Missed Chances

An Essay on the Interplay between Mathematics and the Real World

Richard D. Gill*

Mathematical Institute, University of Utrecht
P.O. Box 80.010, 3508 TA Utrecht, The Netherlands

We consider the familiar antithesis between pure and applied mathematics as it appears in statistics: the 'pure' theory of mathematical statistics on the one hand and applied statistical methodology on the other. Three examples (de Witt's study of annuities, the Delta Committee's work on the Dutch storm-flood disaster of 1953, and modern developments in survival analysis) show that the interplay is far more subtle than the facile distinction pure versus applied suggests.

PREFACE

This paper started life as an inaugural lecture at the University of Leiden. This formal context explains both style and some of the content of the paper: de Witt, one of a generation of outstanding Dutch mathematicians trained in Leiden (together with Huygens and Hudde), was also an outstanding politician and in fact now something of a Dutch national hero. The other ingredient is the surprise of the author - a *mathematical* statistician - at being appointed to a special chair in the *applications* of mathematics.

INTRODUCTION

I want to examine the classical antithesis between pure and applied mathematics, emphasising of course the position of statistics and statisticians. I'd like to underline the existence of this antithesis, the fact that there can be a paradox in putting mathematical statistics amongst the applications of mathematics, by quoting the famous pure mathematician Paul Halmos [1] who said : 'applied mathematics is bad mathematics'. Such a statement coming from that corner may not be surprising; but what to think of the fact that one of my colleagues in mathematical statistics quoted it recently in his inaugural speech [2] without any sign of condemnation?

Of course this is rather an exaggeration. One could deduce from this that there are two easily discernable kinds of statisticians, while many of us can behave differently on different occasions. Sometimes this results in a split personality, occasionally in an enriched one, but always a certain tension between

* The author was formerly head of the department of mathematical statistics, CWI.

theory and practice is present. I hope to show you that the application of statistics in the real world, and its interaction with statistical theory, and even more broadly with mathematics as a whole, is a much more subtle matter than one would think from the facile distinction between theory and practice. Mathematical and applied statistics are much closer than is usually recognised. One could even speak of a close embrace - maybe this is the reason for the special fascination of the discipline.

Let's now examine some applications of statistics in order to look at the role played there by mathematics, clarifying at the same time the proposition I outlined above. To do this I'll first have to tell you a bit more about statistics and its aims.

Statistics in the news

We are confronted daily in the newspapers and on television with pronouncements in which rightly or wrongly statistics are used. Statistics is used for drawing conclusions from observations or data in situations in which chance or accident played so big a role that an equivocal conclusion is not possible. It is only possible to draw inexact conclusions. Statisticians are now able to characterise and minimalise the level of uncertainty. (Today I leave out the important sub-fields 'descriptive statistics' and 'data-analysis' in which one can or wants to do without pronouncements about chance.) The fact that according to one expert the Challenger disaster had a chance of one in every thousand launches while according to another this chance was as big as one in twenty might have something to do with the margins of uncertainty which are undoubtedly present in these two statements. On the other hand I certainly don't want to suggest that in a situation of such complexity a solid statistical solution is possible at all. One should not only have at one's disposal the data relevant to the disaster, but also a model that makes this relationship explicit and quantifiable; the difference between the two statements may be due to the fact that different models were used.

The importance of a model - a framework for thought if you like - can also be illustrated by the reports in the newspapers of the number of extra deaths to be expected as a consequence of the accident in a nuclear reactor in Tsjernobyl. This number has no meaning whatsoever or, to put it kindly, it can be interpreted in many ways. Everyone happens to die exactly once, disaster or no disaster. Does one mean really: an additional number of deaths within a certain (relatively short) period? Or does this number apply to the total number of people who die earlier than would have been the case had the accident not happened? And wouldn't it be relevant in that case *how* much sooner that would be; i.e. wouldn't the 'amount of lost man-years' be a more sensible quantity to use?

I hope to make clear to you later that statistics cannot only be used to draw some conclusions from a collection of data; it also has a much more important use, because it enables us by employing mathematical models to find out whether the available data really are relevant to answer the questions put to the statistician. As often as not this is not the case unless we add all kinds of extra assumptions. A statistical model makes these assumptions explicit and thereby open to discussion, sometimes even disputable.

