

4

Exchangeability and regression models

Peter McCullagh

4.1 Introduction

Sir David Cox's statistical career and his lifelong interest in the theory and application of stochastic processes began with problems in the wool industry. The problem of drafting a strand of wool yarn to near-uniform width is not an auspicious starting point, but an impressive array of temporal and spectral methods from stationary time series were brought to bear on the problem in Cox (1949). His ability to extract the fundamental from the mundane became evident in his discovery or construction of the eponymous Cox process in the counting of neps in a sample of wool yarn Cox (1955*a*). Subsequent applications include hydrology and long-range dependence (Davison and Cox 1989, Cox 1991), models for rainfall (Cox and Isham 1988, Rodriguez-Iturbe *et al.* 1987, 1988), and models for the spread of infectious diseases (Anderson *et al.* 1989).

At some point in the late 1950s, the emphasis shifted to statistical models for dependence, the way in which a response variable depends on known explanatory variables or factors. Highlights include two books on the planning and analysis of experiments, seminal papers on binary regression, the Box–Cox transformation, and an oddly titled paper on survival analysis. This brief summary is a gross simplification of Sir David's work, but it suits my purpose by way of introduction because the chief goal of this chapter is to explore the relation between exchangeability, a concept from stochastic processes, and regression models in which the observed process is modulated by a covariate.

A stochastic process is a collection of random variables, Y_1, Y_2, \dots , usually an infinite set, though not necessarily an ordered sequence. What this means is that \mathcal{U} is an index set on which Y is defined, and for each finite subset $S = \{u_1, \dots, u_n\}$ of elements in \mathcal{U} , the value $Y(S) = (Y(u_1), \dots, Y(u_n))$ of the process on S has distribution P_S on \mathbb{R}^S . This chapter is a little unconventional in that it emphasizes probability distributions rather than random variables. A real-valued process is thus a consistent assignment of probability distributions to observation spaces such that the distribution P_n on \mathbb{R}^n is the marginal distribution of P_{n+1} on \mathbb{R}^{n+1} under deletion of the relevant co-ordinate. A notation such as \mathbb{R}^n that puts undue emphasis on the incidental, the dimension of the observation space, is not entirely satisfactory. Two samples of equal size need not have the same distribution, so we write \mathbb{R}^S rather than \mathbb{R}^n for the set of real-valued functions on the sampled units, and P_S for the distribution.

A process is said to be exchangeable if each finite-dimensional distribution is symmetric, or invariant under co-ordinate permutation. The definition suggests that exchangeability can have no role in statistical models for dependence, in which the distributions are overtly non-exchangeable on account of differences in covariate values. I argue that this narrow view is mistaken for two reasons. First, every regression model is a set of processes in which the distributions are indexed by the finite restrictions of the covariate, and regression exchangeability is defined naturally with that in mind. Second, regression exchangeability has a number of fundamental implications connected with lack of interference (Cox 1958*a*) and absence of unmeasured covariates (Greenland *et al.* 1999). This chapter explores the role of exchangeability in a range of regression models, including generalized linear models, biased-sampling models (Vardi 1985), block factors and random-effects models, models for spatial dependence, and growth-curve models. The fundamental distinction between parameter estimation and sample-space prediction is a recurring theme; see Examples 5 and 6 below.

Apart from its necessity for asymptotic approximations, the main reason for emphasizing processes over distributions is that the unnatural distinction between estimation and prediction is removed. An estimated variety contrast of 50 ± 15 kg/ha is simply a prediction concerning the likely difference of yields under similar circumstances in future seasons. Although the theory of estimation could be subsumed under a theory of prediction for statistical models, there are compelling reasons for maintaining the separation. On a purely theoretical point, estimation in the sense of inference concerning the model parameter may be governed by the likelihood principle, whereas inference in the sense of prediction is not: see Section 4.5.1 below. Second, apart from convenience of presentation and linguistic style, parameter estimation is the first step in naive prediction. The second step, frequently trivial and therefore ignored, is the calculation of conditional distributions or conditional expectations, as in prediction for processes in the standard probabilistic sense. Finally, parameter estimation is equivariant under non-linear transformation: the evidence from the data in favour of $g(\theta) \in S$ is the same as the evidence for $\theta \in g^{-1}S$. Pointwise prediction, in the sense of the conditional mean of the response distribution on a new unit, is not equivariant under non-linear response transformation: the mean of $g(Y)$ is not $g(EY)$.

We do not aim to contribute to philosophical matters such as where the model comes from, nor to answer practical questions such as how to select a model within a class of models, how to compute the likelihood function, or how to decide whether a model is adequate for the task at hand. In addition, while the mathematical interpretation is clear, any physical interpretation of the model requires a correspondence between the mathematical objects, such as units, covariates and observations, and the physical objects that they represent. This correspondence is usually implicit in most discussions of models. Despite its importance, the present chapter has little to say on the matter.

4.2 Regression models

4.2.1 Introduction

We begin with the presumption that every statistical model is a set of processes, one process for each parameter value $\theta \in \Theta$, aiming to explore the consequences of that condition for regression models. The reason for emphasizing processes over distributions is that a process permits inferences in the form of predictions for the response on unsampled units, including point predictions, interval predictions and distributional predictions. Without the notion of a process, the concept of further units beyond those in the sample does not exist as an integral part of the mathematical construction, which greatly limits the possibilities for prediction and inference. Much of asymptotic theory, for instance, would be impossible in the absence of a process or set of processes.

To each potential sample, survey or experiment there corresponds an observation space, and it is the function of a process to associate a probability distribution with each of these spaces. For notational simplicity, we restrict our attention to real-valued processes in which the observation space corresponding to a sample of size n is the n -dimensional vector space of functions on the sampled units. The response value is denoted by $Y \in \mathbb{R}^n$. Other processes exist in which an observation is a more complicated object such as a tree or partition (Kingman 1978), but in a regression model where we have one measurement on each unit, the observation space is invariably a product set, such as \mathbb{R}^n or $\{0, 1\}^n$, of responses or functions on the sampled units.

The processes with which we are concerned here are defined on the set \mathcal{U} of statistical units and observed on a finite subset called the sample. The entire set or population \mathcal{U} is assumed to be countably infinite, and the sample $S \subset \mathcal{U}$ is a finite subset of size n . The term sample does not imply a random sample: in a time series the sample units are usually consecutive points, and similar remarks apply to agricultural field experiments where the sample units are usually adjacent plots in the same field. In other contexts, the sample may be stratified as a function of the covariate or classification factor. A process P is a function that associates with each finite sample $S \subset \mathcal{U}$ of size n a distribution P_S on the observation space \mathbb{R}^S of dimension n . Let $S \subset S'$ be a subsample, and let $P_{S'}$ be the distribution on $\mathbb{R}^{S'}$ determined by the process. If logical contradictions are to be avoided, P_S must be the marginal distribution of $P_{S'}$ under the operation of co-ordinate deletion, that is, deletion of those units not in S' . A process is thus a collection of mutually compatible distributions of this sort, one distribution on each of the potential observation spaces.

An exchangeable process is one for which each distribution P_S is symmetric, or invariant under co-ordinate permutation. Sometimes the term infinitely exchangeable process is used, but the additional adjective is unnecessary when it is understood that we are dealing with a process defined on a countably infinite set. Exchangeability is a fundamental notion, and much effort has been devoted to the characterization of exchangeable processes and partially exchangeable

processes (De Finetti 1974, Aldous 1981, Kingman 1978). Despite the attractions of the theory, the conventional definition of exchangeability is too restrictive to be of much use in applied work, where differences among units are frequently determined by a function x called a covariate.

Up to this point, we have talked of a process in terms of distributions, not in terms of a random variable or sequence of random variables. However, the Kolmogorov extension theorem guarantees the existence of a random variable Y taking values in $\mathbb{R}^{\mathcal{U}}$ such that the finite-dimensional distributions are those determined by P . As a matter of logic, however, the distributions come first and the existence of the random variable must be demonstrated, not the other way round. For the most part, the existence of the random variable poses no difficulty, and all distributional statements may be expressed in terms of random variables. The process Y is a function on the units, usually real-valued but possibly vector-valued, so there is one value for each unit.

To avoid misunderstandings at this point, the statistical units are the objects on which the process is defined. It is left to the reader to interpret this in a suitable operational sense depending on the application at hand. By contrast, the standard definition in the experimental design literature holds that a unit is ‘the smallest division of the experimental material such that two units may receive different treatments’ (Cox 1958*a*). The latter definition implies random or deliberate assignment of treatment levels to units. At a practical level, the operational definition is much more useful than the mathematical definition. While the two definitions coincide in most instances, Example 3 shows that they may differ.

4.2.2 Regression processes

A covariate is a function on the units. It may be helpful for clarity to distinguish certain types of covariate. A quantitative covariate is a function $x: \mathcal{U} \rightarrow \mathbb{R}$ or $x: \mathcal{U} \rightarrow \mathbb{R}^p$ taking values in a finite-dimensional vector space. This statement does not exclude instances in which x is a bounded function or a binary function. A qualitative covariate or factor is a function $x: \mathcal{U} \rightarrow \Omega$ taking values in a set Ω called the set of levels or labels. These labels may have no additional structure, in which case the term nominal scale is used, or they may be linearly ordered or partially ordered or they may constitute a tree or a product set. The exploitation of such structure is a key component in successful model construction, but that is not the thrust of the present work. For the moment at least, a covariate is a function $x: \mathcal{U} \rightarrow \Omega$ taking values in an arbitrary set Ω . Ordinarily, of course, the values of x are available only on the finite sampled subset $S \subset \mathcal{U}$, but we must not forget that the aim of inference is ultimately to make statements about the likely values of the response on unsampled units whose x -value is specified. If statistical models have any value, we must be in a position to make predictions about the response distribution on such units, possibly even on units whose covariate value does not occur in the sample.

At this point the reader might want to draw a distinction between estimation and prediction, but this distinction is more apparent than real. If a variety

contrast is estimated as 50 ± 15 kg/ha, the prediction is that the mean yield for other units under similar conditions will be 35–65 kg/ha higher for one variety than the other, a prediction about the difference of infinite averages. Without the concept of a process to link one statistical unit with another, it is hard to see how inferences or predictions of this sort are possible. Nonetheless, Besag (2002, p. 1271) makes it clear that this point of view is not universally accepted. My impression is that the prediction step is so obvious and natural that its mathematical foundation is taken for granted.

Let $x: \mathcal{U} \rightarrow \Omega$ be a given function on the units. Recall that a real-valued process is a function P that associates with each finite subset $S \subset \mathcal{U}$ a distribution P_S on \mathbb{R}^S , and that these distributions are mutually compatible with respect to subsampling of units. A process having the following property for every integer n is called regression exchangeable or exchangeable modulo x .

(RE) *Two finite samples $S = \{i_1, \dots, i_n\}$ and $S' = \{j_1, \dots, j_n\}$ of equal size, ordered such that $x(i_r) = x(j_r)$ for each r , determine the same distribution $P_S = P_{S'}$ on \mathbb{R}^n .*

Exchangeability modulo x is the condition that if x takes the same value on two samples, the distributions are also the same. Any distinction between units, such as name or identification number that is not included as a component of x , has no effect on the response distribution. The majority of models that occur in practical work have this property, but Example 3 below shows that there are exceptions. The property is a consequence of the definition of a statistical model as a functor on a certain category, the injective maps, in the sense of McCullagh (2002) or Brøns (2002), provided that the parameter space is a fixed set independent of the design.

Exchangeability in the conventional sense implies that two samples of equal size have the same response distribution regardless of the covariate values, so exchangeability implies regression exchangeability. For any function g defined on Ω , exchangeability modulo $g(x)$ implies exchangeability modulo x , and $g(x) \equiv 0$ reduces to the standard definition of exchangeability. Exchangeability modulo x is not to be confused with partial exchangeability as defined by Aldous (1981) for random rectangular matrices.

The first consequence of exchangeability, that differences between distributions are determined by differences between covariate values, is related, at least loosely, to the assumption of ‘no unmeasured confounders’ (Greenland *et al.* 1999). This is a key assumption in the literature on causality. At this stage, no structure has been imposed on the set \mathcal{U} , and no structure has been ruled out. In most discussions of causality the units have a temporal structure, so the observation on each unit is a time sequence, possibly very brief, and the notion of ‘the same unit at a later time’ is well defined (Lindley 2002, Singpurwalla 2002, Pearl 2002). The view taken here is that causality is not a property of a statistical model but of its interpretation, which is very much context dependent. For example, Brownian motion as a statistical model has a causal interpretation

in terms of thermal molecular collisions, and a modified causal interpretation may be relevant to stock-market applications. In agricultural field work where the plots are in linear order, Brownian motion may be used as a model for one component of the random variation, with no suggestion of a causal interpretation. This is not fundamentally different from the use of time-series models and methods for non-temporal applications (Cox 1949). Thus, where the word causal is used, we talk of a causal interpretation rather than a causal model. We say that any difference between distributions is associated with a difference between covariate values in the ordinary mathematical sense without implying a causal interpretation.

The second consequence of regression exchangeability, that the distribution of Y_i depends only on the value of x on unit i , and not on the values on other units, is a key assumption in experimental design and in clinical trials called lack of interference (Cox 1958a, p.19). It is known, however, that biological interference can and does occur. Such effects are usually short range, such as root interference or fertilizer diffusion, so typical field trials incorporate discard strips to minimize the effect. This sort of interference can be accommodated within the present framework by including in x the necessary information about nearby plots. Interference connected with carry-over effects in a crossover trial can be accommodated by defining the statistical units as subjects rather than subjects at specific time points; see Example 3.

4.2.3 Interaction

Suppose the model is such that responses on different units are independent, and that $x = (v_0, v)$ with v_0 a binary variable indicating treatment level, and v a baseline covariate or other classification factor. The response distributions at the two treatment levels are $P_{0,v}$ and $P_{1,v}$. It is conventional in such circumstances to define ‘the treatment effect’ by a function or functional of the two distributions,

$$\text{Treatment effect} = T(P_{1,v}) - T(P_{0,v}) = \tau(v),$$

such that $T(P_{1,v}) = T(P_{0,v})$ implies $P_{1,v} = P_{0,v}$ for all model distributions. For example, T might be the difference between the two means (Cox 1958a), the difference between the two log-transformed means (Box and Cox 1964), the difference between the two medians, the log ratio of the two means, the log odds ratio (Cox 1958c), the log hazard ratio (Cox 1972), or the log variance ratio. In principle, T is chosen for ease of expression in summarizing conclusions and in making predictions concerning differences to be expected in future. Ideally, T is chosen so that, under the model in its simplest form, the treatment effect is constant over the values of v , in which case we say that there is no interaction between treatment and other covariates. In practice, preference is given to scalar functions because these lead to simpler summaries, but the definition does not require this.

If $P_{1,v} = P_{0,v}$ for each v , the model distributions do not depend on the treatment level, and the treatment effect $\tau(v)$ is identically zero. Conversely, a

zero treatment effect in the model implies equality of distributions. If $\tau(v)$ is identically zero, we say that there is no treatment effect. The process is then exchangeable modulo the baseline covariates v , i.e. ignoring treatment. This is a stronger condition than regression exchangeability with treatment included as a component of x .

If the treatment effect is constant and independent of v , we say that there is no interaction, and the single numerical value suffices to summarize the difference between distributions. Although the process is not now exchangeable in the sense of the preceding paragraph, the adjustment for treatment is usually of a very simple form, so much of the simplicity of exchangeability remains. If the treatment effect is not constant, we say that there is interaction. By definition, non-zero interaction implies a non-constant treatment effect, so a zero treatment effect in the presence of non-zero interaction is a logical contradiction.

It is possible to define an average treatment effect $\text{ave}\{\tau(v)\}$, averaged with respect to a given distribution on v , and some authors refer to such an average as the ‘main effect of treatment’. Such averages may be useful in limited circumstances as a summary of the treatment effect in a specific heterogeneous population. However, if the interaction is appreciable, and in particular if the sign of the effect varies across subgroups, we would usually want to know the value in each of the subgroups. A zero value of the average treatment effect does not imply exchangeability in the sense discussed above, so a zero average rarely corresponds to a hypothesis of mathematical interest. Nelder (1977) and Cox (1984a) argue that statistical models having a zero average main effect in the presence of interaction are seldom of scientific interest. McCullagh (2000) reaches a similar conclusion using an argument based on algebraic representation theory in which selection of factor levels is a permissible operation.

4.3 Examples of exchangeable regression models

The majority of exchangeable regression models that occur in practice have independent components, in which case it is sufficient to specify the marginal distributions for each unit. The first four examples are of that type, but the fifth example shows that the component variables in an exchangeable regression process need not be independent or conditionally independent.

Example 1: Classical regression models In the classical multiple regression model, the covariate x is a function on the units taking values in a finite-dimensional vector space \mathcal{V} , which we call the covariate space. Each point $\theta = (\beta, \sigma)$ in the parameter space consists of a linear functional $\beta \in \mathcal{V}'$, where \mathcal{V}' is the space dual to \mathcal{V} , plus a real number σ , and the parameter space consists of all such pairs. If the value of the linear functional β at $v \in \mathcal{V}$ is denoted by $v^T\beta$, the value on unit i is $x_i^T\beta$. In the classical linear regression model the response distribution for unit i is normal with mean equal to $x_i^T\beta$ and variance σ^2 . The model may be modified in a number of minor ways, for example by restricting σ to be non-negative to ensure identifiability.

From the point of view of regression exchangeability, generalized linear models or heavy-tailed versions of the above model are not different in any fundamental way. For example, the linear logistic model in which $\eta_i = x_i^T \beta$ and Y_i is Bernoulli with parameter $1/\{1 + \exp(-\eta_i)\}$ is an exchangeable regression model in which the parameter space consists of linear functionals on \mathcal{V} .

Example 2: Treatment and classification factors A treatment or classification factor is a function x on the units taking values in a set, usually a finite set, called the set of levels. It is conventional in applied work to draw a strong distinction between a treatment factor and a classification factor (Cox 1984a). The practical distinction is an important one, namely that the level of a treatment factor may, in principle at least, be determined by the experimenter, whereas the level of a classification factor is an immutable property of the unit. Age, sex and ethnic origin are examples of classification factors: medication and dose are examples of treatment factors. I am not aware of any mathematical construction corresponding to this distinction, so the single definition covers both. A block factor as defined in Section 4.5 is an entirely different sort of mathematical object from which the concept of a set of levels is missing.

Let Ω be the set of treatment levels, and let $\tau: \Omega \rightarrow \mathbb{R}$ be a function on the levels. In conventional statistical parlance, τ is called the vector or list of treatment effects, and differences such as

$$\tau(M) - \tau(F) \quad \text{or} \quad \tau(\text{Kerr's pinks}) - \tau(\text{King Edward})$$

are called contrasts. We note in passing that the parameter space \mathbb{R}^Ω has a preferred basis determined by the factor levels, and a preferred basis is essential for the construction of an exchangeable prior process for the effects should this be required. Without a preferred basis, no similar construction exists for a general linear functional β in a regression model.

In the standard linear model with independent components, the distribution of the response on unit i is Gaussian with mean $\tau\{x(i)\}$ and variance σ^2 . The parameter space is the set of all pairs (τ, σ) in which τ is a function on the levels and σ is a real number. For each parameter point, condition (RE) is satisfied by the process. Once again, the extension to generalized linear models presents no conceptual difficulty.

