The Statistician (2003)
52, *Part* 4, *pp.* 423–438

Diversities of gifts, but the same spirit

Peter J. Green

University of Bristol, UK

[*The address of the President, delivered to* The Royal Statistical Society *on Wednesday, June 18th, 2003*]

Summary. This address reviews the great diversity of the discipline of statistics, seeking an essential unity among its various aspects. The role of statistical modelling in underpinning the subject is stressed. To safeguard the discipline in the future, it is seen as vital that bonds between different parts of the discipline are strengthened, to avoid fragmentation among its domains of application. Suggestions for improving the dissemination of new methodology are made, and the part that the Royal Statistical Society can play is discussed.

Keywords: Complexity; Dissemination; Highly structured stochastic systems; Statistical modelling; Statistics in society; Technology transfer

1. Introduction

Becoming President of an organization of such intellectual wealth and vitality is an honour deeply felt. Serving the Society in this way has been a challenge, but a rewarding one, and I hand over the chain and burdens of office with mixed feelings. The job would be quite impossible without the support and commitment of the staff and officers, and I thank them all, most sincerely.

The Royal Statistical Society (RSS) is a membership organization in the fullest sense of that phrase, and one aspect of this is only seen fully by the President, who has frequently to approach members—on behalf of the Council—to take on Society responsibilities. So long as future Presidents enjoy the very high rate of acceptance by ‘volunteers’ that I have, the Society is in good shape!

A corollary of the Society’s vigour is its capacity for disagreement: not only are there debates about philosophical and technical issues, but also differences of view about the directions in which the Society is developing—the latter not always comfortable for those holding the ring. There remains far more that unites us than divides us—this is the essential unity to which my title refers.

In this address, I shall take the opportunity to present my views on some aspects of our discipline, speaking for myself rather than on behalf of the Council or the Society at large. Much of what I have to say is quite personal in tone, but I shall also be trying to give an exposition of the central role of statistical modelling in our activities. The address will be primarily concerned with ‘statistics’ in its singular sense, rather than with numerical facts and their interpretations, without in the slightest meaning to suggest that the latter are not fully part of the discipline also. My comments are intended to apply as much to (for example) social scientists and engineers using statistics as they do to those who would describe themselves as statisticians. This address

Address for correspondence: Peter J. Green, Department of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK.
E-mail: p.j.green@bristol.ac.uk

 0039–0526/03/52423

is not a manifesto for the Society, but a personal view of the state and prospects for its unity and that of the subject.

2. A discipline of diversity

As we all recognize, statistics is an extraordinarily diverse discipline. It reaches out into business and industry, into public life and into most other disciplines. In turn, these interactions have profoundly shaped the subject. It embraces a huge variety of aspects: philosophical foundations, mathematical theory, inferential principles, design, data collection, techniques, computation, modelling, and so on. And, crucially, the conduct of its interactions with the rest of the world is part of the subject itself.

People come to statistics from many different directions. I began as a mathematician, and, in seeking a direction after my first degree, was primarily drawn to statistics because of its relevance to the world, but the intellectual diversity was also a major attraction. Others began as social scientists, medics, biologists, engineers, economists, ..., and with many different motivations.

Among statistical organizations worldwide, the RSS sets a superb example by fully representing this diversity of the discipline, seeking to embrace and support all of its strands. Perhaps more than anyone, the President is confronted by that diversity. You need to recognize both the political heterogeneity of the Society, with all its subgroups with goals and expectations competing for resources, and its intellectual heterogeneity, demanding appreciation, and ideally understanding, of many areas of statistical application and technique, and representation of those involved with them.

In this address, I want to explore the diversity of the discipline, and that of the Society. The RSS is remarkable and unique in its range and inclusivity; it both provides essential shelter and nourishment for the discipline and is a microcosm of it. Like the discipline in general, it faces threats and must survive them. And, as with the discipline, it is vital that its different traditions and subgroups remain connected.

3. Connection or fragmentation?

What holds statistics together? Why does, say, a Government statistician in the Department of Health feel that she is part of the same enterprise as a researcher in signal processing working on mobile phone technology? There is little in common in any aspect of their working lives, including, perhaps especially including, the technical content. Yet each would claim to be doing statistical work. Part of the common view is perhaps purely cultural, to do with an interest in the world outside the boundaries of the subject in which they were trained and a desire, not necessarily articulated, to exploit one's skills to better that world. That culture is certainly not universal, across disciplines and countries, but as an attitude learned by example certainly is strongly ingrained in statistics, at least in what is sometimes called the 'British tradition'. Most of what we do is 'statistics in society', whether we are computing DNA match probabilities for a court case, assessing clinical effectiveness, evaluating genetically modified crop experiments or getting the correct denominator in workforce statistics.

The heterogeneity of the discipline gives us intellectual strength, but structural weakness. In work, we tend to be distributed around our organizations, with 'statistics' often in neither the job title nor department name. There may be many employers of statisticians—but each recruits us only in small numbers. In universities, we are found in several different faculties and seek funding from several different research councils. In the UK, even the Government Statistical Service is spread across many different departments of state. Most of us are accountable to someone for whom statistics is a very small part of his or her concern, and many of us feel that we have no champion in a position of power in our organization.

Pressure to fragment is constantly there. After all, for most of us, the delivery of statistical information and its interpretation is not the main product of our organization: statistics is a means to an end. Many employers are happy to treat statistics as just a skill, as, say, information technology might be regarded.

The broad distribution of statistical activity through the workforce strengthens the impact of the discipline, and provides resilience, but fragmentation of the subject itself would spell its end. Divided out among application areas, we would crucially lose both contact with a core set of unifying principles and the cross-fertilization from one domain of application to another that has been such a potent stimulus to the development of the subject. Conversely, if we can stabilize the discipline demographically and strengthen our interconnections—without in any way turning our backs on the world—statistics can flourish.