FIRST CHANCE: DE WITT AND THE PRICE OF ANNUITIES

To illustrate the aforesaid I'll give you three examples. The first two will be quite brief. The last one will not only be longer but also include some advanced mathematics. The first example takes us to seventeenth century Leiden. In that time in the Low Countries not only a political revolution was being consolidated, but also a mathematical one. It was here Descartes had just developed analytic geometry. He hereby showed algebra and geometry to be one; Euclid's geometry can be constructed as ordinary algebra of ordinary numbers. This quantification of space not only unified mathematics for the first time in Western history, but it also offered new tools with which to apply it. Cartesian mathematics played an essential part in the development of the European mechanistic world-view that culminated in the nineteenth century with Laplace, and which we have inherited. Descartes' new mathematics was only a part of his new and controversial philosophy and physics.

In Leiden classical mathematics, not much changed since Euclid, was practised in the classical way. Prince Maurits altered this in 1600 by founding a chair in applied - in particular military - mathematics. A novelty concerning this private chair was that lectures had to be in Dutch instead of the then current Latin. In 1615 Frans van Schooten was appointed to this chair. When he died his son, also named Frans, succeeded him; this younger Van Schooten now was an ardent follower and popularizer of Descartes. Around this time Descartes himself stayed some years in Leiden.

The influence of the younger van Schooten shows itself clearly in the work of three of his pupils: Johan de Witt, Johannes Hudde, and Christiaan Huygens. The contributions to probability-theory of Huygens are most well-known - his book *De Ratiociniis in Ludu Aleae* published in 1657 was the first and also, thanks to the wide distribution brought about by Van Schooten, during fifty years the standard text on probability-theory. Today however, I'd like to tell you something about the work of De Witt [3]. He applied statistics for the first time in politics. This also made him the first person to take probability-theory out of the domain of games of chance. (De Witt's main contribution to mathematics was something entirely different, namely a treatise on conic sections à la analytic geometry.)

In 1671 De Witt, as Grandpensionary, considered it his task to collect funds quickly to defend the Republic in the impending attack from other major countries (England, France, and the bishoprics of Munster and Cologne). In those days, when a small public service offered little opportunity to economize, a common way of raising money was to sell annuities: for a certain price one could buy a fixed annual income or interest for a certain person; this interest was paid as long as that person lived. The buyer usually nominated a younger member of his family. The current purchase-price was fourteen times the amount one got paid annually, independently of the age of the nominee. Inflation was four percent per annum. If one knows the probability distribution of the remaining lifetime of nominees of a given age and bases oneself on a known level of inflation, one can nowadays very easily calculate the fair purchase-price. In those days this was very complicated; even the language to think in these terms was lacking. And not only the language, but also the relevant data: the first and for a long time only mortality-table had been published only nine years before by Graunt in London (it included mortality due to the plague, and was based on an *estimated* total population size). This work, just as much that of De Witt, represented a breakthrough in the field.

De Witt presented his report about the value of annuities to the States General. The report reads like a mathematical discourse. It starts with Huygens' definition of the expectation of a random variable - Huygens had been the first to define this in mathematical terms. (The definition is formulated in terms of the price of a ticket in an equivalent fair lottery.) The report continues by explaining how to calculate an expectation given the probability distribution of the random variable. Because of the lack of empirical data De Witt constructs the probability distribution of remaining lifetime by making some plausible suppositions. He divides a human lifespan into four periods. Within each of these the chances of the person dying are uniformly distributed; the chances in the second period (from 53 to 63) are bigger by a factor one-and-a-half than the chances before that (from 3 to 53), in the third period (from 63 to 73) by a factor two, and in the fourth (from 73 to 80) by a factor three.

After some arithmetic De Witt calculates the rather large purchase-price of sixteen times the annual income for a nominee of the age of three. He substantiates this controversial proposal by giving many reasons why this calculation would even be on the *lower* side. In this he is helped by the possibility of working through the effect of variations in the parameters he chose. It is a most powerful discourse that uses all means (in particular a small statistical analysis from the registers of annuities of Holland and West-Friesland) to support his proposal.

The proposal was accepted, although De Witt's carreer was nearly over; a year later after his leaving office he and his brother were lynched by the mob in the Hague. His discourse on annuities became renowned all over Europe - it was known to Bernouilli and Leibniz for instance.