Example 3: Crossover design In a two-period crossover design, one observation is made on each subject under different experimental conditions at two times sufficiently separated that carry-over effects can safely be neglected. If we regard the subjects as the statistical units, which we are at liberty to do, the design determines the observation space \mathbb{R}^2 for each unit. The observation space corresponding to a set of n units is $(\mathbb{R}^2)^n$. Let x be the treatment regime, so that (x_{i1}, x_{i2}) is the ordered pair of treatment levels given to subject i . In the

conventional statistical model the response distribution for each unit is bivariate Gaussian with covariance matrix $\sigma^2 I_2$. The mean vector is

$$\mathbb{E} \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = \begin{pmatrix} \alpha_i + \tau_{x_{i1}} \\ \alpha_i + \tau_{x_{i2}} + \delta \end{pmatrix},$$

in which α is a function on the subjects, and δ is a common temporal trend. The parameter space consists of all functions α on the n subjects, all functions τ on the treatment levels, plus the two scalars (δ, σ) , so the effective dimension is $n + 3$ for a design with n subjects and two treatment levels.

This model is not regression exchangeable because two units i, j having the same treatment regime $(x_{i1}, x_{i2}) = (x_{j1}, x_{j2})$ do not have the same response distribution for all parameter values: the difference between the two means is $\alpha_i - \alpha_j$. This model is a little unusual in that the parameter space is not a fixed set independent of the design: the dimension depends on the sample size. Nonetheless, it is a statistical model in the sense of McCullagh (2002).

An alternative Gaussian model, with units and observation spaces defined in the same manner, has the form

$$\mathbb{E} \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = \begin{pmatrix} \tau_{x_{i1}} \\ \tau_{x_{i2}} + \delta \end{pmatrix}, \quad \text{cov} \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix},$$

with a fixed parameter space independent of the design. Two samples having the same covariate values also have the same distribution, so condition (RE) is satisfied. The temporal effect δ is indistinguishable from a carry-over effect that is independent of the initial treatment. If there is reason to suspect a non-constant carry-over effect, the model may be extended by writing

$$\mathbb{E} \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = \begin{pmatrix} \tau_{x_{i1}} \\ \tau_{x_{i2}} + \gamma_{x_{i1}} + \delta \end{pmatrix}, \quad \text{cov} \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}.$$

If there are two treatment levels and all four combinations occur in the design, the difference $\gamma_1 - \gamma_0$ is estimable.

Example 4: Biased sampling We consider a biased-sampling model in which observations on distinct units are independent and, for notational convenience, real-valued. The covariate w associates with the i th unit a bias function w_i such that $w_i(x) \geq 0$ for each real number x . The parameter space is either the set of probability distributions or the set of non-negative measures on the Borel sets in \mathbb{R} such that each integral $\int w_i(x) dF(x)$ is finite (Vardi 1985, Kong *et al.* 2003). For each F in the parameter space, the response distribution on unit i is the weighted distribution such that $dF_i(x) \propto w_i(x) dF(x)$. Thus, to each point in the parameter space the model associates a process with independent but non-identically distributed components. Two units having the same bias function also have the same distribution, so the process is regression exchangeable.

The simplest example is one in which $w_i(x) = 1$ identically for each unit, in which case the maximum likelihood estimator \hat{F} is the empirical distribution at the observations. Size-biased sampling corresponds to $w(x) = |x|$ or some power of $|x|$, a phenomenon that arises in a wide range of applications from the wool industry (Cox 1962, §5.4) to stereology and auditing (Cox and Snell 1979). In general, the maximum likelihood estimator \hat{F} is a distribution supported at the observation points, but with unequal atoms at these points.

Example 5: Prediction and smoothing models Consider the modification of the simple linear regression model in which unit i has covariate value x_i , and, in a conventional but easily misunderstood notation,

$$Y_i = \beta_0 + \beta_1 x_i + \eta(x_i) + \epsilon_i. \quad (4.1)$$

The coefficients (β_0, β_1) are parameters to be estimated, ϵ is a process with independent $N(0, \sigma^2)$ components, and η is a zero-mean stationary process on the real line, independent of ϵ , with covariance function

$$\text{cov}\{\eta(x), \eta(x')\} = \sigma_\eta^2 K(x, x').$$

If η is a Gaussian process, the response distribution for any finite collection of n units may be expressed in the equivalent distributional form

$$Y \sim N(X\beta, \sigma^2 I_n + \sigma_\eta^2 V), \quad (4.2)$$

where $V_{ij} = K(x_i, x_j)$ are the components of a positive semidefinite matrix. For each value of $(\beta, \sigma^2, \sigma_\eta^2)$, two samples having the same covariate values determine the same distribution on the observation space, so this model is regression-exchangeable with non-independent components.

The linear combination $\eta(x_i) + \epsilon_i$ in eqn (4.1) is a convenient way of describing the distribution of the process as a sum of more elementary processes. The treacherous aspect of the notation lies in the possibility that η or ϵ might be mistaken for a parameter to be estimated from the data, which is not the intention. The alternative parametric model with independent components and parameter space consisting of all smooth functions η , is also regression exchangeable, but very different from eqn (4.2).

The simplest way to proceed for estimation and prediction is first to estimate the parameters $(\sigma^2, \sigma_\eta^2, \beta_0, \beta_1)$ by maximum likelihood estimation, or by some closely related procedure such as residual maximum likelihood estimation—REML—for the variance components followed by weighted least squares estimation of the regression parameters. With prediction in mind, Wahba (1985) recommends use of generalized crossvalidation over residual maximum likelihood estimation on the grounds that it is more robust against departures from the stochastic model. Efron (2001) considers a range of estimators and seems to prefer the residual maximum likelihood estimator despite evidence of bias. Suppose that this has been done, and that we aim to predict the response value $Y(i^*)$ for a new unit i^* in the same process whose covariate value is $x^* = x(i^*)$.

Proceeding as if the parameter values were given, the conditional expected value of $Y(i^*)$ given the values on the sampled units is computed by the formula

$$\hat{Y}_{i^*} = E\{Y(i^*) | Y\} = \beta_0 + \beta_1 x^* + k^* \Sigma^{-1}(Y - \mu), \quad (4.3)$$

where μ, Σ are the estimated mean and covariance matrix for the sampled units, and $k_i^* = \sigma_\eta^2 K(x^*, x_i)$ is the vector of covariances. The conditional distribution is Gaussian with mean (4.3) and constant variance independent of Y . Interpolation features prominently in the geostatistical literature, where linear prediction is called kriging (Stein 1999).

If η is Brownian motion with generalized covariance function $-|x - x'|$ on contrasts, the prediction function (4.3) is continuous and piecewise linear: if $K(x, x') = |x - x'|^3$, the prediction function is a cubic spline (Wahba 1990, Green and Silverman 1994). Of course, K is not necessarily a simple covariance function of this type: it could be in the Matérn class (Matérn 1986) or it could be a convex combination of simple covariance functions. The cubic and linear splines illustrated in Figure 4.1 are obtained by fitting model (4.2) to simulated data (Wahba 1990, p. 45), in which $\eta(x)$ is the smooth function shown as the dashed line.

In statistical work, the adjective ‘Bayesian’ usually refers to the operation of converting a probability $\text{pr}(A | B)$ into a probability of the form $\text{pr}(B | A)$ by supplying additional information and using Bayes’s theorem. The transformation from the joint distribution of (Y^*, Y) as determined by the process (4.2), to the conditional distribution $Y^* | Y$, does not involve prior information or Bayes’s theorem. Nonetheless, it is possible to cast the argument leading to eqn (4.3) in a Bayesian mould, so the majority of authors use the term Bayesian or empirical Bayes in this context (Wahba 1990, Efron 2001). A formal Bayesian analysis begins with a prior distribution π on the parameters $(\beta, \sigma^2, \sigma_\eta^2)$, and uses the likelihood function to obtain the posterior distribution. The process is such that

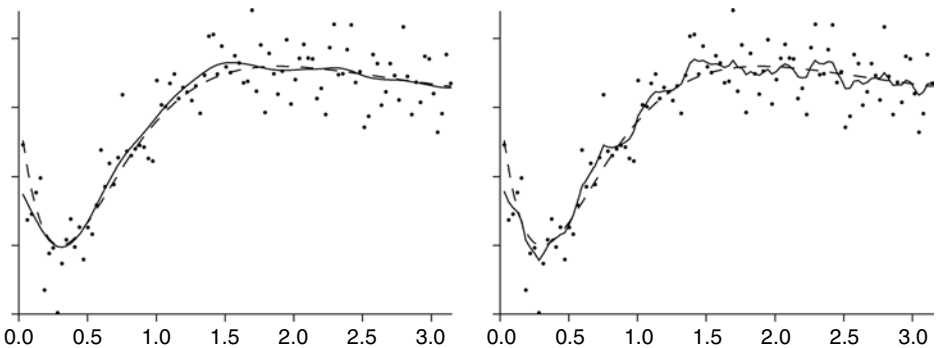


FIG. 4.1. Spline prediction graphs (solid lines) fitted by residual maximum likelihood, cubic on the left, linear on the right. The ideal predictor $\eta(x)$ (dashed line) is taken from Wahba (1990, p. 45).

the predictive distribution for a new unit has mean (4.3) and constant variance depending on the parameters. The Bayesian predictive distribution is simply the posterior average, or mixture, of these distributions. In this context, the adjective ‘Bayesian’ refers to the conversion from prior and likelihood to posterior, not to eqn (4.3).

Since η is not a parameter to be estimated, the introduction of extra-likelihood criteria such as penalty functions or kernel density estimators to force smoothness is, in principle at least, unnecessary. In practice, if the family of covariance functions for η includes a smoothness parameter such as the index in the Matérn class, the likelihood function seldom discriminates strongly. Two covariance functions such as $-|x - x'|$ and $|x - x'|^3$ achieving approximately the same likelihood value, usually produce prediction graphs that are pointwise similar. For the data in Figure 4.1, the maximum log residual likelihood values are 42.2 for the model in which $K(x, x') = -|x - x'|$, 42.1 for the ‘quadratic’ version $|x - x'|^2 \log |x - x'|$, and 41.3 for the cubic. Visually the prediction graphs are very different, so aesthetic considerations may determine the choice for graphical presentation.

To illustrate one crucial difference between an estimator and a predictor, it is sufficient to note that the prediction function (4.3) is not a projection on the observation space in the sense that the least-squares fit is a projection on the observation space. However, it is a projection on a different space, the combined observation-prediction space. Let S_0 be the n sampled units, let S_1 be the m unsampled units for which prediction is required, and let $S = S_0 \cup S_1$ be the combined set. The covariance matrix Σ for the joint distribution in \mathbb{R}^S may be written in partitioned form with components $\Sigma_{00}, \Sigma_{01}, \Sigma_{11}$, and the model matrix may be similarly partitioned, X_0 for the sampled units and X_1 for the unsampled units. The prediction function is linear in the observed value Y_0 and may be written as the sum of two linear transformations,

$$\begin{pmatrix} \hat{Y}_0 \\ \hat{Y}_1 \end{pmatrix} = \begin{pmatrix} P_0 & 0 \\ X_1(X_0^\top \Sigma_{00}^{-1} X_0)^{-1} X_0^\top \Sigma_{00}^{-1} & 0 \end{pmatrix} \begin{pmatrix} Y_0 \\ \star \end{pmatrix} + \begin{pmatrix} Q_0 & 0 \\ \Sigma_{10} \Sigma_{00}^{-1} Q_0 & 0 \end{pmatrix} \begin{pmatrix} Y_0 \\ \star \end{pmatrix}, \quad (4.4)$$

where $P_0 = X_0(X_0^\top \Sigma_{00}^{-1} X_0)^{-1} X_0^\top \Sigma_{00}^{-1}$, $Q_0 = I - P_0$, and \star is the unobserved value. Evidently $\hat{Y}_0 = Y_0$ as it ought. The first transformation is the least-squares projection $P: \mathbb{R}^S \rightarrow \mathbb{R}^S$ onto $\mathcal{X} \subset \mathbb{R}^S$ of dimension p . The second transformation is a projection $T: \mathbb{R}^S \rightarrow \mathbb{R}^S$, self-adjoint with respect to the inner product Σ^{-1} , and thus an orthogonal projection. Its kernel consists of all vectors of the form (x, \star) , i.e. all vectors in $\mathcal{X} + \mathbb{R}^{S_1}$ of dimension $m + p$, and the image is the orthogonal complement. Direct calculation shows that $PT = TP = 0$ so the sum $P + T$ is also a projection.

Most computational systems take $X_1 = X_0$, so the predictions are for new units having the same covariate values as the sampled units. The first component in eqn (4.4) is ignored, and the prediction graphs in Figure 4.1 show the conditional mean \hat{Y}_1 as a function of x , with Σ estimated in the conventional way by marginal maximum likelihood based on the residuals.

The preceding argument assumes that K is a proper covariance function, so Σ is positive definite, which is not the case for the models illustrated in Figure 4.1. However, the results apply to generalized covariance functions under suitable conditions permitting pointwise evaluation provided that the subspace \mathcal{X} is such that K is positive semidefinite on the contrasts in \mathcal{X}^0 (Wahba 1990).

Example 6: Functional response model Consider a growth-curve model in which the stature or weight of each of n subjects is measured at a number of time points over the relevant period. To keep the model simple, the covariate for subject i is the schedule of measurement times alone. If we denote by $Y_i(t)$ the measured height of subject i at time t , the simplest sort of additive growth model may be written in the form

$$Y_i(t) = \alpha_i + m(t) + \eta_i(t) + \epsilon_i(t),$$

in which α is an exchangeable process on the subjects, m is a smooth random function of time with mean μ , η is a zero-mean process continuous in time and independent for distinct subjects, and ϵ is white noise. All four processes are assumed to be independent and Gaussian. The distributions are such that Y is Gaussian with mean $E\{Y_i(t)\} = \mu(t)$ and covariance matrix/function

$$\text{cov}\{Y_i(t), Y_j(t')\} = \sigma_\alpha^2 \delta_{ij} + \sigma_m^2 K_m(t, t') + \sigma_\eta^2 K_\eta(t, t') \delta_{ij} + \sigma_\epsilon^2 \delta_{ij} \delta_{t-t'}.$$

If the functions μ , K_m and K_η are given or determined up to a small set of parameters to be estimated, all parameters can be estimated by maximum likelihood or by marginal maximum likelihood. The fitted growth curve, or the predicted growth curve for a new subject from the same process, can then be obtained by computing the conditional expectation of $Y_{i^*}(t)$ given the data. Note that if σ_m^2 is positive, $Y_{i^*}(t)$ is not independent of the values on other subjects, so the predicted value is not the fitted mean $\mu(t)$.

The model shown above satisfies condition (RE), so it is regression exchangeable even though the components are not independent. It is intended to illustrate the general technique of additive decomposition into simpler processes followed by prediction for unobserved subjects. It is ultimately an empirical matter to decide whether such decompositions are useful in practice, but real data are likely to exhibit departures of various sorts. For example, the major difference between subjects may be a temporal translation, as in the alignment of time origins connected with the onset of puberty. Further, growth measurements are invariably positive and seldom decreasing over the interesting range. In addition, individual and temporal effects may be multiplicative, so the decomposition may be more suitable for log transformed process. Finally, there may be covariates, and if a covariate is time dependent, i.e. a function of t , the response distribution at time t could depend on the covariate history.

4.4 Causality and counterfactuals

4.4.1 Notation

It is conventional in both the applied and the theoretical literature to write the linear regression model in the form

$$E(Y_i | x) = x_i^T \beta, \quad \text{var}(Y_i | x) = \sigma^2,$$

when it is understood that the components are independent. The notation suggests that $x^T \beta$ is the conditional mean of the random variable Y and σ^2 is the conditional variance, as if x were a random variable defined on the same probability space as Y . Despite the notation and the associated description, that is not what is meant because x is an ordinary function on the units, not a random variable.

The correct statement runs as follows. First, x is a function on the units taking values in \mathbb{R}^p , and the values taken by x on a finite set of n units may be listed as a matrix X of order $n \times p$ with rows indexed by sampled units. Second, to each parameter point (β, σ) , the model associates a distribution on \mathbb{R}^n by the formula $N(X\beta, \sigma^2 I_n)$, or by a similar formula for generalized linear models. In this way, the model determines a set of real-valued processes, one process for each parameter point. Each process is indexed by the finite restrictions of x , with values in the observation space \mathbb{R}^n . No conditional distributions are involved at any point in this construction. At the same time, the possibility that the regression model has been derived from a bivariate process by conditioning on one component is not excluded.

Even though no conditional distributions are implied, the conventional notation and the accompanying description are seldom seriously misleading, so it would be pedantic to demand that they be corrected. However, there are exceptions or potential exceptions.

The linear regression model associates with each parameter point $\theta = (\beta, \sigma)$ a univariate process: it does not associate a bivariate process with a pair of parameter values. As a consequence, it is perfectly sensible to compare the probability $P_{x, \theta}(E)$ with the probability $P_{x, \theta'}(E)$ for any event $E \subset \mathbb{R}^n$, as in a likelihood ratio. But it makes no sense to compare the random variable Y in the process determined by θ with the random variable in the process determined by θ' . A question such as ‘How much larger would Y_i have been had β_1 been 4.3 rather than 3.4?’ is meaningless within the present construction because the two processes need not be defined on the same probability space. The alternative representation of a linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \sigma \epsilon_i$$

is potentially misleading on this point because it suggests the answer $(4.3 - 3.4)x_i$.

4.4.2 Exchangeability and counterfactuals

A counterfactual question is best illustrated by examples such as ‘If I had taken aspirin would my headache be gone?’ or ‘How much longer would Cynthia Crabb

have survived had she been given a high dose of chemotherapy rather than a low dose?’ The presumption here is that \mathcal{U} consists of subjects or patients, that x is a function representing treatment and other baseline variables, and that Y is the outcome. Formally, $i = \textit{Cynthia Crabb}$ is the patient name, $x(i) = \textit{low}$ is the treatment component of the covariate, $P_{\{i\}}$ is the survival distribution, and $Y: (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}^{\mathcal{U}}$ is a random variable whose i th component $Y_i(\omega)$ is the outcome for Cynthia Crabb. Since the constructed process is a real-valued function on the units, there is only one survival time for each subject. Within this framework, it is impossible to address a question that requires Cynthia Crabb to have two survival times. For a more philosophical discussion, the reader is referred to Dawid (2000) who reaches a similar conclusion, or Pearl (2000) who reaches a different conclusion.

If the preceding question is not interpreted as a literal counterfactual, it is possible to make progress by using regression exchangeability and interpreting the question in a distributional sense as follows. Among the infinite set of subjects on which the process is defined, a subset exists having the same covariate values as Cynthia Crabb except that they have the high dose of chemotherapy. Conceptually, there is no difficulty in supposing that there is an infinite number of such subjects, identical in all covariate respects to Cynthia Crabb except for the dose of chemotherapy. By regression exchangeability, all such subjects have the same survival distribution. Provided that we are willing to interpret the question as a comparison of the actual survival time of Cynthia Crabb with the distribution of survival times for patients in this subset, the mathematical difficulty of finding Cynthia Crabb in two disjoint sets is avoided. The question may now be answered, and the answer is a distribution that may, in principle, be estimated given sufficient data.

It is clear from the discussion of Dawid (2000) that most statisticians are unwilling to forego counterfactual statements. The reason for this seems to be a deep-seated human need to assign credit or blame, to associate causes with observed effects—this sentence being an instance of the phenomenon it describes. My impression is that most practical workers interpret counterfactual matters such as unit-treatment additivity in this distributional sense, sometimes explicitly so (Cox 1958a, 2000). However, Pearl (2000) argues to the contrary, that counterfactual statements are testable and thus not metaphysical.