4. How the discipline develops

Some users of statistical methods from outside the discipline apparently regard statistics as a static thing, a shelf-full of technique and good practice gathering dust, to be consulted only occasionally, and without enthusiasm or much engagement. In a world in which 'information' has become both a global currency and a global product, in which 'uncertainty' is undiminished and its impact more widely appreciated and in which 'quantification' rules, it is astonishing that a discipline whose centre-piece is the quantification of uncertain information should have this image. Promoting our strengths should be a key priority for the discipline in general and the Society in particular.

Statistics is far from static. The modern subject is almost unrecognizable from that of, say, 1963, and remarkably different even from 1983. Partly that is because of internal innovations, but much of it has to do with the impact of outside influences. What are the stimuli for the discipline's growth? With tentacles reaching in so many directions, it is not surprising that the stimuli are equally diverse.

The demands of applications are constant, posing challenging questions that stretch our subject. In public policy, there is the increasing requirement for evidence-based decision-making, and performance measurement. In the legal system, we see opportunities for a scientific evaluation of evidence. In social science, respect for quantification grows, and public archives and national statistics create opportunities that we need tools to exploit. In business, data mining transforms the scale of question that can be asked, and in technology we must deal with uncertainty in telecommunications and image analysers. In science there are statistical challenges on every scale from astronomy to genomics—indeed, perhaps now even down to the quantum level.

From another direction, technological change is enriching the subject. Sensors and instrumentation, data logging capacity, communications and sheer number processing power have changed unrecognizably within the span of our working lives, transforming the quantity and often the quality of the data with which we deal, and enabling highly computer-intensive analysis.

The theoretical side of the discipline has driven much development also. With the forgetting of old quarrels, divisions between the different philosophical schools have been diluted, releasing intellectual energy into more profitable pursuits, and leading to a broader, more inclusive view of the subject's mathematical and logical foundations. A particular bonus of this has been the rehabilitation of Bayesian statistics as a respectable approach to applied problems, which has carried with it a restoration of the role of probability in statistics. One of the most powerful and pervasive developments in statistical understanding of the last 30 years has been the recognition of conditional independence as a key principle in model building, and its representation and visualization through the ideas of graphical modelling. Another important influence has been probabilistic work on stochastic calculus and martingales, on point processes and in many other areas.

These diverse influences often interact constructively. To take just two examples: Bayesian methods have only become mainstream in biostatistics because the substantive questions, the theoretical framework and the computer hardware and software are now all in place. And we can do modern event history analysis because of the symbiotic development of the application area and the mathematical theory.

5. The role of statistical modelling

The idea of statistical modelling underpins all parts of the discipline, to a degree that is not fully recognized: to bend a metaphor wildly, it is a bridge spanning subdisciplines, enabling the transport of ideas in both directions, and it is the bond that ties them all together into a coherent whole.

In many aspects of statistical work, models are quite explicit and formally described, but there will be many statisticians who would deny a role for modelling in what they are doing and say that they are just ‘working with the data’. But it is really always there in the background. The most basic tabulation, or descriptive or exploratory summary of data, involves some conceptualization of what is free to vary, what scale it is measured on and what it might be dependent on. You would not calculate a mean, for example, without determining which data to include, deciding that variation was approximately additive and, probably, considering that the resulting statistic captured some intrinsic property of the source of the data; that all requires conceptualization. A formal model really only differs in terms of the precision or sophistication of the specification.

In a similar way, we usually contrast model-based and design-based approaches to the analysis of experiments, or the construction of complex sampling schemes, but in both cases you need some underlying model to express the design-based approach with sufficient precision to use it. Alternatively, think of a collection of ratios based on sample data, proportions of victims of crime by age group and social class, say: you cannot give an objective overall summary figure from the data alone; you need assumptions and an understanding of the context.

In other parts of the subject, the role of statistical modelling is more clear. Models provide discipline in the creation of methodology (indeed statistics must be unique in being founded on so few operational principles), but combining these with explicitly stated models for data to generate an extraordinary variety of inferential methods. Models also provide the framework for a philosophical study of our foundations, for expressing principles and of course for the provision of flexible and coherent computational tools.

We could make much more use of the language of statistical modelling to communicate ideas around the subject, and to break down some barriers between theory and practice. Occasionally, ideas do develop independently in more than one part of the subject. A nice example is the way that multilevel modelling and (usually, Bayesian) hierarchical modelling have coexisted with little cross-reference between them at either a methodological or application level until quite recently, and with few common users. I do not want to blur possibly important distinctions, or to denigrate the exceptional innovation in either area, but it would surely help to keep statistics as a connected subject if we could use a common language for similar constructs.

This does not mean that models are ever correct. In common with any other kind of model, a statistical model is just a convenient representation of part of reality, an approximation that is sufficiently good for the purpose, and not necessarily so for any other. Properly identifying the purpose is crucial. An important characteristic of statistical models is that they are capable of being criticized quantitatively, within the framework of the discipline itself, and that criticism should reflect many issues, not only the goodness of fit to the data.

My use of the term 'statistical model' in this section has been deliberately imprecise. I include prescriptions that fall short of providing a full probabilistic characterization of the observable data, where this is impractical because of a lack of information, or unnecessary because of the proposed analysis. However, I did intend to restrict attention to what are sometimes called *data models*, i.e. abstractions of the situation or system under study that yield data as random variables, just as the real situation or system yields data as numbers. This can be contrasted with what Breiman (2001) called *algorithmic modelling*, using machine learning techniques such as neural nets and support vector machines, and making great use of cross-validation for model validation. The distinctions are well set out by Breiman and his discussants; my present view is that this new culture has generated excitingly different methodological ideas, but that I still think it important to see them evaluated according to a more familiar culture. See, for example, Hastie *et al.* (2001).