One may wonder if the members of the States General weren't dazzled by so much mathematics; all the more so because at close reading of the report some strange incongruities come to light. The factors one-and-a-half, two and three are derived from an argument concerning *conditional risks*: for a person aged 58 the death-risk or force of mortality in the coming year would be one-and-a-half times as big as that for a person aged 40. However, the calculations are concerned with *unconditional* chances: i.e. the chance of a three-year-old dying in his 58th year would be one-and-a-half times the chance of him dying in his 40th year. What's more, the factors in the calculations have suddenly changed into two-thirds, a half and one third. This means they're the reciprocals of the original ones! For a long time these 'minor flaws' went unnoticed. In a later correspondence with Hudde, then town-mayor of Amsterdam, who analysed new empirical data, De Witt showed he was well aware of the distinction between age-dependent force of mortality and the unconditional probability density of lifetimes.

De Witt's calculations are set up efficiently. One can repeat them easily choosing different values for the parameters; this possibility obviously influenced his choice of model. The original factors in the unconditional probability distribution result in the politically unrealistic price of eighteen annual interests. It rather appears he chose his figures with the wanted result in mind: after first having got an unacceptable result he adapted his figures without in his great haste (let's assume that not only money but also time was scarce) following through the whole argument and adapting it accordingly. The lack of a clear and concise mathematical language too hampered his study. The matter was complicated even more by a mathematical problem: determining the risk-function over the ages up to a constant of proportionality as De Witt wanted to do, does not fix the entire probability distribution. This *does* happen to be the case when determining the probability density up to proportionality.

In the long run this work of De Witt hardly influenced the evolution of statistics. During the decades after his discourse was published the mathematicians of Europe were busy developing the differential and integral calculus. De Witt's solution to the problem of annuities fell into oblivion. It was only rediscovered two centuries later, during which many completely incorrect solutions were used. Probability-theory and statistics only flourished in our century, when they could take root in much more mature mathematics and be stimulated by prestigious scientific applications.

For the modern pursuit of science it's difficult to learn lessons from such a history. We can appreciate how De Witt applied with flourish the abstract theory of games of chance, which had only just been developed, to a political matter of great urgency. The fact that his analysis was not perfect may be defended in the light of an insufficient mathematical language and incomplete conceptual apparatus. Whether he himself fell victim to this opportunity for confusion or whether he used it for political purposes, I happily leave to historians to decide.

Second chance: how high are the dykes?

Let's continue with a more contemporary example: the flooding of the South-West of Holland in 1953. On account of this disaster the Delta Committee was formed. The Mathematical Centre in Amsterdam was assigned the task of determining by a statistical analysis of the levels of the high tides over the past seventy years how high the seadikes should be - the 'basispeil' (basic level) - to make the chance of inundation one in ten thousand per annum, or, to put it differently and make the chance appear less negligible, one-hundredth in every hundred years. Since a statistical analysis in those days was literally pure brain- and handwork, one had to be very sparing and resourceful in constructing and analysing different models. By selecting for the analysis one high tide level for every depression occurring in the observational period, one could rule out mutual dependence in the levels of high tides in close succession. Choosing these depressions however was a time-consuming and somewhat subjective business. After a thorough analysis it was concluded that the observations were essentially exponentially distributed: i.e. given the fact that the waterlevel exceeds x meters, the chance that it rises one more centimeter stays the same whatever the value of x. This particular chance, and the chance a depression occurs can be estimated from the observations. Together they determine the wanted 'basispeil'. These calculations resulted in a 'basispeil' of 5.1 meters, a figure much higher than the 4.5 meters politicians and engineers had come up with. Taking the possible estimation error into account, a 'safe' basic level would be higher still!

In a stormy meeting in the Treveszaal at the Binnenhof in the Hague the mathematicians - D. van Dantzig and J. Hemelrijk - managed to have the original 4.5 meters changed into 5 meters. This greatly surprised many members of the committee, who wondered what on earth these statisticians thought they were doing.

At the present time this investigation is being repeated [4]. It is helped by a now thirty years longer series of observations, much more refined methods of analysis, and supported by modern computer facilities. In particular - partly inspired by such applications - an extensive and more elegant theory has evolved concerning extremes in a series of mutually dependent random quantities, i.e. a stochastic process. In an interim report the scientists use a more complicated model with an extra parameter; as a result of this the estimated basispeil turns out to be much lower - 4.2 meters only (the chances of increases drop with the level already attained, thus making very extreme waterlevels more unlikely). By including this extra parameter however, the possible estimation error has grown a lot (and a safe 'basispeil' stays the same). This may even lead to the following final conclusion: such a demanding extrapolation can't be made from the available data.