The directness and immediacy of counterfactual statements are appealing by way of parameter interpretation and model explanation. Another way of trying to make sense of the notion is to invoke a latent variable, a bivariate process with two survival times for each subject, so that all random variables exist in the mathematical sense. The first component is interpreted as the survival time at low dose, the second component is interpreted as the survival time at high dose, and the difference $Y_{i2} - Y_{i1}$ or ratio Y_{i2}/Y_{i1} is the desired counterfactual difference in survival times. The treatment value serves to reduce the bivariate process to a univariate process by indicating which of the two components is observed. The net result is a univariate process whose distributions determine the

exchangeable regression model described above. From the present point of view the same process is obtained by an indirect route, so nothing has changed. Unless the observation space for some subset of the units is bivariate, the counterfactual variable is unobservable, so counterfactual prediction cannot arise.

Model constructions involving latent or unobserved random processes are used frequently and successfully in statistical models. The net result in such cases is a univariate marginal process defined on the finite restrictions of the covariate. The introduction of the latent component is technically unnecessary, but it is sometimes helpful as a pedagogical device to establish a mechanistic interpretation (Cox 1992, §5.1). Provided that inferences are restricted to estimation and prediction for this marginal process, the technique is uncontroversial. Counterfactual predictions for the latent bivariate survival process are beyond the observation space on which the marginal process is defined, and thus cannot be derived from the marginal model alone. Nonetheless, with heroic assumptions, such as independence of the two survival components that are unverifiable in the marginal process, counterfactual predictions for the bivariate process may be technically possible. However, the absence of a physical counterpart to the mathematical counterfactual makes it hard to understand what such a statement might mean in practice or how it might be checked.

4.4.3 Exchangeability and causality

Why did Cynthia Crabb receive the low dose when other patients with the same non-treatment covariates received the high dose? A statistical model does not address questions of this sort. However, regression exchangeability is a model assumption implying that, in the absence of a treatment effect, the responses for all patients in Cynthia Crabb's baseline covariate class are exchangeable. As a consequence, all patients whose non-treatment covariate values are the same as those of Cynthia Crabb have the same response distribution. Any departure from exchangeability that is associated with treatment assignment may then be interpreted as evidence against the null model of no treatment effect. In applications where it is a feasible option, objective randomization is perhaps the most effective way to ensure that this model assumption is satisfied.

Like all model assumptions, regression exchangeability may prove unsatisfactory in specific applications. If the treatment assignment is done on doctor's advice on the basis of information not available in x , the exchangeability condition may well be violated. Likewise, if the protocol allows patients to select their own dose level, the choice may be based on factors not included in x , in which case there is little assurance that the patients are exchangeable modulo x .

A central theme in much of causal inference is the attempt to deduce or to predict in a probabilistic sense what would have occurred had the design protocol been different from what it actually was. Since the theory of prediction for processes does not extend beyond the index set on which the process is defined, this sort of prediction requires an explicit broader and perhaps fundamentally different foundation. One can envisage a compound doubly randomized design in which the first arm is a conventional randomized experiment, the response on

each individual being five-year survival. In the elective arm, patients are permitted to select the drug or dose, so the response is bivariate. This sort of process, in which the observation space itself depends on the covariate, is certainly unconventional, but it is not fundamentally different from the definition given in Section 4.2. The definition of regression exchangeability is unchanged, but would usually be considered in a modified form in which the conditional distribution of survival times given the chosen treatment is the same as the distribution of survival times for the assigned treatment in the randomized arm. In other words, the survival distribution depends on treatment and other baseline covariates, but not on whether the treatment is randomly assigned or freely selected. With this modified concept of exchangeability in the extended process, it is possible to extrapolate from the randomized experiment, by making predictions for the outcomes under a different protocol.

It is invariably the case in matters of causal assessment that closer inspection reveals additional factors or an intermediate sequence of events that could affect the interpretation of treatment contrasts had they been included in the model, i.e. if a different model had been used. A good example can be found in the paper by Versluis *et al.* (2000) on the sound-causing mechanism used by the snapping shrimp *Alpheus heterochaelis*. Since the shrimp tend to congregate in large numbers, the combined sound is appreciable and can interfere with naval sonar. The loud click is caused by the extremely rapid closure of the large snapper claw, in the sense that a sound is heard every time the claw snaps shut and no sound is heard otherwise. It had been assumed that the sound was caused by mechanical contact between hard claw surfaces, and the preceding statement had been universally interpreted in that way. However, closer inspection reveals a previously unsuspected mechanism, in which the claw is not the source of the sound. During the rapid claw closure a high-velocity water jet is emitted with a speed that exceeds cavitation conditions, and the sound coincides with the collapse of the cavitation bubble not with the closure of the claw.

In light of this information, what are we to say of causal effects? The initial statement that the rapid closure of the claw causes the sound, is obviously correct in most reasonable senses. It satisfies the conditions of reproducibility, consistency with subject-matter knowledge, and predictability by well-established theory, as demanded by various authors such as Bradford Hill (1937), Granger (1988), and Cox (1992). However, its very consistency with subject-matter knowledge invites an interpretation that is now known to be false. Whether or not the statement is legally correct, it is scientifically misleading, and is best avoided in applications where it might lead to false conclusions. For example, the observation that a shrimp can stun its prey without contact, simply by clicking its claw, might lead to the false conclusion that snails are sensitive to sonar. The complementary statement that the closure of the claw does not cause the sound, although equally defensible, is certainly not better.

Rarely, if ever, does there exist a most proximate cause for any observed phenomenon, so the emphasis must ultimately be on processes and mechanisms (Cox 1992). Confusion results when the words ‘cause’ or ‘causal’ are used with one

mechanism, or no specific mechanism, in mind, and interpreted in the context of a different mechanism. For clinical trials where biochemical pathways are complicated and unlikely to be understood in sufficient detail, the word mechanism is best replaced by protocol. The natural resolution is to avoid the term ‘causal’ except in the context of a specific mechanism or protocol, which might, but need not, involve manipulation or intervention. Thus, the closure of the claw causes the sound through an indirect mechanism involving cavitation. This statement does not exclude the possibility that cavitation itself is a complex process with several stages.

Unless the protocol is well defined, an unqualified statement concerning the causal effect of a drug or other therapy is best avoided. Thus, following a randomized trial in which a drug is found to increase the five-year survival rate, the recommendation that it be approved for general use is based on a model assumption, the prediction that a similar difference will be observed on average between those who elect to use the drug and those who elect not to use it. Equality here is a model assumption, a consequence of regression exchangeability in the modified sense discussed above. As with all model assumptions, this one may prove to be incorrect in specific applications. Unlike counterfactuals, the assumption can be checked in several ways, by direct comparison in a compound doubly randomized experiment, by comparisons within specific subgroups or by comparing trial results with subsequent performance. In the absence of exchangeability, there is no mathematical reason to suppose that the five-year survival rate among those who elect to use the drug should be similar to the rate observed in the randomized experiment. It is not difficult to envisage genetic mechanisms such that those who elect not to use the drug have the longer five-year survival, but all such mechanisms imply non-exchangeability or the existence of potentially identifiable subgroups.

4.5 Exchangeable block models

4.5.1 Block factor

The distinction between a block factor and a treatment factor, closely related to the distinction between fixed and random effects, is a source of confusion and anxiety for students and experienced statisticians alike. As a practical matter, the distinction is not a rigid one. The key distinguishing feature is the anonymous or ephemeral nature of the levels of a block factor. Cox (1984a) uses the term non-specific, while Tukey (1974) prefers the more colourful phrase ‘named and faceless values’ to make a similar distinction.

Even if it is more rigid and less nuanced, a similar distinction can be made in the mathematics. A block factor B is defined as an equivalence relation on the units, a symmetric binary function $B: \mathcal{U} \times \mathcal{U} \rightarrow \{0, 1\}$ that is reflexive and transitive. Equivalently, but more concretely, B is a partition of the units into disjoint non-empty subsets called blocks such that $B(i, j) = 1$ if units i, j are in the same block and zero otherwise. The number of blocks may be finite or

infinite. For the observed set of n units, B is a symmetric positive semidefinite binary matrix whose rank is the number of blocks in the sample.

A treatment or classification factor $x: \mathcal{U} \rightarrow \Omega$ is a list of levels, one for each unit. It may be converted into a block factor by the elementary device of ignoring factor labels, a forgetful transformation defined by

$$B(i, j) = \begin{cases} 1, & \text{if } x(i) = x(j), \\ 0, & \text{otherwise.} \end{cases}$$

If X is the incidence matrix for the treatment factor on the sampled units, each column of X is an indicator function for the units having that factor level, and $B = XX^T$ is the associated block factor. It is not possible to convert a block factor into a treatment factor because the label information, the names of the factor levels, is not contained in the block factor.

Since the information in the block factor B is less than the information in x , exchangeability modulo B is a stronger condition than exchangeability modulo x . A process is called block-exchangeable if the following condition is satisfied for each n . Two samples $\{i_1, \dots, i_n\}$ and $\{j_1, \dots, j_n\}$, ordered in such a way that $B(i_r, i_s) = B(j_r, j_s)$ for each r, s , determine the same distribution on \mathbb{R}^n . Block exchangeability implies that the label information has no effect on distributions. All one-dimensional marginal distributions are the same, and there are only two distinct two-dimensional marginal distributions depending on whether $B(i, j)$ is true or false (one or zero). More generally, the n -dimensional distribution is invariant under those coordinate permutations that preserve the block structure, i.e. permutations π such that $B(i, j) = B(\pi_i, \pi_j)$ for all i, j ,

For a sample of size n , the image or range of X is the same subspace $\mathcal{X} \subset \mathbb{R}^n$ as the range of B in \mathbb{R}^n , the set of functions that are constant on each block. In the following linear Gaussian specifications, the distribution is followed by a description of the parameter space:

- (i) $Y \sim N(X\beta, \sigma^2 I_n), \quad \beta \in \mathbb{R}^\Omega, \sigma > 0,$
- (ii) $Y \sim N(B\gamma, \sigma^2 I_n), \quad \gamma \in \mathcal{X}, \sigma > 0,$
- (iii) $Y \sim N(\mu, \sigma^2 I_n + \sigma_b^2 B), \quad \mu \in \mathbf{1}, \sigma > 0, \sigma_b \geq 0,$

in which $\mathbf{1} \subset \mathbb{R}^n$ is the one-dimensional subspace of constant functions. In the sense that they determine precisely the same set of distributions on the observation space, the first two forms are equivalent up to reparameterization. Even so, the models are very different in crucial respects.

By definition in (i), $\beta \in \mathbb{R}^\Omega$ is a function on the treatment levels, so inference for specific levels or specific contrasts is immediate. In (ii), one can transform from γ to $\beta = X^T \gamma$ only if the block labels are available. Block labels are not used in (ii), so the formulation does not imply that this information is available. Nonetheless, in most applications with which I am familiar, the function $x: \mathcal{U} \rightarrow \Omega$ would be available for use, in which case the two formulations are equivalent. Both are regression exchangeable but neither formulation is block exchangeable.

The third form, the standard random-effects model with independent and identically distributed block effects, is different from the others: it is block exchangeable. The expression may be regarded as defining a process on the finite restrictions of x , or a process on the finite restrictions of the block factor B . In that sense (iii) is ambiguous, as is (ii). In practice, it would usually be assumed that the block names are available for use if necessary, as for example in animal-breeding experiments (Robinson 1991). Given that x is available, it is possible to make inferences or predictions about contrasts among specific factor levels using model (iii). The conditional distribution of the response on a new unit with $x(\cdot) = 1$ is Gaussian with mean and variance

$$\frac{\sigma^2\mu + n_1\sigma_b^2\bar{y}_1}{\sigma^2 + n_1\sigma_b^2}, \quad \sigma^2 \left(1 + \frac{\sigma_b^2}{n\sigma_b^2 + \sigma^2} \right), \quad (4.5)$$

where n_1 is the number of units in the sample for which $x(\cdot) = 1$ and \bar{y}_1 is the average response. In practice, the parameter values must first be estimated from the available data. If the function x is unavailable, the blocks are unlabelled so inference for specific factor levels or specific contrasts is impossible. Nonetheless, since each new unit i^* comes with block information in the form of the extended equivalence relation B , it is possible to make predictive statements about new units such that $B(i^*, 4) = 1$, i.e. new units that are in the same block as unit 4 in the sample. The formula for the conditional distribution is much the same as that given above, so the mathematical predictions have a similar form except that the block does not have a name. It is also possible, on the basis of the model, to make predictive statements about new units that are not in the same block as any of the sample units. Whether such predictions are reliable is a matter entirely dependent on specific details of the application.

An alternative version of (iii) may be considered, in which the inverse covariance matrix, or precision matrix, is expressed as a linear combination of the same two matrices, I_n and B . If the blocks are of equal size, the inverse of $\sigma_0^2 I_n + \sigma_1^2 B$ is in fact a linear combination of the same two matrices, in which case the two expressions determine the same set of distributions on \mathbb{R}^n , and thus the same likelihood function after reparameterization. However, the second formulation does not determine a process because the marginal $(n-1)$ -dimensional distribution after deleting one component is not expressible in a similar form, with a precision matrix that is a linear combination of I_{n-1} and the restriction of B . The absence of a process makes prediction difficult, if not impossible.

It is worth remarking at this point that, for a balanced design, the sufficient statistic for model (iii) is the sample mean plus the between- and within-block mean squares. Even for an unbalanced design, an individual block mean such as \bar{y}_1 is not a function of the sufficient statistic. Accordingly, two observation points $y \neq y'$ producing the same value of the sufficient statistic will ordinarily give rise to different predictions in eqn (4.3) or eqn (4.5). In other words, the conclusions are not a function of the sufficient statistic. One of the subtleties of the likelihood principle as stated, for example, by Cox and Hinkley (1974, p. 39)

or Berger and Wolpert (1988, p. 19) is the clause ‘conclusions about θ ’, implying that it is concerned solely with parameter estimation. Since eqns (4.3) and (4.5) are statements concerning events in the observation space, not estimates of model parameters or statements about θ , there can be no violation of the likelihood principle. On the other hand, a statement about θ is a statement about an event in the tail σ -field of the process, so it is not clear that there is a clear-cut distinction between prediction and parametric inference.

4.5.2 Example: homologous factors

We consider in this section a further, slightly more complicated, example of an exchangeable block model in which the covariate $x = (x_1, x_2)$ is a pair of homologous factors taking values in the set $\Omega = \{1, \dots, n\}$ (McCullagh 2000). If there is only a single replicate, the observation Y is a square matrix of order n with rows and columns indexed by the same set of levels. More generally, the design is said to be balanced with r replicates if, for each cell (i, j) there are exactly r units u for which $x(u) = (i, j)$. For notational simplicity, we sometimes assume $r = 1$, but in fact the design need not be balanced and the assumption of balance can lead to ambiguities in notation.

The following models are block exchangeable:

$$Y_{ij} = \mu + \eta_i + \eta_j + \epsilon_{ij},$$

$$Y_{ij} = \mu + \eta_i - \eta_j + \epsilon_{ij},$$

$$Y_{ij} = \eta_i - \eta_j + \epsilon'_{ij}.$$

In these expressions η/σ_η and ϵ/σ_ϵ are independent standard Gaussian processes, so the parameter space for the first two consists of the three components $(\mu, \sigma_\eta^2, \sigma_\epsilon^2)$. In the third model, $\epsilon'_{ij} = -\epsilon'_{ji}$, so the observation matrix Y is skew-symmetric.

These expressions suggest that Y is a process indexed by ordered pairs of integers, and in this respect the notation is misleading. The ‘correct’ version of the first model is

$$Y(u) = \mu + \eta_{x_1(u)} + \eta_{x_2(u)} + \eta'_{x(u)} + \epsilon(u), \quad (4.6)$$

making it clear that Y and ϵ are processes indexed by the units, and there may be several units such that $x(u) = (i, j)$. In plant-breeding experiments, the units such that $x(u) = (i, i)$ are called self-crosses; in round-robin tournaments, self-competition is usually meaningless, so there are no units such that $x(u) = (i, i)$. In the absence of replication, the interaction process η' and the residual process ϵ are not separately identifiable: only the sum of the two variances is estimable. However, absence of replication in the design does not imply absence of interaction in the model. To put it another way, two distinct models may give rise to the same set of distributions for a particular design. Aliasing of interactions in a fractional factorial design is a well-known example of the phenomenon.

If there is a single replicate, the three models may be written in the equivalent distributional form as follows:

$$\begin{aligned} Y &\sim N(\mu \mathbf{1}, \sigma_\eta^2 K + \sigma_\epsilon^2 I_{n^2}), \\ Y &\sim N(\mu \mathbf{1}, \sigma_\eta^2 K' + \sigma_\epsilon^2 I_{n^2}), \\ Y &\sim N(0, \sigma_\eta^2 K' + \sigma_\epsilon^2 I'_{n^2}). \end{aligned}$$

The matrices K, K', I' are symmetric of order $n^2 \times n^2$ and are given by

$$\begin{aligned} K_{ij,kl} &= \delta_{ik} + \delta_{jl} + \delta_{il} + \delta_{jk}, \\ K'_{ij,kl} &= \delta_{ik} + \delta_{jl} - \delta_{il} - \delta_{jk}, \\ I'_{ij,kl} &= \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}. \end{aligned}$$

Note that δ_{ik} is the block factor for rows, δ_{jl} is the block factor for columns, and the remaining terms δ_{il}, δ_{jk} are meaningless unless the two factors have the same set of levels. Each of the three model distributions is invariant under permutation, the same permutation being applied to rows as to columns. Accordingly, the models depend on the rows and columns as block factors, not as classification factors.

In the standard Bradley–Terry model for ranking competitors in a tournament, the component observations are independent, and the competitor effect $\{\eta_i\}$ is a parameter vector to be estimated (Agresti 2002, p. 436). Such models are closed under permutation of factor levels and under restriction of levels, but they are not invariant, and thus not block exchangeable. By contrast, all three models shown above are block exchangeable, and competitor effects do not occur in the parameter space. To predict the outcome of a match between competitors i, j , we first estimate the variance components by maximum likelihood. In the second stage, the conditional distribution of $Y(u^*)$ given Y for a new unit such that $x(u^*) = (i, j)$ is computed by the standard formulae for conditional distributions, and this is the basis on which predictions are made. This exercise is straightforward provided that the variance components required for prediction in eqn (4.6) are identifiable at the design. An allowance for errors of estimation along the lines of Barndorff-Nielsen and Cox (1996) is also possible.

4.6 Concluding remarks

Kolmogorov's definition of a process in terms of compatible finite-dimensional distributions is a consequence of requiring probability distributions to be well behaved under subsampling of units. Exchangeability is a different sort of criterion based on egalitarianism, the assumption that distributions are unaffected by permutation of units. Regression exchangeability is also based on egalitarianism, the assumption that two sets of units having the same covariate values also have the same response distribution. The range of examples illustrated in Section 4.3 shows that the assumption is almost, but not quite, universal in parametric statistical models.

Regression exchangeability is a fairly natural assumption in many circumstances, but the possibility of failure is not to be dismissed. Failure means that there exist two sets of subjects having the same covariate values that have different distributions, presumably due to differences not included in the covariate. If the differences are due to an unmeasured variable, and if treatment assignment is determined in part by such a variable, the apparent treatment effect is a combination of two effects, one due to the treatment and the other due to the unrecorded variable. Randomization may be used as a device to guard against potential biases of this sort by ensuring that the exchangeability assumption is satisfied, at least in the unconditional sense.