6. Structured systems

I want to use this slightly more technical section to try to convey a flavour of the kind of work that I do; it is relevant here since the approach that I am referring to is not a particular kind of model, but a framework for building models, especially probabilistic models, for empirical data. What is more, it is an approach that has already found a very wide range of application in social science, medicine, science and technology. A recently published edited volume in which I had a hand (Green *et al.*, 2003) calls the area *highly structured stochastic systems*; as ingredients, it includes the more familiar concepts of *graphical models* and *hierarchical modelling*.

The key idea is that of understanding a complex system—whether it is a physical system, such as a complicated piece of instrumentation, or a social system, such as a collection of attributes defined on a structured human population—by modelling it globally. The key to building a global model is to assemble it from small, comprehensible, pieces or modules, each involving only a few variables. That modular structure is the basis for understanding the real system, capturing its important characteristics in statistical terms, defining appropriate inferential methods, conducting the computations to fit the model to data and interpreting the results (predictions, inferences about parameters, consequences for understanding of underlying processes, etc.). It provides a formal discipline for borrowing strength—distributing information from observed data across all the unknowns in the model. Although such techniques are still in their infancy, it includes methods of model criticism, which—like everything else—is expressed locally, enabling revision of local assumptions in an iterative cycle of model improvement.

The formal basis for this modular structure is conditional independence (see Dawid (1979)). This is a familiar idea from elementary probability but can be understood in even more down-to-earth terms: if we have three variables X, Y and Z, then X and Z are conditionally independent given Y if, knowing the value of Y, discovering the value of Z tells you nothing more about X. In a mathematically precise way, this property expresses the extent to which you can break a big system down into smaller pieces, sufficiently small that a statistical model can be proposed for each piece, on the basis of scientific theory, elicitation from experts or simply from empirical observation.

Models of this kind are often visualized graphically, in a diagram in which nodes represent instances of variables, which are linked by lines (and arrows). Roughly speaking, a line is drawn between two nodes if and only if the corresponding variables are *not* conditionally independent, given all others. See Lauritzen (1996) for the complete story. For example, in a genetics application, in which the variables are the genotypes at a particular locus among members of a family, the pattern of conditional independence is determined by the laws of genetics, and the resulting diagram looks like a family tree. In general, the diagram almost always provides a vivid picture of the structure of the system and gives mental cues to further stages in the model building. I regularly doodle such a picture in the early stages of discussing a new project with a colleague from another discipline; it is an effective and concise summary of the structure of the problem, and an extremely efficient aid to elicitation of the quantitative relationships between the variables.

This view of model structure makes sense in both Bayesian and frequentist modes—for me and many others, the Bayesian view is most satisfying intellectually here, in part because all nodes in the graph are treated symmetrically. Whether they are parameters, data, missing data, latent variables or random effects, the only important distinction is whether or not their values are observed. Further, all uncertainties are expressed in a common language, of conditional probability. But these points are admittedly partly matters of taste, and there are plenty of non-Bayesian uses for the graphical representation of conditional independence, and a structured approach to modelling; see, for example, Cox and Wermuth (1996), which has a particular emphasis on using graphical ideas to uncover structure in data from observational studies in the social sciences.

A key property of the Bayesian approach to structured modelling is that it automatically integrates out all sources of uncertainty, properly accounting for variability at all levels in the model. This includes, in principle, uncertainty in the model structure itself. It thus avoids the overoptimistic estimates of precision, and potential biases, that are characteristic of plug-in estimation. Non-Bayesian approaches can do this also, with extra effort and, arguably, in a more arbitrary way.

As an approach to model building, these structured systems have one obvious danger and one major limitation. Both have to do with complexity.

The approach effectively removes any limit to the complexity of the models that we can build and—in principle at least—fit to data. At one level, that is good—most of nature really is very complex, and, if a model does not recognize that, is it sufficiently flexible? Overfitting indeed leads to high variance and poor prediction, but that is at least overt, whereas bias from underfitting is usually hidden. In conventional practice with 'ordinary' statistical models, limitations of our imagination and a lack of appetite for algebraic complexity help to keep statistical models sufficiently small that they are commensurate with the amount of data, and we avoid overfitting. However, it is overoptimistic to imagine that these informal considerations should turn out to strike exactly the right balance—e.g. between bias and variance—appropriate in every circumstance.

The danger of unjustified complexity in highly structured stochastic modelling is mitigated in various ways. Firstly, what is important is the effect of overfitting on parameters of interest in the model—including, of course, the question of whether they are even identifiable. Quantities of real interest sometimes comprise only a small fraction of all the unknowns. We have long been happy to use traditional statistical models with random effects that may almost outnumber the observations; we just learn to count degrees of freedom properly, and the same idea holds here. Secondly, and more generally, measures of complexity of a fitted model are being developed that properly account for all the shrinkage and borrowing of strength that happen automatically in a Bayesian hierarchical model; see, for example, Spiegelhalter *et al.* (2002). Thirdly, the style of modelling and the associated computational techniques genuinely invite and facilitate thorough sensitivity analysis, in which assumptions are varied, model elaborations added and removed, and so on, and the effect on quantities of interest noted. However, there is a real need for further theoretical research to try to understand generically how local assumptions affect a model's global behaviour, and so provide guidance to practitioners.

The major limitation of complex systems of this kind is that they nearly always must rely on Markov chain Monte Carlo (MCMC) computation, at present, which rules out models with a very large number of variables. An active area of research to beat this limitation includes work on better MCMC methods, but we also need to invest effort in discovering generic approximation methods. This raises the interesting possibility that MCMC sampling, which has only quite recently become widely accepted as sufficiently accurate, properly used, may come to define a gold standard.