Third chance: survival analysis?

My third and somewhat more detailed example concerns an area in modern statistics in which I myself have been involved, namely so-called survival analysis. This term - a eufemism as you'll realise later - is the collective name for statistical methods that can be used for the analysis of observed lengths of time between the beginning of a medical treatment and the failing of that treatment; such data are collected for instance to compare a new treatment with an old clinical treatment of certain kinds of cancer. One should think of a clinical trial in a hospital in which during say five years maybe one to two hundred new patients suffering from a certain illness are admitted and treated. For every one of these patients chance is made to decide whether they'll be treated according to the new or the old therapy. As this eliminates the effects of other factors, one is thus enabled to make as unbiased a comparison as possible. We are thinking of illnesses that can't be cured now, but can be kept under control for a number of years. This means that the important patient-response to be compared is not a yes/no variable, but the time-span between admission and treatment to the end of remission, i.e. to recurrence of the illness. These time spans are very variable. The point is to compare the distributions of 'survival-times' in the two groups.

In the fifties and sixties cancer research soared. Under Kennedy for instance the fight against cancer was taken up just as thoroughly as the race to be the first to have a man on the moon. Not only medical scientists occupied themselves with this research, but in their wake many statisticians got involved too. This confrontation with a new kind of statistical problems led to a new flourishing of statistical theory.

Comparing the distribution of a quantity in two populations, from each of which a random sample has been taken, is a standard exercise for an applied statistician. Anyone who at any time has taken a course in elementary statistics, be it for physicists, psychologists or whatever, will have been confronted with the parametric 'Student's t-test' and its non-parametric counterpart the 'two-sample test' of Wilcoxon. The qualifications 'parametric' and 'non-parametric' point to the fact that the first test rests on heavy assumptions of normal distributions in the two populations to be compared, i.e. distributions of a specified shape so that only two numerical parameters (mean and variance) suffice to fix them completely. The second procedure however is correct under minimal assumptions. In the kind of medical research I'm talking about here, usually non-parametric methods are used. One wants to draw as convincing as possible a conclusion that the new treatment is better or worse as the case may be.

Thus far this does not seem to offer the statistician a new challenge. What I haven't mentioned yet is the complicating phenomenom called *censoring*. Obviously one wants to make a decision about the relative merit of each treatment as soon as possible. This implies that at the point in time at which one has to analyse the data, quite a number of patients will still be in remission; the better the new treatment the larger that number will be. It's also possible that patients who are in remission withdraw from the trial or die from a totally independent cause. The observation of the survival time of these patients is *censored*. At a certain (observed) moment in time a veil is drawn over their further history.

Just leaving all these cases out of one's analysis is inefficient at best, at worst completely misleading. It's most important to include all data, censored or not, recognising the difference.

Initially - I'm thinking of the fifties here - many ad hoc adaptations of the classical statistical methods of analysis mentioned above were devised. The only positive thing about these methods was that they supplied something that could be used. A breakthrough was brought about only by essentially new methods that fully recognised the dynamics of the situation. I must stress immediately that it was applied statisticians, N. Mantel and W. Haenszel in particular, using elementary mathematics, but with a strong and healthy intuition, who introduced these new methods [5].

The basic idea is this. It's no use comparing the number of patients in remission (surviving) in the two groups at a certain length of time after admission for these numbers not only result from the ending of the remission, but also from censoring: the termination of the trial and other causes. Let's suppose we look at a certain point in time within the two groups of survivors at that moment. If the censoring really is independent, each patient in these groups has the same conditional chance of leaving remission in a subsequent small time-interval, given the fact that he or she belongs to the group of survivors, as would have been the case without censoring. So it is possible to make a fair comparison between the two treatments for every such time-interval, at least as far as the conditional chance of leaving remission given the event that the patient has been in remission till then. Finally one has to combine this whole series of comparisons, working all the time conditionally on what has happened up to a certain point in time while watching the developments in the following infinitesimal time interval.