As a function on the units, a covariate serves to distinguish one unit from another, and the notion in an exchangeable regression model is that differences between distributions must be explained by differences between covariate values. However, a covariate is not the only sort of mathematical object that can introduce inhomogeneities or distributional differences. The genetic relationship between subjects in a clinical trial is a function on pairs of subjects. It is not a covariate, nor is it an equivalence relation, but it may affect the distribution of pairs as described in Section 4.5. Two pairs having the same covariate value may have different joint distributions if their genetic relationships are different. The relevant notion of exchangeability in this context is that two sets of units having the same covariate values and the same relationships also have the same joint distribution.

Exchangeability is a primitive but fundamental concept with implications in a wide range of applications. Even in spatial applications, if we define the relationship between pairs of units to be their spatial separation, the definition in Section 4.5 is satisfied by all stationary isotropic processes. The concept is not especially helpful or useful in the practical sense because it does not help much in model construction or model selection. Nonetheless, there are exceptions, potential areas of application in which notions of exchangeability may provide useful insights. The following are four examples.

In connection with factorial decomposition and analysis of variance, Cox (1984*a*, §5.5) has observed that two factors having large main effects are more likely to exhibit interaction than two factors whose main effects are small. To mimic this phenomenon in a Bayesian model, it is necessary to construct a partially exchangeable prior process in the sense of Aldous (1981) that exhibits the desired property. Does exchangeability allow this? If so, describe such a process and illustrate its use in factorial models.

Given a regression-exchangeable process, one can duplicate an experiment in the mathematical sense by considering a new set of units having the same covariate values as the given set. For a replicate experiment on the same process, the test statistic $T(Y^*)$ may or may not exceed the value $T(Y)$ observed in the original experiment: the two statistics are exchangeable and thus have the same distribution. The exceedent probability or p -value is a prediction on the combined sample space $\text{pr}\{T(Y^*) \geq T(Y) \mid Y\}$, and as such is not subject to the likelihood

principle. The subsequent inference, that a small p -value is evidence against the model or null hypothesis, if interpreted as evidence in favour of specific parameter points in a larger parameter space, is an inference potentially in violation of the likelihood principle. Bearing in mind the distinction between estimation and prediction, clarify the nature of the likelihood-principle violation (Berger and Wolpert 1988).

A mixture of processes on the same observation spaces is a process, and a mixture of exchangeable processes is an exchangeable process. An improper mixture of processes is not a process in the Kolmogorov sense. The fact that improper mixtures are used routinely in Bayesian work raises questions connected with definitions. What sort of process-like object is obtained by this non-probabilistic operation? Is it feasible to extend the definition of a process in such a way that the extended class is closed under improper mixtures? For example, the symmetric density functions

$$f_n(x_1, \dots, x_n) = n^{-1/2} \Gamma\{(n - \nu)/2\} \pi^{-n/2} \left\{ \sum_{i=1}^n (x_i - \bar{x}_n)^2 \right\}^{-(n-\nu)/2}$$

are Kolmogorov-compatible in the sense that the integral of f_{n+1} with respect to x_{n+1} gives f_n . For $n \geq 2$, the ratio f_{n+1}/f_n is a transition density, in fact Student's t on $n - \nu$ degrees of freedom centred at \bar{x}_n . In symbols, for $n \geq 2$,

$$X_{n+1} = \bar{x}_n + \left\{ \frac{(n^2 - 1)s_n^2}{n(n - \nu)} \right\}^{1/2} \epsilon_n,$$

in which s_n^2 is the sample variance of the first n components, and the components of ϵ are independent with distributions $\epsilon_n \sim t_{n-\nu}$. However, f_n is not integrable on \mathbb{R}^n , so these functions do not determine a process in the Kolmogorov sense, and certainly not an exchangeable process. Nonetheless, the transition densities permit prediction, either for one value or averages such as $\bar{X}_\infty = \bar{x}_n + s_n \epsilon_n / [(n - 1)/\{n(n - \nu)\}]^{1/2}$.

In the preceding example, the transition density $p_n(x; t) = f_{n+1}(x, t)/f_n(x)$ is a function that associates with each point $x = (x_1, \dots, x_n)$ a probability density on \mathbb{R} . In other words, a transition density is a density estimator in the conventional sense. By necessity, the joint two-step transition density $p_n^2(x; t_1, t_2) = f_{n+2}(x, t_1, t_2)/f_n(x)$ is a product of one-step transitions

$$p_n^2(x; t_1, t_2) = p_n(x; t_1) p_{n+1}\{(x, t_1), t_2\}.$$

For an exchangeable process, this density is symmetric under the interchange $t_1 \leftrightarrow t_2$. Both marginal distributions of p_n^2 are equal to p_n , so two-step-ahead prediction is the same as one-step prediction, as is to be expected in an exchangeable process. Ignoring matters of computation, a sequence of one-step predictors determines a two-step predictor, so a one-step density estimator determines a two-step density estimator. For commercial-grade kernel-type density estimators, it appears that these estimators are not the same, which prompts a

number of questions. Is the difference between the two estimators an indication that density estimation and prediction are not equivalent activities? If so, what is the statistical interpretation of the difference? Is the difference appreciable or a matter for concern? If it were feasible to compute both estimators, which one would be preferred, and for what purpose?

Acknowledgement

This research was supported in part by the US National Science Foundation.

Bibliography

- Abbott, A. and Hrycak, A. (1990), ‘Measuring resemblance in sequence data: An optimal matching analysis of musicians’ careers’, *American Journal of Sociology* **96**, 144–185.
- Abramowitz, M. and Stegun, I. A., (eds) (1964), *Handbook of Mathematical Functions*, Dover, New York.
- Agresti, A. (2002), *Categorical Data Analysis*, 2nd, Wiley, New York.
- Alcock, C., Alves, D. R., Becker, A., Bennett, D., Cook, K. H., Drake, A., Freeman, K., Geha, M., Griest, K., Kovacs, G., Lehner, M., Marshall, S., Minniti, D., Nelson, C., Peterson, B., Popowski, P., Pratt, M., Quinn, P., Rodgers, A., Stubbs, C., Sutherland, W., Vandehei, T. and Welch, D. L. (2003), ‘The Macho project Large Magellanic Cloud variable star inventory. XI. Frequency analysis of the fundamental-mode RR Lyrae stars’, *Astrophysical Journal* **598**, 597–609.
- Aldous, D. J. (1981), ‘Representations for partially exchangeable arrays of random variables’, *Journal of Multivariate Analysis* **11**, 581–598.
- Aldous, D. J. and Eagleson, G. K. (1978), ‘On mixing and stability of limit theorems’, *Annals of Probability* **6**, 325–331.
- Andersen, T. G. and Bollerslev, T. (1997), ‘Intraday periodicity and volatility persistence in financial markets’, *Journal of Empirical Finance* **4**, 115–158.
- Andersen, T. G. and Bollerslev, T. (1998), ‘Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies’, *Journal of Finance* **53**, 219–265.
- Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2003c), Some like it smooth, and some like it rough: Untangling continuous and jump components in measuring, modeling and forecasting asset return volatility. Unpublished paper: Economics Dept, Duke University.
- Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2005a), Parametric and nonparametric measurement of volatility, in *Handbook of Financial Econometrics*, North Holland, Amsterdam. To appear.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001), ‘The distribution of exchange rate volatility’, *Journal of the American Statistical Association* **96**, 42–55. Correction published in 2003, volume 98, page 501.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003a), ‘Modeling and forecasting realized volatility’, *Econometrica* **71**, 579–625.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Vega, C. (2003b), ‘Micro effects of macro announcements: Real-time price discovery in foreign exchange’, *American Economic Review* **93**, 38–62.
- Andersen, T. G., Bollerslev, T. and Meddahi, N. (2005b), ‘Correcting the errors: A note on volatility forecast evaluation based on high-frequency data and realized volatilities’, *Econometrica* **73**, 279–296.
- Anderson, R. M., Cox, D. R. and Hillier, H. C. (1989), ‘Epidemiological and statistical aspects of the AIDS epidemic: Introduction’, *Philosophical Transactions of the Royal Society of London, series B* **325**, 39–44.

- Anderson, R. M. and May, R. M. (1991), *Infectious Diseases of Humans: Dynamics and Control*, Clarendon Press, Oxford.
- Anderson, R. M., Medley, G. F., May, R. M. and Johnson, A. M. (1986), 'A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS', *IMA Journal of Mathematics Applied in Medicine and Biology* **3**, 229–263.
- Andersson, H. (1998), 'Limit theorems for a random graph epidemic model', *Annals of Applied Probability* **8**, 1331–1349.
- Andersson, H. and Britton, T. (2000), *Stochastic Epidemic Models and their Statistical Analysis*, Springer, New York. Lecture Notes in Statistics, no. 151.
- Andersson, H. and Djehiche, B. (1998), 'A threshold limit theorem for the stochastic logistic epidemic', *Journal of Applied Probability* **35**, 662–670.
- Andrews, D. W. K. (1991), 'Asymptotic normality of series estimators for non-parametric and semiparametric models', *Econometrica* **59**, 307–345.
- Anonymous (1984), 'Review of mortality results in randomized trials in early breast cancer', *Lancet* **ii**, 1205.
- Armitage, P. (2003), 'Fisher, Bradford Hill, and randomization', *International Journal of Epidemiology* **32**, 925–928.
- Arrigoni, F. and Pugliese, A. (2002), 'Limits of a multi-patch SIS epidemic model', *Journal of Mathematical Biology* **45**, 419–440.
- Athreya, J. S. and Fidkowski, L. M. (2002), 'Number theory, balls in boxes, and the asymptotic uniqueness of maximal discrete order statistics', *Integers* **A3**, 5 pp. (electronic).
- Austin, D. J., Bonten, M. J. M., Weinstein, R. A., Slaughter, S. and Anderson, R. M. (1999), 'Vancomycin-resistant enterococci in intensive-care hospital settings: Transmission dynamics, persistence and the impact of infection control programs', *Proceedings of the National Academy of Science* **96**, 6908–6913.
- Back, K. (1991), 'Asset pricing for general processes', *Journal of Mathematical Economics* **20**, 371–395.
- Bailey, N. T. J. (1975), *The Mathematical Theory of Infectious Diseases and its Applications*, Griffin, London.
- Ball, F. (1999), 'Stochastic and deterministic models for SIS epidemics among a population partitioned into households', *Mathematical Biosciences* **156**, 41–67.
- Ball, F. and Clancy, D. (1993), 'The final size and severity of a generalised stochastic multitype epidemic model', *Advances in Applied Probability* **25**, 721–736.
- Ball, F. and Clancy, D. (1995), 'The final outcome of an epidemic model with several different types of infective in a large population', *Journal of Applied Probability* **32**, 579–590.
- Ball, F. and Lyne, O. (2001), 'Stochastic multi-type SIR epidemics among a population partitioned into households', *Advances in Applied Probability* **33**, 99–123.
- Ball, F. and Lyne, O. (2002), 'Optimal vaccination policies for stochastic epidemics among a population of households', *Mathematical Biosciences* **177**, 333–354.
- Ball, F., Mollison, D. and Scalia-Tomba, G. (1997), 'Epidemics with two levels of mixing', *Advances in Applied Probability* **7**, 46–89.
- Ball, F. and Neal, P. (2002), 'A general model for stochastic SIR epidemics with two levels of mixing', *Mathematical Biosciences* **180**, 73–102.
- Ball, F. and Neal, P. (2003), 'The great circle epidemic model', *Stochastic Processes and their Applications* **107**, 233–268.

- Bandi, F. M. and Russell, J. R. (2003), Microstructure noise, realized volatility, and optimal sampling. Unpublished paper presented at the Realized Volatility Conference, Montreal, 8th November, 2003.
- Banks, S. C. (2002), ‘Agent-based modeling: A revolution?’, *Proceedings of the National Academy of Sciences* **99**, 7199–7200.
- Barabasi, A.-L. and Albert, R. (1999), ‘Emergence of scaling in random networks’, *Science* **286**, 509–512.
- Barabasi, A.-L., Albert, R. and Jeong, H. (1999), ‘Mean-field theory for scale-free random networks’, *Physica A* **272**, 173–187.
- Barbour, A. D. (1972), ‘The principle of diffusion of arbitrary constants’, *Journal of Applied Probability* **9**, 519–541.
- Barbour, A. D. (1974), ‘On a functional central limit theorem for Markov population processes’, *Advances in Applied Probability* **6**, 21–39.
- Barbour, A. D. (1975), ‘The duration of the closed stochastic epidemic’, *Biometrika* **62**, 477–82.
- Barbour, A. D. and Kafetzaki, M. (1993), ‘A host-parasite model yielding heterogeneous parasite loads’, *Journal of Mathematical Biology* **31**, 157–176.
- Barbour, A. D. and Mollison, D. (1990), Epidemics and random graphs, in J.-P. Gabriel, C. Lefèvre and P. Picard, (eds) *Stochastic Processes in Epidemic Theory. Lecture Notes in Biomathematics*, vol. 86, Springer, Berlin, pp. 86–89.
- Barbour, A., Heesterbeek, J. and Luchsinger, C. (1996), ‘Threshold and initial growth rates in a model of parasitic infection’, *Advances in Applied Probability* **6**, 1045–1074.
- Barchielli, A. and Holevo, A. S. (1995), ‘Constructing quantum stochastic measurement processes via classical stochastic calculus’, *Stochastic Processes and their Applications* **58**, 293–317.
- Barndorff-Nielsen, O. E. (1986), ‘Inference on full or partial parameters based on the standardized signed log likelihood ratio’, *Biometrika* **73**, 307–322.
- Barndorff-Nielsen, O. E. (1997), ‘Superposition of Ornstein–Uhlenbeck type processes’, *Theory of Probability and its Applications* **46**, 175–194.
- Barndorff-Nielsen, O. E. (1998a), Probability and statistics: Self-decomposability, finance and turbulence, in *Probability towards 2000 (New York, 1995)*, Springer, New York, pp. 47–57.
- Barndorff-Nielsen, O. E. (1998b), ‘Processes of normal inverse Gaussian type’, *Finance and Stochastics* **2**, 41–68.
- Barndorff-Nielsen, O. E., Blæsild, P. and Schmiegel, J. (2004e), ‘A parsimonious and universal description of turbulent velocity increments’, *European Physical Journal, series B* **41**, 345–363.
- Barndorff-Nielsen, O. E. and Cox, D. R. (1979), ‘Edgeworth and saddle-point approximations with statistical applications (with discussion)’, *Journal of the Royal Statistical Society, series B* **41**, 279–312.
- Barndorff-Nielsen, O. E. and Cox, D. R. (1989), *Asymptotic Techniques for Use in Statistics*, Chapman & Hall, London.
- Barndorff-Nielsen, O. E. and Cox, D. R. (1994), *Inference and Asymptotics*, Chapman & Hall, London.
- Barndorff-Nielsen, O. E. and Cox, D. R. (1996), ‘Prediction and asymptotics’, *Bernoulli* **2**, 319–340.
- Barndorff-Nielsen, O. E., Gill, R. D. and Jupp, P. E. (2001c), On quantum statistical inference. Research Report 2001–19. MaPhySto, University of Aarhus.

- Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J., Podolskij, M. and Shephard, N. (2004a), A central limit theorem for realised power and bipower variations of continuous semimartingales. Festschrift in Honour of A. N. Shiryaev. Forthcoming.
- Barndorff-Nielsen, O. E., Jensen, J. L. and Sørensen, M. (1990), 'Parametric modelling of turbulence', *Philosophical Transactions of the Royal Society of London, series A* **332**, 439–455.
- Barndorff-Nielsen, O. E. and Lindner, A. (2004), Some aspects of Lévy copulas. Research Report (Submitted).
- Barndorff-Nielsen, O. E., Maejima, M. and Sato, K. (2004c), Infinite divisibility and time change. *Journal of Theoretical Probability*, forthcoming.
- Barndorff-Nielsen, O. E., Maejima, M. and Sato, K. (2004d), Some classes of multivariate infinitely divisible distributions admitting stochastic integral representation. *Bernoulli*, forthcoming.
- Barndorff-Nielsen, O. E., Mikosch, T. and Resnick, S. I., (eds) (2001a), *Lévy Processes: Theory and Applications*, Birkhäuser, Boston, MA.
- Barndorff-Nielsen, O. E., Nicolato, E. and Shephard, N. (2002), 'Some recent developments in stochastic volatility modelling', *Quantitative Finance* **2**, 11–23.
- Barndorff-Nielsen, O. E., Pedersen, J. and Sato, K. (2001b), 'Multivariate subordination, self-decomposability and stability', *Advances in Applied Probability* **33**, 160–187.
- Barndorff-Nielsen, O. E. and Peres-Abreu, V. (2005), Matrix subordinators and Upsilon transformation. (In preparation).
- Barndorff-Nielsen, O. E. and Schmiegel, J. (2004), 'Lévy-based tempo-spatial modelling; with applications to turbulence.', *Russian Mathematical Surveys (Uspekhi Acad. Nauk)* **59**, 65–90.
- Barndorff-Nielsen, O. E. and Shephard, N. (2001a), Modelling by Lévy processes for financial econometrics, in O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick, (eds), *Lévy Processes: Theory and Applications*, Birkhäuser, Boston, MA, pp. 283–318.
- Barndorff-Nielsen, O. E. and Shephard, N. (2001b), 'Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion)', *Journal of the Royal Statistical Society, series B* **63**, 167–241.
- Barndorff-Nielsen, O. E. and Shephard, N. (2002), 'Econometric analysis of realised volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society, series B* **64**, 253–280.
- Barndorff-Nielsen, O. E. and Shephard, N. (2003), Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, forthcoming.
- Barndorff-Nielsen, O. E. and Shephard, N. (2004a), 'Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics', *Econometrica* **72**, 885–925.
- Barndorff-Nielsen, O. E. and Shephard, N. (2004b), A feasible central limit theory for realised volatility under leverage. Unpublished discussion paper: Nuffield College, Oxford.
- Barndorff-Nielsen, O. E. and Shephard, N. (2004c), 'Power and bipower variation with stochastic volatility and jumps (with discussion)', *Journal of Financial Econometrics* **2**, 1–48.
- Barndorff-Nielsen, O. E. and Shephard, N. (2005), *Continuous Time Approach to Financial Volatility*, Cambridge University Press, Cambridge.

