbol{Q}\cdot\boldsymbol{A}\cdot\boldsymbol{X} = \boldsymbol{Q}\cdot\boldsymbol{Y};$$

the left hand side does not change if we insert in it the product $\boldsymbol{Q}^{-1}\cdot\boldsymbol{Q}$:

$$\boldsymbol{Q}\cdot\boldsymbol{A}\cdot\boldsymbol{Q}^{-1}\cdot\boldsymbol{Q}\cdot\boldsymbol{X} = \boldsymbol{Q}\cdot\boldsymbol{Y};$$

this last equation proves that matrix $\boldsymbol{Q}\cdot\boldsymbol{A}\cdot\boldsymbol{Q}^{-1} = \boldsymbol{B}$ transforms vector $\boldsymbol{Q}\cdot\boldsymbol{X}$ into vector $\boldsymbol{Q}\cdot\boldsymbol{Y}$. But these last two vectors represent vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ in different coordinates; therefore similar matrices represent the same transformation in different coordinates.

D.6. EIGENVALUES AND EIGENVECTORS

Given the square matrix $\boldsymbol{A}$, the equation

$$\boldsymbol{A}\cdot\boldsymbol{X} = \lambda\cdot\boldsymbol{X},$$

where $\boldsymbol{X}$ is a vector and λ a scalar, has only the obvious solution $\boldsymbol{X} = \boldsymbol{0}$, unless matrix $\boldsymbol{A}\cdot\boldsymbol{X} - \lambda\cdot\boldsymbol{X}$ is singular, i.e.

$$\det[\boldsymbol{A} - \lambda\cdot\boldsymbol{I}] = 0. \tag{3}$$

This last equation is called the *characteristic equation* of matrix $\boldsymbol{A}$. The values of λ that satisfy equation (3) are called the *eigenvalues* of $\boldsymbol{A}$; the vectors $\boldsymbol{X}$ that satisfy equation (3) in correspondence to each eigenvalue are called the *eigenvectors* of $\boldsymbol{A}$.

Given matrix

$$\boldsymbol{A} = \begin{pmatrix} 0 & 3 \\ 1 & -2 \end{pmatrix},$$

its characteristic equation is

$$\begin{vmatrix} -\lambda & 3 \\ 1 & -2-\lambda \end{vmatrix} = 0,$$

i.e.,

$$\lambda^2 + 2\lambda - 3 = 0;$$

there are two eigenvalues, –3 and +1. To the eigenvalue –3 corresponds equation

$$\begin{pmatrix} 3 & 3 \\ 1 & -2+3 \end{pmatrix}\cdot\boldsymbol{X} = \boldsymbol{0},$$

thence the eigenvector

$$\boldsymbol{X} = \begin{pmatrix} +1 \\ -1 \end{pmatrix}.$$

To the eigenvalue +1 corresponds equation

$$\begin{pmatrix} -1 & 3 \\ 1 & -2-1 \end{pmatrix}\cdot\boldsymbol{X} = \boldsymbol{0},$$

thence the eigenvector

$$\boldsymbol{X} = \begin{pmatrix} +3 \\ +1 \end{pmatrix}.$$

D.7. DIAGONALIZATION OF A MATRIX

Suppose that matrix $\boldsymbol{A}$ of dimension $n \times n$ has n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, with the corresponding eigenvectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$. Equation

$$(\boldsymbol{A} - \lambda \cdot \boldsymbol{I}) \cdot \boldsymbol{X} = \boldsymbol{0},$$

i.e.,

$$\boldsymbol{A} \cdot \boldsymbol{X}_i = \lambda \cdot \boldsymbol{X}_i, \quad i = 1, 2, \ldots, n$$

hold.

With the eigenvalues $\boldsymbol{X}_i$ form the matrix

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_n \end{pmatrix};$$

it follows,

$$\begin{aligned} \boldsymbol{A} \times \boldsymbol{P} &= \boldsymbol{A} \cdot \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{A} \cdot \boldsymbol{X}_1 & \boldsymbol{A} \cdot \boldsymbol{X}_2 & \cdots & \boldsymbol{A} \cdot \boldsymbol{X}_n \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 \cdot \boldsymbol{X}_1 & \lambda_2 \cdot \boldsymbol{X}_2 & \cdots & \lambda_n \cdot \boldsymbol{X}_n \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_n \end{pmatrix} \cdot \boldsymbol{\Lambda} = \boldsymbol{P} \cdot \boldsymbol{\Lambda}, \end{aligned}$$

i.e.,

$$\boldsymbol{A} \cdot \boldsymbol{P} = \boldsymbol{P} \cdot \boldsymbol{\Lambda}, \tag{4}$$

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

is a diagonal matrix formed with the eigenvalues of $\boldsymbol{A}$.