I have to go into some technical details now to show that the solution of Mantel and Haenszel entails a few quite unusual elements. Consider a time-interval during which a certain patient leaves remission which is so small that in this interval this is the only event occuring. Classify all patients who are present at the beginning of the interval in a two way table: per row, treatment group 1 or 2; per column, does or does not leave remission in this period. Calculate the (non squared) contribution of one of the four cells to the well-known chi-square test-statistic of independence (observed minus expected number), add these contributions over all the available points in time, square, divide by the sum of the expected numbers, and compare with the chi-square distribution with one degree of freedom.

To anyone acquainted with statistical theory this procedure makes a mysterious impression, however healthy the initial philosophy was. In each of our two by two tables one column has a total of one - one person leaving remission. We learn however the practical rule that the asymptotics - the 'large sample approximation' - of the chi-square test-statistic only works if at least five individuals per cell occur under the hypothesis of independence of treatment-group and the leaving of remission. Another objection is the fact that all these comparisons are interdependent; the more patients leave remission *now*, the fewer patients are left to compare at a later point in time.

Even though people weren't entirely happy with the justification of these methods, they put up with them, especially after from completely different, but just as bizarre reasonings (Fisher score test based on the marginal density of the rank numbers of the observations in an imaginary but intuitively comparable experimental design [6]) the same solution amazingly emerged. To tackle all kinds of variations on the basic problem described here, more and more refined methods of analysis were devised via ever more daring heuristics and intuition. The climax was D. R. Cox's regression model [7], introduced in 1972, and for which he only three years later (informally) pin-pointed the underlying idea. All this took place mainly in the medical-statistical and applied statistical specialist literature. Mathematical statisticians on the whole either ignored all this or viewed it with suspicion, though some of them studied the new statistical methods by using classical mathematical-statistical techniques. In a classic case [8] they discovered after gigantic calculations during which huge formulas canceled each other out as if by magic, that the final formulas were not only compact and attractive, but also the same as the ones suggested by the daring heuristics.

In this informal mathematics attention has shifted away from the traditional counting over individuals to the counting over points in time. The formal mathematics - in this case probability theory - needed to study this will have to make this transition too. A similar shift has taken place for medical-statistical practice: it is easier to apply statistics if we're prepared in the first place to investigate the time dependent conditional risk of leaving remission, rather than concentrating on the unconditional chances of leaving remission at different times.

Only in 1975 was the mathematical-statistical theory discovered that was needed for a satisfactory, elegant and complete account of these methods. This was given in the Berkeley thesis [9] of the Norwegian statistician Odd Olai Aalen. It took at least another ten years to fully exploit and develop and it gave me great pleasure to participate in this exciting process. Aalen arrived from Oslo with practical experience in analysing survival times: his master's thesis was concerned with an investigation of the survival time of the intra-uterine device. He was backed by a sound knowledge of the theory of stochastic processes ('one bloody thing after another' as the famous English statistician R.A. Fisher once explained to a journalist), in particular Markov processes, as used in statistical applications in demography and pervading modern probability theory. In Berkeley he came into contact with a group of mathematicians, the Frenchmen Bremaud and Jacod among others, who were busy applying a new theory of stochastic integrals - a stochastic infinitesimal calculus - to problems of controlling and filtering counting processes. It had become evident that the idea of the time-dependent, conditional intensity of new events in a random process was closely connected with the basic idea in this theory: the Doob-Meyer decomposition of a nice stochastic process into a systematic (predictable) part and a so-called *martingale*. Aalen recognised these same elements in survival analysis and discovered also that all kinds of statistically interesting quantities could be described in a simple way in terms of this theory, i.e. as stochastic integrals of predictable processes with respect to martingales.

I'll attempt to clarify these terms somewhat. The term martingale stems from Monte Carlo: it is a gambling system in which one supposes that if in roulette the ball has fallen on red in less than half of the rounds, the chances that this will happen in the next rounds will be bigger. In probability theory however a martingale is the abstraction of the cumulative gain (as function of time) in a fair game of chance. The average of this is zero, independent of gambling system or rule for setting stakes. In our example the connection is that if the two treatments are equally good, and if censoring is independent of survival, the result at each point in time of who, if anyone, leaves remission, is pure chance just as in roulette. The test-statistic of Mantel and Haenszel is the final gain in this game when using a certain rule of setting stakes. If one treatment is better than the other, it isn't a fair game and (on average) a cumulative gain (or loss) will occur. From martingale theory one can see how big a gain must be, to make it unlikely to be just the result of chance, i.e. to make a systematic difference plausible. This theory certainly isn't trivial especially in the 'continuous-time' version we need here.