- Barndorff-Nielsen, O. E. and Shiryaev, A. N. (2005), 'Change of time and change of measure'. (To appear.).
- Barndorff-Nielsen, O. E. and Thorbjørnsen, S. (2002a), 'Lévy laws in free probability', *Proceedings of the National Academy of Sciences* **99**, 16568–16575.
- Barndorff-Nielsen, O. E. and Thorbjørnsen, S. (2002b), 'Lévy processes in free probability', *Proceedings of the National Academy of Sciences* **99**(26), 16576–16580.
- Barndorff-Nielsen, O. E. and Thorbjørnsen, S. (2002c), 'Self-decomposability and Lévy processes in free probability', *Bernoulli* **8**, 323–366.
- Barndorff-Nielsen, O. E. and Thorbjørnsen, S. (2004a), 'A connection between free and classical infinite divisibility', *Infinite Dimensional Analysis, Quantum Probability, and Related Topics* **7**, 573–590.
- Barndorff-Nielsen, O. E. and Thorbjørnsen, S. (2004b), 'The Lévy–Itô decomposition in free probability', *Probability Theory and Related Fields*. (To appear).
- Barndorff-Nielsen, O. E. and Thorbjørnsen, S. (2004c), Regularising mappings of Lévy measures. (Submitted).
- Barndorff-Nielsen, O. E., Vedel Jensen, E. B., Jónsdóttir, K. Ýr and Schmiegel, J. (2004), Spatio-temporal modelling—with a view to biological growth. Research Report 2004–12, T. N. Thiele Centre, Aarhus.
- Bartlett, M. S. (1949), 'Some evolutionary stochastic processes', *Journal of the Royal Statistical Society, series B* **11**, 211–229.
- Bartlett, M. S. (1956), Deterministic and stochastic models for recurrent epidemics, in 'Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability', Vol. 4, University of California Press, Berkeley, pp. 81–109.
- Bartlett, M. S. (1957), 'Measles periodicity and community size', *Journal of the Royal Statistical Society, series A* **120**, 48–70.
- Bartlett, M. S. (1960), *Stochastic Population Models in Ecology and Epidemiology*, Methuen, London.
- Bartlett, M. S. (1990), 'Chance or chaos (with discussion)', *Journal of the Royal Statistical Society, series A* **153**, 321–348.
- Bates, D. M. and Watts, D. G. (1988), *Nonlinear Regression Analysis and Its Applications*, Wiley, New York.
- Beck, N. (2000), 'Political methodology: a welcoming discipline', *Journal of the American Statistical Association* **95**, 651–654.
- Becker, N. G. (1989), *Analysis of Infectious Disease Data*, Chapman & Hall, London.
- Becker, N. G. and Britton, T. (1999), 'Statistical studies of infectious disease incidence', *Journal of the Royal Statistical Society, series B* **61**, 287–308.
- Becker, N. G., Britton, T. and O'Neill, P. D. (2003), 'Estimating vaccine effects on transmission of infection from household outbreak data', *Biometrics* **59**, 467–475.
- Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988), *The New S Language: A Programming Environment for Data Analysis and Graphics*, Pacific Grove, CA: Wadsworth and Brooks/Cole.
- Begun, J. M., Hall, W. J., Huang, W. M. and Wellner, J. (1983), 'Information and asymptotic efficiency in parametric-nonparametric models', *Annals of Statistics* **11**, 432–452.
- Behnke, J. M., Bajer, A., Sinski, E. and Wakelin, D. (2001), 'Interactions involving intestinal nematodes of rodents: experimental and field studies', *Parasitology* **122**, 39–49.

- Ben Ghorbal, A. and Schürmann, M. (2002), 'Non-commutative notions of stochastic independence', *Mathematical Proceedings of the Cambridge Philosophical Society* **133**, 531–561.
- Beran, R. J. (1974), 'Asymptotically efficient adaptive rank estimators in location models', *Annals of Statistics* **2**, 63–74.
- Bercovici, H. and Pata, V. (1999), 'Stable laws and domains of attraction in free probability', *Annals of Mathematics* **149**, 1023–1060.
- Berger, J. O. and Wolpert, R. L. (1988), *The Likelihood Principle*, Vol. 6 of *Lecture Notes—Monograph Series*, 2nd edn, Institute of Mathematical Statistics, Hayward, California.
- Berkson, J. (1950), 'Are there two regressions?', *Journal of the American Statistical Association* **45**, 164–180.
- Bernardo, J. M. and Smith, A. F. M. (1994), *Bayesian Theory*, Wiley, New York.
- Bernoulli, D. (1760), 'Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'incubation pour la prévenir', *Mémoires Mathématiques et Physiques de l'Académie Royale de Sciences de Paris*, pp. 1–45.
- Bertoin, J. (1996), *Lévy Processes*, Cambridge University Press, Cambridge.
- Bertoin, J. (1997), Subordinators: Examples and applications, in 'Lectures on Probability Theory and Statistics', Springer, Berlin, pp. 1–91. Ecole d'Eté de Probabilités de Saint-Flour XXVII.
- Besag, J. E. (1974), 'Spatial interaction and the statistical analysis of lattice systems (with discussion)', *Journal of the Royal Statistical Society, series B* **34**, 192–236.
- Besag, J. E. (2002), 'Discussion of 'What is a statistical model?' by P. McCullagh', *Annals of Statistics* **30**, 1267–1277.
- Besag, J. E., York, J. and Mollié, A. (1991), 'Bayesian image restoration, with two applications in spatial statistics (with discussion)', *Annals of the Institute of Statistical Mathematics* **43**, 1–59.
- Bickel, P. J. (1982), 'On adaptive estimation', *Annals of Statistics* **10**, 647–671.
- Bickel, P. J. (1993), Estimation in semiparametric models, in C. R. Rao, (ed.), *Multivariate Analysis: Future Directions*, North-Holland, Amsterdam, pp. 55–73.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore.
- Bickel, P. J. and Kwon, J. (2001), 'Inference in semiparametric models: Some questions and one answer', *Statistica Sinica* **11**, 863–960.
- Bickel, P. J. and Ritov, Y. (1988), 'Estimating squared density derivatives: Sharp best order of convergence estimates', *Sankhya* **50**, 136–146.
- Bickel, P. J. and Ritov, Y. (2000), 'Non- and semiparametric statistics: Compared and contrasted', *Journal of Statistical Planning and Inference* **91**, 209–228.
- Bickel, P. J., Ritov, Y. and Stoker, T. M. (2004), Tailor-made tests for goodness-of-fit to semiparametric hypotheses, Technical report, Department of Statistics, The Hebrew University of Jerusalem.
- Billard, L., Medley, G. F. and Anderson, R. M. (1990), The incubation period for the AIDS virus, in J.-P. Gabriel, C. Lefèvre and P. Picard, (eds), *Stochastic Processes in Epidemic Theory. Lecture Notes in Biomathematics*, vol. 86', Springer, Berlin, pp. 21–35.

- Bird, S. M., Cox, D. R., Farewell, V. T., Goldstein, H., Holt, T. and Smith, P. C. (2005), 'Performance indicators: Good, bad and ugly', *Journal of the Royal Statistical Society, series A* **168**, 1–27.
- Blanks, R. G., Moss, S. M., McGahan, C. E., Quinn, M. J. and Babb, P. J. (2000), 'Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990–8: Comparison of observed with predicted mortality', *British Medical Journal* **321**, 665–669.
- Blossfeld, H.-P., Hamerle, A. and Mayer, K. U. (1989), *Event History Analysis. Statistical Theory and Applications in the Social Sciences*, Hillsday, NJ: Erlbaum.
- Blundell, R. and Powell, J. L. (2003), Endogeneity in nonparametric and semiparametric regression models, in M. Dewatripont, L.-P. Hansen and S. J. Turnovsky, (eds), *Advances in Economics and Econometrics: Theory and Applications*, Vol. 2, Cambridge University Press, Cambridge, pp. 312–357.
- Bochner, S. (1949), 'Diffusion equation and stochastic processes', *Proceedings of the National Academy of Sciences* **85**, 369–370.
- Bochner, S. (1955), *Harmonic Analysis and the Theory of Probability*, University of California Press.
- Bodmer, W. (2003), 'R.A. Fisher, statistician and geneticist extraordinary: A personal view', *International Journal of Epidemiology* **32**, 938–942.
- Bolker, B. M. (2003), 'Combining endogenous and exogenous spatial variability in analytical population models', *Theoretical Population Biology* **64**, 255–270.
- Bollobás, B. (1985), *Random Graphs*, Academic Press, London.
- Bondesson, L. (1981), 'Classes of infinitely divisible distributions and densities', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte gebiete* **57**, 39–71.
- Bondesson, L. (1992), *Generalized Gamma Distributions and Related Classes of Distributions and Densities*, Springer, Berlin.
- Borg, I. and Groenen, P. (1997), *Modern Multidimensional Scaling. Theory and Applications*, Springer-Verlag, New York.
- Bottomley, C., Isham, V. and Basáñez, M.-G. (2005) Population biology of multispecies helminth infection: Interspecific interactions and parasite distribution. *Parasitology*, in press.
- Bouton, C., Henry, G., Colantuoni, C. and Pevsner, J. (2003), Dragon and Dragon View: Methods for the annotation, analysis and visualization of large-scale gene expression data, in G. Parmigiani, E. S. Garrett, R. A. Irizarry and S. L. Zeger, (eds), *The Analysis of Gene Expression Data*, Springer, New York.
- Box, G. E. P. and Cox, D. R. (1964), 'An analysis of transformations (with discussion)', *Journal of the Royal Statistical Society, series B* **26**, 211–246.
- Box, J. F. (1978), *R. A. Fisher. The Life of a Scientist*, Wiley, New York.
- Bradford Hill, A. (1937), *Principles of Medical Statistics*, Arnold, London.
- Bradford Hill, A. (1965), 'The environment and disease: Association or causation?', *Proceedings of the Royal Society of Medicine* **58**, 295–300.
- Bradley, K. M., Bydder, G. M., Budge, M. M., Hajnal, J. V., White, S. J., Ripley, B. D. and Smith, A. D. (2002), 'Serial brain MRI at 3–6 month intervals as a surrogate marker for Alzheimer's disease.', *British Journal of Radiology* **75**, 506–513.
- Brazzale, A. R. (2000), Practical Small-Sample Parametric Inference, PhD thesis, Department of Mathematics, Ecole Polytechnique Fédérale de Lausanne.
- Breiman, L. (2001), 'Statistical modeling: The two cultures (with discussion)', *Statistical Science* **16**, 199–215.

- Breiman, L. and Friedman, J. (1997), 'Predicting multivariate responses in multiple linear regression (with discussion)', *Journal of the Royal Statistical Society, series B* **59**, 3–37.
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984), *Classification and Regression Trees*, Wadsworth and Brooks/Cole, Pacific Grove, California.
- Breslow, N. E. and Clayton, D. G. (1993), 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association* **88**, 9–25.
- Brillinger, D. R. (1983), A generalized linear model with Gaussian regressor variables, in P. J. Bickel, K. A. Doksum and J. L. Hodges, (eds), *A Festschrift for Erich L. Lehmann*, Wadsworth and Brooks/Cole, Pacific Grove, California, pp. 97–114.
- Brillinger, D. R. (1986), 'The natural variability of vital rates and associated statistics (with discussion)', *Biometrics* **42**, 693–734.
- Brøns, H. (2002), 'Discussion of "What is a statistical model?" by P. McCullagh', *Annals of Statistics* **30**, 1279–1283.
- Browne, M. W. (2000), 'Psychometrics', *Journal of the American Statistical Association* **95**, 661–665.
- Bruce, M. C., Donnelly, C. A., Alpers, M. P., Galinski, M. R., Barnwell, J. W., Walliker, D. and Day, K. P. (2000), 'Cross-species interactions between malaria parasites in humans', *Science* **287**, 845–848.
- Carr, P., Chang, E. and Madan, D. B. (1998), 'The variance gamma process and option pricing', *European Finance Review* **2**, 79–105.
- Carr, P., Geman, H., Madan, D. B. and Yor, M. (2003), 'Stochastic volatility for Lévy processes', *Mathematical Finance* **13**, 345–382.
- Carr, P. and Wu, L. (2004), 'Time-changed Lévy processes and option pricing', *Journal of Financial Economics* pp. 113–141.
- Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A. and Poulton, R. (2003), 'Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene', *Science* **301**, 386–389.
- Chalmers, I. (2003), 'Fisher and Bradford Hill: Theory and pragmatism?', *International Journal of Epidemiology* **32**, 922–924.
- Chamberlain, G. (1986), 'Asymptotic efficiency in semiparametric models with censoring', *Journal of Econometrics* **32**, 189–218.
- Chamberlain, G. (1987), 'Asymptotic efficiency in estimation with conditional moment restrictions', *Journal of Econometrics* **34**, 305–334.
- Chambers, J. M. (1998), *Programming with Data. A Guide to the S Language*, Springer-Verlag, New York.
- Chambers, J. M. and Hastie, T. J., (eds) (1992), *Statistical Models in S*, Chapman & Hall, New York.
- Chambers, R. L. and Skinner, C. J. (2003), *Analysis of Survey Data*, Chichester: Wiley.
- Chan, M. S. and Isham, V. S. (1998), 'A stochastic model of schistosomiasis immunoparasitology', *Mathematical Biosciences* **151**, 179–198.
- Chen, H. (1988), 'Convergence rates for parametric components in a partly linear model', *Annals of Statistics* **16**, 136–146.
- Chetelat, G. and Baron, J. C. (2003), 'Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging', *Neuroimage* **2**, 525–541.
- Chib, S. (1995), 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association* **90**, 1313–21.

- Chib, S. and Jeliazkov, I. (2001), 'Marginal likelihood from the Metropolis–Hastings output', *Journal of the American Statistical Association* **96**, 270–281.
- Choi, E. and Hall, P. (1999), 'Nonparametric approach to analysis of space-time data on earthquake occurrences', *Journal of Computational and Graphical Statistics* **8**, 733–748.
- Christensen, B. J. and Prabhala, N. R. (1998), 'The relation between implied and realized volatility', *Journal of Financial Economics* **37**, 125–150.
- Chung, K. L. and Zang, J.-C. (2003), *Introduction to Random Time and Quantum Randomness*, World Scientific, Singapore.
- Claeskens, G. (2004), 'Restricted likelihood ratio lack of fit tests using mixed spline models', *Journal of the Royal Statistical Society, series B* **66**, 909–926.
- Clancy, D. (1994), 'Some comparison results for multitype epidemic models', *Journal of Applied Probability* **31**, 9–21.
- Clancy, D. (1996), 'Carrier-borne epidemic models incorporating population mobility', *Mathematical Biosciences* **132**, 185–204.
- Clancy, D. (1999a), 'Optimal interventions for epidemic models with general infection and removal rate functions', *Journal of Mathematical Biology* **39**, 309–331.
- Clancy, D. (1999b), 'Outcomes of epidemic models with general infection and removal rate functions at certain stopping times', *Journal of Applied Probability* **36**, 799–813.
- Clancy, D. and Pollett, P. K. (2003), 'A note on the quasi-stationary distributions of birth-death processes and the SIS logistic epidemic', *Journal of Applied Probability* **40**, 821–825.
- Clapp, T. C. and Godsill, S. J. (1999), Fixed-lag smoothing using sequential importance sampling, in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, (eds), *Bayesian Statistics 6*, Oxford University Press, Oxford, pp. 743–752.
- Clayton, D. G. (1996), Generalized linear mixed models, in W. R. Gilks, S. Richardson and D. J. Spiegelhalter, (eds), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London, pp. 275–301.
- Clayton, D. G. and Kaldor, J. (1987), 'Empirical Bayes estimates of age-standardised relative risks for use in disease mapping', *Biometrics* **43**, 671–681.
- Clogg, C. C. (1992), 'The impact of sociological methodology on statistical methodology (with discussion)', *Statistical Science* **7**, 183–207.
- Colantuoni, C., Henry, G., Bouton, C., Zeger, S. L. and Pevsner, J. (2003), Snomad: Biologist-friendly web tools for the standardization and normalization of microarray data, in G. Parmigiani, E. S. Garrett, R. A. Irizarry and S. L. Zeger, (eds), *The Analysis of Gene Expression Data*, Springer, New York.
- Collins, F. S., Green, E. D., Guttmacher, A. E. and Guyer, M. S. (2003), 'A vision for the future of genomics research', *Nature* **422**, 835–847.
- Cont, R. and Tankov, P. (2004), *Financial Modelling with Jump Processes*, Chapman & Hall/CRC, London.
- Coombs, C. H. (1964), *A Theory of Data*, Wiley, New York.
- Cornell, S. J. and Isham, V. S. (2004), 'Ultimate extinction in a bisexual, promiscuous Galton–Watson metapopulation', *Australian and New Zealand Journal of Statistics* **46**, 87–98.
- Cornell, S. J., Isham, V. S. and Grenfell, B. T. (2004), 'Stochastic and spatial dynamics of nematode parasites in ruminants', *Proceedings of the Royal Society of London, Series B* **271**, 1243–1250.

- Cornell, S. J., Isham, V. S., Smith, G. and Grenfell, B. T. (2003), 'Spatial parasite transmission, drug resistance and the spread of rare genes', *Proceedings of the National Academy of Science* **100**, 7401–7405.
- Cornfield, J. (1951), 'A method for estimating comparative rates from clinical data: application to cancer of the lung, breast and cervix', *Journal of the National Cancer Institute* **11**, 1269–1275.
- Corsi, F., Zumbach, G., Müller, U. and Dacorogna, M. (2001), 'Consistent high-precision volatility from high-frequency data', *Economic Notes* **30**, 183–204.
- Cox, D. R. (1949), 'Theory of drafting of wool slivers', *Proceedings of the Royal Society of London, series A* **197**, 28–51.
- Cox, D. R. (1955a), 'Some statistical methods connected with series of events (with discussion)', *Journal of the Royal Statistical Society, series B* **17**, 129–164.
- Cox, D. R. (1955b), 'A use of complex probabilities in the theory of stochastic processes', *Proceedings of the Cambridge Philosophical Society* **51**, 313–319.
- Cox, D. R. (1958a), *Planning of Experiments*, Wiley, New York.
- Cox, D. R. (1958b), 'Some problems connected with statistical inference', *Annals of Mathematical Statistics* **29**, 359–372.
- Cox, D. R. (1958c), 'The regression analysis of binary sequences (with discussion)', *Journal of the Royal Statistical Society, series B* **20**, 215–242.
- Cox, D. R. (1961), Tests of separate families of hypotheses, in 'Proceedings of the 4th Berkeley Symposium on Mathematics, Probability and Statistics', Vol. 1, University of California Press, Berkeley, pp. 105–123.
- Cox, D. R. (1962), *Renewal Theory*, Chapman & Hall, London.
- Cox, D. R. (1966), 'Notes on the analysis of mixed frequency distributions', *British Journal of Mathematical Psychology* **19**, 39–47.
- Cox, D. R. (1972), 'Regression models and life tables (with discussion)', *Journal of the Royal Statistical Society, series B* **34**, 187–220.
- Cox, D. R. (1972a), The statistical analysis of point processes, in P. A. Lewis, (ed.), *Stochastic Point Processes*, Wiley, New York.
- Cox, D. R. (1975), 'Partial likelihood', *Biometrika* **62**, 269–76.
- Cox, D. R. (1981), 'Statistical analysis of time series: Some recent developments (with discussion)', *Scandinavian Journal of Statistics* **8**, 93–115.
- Cox, D. R. (1984a), 'Interaction (with discussion)', *International Statistical Review* **52**, 1–31.
- Cox, D. R. (1984b), Long range dependence: A review, in H. A. David and H. T. David, (eds), *Statistics: An Appraisal*, Iowa State University Press, pp. 55–74.
- Cox, D. R. (1986), 'Comment on Holland, W., Statistics and causal inference', *Journal of the American Statistical Association* **81**, 945–960.
- Cox, D. R. (1990), 'Role of models in statistical analysis', *Statistical Science* **5**, 169–174.
- Cox, D. R. (1991), 'Long-range dependence, non-linearity and time irreversibility', *Journal of Time Series Analysis* **12**, 329–335.
- Cox, D. R. (1992), 'Causality: Some statistical aspects', *Journal of the Royal Statistical Society, series A* **155**, 291–301.
- Cox, D. R. (1995), 'The relation between theory and application in statistics (with discussion)', *Test* **4**, 207–261.
- Cox, D. R. (1997), 'The current position of statistics: A personal view, (with discussion)', *International Statistical Review* **65**, 261–276.