If we multiply both sides of this equation by $\boldsymbol{P}^{-1}$ on either side,

$$\boldsymbol{A} = \boldsymbol{P} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{P}^{-1},$$

$$\boldsymbol{P}^{-1} \cdot \boldsymbol{A} \cdot \boldsymbol{P} = \boldsymbol{\Lambda}.$$

An immediate consequence of these two equations is that similar matrices have the same eigenvalues (but not the same eigenvectors). In fact, given

$$\boldsymbol{A} = \boldsymbol{Q} \cdot \boldsymbol{B} \cdot \boldsymbol{Q}^{-1},$$

it follows

$$\boldsymbol{\Lambda} = \boldsymbol{P}^{-1} \cdot \boldsymbol{A} \cdot \boldsymbol{P} = \boldsymbol{P}^{-1} \cdot \boldsymbol{Q} \cdot \boldsymbol{B} \cdot \boldsymbol{Q}^{-1} \cdot \boldsymbol{P}$$

$$\boldsymbol{\Lambda} = (\boldsymbol{P}^{-1} \cdot \boldsymbol{Q}) \cdot \boldsymbol{B} \cdot (\boldsymbol{Q}^{-1} \cdot \boldsymbol{P}),$$

where matrices $\boldsymbol{P}^{-1} \cdot \boldsymbol{Q}$ and $\boldsymbol{Q}^{-1} \cdot \boldsymbol{P}$ are formed with the eigenvectors of $\boldsymbol{B}$, and are one the inverse of the other.

D.8. MULTIPLE EIGENVALUES

The statement of the previous section is not valid if the eigenvalues are not all distinct. Consider matrix

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix};$$

its characteristic equation is

$$\begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = \lambda^2 - (a+d)\lambda + ad - bc = 0$$

and its eigenvalues are

$$\lambda_1 = \frac{1}{2}\left(-b + \sqrt{(a-d)^2 + 4bc}\right),$$
$$\lambda_2 = \frac{1}{2}\left(-b - \sqrt{(a-d)^2 + 4bc}\right).$$

If $(a-d)^2 + 4bc = 0$, the two eigenvalues are identical and we cannot put matrix $\boldsymbol{A}$ in diagonal form as shown in equation (4), because there is no matrix $\boldsymbol{P}$ such that

$$\boldsymbol{A} \cdot \boldsymbol{P} = \boldsymbol{P} \cdot \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{pmatrix}.$$

In fact, for any matrix $\boldsymbol{P}$, it is always

$$\boldsymbol{P} \cdot \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{pmatrix} \cdot \boldsymbol{P}^{-1} = \lambda_1 \cdot \boldsymbol{I}.$$

What we can do, though, is to find a matrix $\boldsymbol{P}$ such that

$$\boldsymbol{A} \cdot \boldsymbol{P} = \boldsymbol{P} \cdot \begin{pmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{pmatrix}, \tag{5}$$

where this last matrix is called a *Jordan matrix.* If we write

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

from equation (5) we get

$$ap_{11} + bp_{21} = \lambda_1 p_{11},$$
$$cp_{11} + dp_{21} = \lambda_1 p_{21},$$
$$ap_{12} + bp_{22} = p_{11} + \lambda_1 p_{12},$$
$$cp_{12} + dp_{22} = p_{21} + \lambda_1 p_{22};$$

the first two equations have always a solution in p_{11} and p_{21} as long as λ_1 is an eigenvalue of A; they can be written in the form

$$A \cdot \begin{pmatrix} p_{11} \\ p_{21} \end{pmatrix} = \lambda_1 \begin{pmatrix} p_{11} \\ p_{21} \end{pmatrix}$$

and show that the first column of P is an eigenvector of A; the last two equations can be written in the form

$$A \cdot \begin{pmatrix} p_{12} \\ p_{22} \end{pmatrix} = \begin{pmatrix} p_{11} \\ p_{21} \end{pmatrix} + \lambda_1 \begin{pmatrix} p_{12} \\ p_{22} \end{pmatrix}$$

and show that there is always a solution for p_{12} and p_{22}; we can summarize these results by writing

$$\left(A - \lambda_1 I\right) \cdot P_1 = 0,$$
$$\left(A - \lambda_1 I\right) \cdot P_2 = P_1,$$

where P_1 and P_2 are the first and second column, respectively, of P.

D.9. COMPLEX EIGENVALUES

The statement of section D.7 is not valid if the eigenvalues are not all real. Consider matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix};$$

if $(a-d)^2 + 4bc < 0$ its eigenvalues are complex conjugate. In this case we cannot put matrix A in diagonal form with real eigenvalues as shown in equation (4).