All this amounted to the fact that, almost ready-made, exactly the right mathematical theory became available that translated the intuition of the first appliers into tough, precise mathematical theorems. Where necessary it also readjusted this intuition: certain aspects of it had led them astray. This mathematical theory had just been formed by Paul-André Meyer and his school in Strasbourg. For a long time it was quite inaccessible, made inpenetrable by so much French abstraction and details. All the more so as it had been designed as pure mathematics - l'art pour l'art - and rested on deep and abstract results from the potential theory of Choquet. Only gradually, and helped by the applications, the essentials of this theory became extracted.

Through the years more theory has been added that time and again could immediately be used in the applied field. Here I am thinking in particular of the martingale central limit theorems of the Chilean R. Rebolledo and of members of the Russian school around A.N. Shiryayev. With these one can give conditions under which all kinds of statistical quantities have an approximately normal distribution. The use of these theorems belongs to the most indispensable part of daily statistical practice (in particular the application of the method of Mantel and Haenszel).

Looking back it is not difficult to find all kinds of forerunners and indications of this theory in the applied literature. An interesting and very explicit example is offered by a paper [10] in a famous English biostatistical journal in which the authors *left out* a justification using martingal theory because this according to the editors would have been too difficult for the ordinary readers, who of course were mainly applied statisticians. This was a missed chance indeed. At the same time this paper also gives an example of the *derailing* of healthy practical intuition - I mention this to stress the fact that beautiful but difficult mathematics is not just a game, but on the contrary, absolutely vital to supply clarity, precision, and the firm base for the next soaring of intuition.

The application of this theory of martingals, counting processes and stochastic integrals has, via a number of brilliant successes, led to a uniform treatment of a whole range of methods in survival analysis and to the clear demarcation of their applicability. It also led to a pruning of the uncontrolled growth of partially useful, partially barren concepts and theories. Now one can with the greatest ease study new methods in an existing theoretical framework; it is also possible to transplant ideas from the limited area of survival-analysis to all kinds of different situations where the intensity of events in time is studied. I could give examples from demography, epidemiology, ethology, psychology, and econometrics. This baroque and inaccessible abstract theory has now become an extremely strong and intuitively quite manageable calculus.

I'd like to make some additional remarks concerning this last example. Firstly, all this activity has stimulated other areas within statistics very much indeed. Thus the applied field offered a number of examples of *semiparametric* *models*: statistical models in which infinite dimensional parameters occur. I won't explain what this means here, but the point is that exactly at the right moment survival analysis offered a number of specific and interesting examples and thereby gave a strong impetus to a just evolving synthesis of parametric and non-parametric methods in the very heart of mathematical statistics: phenomena arose that demanded admittance into a general theory; an experimental laboratory was available the findings of which any selfrespecting theory ought to be able to predict and explain.

As I mentioned before all this also has important consequences for other areas of applications. In particular it's now becoming possible to investigate in a sound way the interaction between observational frame and studied object (i.e. life-histories) in demography and epidemiology. I am thinking here of the fact that in these disciplines one is more often than not compelled to work with observational material that is retrospective in character: collected after the events, and depending on the random developments one actually studies included in the sample or not, with the resulting consequences for bias.

These developments also had their effect on the statistical tradition in certain countries, in particular France and the Soviet Union. These countries have very strong probabilistic schools, but are relatively weak as far as applied statistics are concerned. These new possibilities to apply their 'own' pure mathematics caused a renewed interest in applied statistics. In both countries mathematicians of stature became involved in it.

Of course it is appropriate to say something here about the consequences for cancer-research, which after all was the subject of this example. We must recognise the fact that statistics plays such a vital role here precisely because of the failure to make a real medical breakthrough. It is indeed a fact that statistics is at its best under circumstances like that. For the time being small improvements in 'average' survival time have to be traced and proven bit by bit with great difficulty. This way differences in optimal treatment for different patients have been established. It is true that other factors, 'quality of life' for instance sometimes undo these small gains.