- Cox, D. R. (2000), 'Comment on "Causal inference without counterfactuals" by A. P. Dawid', *Journal of the American Statistical Association* **95**, 424–425.
- Cox, D. R. (2004), 'The accidental statistician', *Significance* **1**, 27–29.
- Cox, D. R., Fitzpatrick, R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J. and Jones, D. R. (1992), 'Quality-of-life assessment: Can we keep it simple? (with discussion)', *Journal of the Royal Statistical Society, series A* **155**, 353–393.
- Cox, D. R. and Hall, P. (2002), 'Estimation in a simple random effects model with nonnormal distributions', *Biometrika* **89**, 831–840.
- Cox, D. R. and Hinkley, D. V. (1974), *Theoretical Statistics*, Chapman & Hall, London.
- Cox, D. R. and Isham, V. (1986), 'The virtual waiting-time and related processes', *Advances in Applied Probability* **18**, 558–573.
- Cox, D. R. and Isham, V. S. (1988), 'A simple spatial-temporal model of rainfall', *Proceedings of the Royal Society of London, series A* **415**, 317–328.
- Cox, D. R. and Miller, H. D. (1965), *The Theory of Stochastic Processes*, Chapman & Hall, London.
- Cox, D. R. and Reid, N. (1987), 'Parameter orthogonality and approximate conditional inference (with discussion)', *Journal of the Royal Statistical Society, series B* **49**, 1–39.
- Cox, D. R. and Reid, N. (2004), 'A note on pseudolikelihood constructed from marginal densities', *Biometrika* **91**, 729–737.
- Cox, D. R. and Snell, E. J. (1979), 'On sampling and the estimation of rare errors (Corr: **69**, 491)', *Biometrika* **66**, 125–132.
- Cox, D. R. and Snell, E. J. (1981), *Applied Statistics: Principles and Examples*, Chapman & Hall, London.
- Cox, D. R. and Solomon, P. J. (2003), *Components of Variance*, Chapman & Hall, London.
- Cox, D. R. and Wermuth, N. (1996), *Multivariate Dependencies*, Chapman & Hall, London.
- Cox, D. R. and Wermuth, N. (2001), 'Some statistical aspects of causality', *European Sociological Review* **17**, 65–74.
- Cox, D. R. and Wermuth, N. (2004), 'Causality: a statistical view', *International Statistical Review* **72**, 285–305.
- Cox, N. J., Frigge, M., Nicolae, D. L., Concanno, P., Harris, C. L., Bell, G. I. and Kong, A. (1999), 'Loci on chromosome 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans', *Nature Genetics* **21**, 213–215.
- Cox, T. F. and Cox, M. A. A. (2001), *Multidimensional Scaling*, 2nd edn, Chapman & Hall/CRC.
- Dacorogna, M. M., Gencay, R., Müller, U. A., Olsen, R. B. and Pictet, O. V. (2001), *An Introduction to High-Frequency Finance*, Academic Press, San Diego.
- Daley, D. J. and Gani, J. (1999), *Epidemic Modelling*, Cambridge University Press, Cambridge.
- Daley, D. J. and Kendall, D. G. (1965), 'Stochastic rumours', *Journal of Institute of Mathematics and its Applications* **1**, 42–55.
- Damien, P. and Walker, S. (2002), 'A Bayesian non-parametric comparison of two treatments', *Scandinavian Journal of Statistics* **29**, 51–56.
- Daniels, H. E. (1991), A look at perturbation approximations for epidemics, in I. V. Basawa and R. L. Taylor, (eds), *Selected Proceedings of the Sheffield Symposium on Applied Probability*, IMS Monograph Series, Vol. 18, Hayward, CA, pp. 48–65.

- Darby, S. C., McGale, P., Peto, R., Granath, F., Hall, P. and Ekbom, A. (2003), 'Mortality from cardiovascular disease more than 10 years after radiotherapy for breast cancer: nationwide cohort study of 90 000 Swedish women', *British Medical Journal* **326**, 256–257.
- Darroch, J. and Seneta, E. (1967), 'On quasi-stationary distributions in absorbing continuous-time finite Markov chains', *Journal of Applied Probability* **4**, 192–196.
- Datta, S., Longini, I. M. and Halloran, M. E. (1999), 'Efficiency of estimating vaccine efficacy for susceptibility and infectiousness: Randomization by individual versus household', *Biometrics* **55**, 792–798.
- Davey Smith, G., Dorling, D., Mitchell, R. and Shaw, M. (2002), 'Health inequalities in Britain: continuing increases up to the end of the 20th century', *Journal of Epidemiology and Community Health* **56**, 434–435.
- Davies, E. B. (1976), *Quantum Theory of Open Systems*, Academic Press, London.
- Davies, E. B. and Lewis, J. T. (1970), 'An operational approach to quantum systems', *Communications in Mathematical Physics* **17**, 239–260.
- Davison, A. C. (2003), *Statistical Models*, Cambridge University Press, Cambridge.
- Davison, A. C. and Cox, D. R. (1989), 'Some simple properties of sums of random variables having long-range dependence', *Proceedings of the Royal Society of London, series A* **424**, 255–262.
- Davison, A. C. and Wang, S. (2002), 'Saddlepoint approximations as smoothers', *Biometrika* **89**, 933–938.
- Dawid, A. P. (2000), 'Causality without counterfactuals (with discussion)', *Journal of the American Statistical Association* **95**, 407–448.
- De Finetti, B. (1974), *Theory of Probability: Volumes 1 and 2*, Wiley, New York.
- De la Cal, J. (1989), 'On the three series theorem in number theory', *Annals of Probability* **17**, 357–361.
- Deeming, T. J. (1975), 'Fourier analysis with unequally-spaced data', *Astrophysics and Space Science* **36**, 137–158.
- Delattre, S. and Jacod, J. (1997), 'A central limit theorem for normalized functions of the increments of a diffusion process in the presence of round off errors', *Bernoulli* **3**, 1–28.
- Denison, D. G. T., Holmes, C. C., Mallick, B. K. and Smith, A. F. M. (2002), *Bayesian Methods for Nonlinear Classification and Regression*, Wiley, Chichester.
- DiBona, C., Ockman, S. and Stone, M., (eds) (1999), *Open Sources. Voices from the Open Source Revolution*, O'Reilly, Sebastopol, CA.
- Dieckmann, U. and Law, R. (2000), Relaxation projections and the method of moments, in U. Dieckmann, R. Law and J. A. J. Metz, (eds), *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, Cambridge University Press, Cambridge, pp. 412–455.
- Diekmann, O. and Heesterbeek, J. A. P. (2000), *Mathematical Epidemiology of Infectious Diseases*, Wiley, Chichester.
- Dietz, K. (1976), The incidence of infectious diseases under the influence of seasonal fluctuations, in *Lecture Notes in Biomathematics*, vol. 11, Springer, Berlin, pp. 1–15.
- Dietz, K. (1988), 'The first epidemic model: A historical note on P. D. En'ko', *Australian Journal of Statistics* **A3**, 56–65.
- Dietz, K. (1989), 'A translation of the paper "On the course of epidemics of some infectious diseases" by P. D. En'ko', *International Journal of Epidemiology* **18**, 749–755.

- Dietz, K. and Schenzle, D. (1985), Mathematical models for infectious disease statistics., in A. C. Atkinson and S. E. Fienberg, (eds), *A Celebration of Statistics: The ISI centenary volume*. Springer, New York, pp. 167–204.
- Diggle, P. J., Heagerty, P., Liang, K.-Y. and Zeger, S. L. (2002), *Analysis of Longitudinal Data*, 2nd edn, Oxford University Press, Oxford.
- Diggle, P. J., Moyeed, R. A. and Tawn, J. A. (1998), ‘Model-based geostatistics (with discussion)’, *Applied Statistics* **47**, 299–350.
- Dobrushin, R. L. (1968), ‘The description of a random field by means of conditional probabilities and conditions of its regularity’, *Theory of Probability and its Applications* **13**, 197–224.
- Dodds, S. and Watts, D. J. (2005), ‘A generalized model of social and biological contagion’, *Journal of Theoretical Biology*, **232**, 587–604.
- Doll, R. (2003), ‘Fisher and Bradford Hill: Their personal impact’, *International Journal of Epidemiology* **32**, 929–931.
- Dolton, P. (2002), ‘Evaluation of economic and social policies’, *Journal of the Royal Statistical Society, series A* **165**, 9–11.
- Dominici, F., Daniels, M., Zeger, S. L. and Samet, J. (2002), ‘Air pollution and mortality: estimating regional and national dose-response relationships’, *Journal of the American Statistical Association* **97**, 100–111.
- Dominici, F., McDermott, A., Zeger, S. L. and Samet, J. M. (2003), ‘National maps of the effects of PM on mortality: Exploring geographical variation’, *Environmental Health Perspectives* **111**, 39–43.
- Donald, S. G. and Newey, W. (1994), ‘Series estimation of semilinear models’, *Journal of Multivariate Analyses* **50**, 30–40.
- Donnelly, C. A., Woodroffe, R., Cox, D. R., Bourne, J., Gettinby, G., Le Fevre, A. M., McNerney, J. P. and Morrison, W. I. (2003), ‘Impact of localised badger culling on tuberculosis incidence in British cattle’, *Nature* **426**, 834–837.
- Donoho, D. L. (1995), ‘Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition’, *Applied and Computational Harmonic Analysis* **2**, 101–126.
- Doornik, J. A. (2001), *Ox: Object Oriented Matrix Programming, 3.0*, Timberlake Consultants Press, London.
- Doucet, A., de Freitas, N. and Gordon, N. J., (eds) (2001), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York.
- Drost, F. C., Klaassen, C. A. and Werker, B. J. M. (1994), ‘Adaptiveness in time series models’, *Asymptotic Statistics: Proceedings of the Fifth Prague Symposium* **5**, 203–211.
- Duan, N. and Li, K. C. (1987), ‘Distribution-free and link-free estimation for the sample selection model’, *Journal of Econometrics* **35**, 25–35.
- Duan, N. and Li, K. C. (1991), ‘Slicing regression: A link-free regression method’, *Annals of Statistics* **19**, 505–530.
- Duerr, H. P., Dietz, K. and Eichner, M. (2003), ‘On the interpretation of age-intensity profiles and dispersion patterns in parasitological surveys’, *Parasitology* **126**, 87–101.
- Duerr, H. P., Dietz, K., Schulz-Key, H., Büttner, D. W. and Eichner, M. (2002), ‘Infection-associated immunosuppression as an important mechanism for parasite density regulation in onchocerciasis’, *Transactions of the Royal Society of Tropical Hygiene and Medicine* **97**, 242–250.

- Durrett, R. (1995), Spatial epidemic models, in D. Mollison, (ed.), *Epidemic Models: their Structure and Relation to Data*, Cambridge University Press, Cambridge, pp. 187–201.
- Durrett, R. and Levin, S. (1994), ‘Stochastic spatial models: a user’s guide to ecological applications’, *Philosophical Transactions of the Royal Society of London* **B343**, 329–350.
- EBCTCG (1988), ‘The effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer: an overview of 61 randomized trials among 28,859 women’, *New England Journal of Medicine* **319**, 1681–1692.
- EBCTCG (1990), *Treatment of Early Breast Cancer*, Vol. 1: Worldwide Evidence, Oxford University Press, Oxford.
- EBCTCG (1992), ‘Systemic treatment of early breast cancer by hormonal, cytotoxic or immune therapy: 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women’, *Lancet* **339**, 1–15 and 71–85.
- EBCTCG (1995), ‘Effects of radiotherapy and surgery in early breast cancer: An overview of the randomized trials’, *New England Journal of Medicine* **333**, 1444–1455.
- EBCTCG (1996), ‘Ovarian ablation in early breast cancer: Overview of the randomised trials’, *Lancet* **348**, 1189–1196.
- EBCTCG (1998a), ‘Polychemotherapy for early breast cancer: An overview of the randomised trials’, *Lancet* **352**, 930–942.
- EBCTCG (1998b), ‘Tamoxifen for early breast cancer: An overview of the randomised trials’, *Lancet* **351**, 1451–1467.
- EBCTCG (2000), ‘Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: An overview of the randomised trials’, *Lancet* **355**, 1757–1770.
- EBCTCG (2005), ‘Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials’, *Lancet* **365**, 1687–1717.
- Eberlein, E., Kallsen, J. and Kristen, J. (2001), Risk management based on stochastic volatility. FDM Preprint 72, University of Freiburg.
- Eberlein, E. and Özkan, F. (2003), ‘Time consistency of Lévy models’, *Quantitative Finance* **3**, 40–50.
- Ederington, L. and Lee, J. H. (1993), ‘How markets process information: News releases and volatility’, *Journal of Finance* **48**, 1161–1191.
- Efromovich, S. (1994), Nonparametric curve estimation from indirect observations, in J. Sall and A. Lehman, (eds), *Computing Science and Statistics, Vol. 26: Computationally Intensive Statistical Methods*, Interface Foundation of America, pp. 196–200.
- Efromovich, S. (1999), *Nonparametric Curve Estimation: Methods, Theory, and Applications*, Springer, New York.
- Efron, B. (2001), ‘Selection criteria for scatterplot smoothers’, *Annals of Statistics* **29**, 470–504.
- Elmore, R., Hall, P. and Neeman, A. (2003), ‘A birational map out of the symmetrical power of A^{n+1} ’, Manuscript.
- Engle, R. F., Granger, C. W. J., Rice, J. and Weiss, A. (1984), ‘Semiparametric estimates of the relation between weather and electricity sales’, *Journal of the American Statistical Association* **81**, 310–320.
- Eraker, B., Johannes, M. and Polson, N. G. (2003), ‘The impact of jumps in returns and volatility’, *Journal of Finance* **53**, 1269–1300.

- Erikson, R. and Goldthorpe, J. H. (1992), *The Constant Flux: A Study of Class Mobility in Industrial Societies*, Oxford: Clarendon Press.
- Erwin, E., Obermayer, K. and Schulten, K. (1992), 'Self-organizing maps: ordering, convergence properties and energy functions', *Biological Cybernetics* **67**, 47–55.
- Estevez-Perez, G., Lorenzo-Cimadevila, H. and Quintela-Del-Rio, A. (2002), 'Nonparametric analysis of the time structure of seismicity in a geographic region', *Annals of Geophysics* **45**, 497–511.
- Evans, M. and Swartz, T. (2000), *Approximating Integrals via Monte Carlo and Deterministic Methods*, Oxford: Oxford University Press.
- Feinberg, A. P. (2001), 'Cancer epigenetics take center stage', *Proceedings of the National Academy of Sciences* **98**, 392–394.
- Feller, W. (1971), *An Introduction to Probability Theory*, 2nd edn, Wiley, New York. Volume II.
- Ferguson, N. M., Donnelly, C. A. and Anderson, R. M. (2001a), 'The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions', *Science* **292**, 1155–1160.
- Ferguson, N. M., Donnelly, C. A. and Anderson, R. M. (2001b), 'Transmission intensity and impact of control policies on the foot-and-mouth epidemic in Great Britain', *Nature* **413**, 542–548.
- Ferguson, N. M., Galvani, A. P. and Bush, R. M. (2003), 'Ecological and immunological determinants of influenza evolution', *Nature* **422**, 428–433.
- Fienberg, S. E. (1989), Comments on 'Modeling considerations from a modeling perspective', in D. Kasprzyk, G. Duncan, G. Kalton and M. P. Singh, (eds), *Panel Surveys*, New York: Wiley, pp. 566–574.
- Fienberg, S. E. (2000), 'Contingency tables and log-linear models: Basic results and new developments', *Journal of the American Statistical Association* **95**, 643–647.
- Finkenstädt, B. F., Bjørnstad, O. and Grenfell, B. T. (2002), 'A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks', *Biostatistics* **3**, 493–510.
- Firth, D. and Bennett, K. E. (1998), 'Robust models in probability sampling (with discussion)', *Journal of the Royal Statistical Society, series B* **60**, 3–21.
- Fisher, B., Redmond, C., Legault-Poisson, S., Dimitrov, N.-V., Brown, A.-M., Wickerham, D. L., Wolmark, N., Margolese, R. G., Bowman, D. and Glass, A. G. (1990), 'Postoperative chemotherapy and tamoxifen compared with tamoxifen alone in the treatment of positive-node breast cancer patients aged 50 years and older with tumors responsive to tamoxifen: Results from the National Surgical Adjuvant Breast and Bowel Project B-16', *Journal of Clinical Oncology* **8**, 1005–1018.
- Fisher, R. A. (1935), *The Design of Experiments*, Oliver and Boyd, Edinburgh.
- Fox, J. (2002), *An R and S-Plus Companion to Applied Regression*, Thousand Oaks, CA: Sage.
- Frangakis, C. E. and Rubin, D. B. (1999), 'Addressing complications of intention-to-treat analysis in the presence of all-or-none treatment-noncompliance and subsequent missing outcomes', *Biometrika* **86**, 365–379.
- Frank, O. and Strauss, D. (1986), 'Markov graphs', *Journal of the American Statistical Association* **81**, 832–842.
- Franklin, R. and Gosling, R. G. (1953), 'Molecular configuration in sodium thymonucleate', *Nature* **171**, 740–741.

- Franz, U. (2003), 'Unification of Boolean, monotone, anti-monotone and tensor independence and Lévy processes', *Mathematische Zeitschrift* **243**, 779–816.
- Fraser, D. A. S. (1990), 'Tail probabilities from observed likelihoods', *Biometrika* **77**, 65–76.
- Fraser, D. A. S. (2003), 'Likelihood for component parameters', *Biometrika* **90**, 327–339.
- Fraser, D. A. S. and Reid, N. (1995), 'Ancillaries and third order significance', *Utilitas Mathematica* **47**, 33–53.
- Fraser, D. A. S., Reid, N. and Wu, J. (1999), 'A simple general formula for tail probabilities for frequentist and Bayesian inference', *Biometrika* **86**, 249–264.
- Friedman, J. (1991), 'Multivariate adaptive regression splines (with discussion)', *Annals of Statistics* **19**, 1–141.
- Friedman, J. H. and Stuetzle, W. (1981), 'Projection pursuit regression', *Journal of the American Statistical Association* **76**, 817–823.
- Friedman, J. H., Stuetzle, W. and Schroeder, A. (1984), 'Projection pursuit density estimation', *Journal of the American Statistical Association* **79**, 599–608.
- Frisch, U. (1995), *Turbulence: The Legacy of A. N. Kolmogorov*, Cambridge University Press, Cambridge.
- Furnival, G. M. and Wilson Jnr., R. W. (1974), 'Regressions by leaps and bounds', *Technometrics* **16**, 499–511.
- Gabriel, J.-P., Lefevre, C. and Picard, P., (eds) (1990), *Stochastic Processes in Epidemic Theory*, number 86 in *Lecture Notes in Biomathematics*, Springer, Berlin.
- Geman, H., Madan, D. B. and Yor, M. (2002), 'Stochastic volatility, jumps and hidden time changes', *Finance and Stochastics* **6**, 63–90.
- Gentleman, R. and Carey, V. (2003), Visualization and annotation of genomic experiments, in G. Parmigiani, E. S. Garrett, R. A. Irizarry and S. L. Zeger, (eds), *The Analysis of Gene Expression Data*, Springer, New York.
- Ghosh, J. K. and Ramamoorthi, R. V. (2003), *Bayesian Nonparametrics*, Springer, New York.
- Ghosh, M., Natarajan, K., Stroud, T. W. F. and Carlin, B. P. (1998), 'Generalized linear models for small-area estimation', *Journal of the American Statistical Association* **93**, 273–282.
- Ghysels, E., Harvey, A. C. and Renault, E. (1996), Stochastic volatility, in C. R. Rao and G. S. Maddala, (eds), *Statistical Methods in Finance*, North-Holland, Amsterdam, pp. 119–191.
- Gibson, G. J. and Renshaw, E. (2001), 'Likelihood estimation for stochastic compartmental models using Markov chain methods', *Statistics and Computing* **11**, 347–358.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J., (eds) (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall/CRC.
- Gleason, A. M., Greenwood, E. and Kelly, L. M. (1980), *The William Lowell Putnam Mathematical Competition*, Mathematical Association of America, Washington, DC.
- Glonek, G. F. V. and Solomon, P. J. (2004), 'Factorial and time course designs for cDNA microarray experiments', *Biostatistics* **5**, 89–111.
- Goldstein, H. (1986), 'Multilevel mixed linear model analysis using iterative generalised least squares', *Biometrika* **73**, 43–56.
- Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993), 'A novel approach to nonlinear and non-Gaussian Bayesian state estimation', *IEEE-Proceedings F* **140**, 107–113.