Notwithstanding these concerns, highly structured stochastic systems already present very many success stories, and there is only space to give a very partial list of general areas, as examples. In genomics and bioinformatics, there has been work in DNA and protein sequencing, in gene mapping and in evolutionary genetics. In spatial statistics, these ideas have been applied to image analysis and geographical epidemiology, and, in temporal problems, to longitudinal data analysis, financial time series and signal processing. This is a heterogeneous list of topics and encompasses all modelling styles from mechanistic to empirical; the common thread is the need to express complex patterns of dependence to extract the information that is required from the data.

7. The methodology gap

There has always been a wide range of different kinds of statistics, with clusters of like-minded people forming loosely defined subdisciplines. Such groupings develop their own common ideas and jargon and quite naturally there is weaker communication between groups than within groups. I sense that such problems of a lack of communication are growing. As you would hope, there is little evidence of this in the Society's journals, but those of us who regularly referee papers for subject-matter journals are well aware of the wide use of outdated and inappropriate statistical techniques in some areas. The pressures are certainly there: the pace of working life discourages reading around the subject and increases specialization, and places statisticians under pressure to produce quick approximate answers without the chance to assess whether the approximations are adequate for the purpose. Training becomes more focused so that theoreticians do not learn about data collection, let alone help to formulate a research hypothesis, whereas applied statisticians may not be given enough mathematical grounding to understand the limitations of a piece of software.

Most of us are guilty of allowing these gaps to develop. Even though the Society provides something for almost every specialism, how many of us exploit that? Many of us have, I hope, been to a Section meeting in the last year. But how long is it since you last went to a meeting of a *different* Section? I suggest that this is something we all need to work on; seeking greater understanding of other parts of statistics brings a personal benefit but also works to keep the discipline together for the longer-term good.

Our own journals do demonstrate some excellent practice, and without claiming any representativeness in what is a rather arbitrary selection—and with apologies to those omitted—I shall illustrate this point with a brief mention of four papers from recent volumes of *Applied Statistics*.

(a) Raab and Donnelly (1999) described an investigation into non-response bias in the context of a study of students' sexual behaviour. Both likelihood and Bayesian methods were used, with statistical models strongly informed by the structure of the data and knowledge of the subject-matter; the need to allow for non-ignorable non-response was demonstrated. Sources of real prior information were carefully categorized, and prior sensitivity was assessed.

(b) Prado *et al.* (2001) developed non-stationary time series models for multiple electroencephalographic series from patients in a neuropsychiatric clinic. Models were motivated using both considerations of data analysis and scientific understanding. Complex time-varying spatial structure was uncovered in the data, and its relevance to neuroscience was discussed, as were the limitations of the statistical models proposed.

(c) Diggle *et al.* (2002) presented a case-study describing the variation in the prevalence of malaria among children in villages in the Gambia. They fitted both child-specific and village level covariates with public health relevance and included both structured and unstructured spatial components to adjust for the effects of unmeasured covariates. They gave an analysis of problems of both modelling and implementation.

(d) Myles *et al.* (2002) constructed a complex model for the longitudinal analysis of data from a breast cancer screening study. With many short time series of data, there was a clear need for borrowing strength—sharing information across patients—to a degree that is determined by the data themselves. Data quality, modelling issues and subject-matter considerations were all carefully discussed.

Each of these papers seems to me exemplary for its full engagement with an issue of importance, through appropriate statistical modelling, with implications for both the application and the methodology fully examined.

8. Making more of methodology

Those developing statistical methodology have a curious image problem. From the more applied boundaries of the subject, they are sometimes categorized as either out of touch or too clever by half. Yet relevance to applications is nearly always the main stimulus to and justification for research in methodology, and, for example, when competing for research funding with other mathematical scientists, work of this kind is often dismissed as failing to be sufficiently mathematical!

There are layers of principles here, and a continuum of shades of generality, but I would argue that for the vigour of the subject and cross-fertilization between its domains of application there is a vital place for 'generic methodology' existing and properly supported, between mathematical statistics and specific applications. Without that provision, how would, say, the framework of generalized linear models have been developed? The facts that the generalized likelihood ratio test and the Fisher scoring method already existed, as did the practice of fitting dose–response relationships by maximum likelihood in toxicology, do not diminish the achievement of identifying an underpinning framework for the whole class of models, and developing the general methodology of model specification, computation and inference (Nelder and Wedderburn, 1972). The framework and methodology have in turn generated application-specific techniques, and helped to promote good practice generally.

This is just one example of very many that could be cited, of course. We must ensure the continued existence of an intellectual environment in which future such innovations can be made. Arguments of this kind need to be made repeatedly to all with responsibility for strategic planning in higher education, especially funding agencies, which tend to make naïve distinctions between core and applied work: the one potentially failing tests of 'relevance' and the other treated as another funder's problem.

A good research contribution in statistical methodology has many hurdles to surmount before seeing publication. A paper typically must cover the philosophical principles underpinning the work, some mathematical development, statistical modelling of some real process or phenomenon, computational implementation, data analysis, model criticism and interpretation, both of resulting inferences and of the performance of the method in both its statistical and computational aspects. There will commonly be innovative aspects to all these ingredients, as well as a need for comparisons with existing techniques at each stage. The completed work then must face the commendably high standards (and lamentably slow pace) of refereeing that are typical in statistics, and so it is hardly surprising that statisticians' publications lists are shorter than those of most other researchers!