I'd like to make a second remark to give you a proper perspective. Let's ask ourselves: what did the application of the martingale theory in this case really consist of? What was applied? Apart from the *modelling* of certain phenomena, and the *motivation* of certain methods of data analysis (both certainly rather important), this beautiful mathematics mainly supplied approximations the accuracy of which cannot (that is, not yet) be determined from the theory: the fact for instance that the test-statistic is *asymptotically* normally distributed, and in a certain sense *asymptotically* optimal. I don't want to suggest that such results can't be applied, on the contrary; but we must realise that this application isn't properly completed yet. In all situations of some complexity we'll always have to content ourselves with approximate solutions to the problems really posed. The theory has to be supplemented with empirical research and practical experience.

CONCLUSIONS

This last example illustrates the fact that the mathematical theorems of mathematical statistics, however beautiful and profound, are hardly ever applied in the narrow sense. (In my opinion the same applies to many branches of mathematics if more than trivial applications are at stake.) Mathematical statistics offers thinking-apparatus, tools with which to analyse complex problems and to reduce them by approximations to subproblems that can be understood well and thus solved. The varying demands of reality compels the statistician to find an ever changing compromise between analysability and realism of the mathematical modelling. The fact that there never is *one* right solution gives the mathematician the liberty to find the most far-reaching, the most enlightening solution. The inferential statistics I talked about comes into its own on the borderline between determinism and chaos, where patterns can only just be discerned. Statistics gets really interesting only if so much chance is involved, that a *definite* conclusion can't be drawn.

Let me finally quote Dirichlet [11], who in 1852 thought the direction of the then modern analysis to be: 'the supplanting of calculations by ideas'. In my opinion this should be the task of mathematical statistics even now. The distinction between pure and applied mathematics with which I started my discourse, can instinctively be characterised by beautiful ideas on the one hand, and possibly usefull but dull and endless calculations on the other. I hope to have shown you that this distinction is misplaced. The point is to find in applied mathematics the ideas that make the calculations selfevident and clear, and to shift the limits of ability and knowledge as far as possible.

REFERENCES

1. P.R. HALMOS (1981). Applied mathematics is bad mathematics. L.A. STEEN (ed.). *Mathematics Tomorrow,* Springer, New York, 9-20.
2. W. ALBERS (1985). *De Ongrijpbare Zekerheid* (Unattainable Certainty), Inaugural lecture, University of Limburg, Maastricht.
3. J. DE WITT (1671). *Waerdye van Lyf-Renten naer Proportie van Los-Renten,* The Hague. See also: A. HALD (in preparation), *A History of Probability, Statistics and Insurance Mathematics before 1880.*
4. Cooperative research group 'basispeilen' (1986). *Onderzoek Basispeilen. Interim Report, June 1986,* Report GWIO-86.008, Ministry of Public Works, Tidal waters section, The Hague. See also: L. DE HAAN (1986). The arch-enemy attacked with mathematics. M. HAZEWINKEL, J.K. LENSTRA, L.G.L.T. MEERTENS (eds.). *Mathematics and Computer Science* II, CWI Monograph 4, North-Holland, 51-60.
5. N. MANTEL, W. HAENSZEL (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst. 22*, 719-748. See also: N. MANTEL (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports 50*, 163-170.
6. R. PETO, J. PETO (1972). Asymptotically efficient rank invariant test procedures (with discussion). *J. Roy. Statist. Soc. (A) 135*, 185-206.

7. D. R. Cox (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. (B) 34*, 187-200.
8. J. Crowley, D.R. Thomas (1975). *Large Sample Theory for the Log Rank Test,* Tech. Report 415, Dept. of Statistics, Univ. of Wisconsin-Madison.
9. O. O. Aalen (1975). *Statistical Theory for a Family of Counting Processes,* Ph.D. thesis, University of California, Berkeley; reprint (1976), Institute of Mathematical Statistics, University of Copenhagen.
10. R.E Tarone, J. Ware (1977). On distribution free tests for the equality of survival distributions. *Biometrika 64*, 156-160.
11. G.P. Lejeune Dirichlet (1852). Gedächtenisrede auf C.G.J. Jacobi [p. 19], *Abh. Preuss. Akad. Wiss.* 1-26. See also: E.E. Kummer (1860). Gedächtenisrede auf G.P. Lejeune Dirichlet [p. 33], *Abh. König. Akad. Wiss. Berlin* 1-36; and: N. Bourbaki (1950). The architecture of mathematics [p.231], *Amer. Math. Monthly 57*, 221-231.