- Granger, C. W. J. (1988), 'Some recent developments in a concept of causality', *Journal of Econometrics* **39**, 199–211.
- Green, P. J. and Richardson, S. (2002), 'Hidden Markov models and disease mapping', *Journal of the American Statistical Association* **97**, 1055–1070.
- Green, P. J. and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman & Hall, London.
- Greenland, S., Robins, J. M. and Pearl, J. (1999), 'Confounding and collapsibility in causal inference', *Statistical Science* **14**, 29–46.
- Greenwood, P. E. and Wefelmeyer, W. (1995), 'Efficiency of empirical estimators for Markov chains', *Annals of Statistics* **23**, 132–143.
- Grenfell, B. T. (1992), 'Chance and chaos in measles dynamics', *Journal of the Royal Statistical Society, series B* **54**, 383–398.
- Grenfell, B. T. and Dobson, A., (eds) (1995), *Ecology of Infectious Diseases in Natural Populations*, Cambridge University Press, Cambridge.
- Guo, D., Wang, X. and Chen, R. (2003), 'Nonparametric adaptive detection in fading channels based on sequential Monte Carlo and Bayesian model averaging', *Annals of the Institute of Statistical Mathematics* **55**, 423–436.
- Gupta, S., Swinton, J. and Anderson, R. M. (1991), 'Theoretical studies of the effects of heterogeneity in the parasite population on the transmission dynamics of malaria', *Proceedings of the Royal Society of London* **B256**, 231–238.
- Gupta, S., Trenholme, K., Anderson, R. M. and Day, K. P. (1994), 'Antigenic diversity and the transmission dynamics of *plasmodium falciparum*', *Science* **263**, 961–963.
- Hajek, J. (1970), 'A characterization of limiting distributions of regular estimates', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte gebiete* **14**, 323–330.
- Halgreen, C. (1979), 'Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte gebiete* **47**, 13–17.
- Hall, P. (2001), Statistical science—evolution, motivation and direction, in B. Engquist and W. Schmid, (eds), *Mathematics Unlimited—2001 and Beyond*, Springer, Berlin, pp. 565–575.
- Hall, P. and Raimondo, M. (1998), 'On global performance of approximations to smooth curves using gridded data', *Annals of Statistics* **26**, 2206–2217.
- Hall, P. and Zhou, X.-H. (2003), 'Nonparametric estimation of component distributions in a multivariate mixture', *Annals of Statistics* **31**, 201–224.
- Halsey, A. H. and Webb, J. (2000), *Twentieth-Century British Social Trends*, Basingstoke: Macmillan.
- Hamer, W. H. (1906), 'Epidemic disease in England', *The Lancet* **1**, 733–739.
- Hampel, F. R. (1974), 'The influence function and its role in robust estimation', *Journal of the American Statistical Association* **62**, 1179–1186.
- Handcock, M. S. (2003), Assessing degeneracy in statistical models of social networks, Technical report, Center for Statistics and the Social Sciences, University of Washington. Working Paper no. 39.
- Hansen, P. R. and Lunde, A. (2003), An optimal and unbiased measure of realized variance based on intermittent high-frequency data. Unpublished paper, presented at the Realized Volatility Conference, Montreal, 7th November, 2003.
- Hart, J. D. (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, Springer, New York.

- Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- Hasegawa, H. and Kozumi, H. (2003), 'Estimation of Lorenz curves: A Bayesian non-parametric approach', *Journal of Econometrics* **115**, 277–291.
- Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized Additive Models*, Chapman & Hall, London.
- Hastie, T., Tibshirani, R., Eisen, M. B., Alizadeh, A., Levy, R., Staudt, L., Chan, W. C., Botstein, D. and Brown, P. (2000), 'Gene shaving as a method for identifying distinct sets of genes with similar expression patterns', *Genome Biology* **1**.
- Hayes, L. J. and Berry, G. (2002), 'Sampling variability of the Kunst-Mackenbach relative index of inequality', *Journal of Epidemiology and Public Health* **56**, 762–765.
- Head, J., Stansfeld, S. A. and Siegrist, J. (2004), 'The psychosocial work environment and alcohol dependence: a prospective study', *Occupational and Environmental Medicine* **61**, 219–214.
- Heagerty, P. J. and Zeger, S. L. (2000), 'Marginalized multilevel models and likelihood inference (with discussion)', *Statistical Science* **15**, 1–26.
- Heckman, J., Ichimura, H. and Todd, P. (1998), 'Matching as an econometric evaluation estimator', *Review of Economic Studies* **65**, 261–294.
- Heckman, N. E. (1986), 'Spline smoothing in a partly linear model', *Journal of the Royal Statistical Society, series B* **48**, 244–248.
- Heesterbeek, J. A. P. (1992), R_0 . *PhD Thesis*, CWI, Amsterdam.
- Herrbert, J. and Isham, V. (2000), 'Stochastic host-parasite population models', *Journal of Mathematical Biology* **40**, 343–371.
- Hesterberg, T. C. (1996), 'Control variates and importance sampling for efficient bootstrap simulations', *Statistics and Computing* **6**, 147–157.
- Hethcote, H. W. (2000), 'The mathematics of infectious diseases', *SIAM Review* **42**, 599–653.
- Hirth, U. M. (1997), 'Probabilistic number theory, the GEM/Poisson-Dirichlet distribution and the arc-sine law', *Combinatorics, Probability and Computing* **6**, 57–77.
- Hoem, J. M. (1989), The issue of weights in panel surveys of individual behavior, in D. Kasprzyk, G. Duncan, G. Kalton and M. P. Singh, (eds), *Panel Surveys*, New York: Wiley, pp. 539–565.
- Hoffmann, M. and Lepski, O. (2002), 'Random rates in anisotropic regression (with discussion)', *Annals of Statistics* **30**, 325–396.
- Holevo, A. S. (2001a), Lévy processes and continuous quantum measurements, in O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick, (eds), *Lévy Processes: Theory and Applications*, Birkhäuser, Boston, MA, pp. 225–239.
- Holevo, A. S. (2001b), *Statistical Structure of Quantum Theory*, Springer, Heidelberg.
- Holland, P. (1986), 'Statistics and causal inference (with discussion)', *Journal of the American Statistical Association* **81**, 945–970.
- Hortobagyi, G. N., Gutterman, J. U. and Blumenschein, G. R. (1978), Chemotherapy of advanced breast cancer with BCG, in W. D. Terry and D. Windhurst, (eds), *Immunotherapy of Cancer: Present Status of Trials in Man*, Raven Press, New York, pp. 155–168.
- Huang, X. and Tauchen, G. (2003), The relative contribution of jumps to total price variation. Unpublished paper: Department of Economics, Duke University.
- Huber, P. J. (1985), 'Projection pursuit (with discussion)', *Annals of Statistics* **13**, 435–525.

- Ibragimov, I. A. and Has'minskii, R. Z. (1981), *Statistical Estimation: Asymptotical Theory*, Springer, New York.
- Ihaka, R. and Gentleman, R. (1996), 'R: A language for data analysis and graphics', *Journal of Computational and Graphical Statistics* **5**, 299–314.
- Indritz, J. (1963), *Methods in Analysis*, Macmillan, New York.
- Irizarry, R. A., Gautier, L. and Cope, L. M. (2003), An R package for analysis of Affymetrix oligonucleotide arrays, in G. Parmigiani, E. S. Garrett, R. A. Irizarry and S. L. Zeger, (eds), *The Analysis of Gene Expression Data*, Springer.
- Isham, V. (1988), 'Mathematical modelling of the transmission dynamics of HIV and AIDS: A review', *Journal of the Royal Statistical Society, series A* **151**, 5–30. With discussion, pp. 120–123.
- Isham, V. (1991), 'Assessing the variability of stochastic epidemics', *Mathematical Biosciences* **107**, 209–224.
- Isham, V. (1993a), Statistical aspects of chaos: A review, in O. E. Barndorff-Nielsen, D. R. Cox, J. L. Jensen and W. S. Kendall, (eds), *Networks and Chaos*, Chapman & Hall, London.
- Isham, V. (1993b), 'Stochastic models for epidemics with special reference to AIDS', *Advances in Applied Probability* **3**, 1–27.
- Isham, V. (1995), 'Stochastic models of host-macroparasite interaction', *Advances in Applied Probability* **5**, 720–740.
- Isham, V. and Medley, G., (eds) (1996), *Models for Infectious Human Diseases: Their Structure and Relation to Data*, Cambridge University Press, Cambridge.
- Jacod, J. (1994), 'Limit of random measures associated with the increments of a Brownian semimartingale'. Unpublished paper: Laboratoire de Probabilités, Université Pierre and Marie Curie, Paris.
- Jacod, J. and Moral, P. D. (2001), Interacting particle filtering with discrete observations, in A. Doucet, N. de Freitas and N. J. Gordon, (eds), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, pp. 43–77.
- Jacod, J., Moral, P. D. and Protter, P. (2001), 'The Monte-Carlo method for filtering with discrete-time observations', *Probability Theory and Related Fields* **120**, 346–368.
- Jacod, J. and Protter, P. (1998), 'Asymptotic error distributions for the Euler method for stochastic differential equations', *Annals of Probability* **26**, 267–307.
- Jacquez, J. A. and Simon, C. P. (1990), 'AIDS: The epidemiological significance of two mean rates of partner change', *IMA Journal of Mathematics Applied in Medicine and Biology* **7**, 27–32.
- Jacquez, J. A. and Simon, C. P. (1993), 'The stochastic SI model with recruitment and deaths. I. Comparisons with the closed SIS model', *Mathematical Biosciences* **117**, 77–125.
- Jacquez, J. A., Simon, C. P. and Koopman, J. (1989), Structured mixing: Heterogeneous mixing by the definition of activity groups, in C. Castillo-Chavez, (ed.), *Mathematical and statistical approaches to AIDS epidemiology. Lecture notes in Biomathematics*, **83**, Springer, Berlin, pp. 301–315.
- Jeong, H., Mason, S. P., Barabási, A.-L. and Oltvai, Z. N. (2001), 'Lethality and centrality in protein networks', *Nature* **411**, 41.
- Johannes, M., Polson, N. G. and Stroud, J. (2002), Nonlinear filtering of stochastic differential equations with jumps. Unpublished paper: Graduate School of Business, Columbia University.

- Johannes, M., Polson, N. and Stroud, J. (2004), Practical filtering with parameter learning, in A. C. Harvey, S. J. Koopman and N. Shephard, (eds), *State Space and Unobserved Component Models: Theory and Applications. Proceedings of a Conference in Honour of James Durbin*, Cambridge University Press, Cambridge, pp. 236–247.
- Johnson, A. M. (1996), Sources and use of empirical observations to characterise networks of sexual behaviour, in V. Isham and G. Medley, (eds), *Models for Infectious Human Diseases: Their Structure and Relation to Data*, Cambridge University Press, Cambridge, pp. 253–262.
- Jones, J. H. and Handcock, M. S. (2003), ‘An assessment of preferential attachment as a mechanism for human sexual network formation’, *Proceedings of the Royal Society of London* **B270**, 1123–1128.
- Jurek, Z. J. and Mason, J. D. (1993), *Operator-Limit Distributions in Probability Theory*, Wiley, New York.
- Kalman, R. E. (1960), ‘A new approach to linear filtering and prediction problems’, *Journal of Basic Engineering* **82**, 35–45.
- Kao, R. R. (2002), ‘The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK’, *Trends in Microbiology* **10**, 279–286.
- Karatzas, I. and Shreve, S. E. (1991), *Brownian Motion and Stochastic Calculus*, Vol. 113 of *Graduate Texts in Mathematics*, 2nd edn, Springer–Verlag, Berlin.
- Karatzas, I. and Shreve, S. E. (1998), *Methods of Mathematical Finance*, Springer–Verlag, New York.
- Kaufman, J. S., Kaufman, S. and Poole, C. (2003), ‘Causal inference from randomized trials in social epidemiology’, *Social Science and Medicine* **7**, 2397–2409.
- Keeling, M. J. (1999a), ‘Correlation equations for endemic diseases: externally imposed and internally generated heterogeneity’, *Proceedings of the Royal Society of London* **B266**, 953–960.
- Keeling, M. J. (1999b), ‘The effects of local spatial structure on epidemiological invasions’, *Proceedings of the Royal Society of London* **B266**, 859–867.
- Keeling, M. J. (2000), ‘Multiplicative moments and measures of persistence in ecology’, *Journal of Theoretical Biology* **205**, 269–281.
- Keeling, M. J. and Grenfell, B. T. (1997), ‘Disease extinction and community size’, *Science* **275**, 65–67.
- Keeling, M. J. and Grenfell, B. T. (1998), ‘Effect of variability in infection period on the persistence and spatial spread of infectious diseases’, *Mathematical Biosciences* **147**, 207–226.
- Keeling, M. J. and Grenfell, B. T. (2000), ‘Individual-based perspectives on R_0 ’, *Journal of Theoretical Biology* **203**, 51–61.
- Keeling, M. J. and Grenfell, B. T. (2002), ‘Understanding the persistence of measles: reconciling theory, simulation and observation’, *Proceedings of the Royal Society of London* **B269**, 335–343.
- Keeling, M. J., Woolhouse, M. E. J., May, R. M., Davies, G. and Grenfell, B. T. (2003), ‘Modelling vaccination strategies against foot-and-mouth disease’, *Nature* **421**, 136–142.
- Keeling, M. J., Woolhouse, M. E. J., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., Cornell, S. J., Kappey, J., Wilesmith, J. and Grenfell, B. T. (2001), ‘Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape’, *Science* **294**, 813–817.
- Kempthorne, O. (1952), *The Design and Analysis of Experiments*, Wiley, New York.

- Kermack, W. O. and McKendrick, A. G. (1927), 'Contributions to the mathematical theory of epidemics, part I.', *Proceedings of the Royal Society of London* **A115**, 700–721.
- Kerr, M. K. and Churchill, G. A. (2001), 'Experimental design for gene expression microarrays', *Biostatistics* **2**, 183–201.
- Khinchine, A. Y. (1964), *Continued Fractions*, University of Chicago Press, Chicago.
- Kilkenny, D., Reed, M. D., O'Donoghue, D., Kawaler, S., Mukadam, A., Kleinman, S., Nitta, A., Metcalfe, T. S., Provencal, J. L., Watson, T. K., Sullivan, D. J., Sullivan, T., Shobbrook, R., Jiang, X. ., Joshi, S., Ashoka, B. N., Seetha, S., Leibowitz, E., Ibbetson, P., Mendelson, H., Meistas, E., Kalytis, R., Alisauskas, D., P., M., van Wyk, F., Stobie, R. S., Marang, F., Zola, S., Krzesinski, J., Ogloza, W., Moskalik, P., Silvotti, R., Piccioni, A., Vauclair, G., Dolez, N., Chevreton, M., Dreizler, S., Schuh, S. L., Deetjen, J. L., Solheim, J. E., Perez, J. M. G., Ulla, A., Ostensen, R., Manteiga, M., Suarez, O., Burleigh, M., Kepler, S. O., Kanaan, A. and Giovannini, O. (2003), 'A whole earth telescope campaign on the pulsating subdwarf B binary system PG 1336-018 (NY Vir)', *Monthly Notices of the Royal Astronomical Society* **343**, 834–846.
- Kim, S., Shephard, N. and Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies* **65**, 361–393.
- Kingman, J. F. C. (1978), 'The representation of partition structures', *Journal of the London Mathematical Society* **18**, 374–380.
- Kious, B. M. (2001), 'The Nuremberg Code: its history and implications', *Princeton Journal of Bioethics* **4**, 7–19.
- Knott, M. and Bartholomew, D. J. (1999), *Latent Variable Models and Factor Analysis* (2nd edn.), London: Edward Arnold.
- Kohonen, T. (1995), *Self-Organizing Maps*, Springer, Berlin.
- Kong, A., McCullagh, P., Meng, X.-L., Nicolae, D. and Tan, Z. (2003), 'A theory of statistical models for Monte Carlo integration (with discussion)', *Journal of the Royal Statistical Society, series B* **65**, 585–618.
- Koopman, J., Simon, C. P., Jacquez, J. A. and Park, T. S. (1989), Selective contact within structured mixing with an application to HIV transmission risk from oral and anal sex, in C. Castillo-Chavez, (ed.), *Mathematical and statistical approaches to AIDS epidemiology. Lecture notes in Biomathematics*, 83, Springer, Berlin, pp. 316–349.
- Koshevnik, Y. A. and Levit, B. Y. (1976), 'On a nonparametric analogue of the information matrix', *Theory of Probability and its Applications* **21**, 738–753.
- Kraus, K. (1971), 'General state changes in quantum theory', *Annals of Physics* **64**, 331–335.
- Kraus, K. (1983), *States, Effects and Operations: Fundamental Notions of Quantum Theory*, Springer, Heidelberg.
- Kretzschmar, M. (1989a), 'Persistent solutions in a model for parasitic infections', *Journal of Mathematical Biology* **27**, 549–573.
- Kretzschmar, M. (1989b), 'A renewal equation with a birth-death process as a model for parasitic infections', *Journal of Mathematical Biology* **27**, 191–221.
- Kruskal, J. B. (1964a), 'Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis', *Psychometrika* **29**, 1–29.
- Kruskal, J. B. (1964b), 'Non-metric multidimensional scaling: a numerical method', *Psychometrika* **29**, 115–129.

- Kunst, A. E. and Mackenbach, J. P. (1995), *Measuring Socio-economic Inequalities in Health*, Copenhagen: World Health Organisation.
- Kurtz, T. G. (1970), 'Solutions of ordinary differential equations as limits of pure jump Markov processes', *Journal of Applied Probability* **7**, 49–58.
- Kurtz, T. G. (1971), 'Limit theorems for sequences of jump Markov processes approximating ordinary differential processes', *Journal of Applied Probability* **8**, 344–356.
- Kurtz, T. G. (1981), *Approximation of Population Processes*, SIAM, Philadelphia.
- Laio, F., Porporato, A., Ridolfi, L. and Rodriguez-Iturbe, I. (2001a), 'On the mean first passage times of processes driven by white shot noise', *Physical Review, series E* **63**, 36105.
- Laio, F., Porporato, A., Ridolfi, L. and Rodriguez-Iturbe, I. (2001b), 'Plants in water-controlled ecosystems: Active role in hydrological processes and response to water stress. II. Probabilistic soil moisture dynamics', *Advances in Water Resources* **24**, 707–723.
- Law, R. and Dieckmann, U. (2000), Moment approximations of individual-based models, in U. Dieckmann, R. Law J. A. J. Metz, (eds), *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, Cambridge University Press, Cambridge, pp. 252–270.
- Lefèvre, C. (1990), Stochastic epidemic models for S-I-R infectious diseases: a brief survey of the recent general theory, in J.-P. Gabriel, C. Lefèvre and P. Picard, eds, *Stochastic Processes in Epidemic Theory. Lecture Notes in Biomathematics*, vol. 86, Springer, Berlin, pp. 1–12.
- Levit, B. Y. (1975), 'On the efficiency of a class of non-parametric estimates', *Theory of Probability and its Applications* **20**, 723–740.
- Levit, B. Y. (1978), 'Infinite dimensional informational inequalities', *Theory of Probability and its Applications* **23**, 371–377.
- Li, H., Wells, M. and Yu, L. (2004), An MCMC analysis of time-changed Lévy processes of stock returns. Unpublished paper: Johnson Graduate School of Management, Cornell University, Ithaca, U.S.A.
- Li, W.-H. (1997), *Molecular Evolution*, Sinauer, Sunderland, MA.
- Liang, K.-Y., Chiu, Y.-F., Beaty, T. H. and Wjst, M. (2001), 'Multipoint analysis using affected sib pairs: Incorporating linkage evidence from unlinked regions', *Genetic Epidemiology* pp. 158–172.
- Liang, K.-Y. and Zeger, S. L. (1995), 'Inference based on estimating functions in the presence of nuisance parameters (with discussion)', *Statistical Science* **10**, 158–172.
- Lindley, D. V. (2002), 'Seeing and doing: The concept of causation (with discussion)', *International Statistical Review* **70**, 191–214.
- Lindley, D. V. and Smith, A. F. M. (1972), 'Bayes estimates for the linear model (with discussion)', *Journal of the Royal Statistical Society, series B* **34**, 1–41.
- Lipsitz, S. R., Ibrahim, J. and Zhao, L. P. (1999), 'A weighted estimating equation for missing covariate data with properties similar to maximum likelihood', *Journal of the American Statistical Association* **94**, 1147–1160.
- Little, R. J. A. and Rubin, D. B. (2002), *Statistical Analysis with Missing Data* (2nd edn.), New York: Wiley.
- Lloyd, A. L. (2001), 'Destabilisation of epidemic models with the inclusion of realistic distributions of infectious periods', *Proceedings of the Royal Society of London* **B268**, 985–993.
- MacLachlan, G. J. and Peel, D. (2002), *Finite Mixture Models*, Wiley, New York.