There is understandable criticism that generic methodology is often illustrated by atypically simple and unconvincing applications, but the presence of all these pressures does mostly explain this. In any case, there may be good reason to keep illustrations simple for the sake of exposition, and where there are several different target application areas. Statistics research may very often be driven by (specific) applications, but it should not necessarily be bound by them.

I hesitate to add to the burdens of methodologists, but perhaps they (we) must go even a little further. Now that methodology is often so complicated and computationally intensive, the standard dissemination vehicle of the 16-page refereed learned journal paper is no longer adequate. Such a paper still serves some of its traditional purposes, of providing a permanent record of the event of the discovery, a signal of quality and a compact introduction to the ideas. But it fails in other respects. Most statistics papers, as published, no longer satisfy the conventional scientific criterion of reproducibility: could a reasonably competent and adequately equipped reader obtain equivalent results if the experiment or analysis were repeated? Typically, the answer is no, both because there is not space to specify sufficient detail and because repetition could involve a huge cost in time and effort in developing computer code to parallel that of the author. Furthermore, the purpose of the research presumably goes well beyond the confines of the illustrations within the text of the paper; the intention is surely that it be used, and not solely by its progenitor. This considerably raises the stakes as to what 'reproducibility' might mean in statistics. Although a typical article will discuss the principles that guide future applications, it is often of limited use when it comes to the practicalities of doing so. I suggest that the author ought to take much more of the responsibility for the 'technology transfer' process of getting the technique into the applied user's toolbox than is currently typical.

A start on addressing these problems can be made by encouraging, and perhaps in the course of time insisting on, the public dissemination of a *reference implementation* of the methodology. This proposal was convincingly articulated by Brian Ripley in his plenary lecture at the RSS 2002 conference, 'Statistical methods *need* software'. A reference implementation is, in his words,

> 'some code which is warranted to give the authors' intended answers in a moderately-sized problem. It need not be efficient, but it should be available to anyone and everyone.'

Normally, this code should be capable of reproducing the examples in the paper. By definition, a computer code implementing a method is a completely precise description of that method and, provided that certain housekeeping matters, such as initialization of random-number streams in simulations, are dealt with, the test of reproducibility is satisfied.

There must be *caveats*. There will be occasions when it is necessary to restrict access to raw data for reasons of confidentiality or commercial sensitivity, and the author may wish to protect intellectual property rights in her code; it may nevertheless be appropriate to publish a description of the underlying principles, although perhaps the criteria for a paper's acceptance might then change to reflect the fact that open dissemination of the methodology is not really the goal.

As Brian Ripley argued, it is desirable for there to be absolutely no constraints on using such reference code, and he advocated that it be written in a freely available open source language, namely R (Ihaka and Gentleman, 1996). Versions of this for most popular computer systems, including all recent versions of Windows, can be downloaded from the Internet (see `http://cran.r-project.org`).

If we accept the idea that the dissemination of statistical methodology research has at least two components—the paper and the reference implementation—then it is a logical next step to adopt a good idea from current practice in the life sciences, that of the Web page supporting the paper, whereon the published paper is available, together typically with a more discursive version of the work, giving more detail than space allows in a journal, and access to the data sets used in illustrations. This is a convenient place to put the reference code and instructions for users.

This extra work would require something of a culture change among researchers, but I think entails a cost that is worth paying. As Ripley (2002) says,

> 'the process of getting methods into the hands of end users is undervalued by academia and the statistical community at large',

and

> 'some statistical methodology is seriously undervalued as a result'.

In return, employers and funders of statistical researchers have a duty to reward properly authors who make a serious effort to make their innovations generally usable, by modifying promotions criteria, grant assessment procedures, etc. Organizations serving the discipline, such as the RSS, also have a part to play in this, as I shall suggest below.

More radical changes to vehicles for disseminating methodology may be coming along. We already have interactivity in browsing journal articles: all the RSS journals are now published on line simultaneously with paper printing, and, through Blackwell's Synergy system, hypertext mark-up language versions provide navigation around a paper, and 'realtime' cross-referencing to cited articles. There is active research into distributed statistical systems such as the Omega project for statistical computing (`http://www.omegahat.org/`) in which data and methods are delivered over the Web. If all this is integrated, fully interactive dissemination of research would become possible.

9. The Society's role in closing gaps

I have argued that we need to bridge some gaps within statistics, for the future health of the discipline. Doing so is thus clearly within the scope of the RSS's objectives—nurturing the discipline should surely involve keeping it alive and in one piece! Furthermore, through its philosophy and history it has the breadth and inclusivity to be able to act successfully on those objectives. A noteworthy example of gap bridging was the merger with the Institute of Statisticians 10 years ago; this has strengthened both of our voices and enriched our activities in a way that amply repays the occasional tensions in operating an organization with a dual personality, simultaneously a Learned Society and a professional body. (I believe that we must keep this structure, supporting both the professional statistician and the general interest member; we are not and could not become sufficiently strong—either in numerical weight or intellectual breadth—with one group alone.) How can the RSS do more to bridge gaps through its activities?

For many, the publications are the most valued benefit of membership of the Society. I believe that the recent publications review, led by Sylvia Richardson, has produced excellent recommendations (which have been accepted by Council and are now being implemented). Given that our whole journal is simply too large to be treated as a single publication, the idea of rebranding the different series as being distinguished essentially by the level of technicality needed in exposition and not field of application should certainly assist in cross-fertilization. The revised remits of all the series reaffirm the role of the whole journal as focusing on statistical developments that are relevant to the real world.

The review advocated an additional role for Series B in publishing review style papers that seek actively to deliver the cutting edge methodologies that are normally published in its pages to a broader range of readers, and I think that Series C could usefully further develop its 'technology transfer' role as well. I also have high expectations for the new magazine, to appear from early 2004, which will be sent to all Fellows as part of their membership benefits. Apart from *RSS News*, which seldom carries very much technical content, this will be the first publication for decades that will be seen by all members, and it provides a tremendous opportunity for us all to learn about other parts of our subject, and unfamiliar domains of application.