We can find a matrix P such that

$$A \cdot P = P \cdot \begin{pmatrix} \lambda & \omega \\ -\omega & \lambda \end{pmatrix},$$

where

$$\lambda = \frac{a+d}{2}, \quad \omega = \sqrt{-\left(\frac{a-d}{2}\right)^2 - bc};$$

in fact matrices

$$\begin{pmatrix} \lambda & \omega \\ -\omega & \lambda \end{pmatrix} \text{ and } \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

have exactly the same eigenvalues.

D.10. EXPONENTIAL OF A MATRIX

In analogy with the series

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

we define the exponential of a matrix with the series

$$e^{\boldsymbol{A}} = \boldsymbol{I} + \boldsymbol{A} + \frac{\boldsymbol{A}^2}{2!} + \frac{\boldsymbol{A}^3}{3!} + \cdots$$

It is easy to prove the following properties:

$$\boldsymbol{A} = \boldsymbol{Q} \cdot \boldsymbol{B} \cdot \boldsymbol{Q}^{-1} \Rightarrow e^{\boldsymbol{A}} = \boldsymbol{Q} \cdot e^{\boldsymbol{B}} \cdot \boldsymbol{Q}^{-1},$$

$$\boldsymbol{A} \cdot \boldsymbol{B} = \boldsymbol{B} \cdot \boldsymbol{A} \Rightarrow e^{\boldsymbol{A}+\boldsymbol{B}} = e^{\boldsymbol{A}} \cdot e^{\boldsymbol{B}},$$

$$e^{-\boldsymbol{A}} = \left(e^{-\boldsymbol{A}}\right)^{-1},$$

$$e^{\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)} = \mathrm{diag}\left(e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n}\right).$$

D.11. HAMILTON-CAYLEY THEOREM

We have seen in section D.6 that equation (3) is called the characteristic equation of matrix $\boldsymbol{A}$. The Hamilton-Cayley theorem states that every square matrix satisfies its own characteristic equation. In other words, if

$$f(\lambda) = \det[\boldsymbol{A} - \lambda \cdot \boldsymbol{I}],$$

then

$$f(\boldsymbol{A}) = \boldsymbol{0}.$$

The proof is quite simple. Form the matrix $\boldsymbol{B}(\lambda)$, whose element of row i and column j is the cofactor of element of row j and column i of matrix $[\boldsymbol{A} - \lambda \cdot \boldsymbol{I}]$; these two matrices are one the inverse of the other (see section D.4) except for a factor $\det[\boldsymbol{A} - \lambda \cdot \boldsymbol{I}]$; therefore

$$\boldsymbol{B}(\lambda) \cdot [\boldsymbol{A} - \lambda \cdot \boldsymbol{I}] = f(\boldsymbol{A}) \cdot \boldsymbol{I}.$$

This expression shows that $\boldsymbol{A} - \lambda \cdot \boldsymbol{I}$ is a factor of $f(\boldsymbol{A}) \cdot \boldsymbol{I}$, q.e.d.

A scalar polynomial $f(\lambda)$ is called an *annihilating polynomial* of a square matrix $\boldsymbol{A}$ if $f(\boldsymbol{A}) = \boldsymbol{0}$. For instance, given matrix

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

the polynomial

$$f(\lambda) = \lambda^3 - (a_{11} + a_{22} + a_{33})\lambda^2 + (a_{11}a_{22} + a_{11}a_{33} + a_{22}a_{33} - a_{12}a_{21} - a_{13}a_{31} - a_{23}a_{32})\lambda$$
$$-a_{11}a_{22}a_{33} - a_{12}a_{23}a_{31} - a_{13}a_{21}a_{32} + a_{11}a_{23}a_{32} + a_{13}a_{22}a_{31} + a_{12}a_{21}a_{33}$$

formed by the determinant

$$\det(A) = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix}$$

is an annihilating polynomial of A.

D.12. PERSYMMETRIC MATRICES

A square matrix such that each line perpendicular to the principal diagonal has all its elements alike is called a *persymmetric matrix.* In a $n \times n$ persymmetric matrix

$$\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_2 & a_3 & a_4 & \cdots & a_{n+1} \\ a_3 & a_4 & a_5 & \cdots & a_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n+1} & a_{n+2} & \cdots & a_{2n-1} \end{pmatrix}$$

there are evidently at most $2n - 1$ distinct elements, i.e., those of the principal diagonal and one adjacent minor diagonal.

From its definition, it follows that in a persymmetric matrix the element of row i and column j is equal to the element of row $i + k$ and column $j - k$, where k is any positive integer smaller than j.

This page intentionally left blank

INDEX