I hope that we can also see more proactive cross-fertilization in other aspects of Society activity. Some recent joint meetings organized between Sections have been very successful at bridging gaps, with the meeting in January 2003 on performance monitoring providing an excellent example, and there is surely much scope for this kind of meeting. The new arrangements for organizing the Society's annual conference via an appointed Programme Chair and committee should work to deliver a more integrated programme. That of next year's Society conference, RSS 2004 in Manchester, co-ordinated by John Fox, has an explicit objective of connecting practice with research, an excellent opportunity to build some new links across the discipline.

The education and training of the next generations of statisticians is and should be an important focus of Society activity. In the context of UK higher education, current pressures make this difficult. There is a serious shortage of statisticians who wish to pursue an academic teaching career, driving down the proportions of statisticians among mathematical science lecturers, with the obvious effect on syllabuses. At the same time, many feel that students are entering higher education less well prepared for degree level mathematics and statistics. Graduates with 'statistics' in their degree title are few, and they are less fully trained than they need to be. This seems to place a greater demand for Master of Science level courses in statistics, at just the time when public funding for such programmes is becoming especially difficult. The model that we may be expected to adopt for professional training involves Masters level modules studied while in work, under company sponsorship, a framework that many feel is totally unsuited to the character of the discipline of statistics, and the very broad range of careers where statistics skills are needed. Some of these observations are admittedly subjective and based on anecdotal evidence; a much clearer and more authoritative picture should emerge from the Society's current study of the recruitment and supply of statisticians, led by Julian Calder and Tony O'Hagan.

We should be realistic about what the Society could possibly achieve within higher education given its limited resources, but I believe that it could play a more proactive role in promoting the use of centrally funded teaching resources (e.g. those produced through the Learning and Teaching Support Network for mathematics, statistics and operations research by the RSS Centre for Statistical Education), and working to raise the profile of our scheme for accrediting university degree courses. Such measures could act to enhance the quality of statistics provision, especially in terms of its relevance, but also make it more efficient, thus giving it some protection against funding cuts and our demographic problems.

At a somewhat more advanced level, short courses, here at Errol Street, as well as part of our conferences, are being proposed as a means of raising revenue. They also can play an important function in broadening perspectives and preserving standards. I hope that the educational purpose of such courses will not be compromised through an undue emphasis on income generation, in view of the very wide variations in the abilities of different kinds of employer to pay for such training. At least, we should price flexibly and seek various sources of funding. In the case of the recently instituted graduate training programme in statistics organized by the Research Section, research council funding is an important ingredient in making them feasible.

A final suggestion to the Society on promoting interconnections between parts of the discipline concerns the honours system. The criteria for the Bradford Hill Medal for medical statistics explicitly include excellence in exposition, alongside development and application, and although we cannot easily rewrite medal criteria generally I would like to see greater recognition of successful dissemination of innovation in the assessment of candidates for awards. And could we award some kind of 'seal of approval' for good practice in making software associated with research papers available to the public?

10. Only connect the prose and the passion. . .

I have tried to convey concern rather than alarm about the health of our discipline. In fact, I am optimistic that it will flourish into the future. To do so, statistics must remain an intellectually exciting field for young people to enter, as well as becoming adequately financially rewarding. And, although it is unlikely that such a diverse collection of people will ever be speaking with only one voice, it is vital that we remain connected—and in this the Society can have a leading role.

Peter Armitage took his Presidential address in part as 'an opportunity to demonstrate the unity in diversity which characterizes our activities' (Armitage, 1983). The theme has not, I think, been quite so overtly revisited by Presidents in the 20 years since then, although by the nature of our subject it is never very far from the discussion. In seconding the vote of thanks to Armitage, Claus Moser commented that

> 'We still suffer from too much diversity and inadequate bridges between the various parts of our subjects'.

As I hope to have made clear, I regard the diversity to be celebrated rather than suffered, but I certainly share Lord Moser's view that the bridges need to be strengthened.

> 'Only connect! That was the whole of her sermon. Only connect the prose and the passion, and both will be exalted . . . .'

Howards End, chapter 22
(E. M. Forster)

This was not written about statistics, and I would not suggest that any of what we do is really prosaic, but perhaps we should all try harder to understand what people are passionate about in other parts of our discipline!

References

Armitage, P. (1983) Trials and errors: the emergence of clinical statistics. *J. R. Statist. Soc.* A, **146**, 321–334.
Breiman, L. (2001) Statistical modeling: the two cultures (with discussion). *Statist. Sci.*, **16**, 199–231.
Cox, D. R. and Wermuth, N. (1996) *Multivariate Dependencies*. London: Chapman and Hall.
Dawid, A. P. (1979) Conditional independence in statistical theory (with discussion). *J. R. Statist. Soc.* B, **41**, 1–31.

Diggle, P., Moyeed, R., Rowlingson, B. and Thomson, M. (2002) Childhood malaria in the Gambia: a case-study in model-based geostatistics. *Appl. Statist.*, **51**, 493–506.
Green, P. J., Hjort, N. L. and Richardson, S. (eds) (2003) *Highly Structured Stochastic Systems*. Oxford: Oxford University Press.
Hastie, T., Tibshirani, R. and Freedman, J. (2001) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer.
Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Statist.*, **5**, 299–314.
Lauritzen, S. (1996) *Graphical Models*. Oxford: Oxford University Press.
Myles, J. P., Duffy, S. W., Prevost, T. C., Day, N. E., Hakama, M. and Salminen, T. (2002) Markov chain Monte Carlo inference applied to a complex model for the longitudinal analysis of breast patterns. *Appl. Statist.*, **51**, 197–207.
Nelder, J. A. and Wedderburn, R. W. M. (1972) Generalized linear models. *J. R. Statist. Soc.* A, **135**, 370–384.
Prado, R., West, M. and Krystal, A. D. (2001) Multi-channel electroencephalographic analyses via dynamic regression models with time-varying lag–lead structure. *Appl. Statist.*, **50**, 95–110.
Raab, G. M. and Donnelly, C. A. (1999) Information on sexual behaviour when some data are missing. *Appl. Statist.*, **48**, 117–133.
Ripley, B. D. (2002) Statistical methods *need* software: a view of statistical computing. *Royal Statistical Society Conf, Plymouth, Sept*. (Available from `http://www.stats.ox.ac.uk`.)
Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.

Vote of thanks

Denise Lievesley (*United Nations Educational, Scientific and Cultural Organisation, Montreal*)
Peter's address has characteristically revealed as much about himself as his subject-matter and his wide sweep has given us much to dwell on. His passion for statistics is revealed in his choice of topic—how can we bring under one umbrella what can often appear to be a large number of disparate disciplines? He lights on statistical models as the common factor and gives us a masterly summary of the essence of modelling processes. I suspect that many of us who do not specialize in Markov chain Monte Carlo methods would have been happy to hear much more about his current research. But, appropriately for a Presidential address, he resists the temptation and turns to his other concerns, namely the future of statistics and the role of the Royal Statistical Society (RSS).

I would like to pick on just a few of Peter's themes for comment and extension. He celebrates the diversity of the profession, focusing on the various aspects of the subject and the theoretical–applied divide. But it seems to me that we should also be concerned with *global diversity*—the contrast between the developed and the developing world—and we must ensure that the international dimension is fully integrated into the Society's work once the new governance and structure framework is in place.

Peter highlights the fact that being President of the RSS forces one to address issues that are outside one's specialist realm. A very positive illustration is Peter's own participation in the National Statistics Working Party where his new perceptions and wise counsel have been so valuable. Peter is apparently an expert in 'reverse jumps'. I hope that he will not do one now as we need his continued participation in social and official statistics.

Pursuing the topic of diversity, he notes the danger that divisions among our application areas, and an increasing emphasis on specialization, might cause us to lose contact with a core set of unifying principles. In his Presidential address to the International Statistic Institute, Peter Jakob Bjerve expressed 'worries about the wide gap existing between official statisticians and academic statisticians' (Bjerve, 1975). A committee, under the chairmanship of Joe Duncan and Jim Durbin, to review this problem urged that the Institute should regard the integration of statistics as one of its major objectives (Duncan and Durbin, 1980). And, in 1992, the Moriguti report concluded that 'history indicates that both specialization and integration have their value for members of the statistics profession' (Moriguti, 1994).

One of the strengths of statistics is that it is an applied science and Peter cites and lauds four papers which develop new methodologies that are applicable to real world phenomena. We could ask what the profession and the RSS can do to strengthen the relevance of statistics. Do we have mechanisms for focusing our efforts on important problems of society? Could we not, for example, peer review how well official statistics address key policy issues? The peer reviewers could be statisticians together with substantive experts in the field. But to be effective we must focus on content and methodology and not solely on protocol and procedure.

As Peter points out, difficulties of communication within the profession are a concomitant of our diversity. But is this unique to statistics, or is it also a problem in other fields as research becomes increasingly specialized? Do not the pressures on our time mean that we can no longer be catholic in our interests? I share Peter's regrets about this. I always gain something of relevance to my work by attending an Ordinary or Section meeting of the Society, however remote the subject might appear to be from my own.

In considering Peter's comments about modelling as a unifying theme, it is interesting to note that official statisticians often use models in areas such as national accounts, balancing of input–output tables, hedonic regressions for quality adjustment, population estimates and estimating purchasing power parities. Furthermore, there are other cases where their models are implicit and involve unarticulated assumptions such as 'non-respondents are missing at random'.

As an interesting example of the use of models in official statistics, the US Academy of Science recommended the application of small area methods to estimate the proportion of people living in poverty rather than relying on out-of-date census data. Validation of the small area estimates was a key part of the study and had a big influence on the findings.

In discussing this research, the Australian Statistician Dennis Trewin concluded that official statisticians should make greater use of modelling but stressed that the underlying models need to be described in terms that users can understand so that their validity can be debated (Trewin, 1999).

Peter discusses issues concerning the publication of papers in learned journals and mentions the use of the Web to complement the paper publication. His suggestion that the Web version should provide implementation details is excellent. This can be an effective way of bridging theoretical and applied statistics.

The current digital revolution is changing the paradigm of scholarly communications. In a talk that I attended in New York recently, Professor Harold Varmus described the Public Library of Science (`http://www.plos.org/`) which reverses the business model by charging authors the cost of the publication and making the results freely available for anyone with Internet access. This could be an attractive model particularly for scientists in developing countries where the cost of subscribing to a large number of journals is punitive. It would of course be a difficult model for the RSS to adopt at present, given that a large proportion of our income derives from the sale of publications. Nevertheless it is important that we keep abreast of such developments.

I would like to congratulate Peter on a most stimulating and panoramic presentation and to propose a vote of thanks to him not only for his paper but also for his vision and commitment in leading the RSS as it seeks to review its structure and governance. I have learnt much from working with him and seeing his analytical approach in action. Particularly memorable was his formula for determining the RSS's reserves which I hope he will submit to the research assessment exercise.

Fostering mutual respect within our profession is the key message of his presentation and he has led by example.

Before handing over to Robert Curnow, my predecessor as President of the RSS, to second the vote, I would just like to reassure those who are new to our annual general meetings that they will not have to hear speeches from *all* the Past-Presidents who are listed on the board on the wall.

In conclusion, I extend my warm congratulations to Peter on becoming a Fellow of the Royal Society.

R. N. Curnow (*University of Reading*)
President, Past-President, Fellows and guests: I first join with Professor Lievesley in congratulating and thanking Professor Green for all that he has achieved during his 2 years as President of this Society. With many other colleagues, I also congratulate him on his recent election to Fellowship of the Royal Society.

Each President expects his or her successor to have an easier time in office than they did. I doubt whether, in recent years, this has ever been true. The Society continues to extend its activities and its influence in society. During the last 2 years under Professor Green's leadership we have seen more progress, particularly in the contributions that the Society makes to the discussion of issues of national importance and in the planning of improvements in the services to be provided to Fellows.

The President has reviewed for us several aspects of the current state of our profession. The expertise and experience underlying his comments on the central role of statistical modelling will ensure a respectful but also, in the traditions of our Society, a detailed and critical assessment.

Professor Green comments that 'the modern subject is almost unrecognizable from that of, say, 1963'. I was around at that time, just, and can say, as in the title of the address, that the spirit was the same. The restrictions in those days to the use of simple and obviously unrealistic empirical models made work with colleagues of other disciplines difficult. Indeed, some of the best scientists rightly rejected our advances!

Many, but not all, of the objections to Bayesian methods centred on the need to assume prior distributions simply because they were conjugate to the data distributions and therefore could be handled analytically. Most of us foresaw the eventual removal of these aggravating restrictions. Few, if any, of us expected the explosion of computing power that occurred and the massive opportunities that it provided for more realistic modelling and analysis. I believe that our profession has good reason to celebrate the advantages that we have taken of the computer revolution through the development of new ideas and new methodology. The abilities to design and analyse experiments by using realistic models and highly structured whole systems have vastly increased our relevance in many areas. Our advice is now much more widely sought. Despite this we still moan about being ignored where we have much to contribute!

What might have been predicted but has not happened? In the late 1950s and early 1960s there was increasing interest in the use of statistical methods to look at the optimal allocation of resources in production processes and in those experiments that lead to decisions, e.g. the planning of whole drug screening processes. The objectives were not to maximize power to negate meaningless null hypotheses nor to achieve a specified accuracy. Rather, the objectives were to maximize the worth of the drugs or processes selected at the end of an entire process. This brings in costs and benefits explicitly and prior distributions of performance. I do not argue that this approach is always appropriate. Increased understanding of mechanisms and processes is nearly always important and often paramount. However, I am surprised that decision theory and the associated problems of estimating costs and benefits have not been developed and applied as successfully as have other areas of statistical inference.

On a similar theme, Government policy making generally involves choosing between alternatives each with their cost and benefit. Obvious recent examples are the decisions about the measles, mumps and rubella vaccine and about the use of UK blood donations and of reusable surgical instruments following the cases of variant Creutzfeld–Jakob disease. Appeals to the so-called precautionary principle are rarely relevant. We need to become even more involved in the way that information is used to influence these decisions. Should committees elicit prior distributions and, if so, how? Should we pool priors or pool individual decisions? Are we sufficiently engaged in how the costs and benefits have been estimated? The ability to model and analyse complex systems has, as Professor Green emphasizes, provided much better information and understanding. Armed with these tools we need to become more involved in the decision-making process itself. I suggest also that in many circumstances the estimated consequences under different assumptions need to be presented rather than the uncertainty increased to represent the sensitivity to assumptions. This highlights sensitivity and hence the importance or otherwise of the assumptions. Where there is significant sensitivity, decision makers may then use their own judgments to reach personal decisions that are then pooled.

I share Professor Green's concern about the increasing but I believe inevitable specialization in areas of application. As he says, this tendency is reducing the appreciation of methodology developed in one area but actually of wider application. He highlights the ways in which the Society plans to reduce this isolation through its publications and meetings. I hope that there will be more meetings with other Societies as well as joint meetings between our Sections. I also share Professor Green's concern about the reduced mathematical background of many recently graduated applied statisticians. Again referring to the 1950s and 1960s, the emphasis then was, for understandable reason, on teaching the mathematical derivations of distributions and the detailed theory of simple and therefore unrealistic stochastic processes. The pendulum has swung too far to insufficient understanding of the theory and structure underlying our discipline. This can obviously lead to an inappropriate use of new techniques and the limitation of new developments.

We need to ensure that our colleagues who become knowledgeable about specific areas of application and can therefore recognize and adapt new advanced methodology are given appropriate credit for their particular skills. I fully agree with the President that the Society must argue against any suggestion that the development of general methodology and its application in new areas are competitors in esteem or in funding. As a Society and a profession we surely agree with Professor Ripley and Professor Green that the public dissemination of 'a reference implementation' of all new methodology through printed material or the Web should be recognized as a vital part of publication.

I have much pleasure in seconding the vote of thanks to Professor Green for his period as President and a stimulating address this evening.

The vote of thanks was passed by acclamation.

References

Bjerve, P. J. (1975) Presidential Address presented to the 40th session of the International Statistical Institute. International Statistical Institute, The Hague.

Duncan, J. W. and Durbin, J. (1980) Report of the International Statistical Institute Committee on the Integration of Statistics. *Int. Statist. Rev.*, **48**.
Moriguti, S. (1994) The Moriguti report on the role of statisticians. *Occasional Paper Series 4*. International Statistical Institute, The Hague.
Trewin, D. (1999) Report on small area statistics conference. In *The Survey Statistician*. Libourne: International Association of Survey Statisticians.