- Mancini, C. (2001), Does our favourite index jump or not? Dipartimento di Matematica per le Decisioni, Università di Firenze.
- Mancini, C. (2003), Estimation of the characteristics of jump of a general Poisson-diffusion process. Dipartimento di Matematica per le Decisioni, Università di Firenze.
- Maras, A. M. (2003), 'Adaptive nonparametric locally optimum Bayes detection in additive non-Gaussian noise', *IEEE Transactions on Information Theory* **49**, 204–220.
- Marks, H. M. (2003), 'Rigorous uncertainty: Why R.A. Fisher is important', *International Journal of Epidemiology* **32**, 932–937.
- Masoliver, J. (1987), 'First-passage times for non-Markovian processes: Shot noise', *Physical Review, series A* **35**, 3918–3928.
- Matérn, B. (1986), *Spatial Variation*, Springer, New York.
- McCullagh, P. (2000), 'Invariance and factorial models (with discussion)', *Journal of the Royal Statistical Society, series B* **62**, 209–256.
- McCullagh, P. (2002), 'What is a statistical model? (with discussion)', *Annals of Statistics* **30**, 1225–1310.
- McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, 2nd edn, Chapman & Hall, London.
- McKendrick, A. G. (1926), 'Applications of mathematics to medical problems', *Proceedings of the Edinburgh Mathematical Society* **44**, 98–130.
- Medical Research Council (2001), *Medical Research Council Research Funding Strategy and Priorities, 2001–2004*, Medical Research Council, London.
- Merton, R. C. (1980), 'On estimating the expected return on the market: An exploratory investigation', *Journal of Financial Economics* **8**, 323–361.
- Milton, J. (2003), 'Spies, magicians, and Enid Blyton: How they can help improve clinical trials', *International Journal of Epidemiology* **32**, 943–944.
- Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998), 'Log Gaussian Cox processes', *Scandinavian Journal of Statistics* **25**, 451–482.
- Møller, J. and Waagepetersen, R. (2003), *Statistical Inference and Simulation for Spatial Point Processes*, Chapman & Hall/CRC, London.
- Mollison, D. (1991), 'Dependence of epidemic and population velocities on basic parameters', *Mathematical Biosciences* **107**, 255–287.
- Mollison, D., ed. (1995), *Epidemic Models: their Structure and Relation to Data*, Cambridge University Press, Cambridge.
- Mollison, D., Isham, V. and Grenfell, B. (1994), 'Epidemics: Models and data (with discussion)', *Journal of the Royal Statistical Society, series A* **157**, 115–149.
- Mølmer, K. and Castin, Y. (1996), 'Monte Carlo wavefunctions', *Coherence and Quantum Optics* **7**, 193–202.
- Moore, G. E. (1965), 'Cramming more components onto integrated circuits', *Electronics* **38**(8), 114–117.
- Morris, M. (1991), 'A loglinear modelling framework for selective mixing', *Mathematical Biosciences* **107**, 349–377.
- Morris, M. (1996), Behaviour change and non-homogeneous mixing, in V. Isham and G. Medley, (eds), *Models for Infectious Human Diseases: their Structure and Relation to Data*, Cambridge University Press, Cambridge, pp. 239–252.
- Morris, M. and Kretzschmar, M. (1989), 'Concurrent partnerships and the spread of HIV', *AIDS* **11**, 641–648.
- MRC Streptomycin in Tuberculosis Trials Committee (1948), 'Streptomycin treatment for pulmonary tuberculosis', *British Medical Journal* pp. 769–782.

- Munch, R. (2003), 'Robert Koch', *Microbes and Infection* **5**, 69–74.
- Muraki, N. (2003), 'The five independencies as natural products', *Infinite Dimensional Analysis, Quantum Probability, and Related Topics* **6**, 337–371.
- Murphy, S. A. and van der Vaart, A. W. (2000), 'Semiparametric likelihood ratio inference', *Annals of Statistics* **25**, 1471–1509.
- Nåsell, I. (1985), *Hybrid Models of Tropical Infections*, Lecture Notes in Biomathematics, no. 59, Springer, Berlin.
- Nåsell, I. (1995), The threshold concept in stochastic epidemic and endemic models, in D. Mollison, (ed.), *Epidemic Models: Their Structure and Relation to Data*, Cambridge University Press, pp. 71–83.
- Nåsell, I. (1996), 'The quasi-stationary distribution of the closed endemic SIS model', *Advances in Applied Probability* **28**, 895–932.
- Nåsell, I. (1999), 'On the time to extinction in recurrent epidemics', *Journal of the Royal Statistical Society, series B* **61**, 309–330.
- Nåsell, I. (2002a), Measles outbreaks are not chaotic, in C. Castillo-Chavez, (ed.), *Mathematical Approaches for Emerging and Reemerging Infectious Diseases II: Models, Methods and Theory. IMA Math. Appl. vol 126*, Springer, New York, pp. 85–114.
- Nåsell, I. (2002b), 'Stochastic models of some endemic infections', *Mathematical Biosciences* **179**, 1–19.
- Nåsell, I. (2003), 'Moment closure and the stochastic logistic model', *Theoretical Population Biology* **63**, 159–168.
- National Institutes of Health (2000), 'Adjuvant Therapy for Breast Cancer', *NIH Consensus Statement* **17**, 1–35.
- Nelder, J. A. (1977), 'A reformulation of linear models (with discussion)', *Journal of the Royal Statistical Society, series A* **140**, 48–77.
- Nelder, J. A. and Wedderburn, R. W. M. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society, series A* **135**, 370–384.
- Newey, W. (1990), 'Semiparametric efficiency bounds', *Journal of Applied Econometrics* **5**, 99–135.
- Newman, M. E. J. (2003), 'Properties of highly clustered networks', *Physical Review* **E68**, 026121.
- Newman, M. E. J., Jensen, I. and Ziff, R. M. (2002), 'Percolation and epidemics in a two-dimensional small world', *Physical Review E* **65**, 021904.
- Newman, M. E. J., Watts, D. J. and Strogatz, S. H. (2002), 'Random graph models of social networks', *Proceedings of the National Academy of Science* **99**, 2566–2572.
- Neyman, J. and Scott, E. L. (1948), 'Consistent estimates based on partially consistent observations', *Econometrica* **16**, 1–32.
- Nicolato, E. and Venardos, E. (2003), 'Option pricing in stochastic volatility models of the Ornstein–Uhlenbeck type', *Mathematical Finance* **13**, 445–466.
- Noble, M., Smith, G., Penhale, B., Wright, G., Dibben, C., Owen, T. and Lloyd, M. (2000), *Measuring Multiple Deprivation at the Small Area Level: The Indices of Deprivation 2000*, London: Department of the Environment, Transport and the Regions, Regeneration Series.
- Ogata, Y. (2001), 'Exploratory analysis of earthquake clusters by likelihood-based trigger models', *Journal of Applied Probability* **38A**, 202–212.
- Oh, H. S., Nychka, D. and Brown, T. (2004), 'Period analysis of variable stars by robust smoothing', *Applied Statistics* **53**, 15–30.

- O'Neill, P. D. (2002), 'A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods', *Mathematical Biosciences* **180**, 103–114.
- O'Neill, P. D. and Roberts, G. O. (1999), 'Bayesian inference for partially observed stochastic epidemics', *Journal of the Royal Statistical Society, series A* **162**, 121–129.
- Owen, A. B. (2001), *Empirical Likelihood*, Chapman & Hall/CRC, Boca Raton.
- Pace, L. and Salvani, A. (1997), *Principles of Statistical Inference from a Neo-Fisherian Perspective*, World Scientific, Singapore.
- Patil, G. P. (1963), 'Minimum variance unbiased estimation and certain problems of additive number theory', *Annals of Mathematical Statistics* **34**, 1050–1056.
- Pattison, P. and Robins, G. (2002), 'Neighbourhood-based models of social networks', *Sociological Methodology* **32**, 301–337.
- Pearl, J. (2000), *Causality: Models, Reasoning and Inference*, Cambridge University Press, Cambridge.
- Pearl, J. (2002), 'Comments on "Seeing and doing: The concept of causation" by D. V. Lindley', *International Statistical Review* **70**, 207–214.
- Pearl, J. (2003), 'Statistics and causal inference: A review', *Test* **12**, 281–345.
- Pedersen, J. (2003), The Lévy–Itô decomposition of independently scattered random measures. Research Report 2003-2, MaPhySto, University of Aarhus.
- Peto, R. (1996), 'Five years of tamoxifen—or more?', *Journal of the National Cancer Institute* **88**, 1791–1793.
- Peto, R., Boreham, J., Clarke, M., Davies, C. and Beral, V. (2000), 'UK and USA breast cancer deaths down 25% at ages 20–69 years', *Lancet* **355**, 1822.
- Pfanzagl, J. and Wefelmeyer, W. (1982), *Contributions to a General Asymptotic Statistical Theory*, Vol. 13 of *Lecture Notes in Statistics*, Springer, New York.
- Pfeffermann, D. (1993), 'The role of sampling weights when modeling survey data', *International Statistical Review* **61**, 317–337.
- Pierce, D. A. and Bellio, R. (2004), The effect of choice of reference set on frequency inferences. Preprint.
- Pierce, D. A. and Peters, D. (1992), 'Practical use of higher order asymptotics for multiparameter exponential families (with discussion)', *Journal of the Royal Statistical Society, series B* **54**, 701–738.
- Pinheiro, J. C. and Bates, D. M. (2000), *Mixed-Effects Models in S and S-PLUS*, Springer-Verlag, New York.
- Pison, G., Rousseeuw, P. J., Filzmoser, P. and Croux, C. (2003), 'Robust factor analysis', *Journal of Multivariate Analysis* **84**, 145–172.
- Pitman, E. J. G. (1937), 'Significance tests which may be applied to samples from any population', *Journal of the Royal Statistical Society, Supplement* **4**, 119–130.
- Pitt, M. K. and Shephard, N. (1999), 'Filtering via simulation: auxiliary particle filter', *Journal of the American Statistical Association* **94**, 590–599.
- Pitt, M. K. and Shephard, N. (2001), Auxiliary variable based particle filters, in N. de Freitas, A. Doucet and N. J. Gordon, (eds), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, pp. 273–293.
- Poole, D. and Raftery, A. E. (2000), 'Inference from deterministic simulation models: The Bayesian melding approach', *Journal of the American Statistical Association* **95**, 1244–1255.
- Porporato, A., Daly, E. and Rodriguez-Iturbe, I. (2004), 'Soil water balance and ecosystem response to climate change', *American Naturalist* **164**, 625–633.

- Porporato, A., Laio, F., Ridolfi, L., Caylor, K. K. and Rodriguez-Iturbe, I. (2003), 'Soil moisture and plant stress dynamics along the Kalahari precipitation gradient', *Journal of Geophysical Research* **108**, 4127–4134.
- Porporato, A., Laio, F., Ridolfi, L. and Rodriguez-Iturbe, I. (2001), 'Plants in water-controlled ecosystems: Active role in hydrological processes and response to water stress. III. Vegetation water stress', *Advances in Water Resources* **24**, 725–744.
- Porporato, A. and Rodriguez-Iturbe, I. (2002), 'Ecohydrology—a challenging multidisciplinary research perspective', *Hydrological Sciences Journal* **47**, 811–822.
- Poterba, J. and Summers, L. (1986), 'The persistence of volatility and stock market fluctuations', *American Economic Review* **76**, 1124–1141.
- Preston, C. J. (1973), 'The generalised Gibbs states and Markov random fields', *Advances in Applied Probability* **5**, 242–261.
- Preston, D. L., Shimizu, Y., Pierce, D. A., Suyama, A. and Mabuchi, K. (2003), 'Studies of mortality of atomic bomb survivors. Report 13: Solid cancer and noncancer disease mortality 1950–1997', *Radiation Research* **160**, 381–407.
- Pritchard, K. I., Paterson, A. H., Fine, S., Paul, N. A., Zee, B., Shepherd, L. E., Abuzahra, H., Ragaz, J., Knowling, M., Levine, M. N., Verma, S., Perrault, D., Walde, P. L., Bramwell, V. H., Poljicak, M., Boyd, N., Warr, D., Norris, B. D., Bowman, D., Armitage, G. R., Weizel, H. and Buckman, R. A. (1997), 'Randomized trial of cyclophosphamide, methotrexate, and fluorouracil chemotherapy added to tamoxifen as adjuvant therapy in postmenopausal women with node-positive estrogen and/or progesterone receptor-positive breast cancer: A report of the National Cancer Institute of Canada Clinical Trials Group. Breast Cancer Site Group', *Journal of Clinical Oncology* **15**, 2302–2311.
- Protter, P. (1992), *Stochastic Integration and Differential Equations*, 2nd edn, Springer-Verlag, New York.
- Quinlan, J. R. (1993), *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA.
- Quinn, B. G. (1999), 'A fast efficient technique for the estimation of frequency: interpretation and generalisation', *Biometrika* **86**, 213–220.
- Quinn, B. G. and Hannan, E. J. (2001), *The Estimation and Tracking of Frequency*, Cambridge University Press, Cambridge.
- Quinn, B. G. and Thompson, P. J. (1991), 'Estimating the frequency of a periodic function', *Biometrika* **78**, 65–74.
- Quinn, M., Babb, P., Brock, A., Kirby, E. and Jones, J. (2001), Cancer Trends in England and Wales, in *Studies on Medical and Population Subjects No. 66*, The Stationery Office, London, pp. 1950–1999.
- Raftery, A. E. (2000), 'Statistics in sociology, 1950–2000', *Journal of the American Statistical Association* **95**, 654–661.
- Raftery, A. E., Tanner, M. A. and Wells, M. T. (2001), *Statistics in the 21st Century*, Chapman & Hall, London.
- Ramsay, J. O. and Silverman, B. W. (1997), *Functional Data Analysis*, Springer, New York.
- Ramsay, J. O. and Silverman, B. W. (2002), *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York.
- Rao, J. N. K. (2003), *Small Area Estimation*, New York: Wiley.
- Read, J. M. and Keeling, M. J. (2003), 'Disease evolution on networks: the role of contact structure', *Proceedings of the Royal Society of London, Series B* **B270**, 699–708.

- Reid, N. (1994), 'A conversation with Sir David Cox', *Statistical Science* **9**, 439–455.
- Reid, N. (2003), 'Asymptotics and the theory of inference', *Annals of Statistics* **31**, 1695–1731.
- Reid, N., Mukerjee, R. and Fraser, D. A. S. (2003), Some aspects of matching priors, in M. Moore, S. Froda and C. Léger, (eds), *Mathematical Statistics and Applications: Festschrift for C. van Eeden*, Vol. 42 of *Lecture Notes – Monograph Series*, Institute of Mathematical Statistics, Hayward, CA, pp. 31–44.
- Renshaw, E. (1991), *Modelling Biological Populations in Space and Time*, Cambridge University Press, Cambridge.
- Rényi, A. (1963), 'On stable sequences of events', *Sankhyā, Series A* **25**, 293–302.
- Rice, J. A. (1986), 'Convergence rates for partially splined estimates', *Statistics and Probability Letters* **4**, 203–208.
- Rice, J. A. and Silverman, B. W. (1991), 'Estimating the mean and covariance structure nonparametrically when the data are curves', *Journal of the Royal Statistical Society, series B* **53**, 233–243.
- Richardson, M. P. and Stock, J. H. (1989), 'Drawing inferences from statistics based on multi-year asset returns', *Journal of Financial Economics* **25**, 323–48.
- Riley, S., Fraser, C., Donnelly, C. A., Ghani, A. C., Abu-Raddad, L. J., Hedley, A. J., Leung, G. M., Ho, L.-M., Lam, T.-H., Thach, T. Q., Chau, P., Chan, K.-P., Lo, S.-V., Leung, P.-Y., Tsang, T., Ho, W., Lee, K.-H., Lau, E. M. C., Ferguson, N. M. and Anderson, R. M. (2003), 'Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions', *Science* **300**, 1961–66.
- Ripley, B. D. (1993), Statistical aspects of neural networks, in O. E. Barndorff-Nielsen, J. L. Jensen and W. S. Kendall, (eds), *Networks and Chaos—Statistical and Probabilistic Aspects*, Chapman & Hall, London, pp. 40–123.
- Ripley, B. D. (1994), 'Neural networks and related methods for classification (with discussion)', *Journal of the Royal Statistical Society, series B* **56**, 409–456.
- Ripley, B. D. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge.
- Ritov, Y. and Bickel, P. J. (1990), 'Achieving information bounds in non- and semi-parametric models', *Annals of Statistics* **18**, 925–938.
- Ritter, H., Martinetz, T. and Schulten, K. (1992), *Neural Computation and Self-Organizing Maps. An Introduction*, Addison-Wesley, Reading, MA.
- Robins, J. M. (2004), Optimal structural nested models for optimal decisions, in D. Y. Lin and P. Heagerty, (eds), *Proceedings of the 2nd Seattle Symposium on Survival Analysis*, Springer, New York.
- Robins, J. M., Mark, S. D. and Newey, W. K. (1992), 'Estimating exposure effects by modelling the expectation of exposure conditional on confounders', *Biometrics* **48**, 479–495.
- Robins, J. M. and Ritov, Y. (1997), 'Toward a course of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models', *Statistics in Medicine* **26**, 285–319.
- Robins, J. M. and Rotnitzky, A. (1992), Recovery of information and adjustment for dependent censoring using surrogate markers, in N. Jewell, K. Dietz and V. Farewell, (eds), *AIDS Epidemiology-Methodological Issues*, V. Farewell, Boston, USA, pp. 297–331.
- Robins, J. M. and Rotnitzky, A. (2001), 'Discussion of: "Inference in semiparametric models: some questions and one answer"', *Statistica Sinica* **11**, 863–960.

- Robins, J. M., Rotnitzky, A. and van der Laan, M. (2000), ‘Discussion of: “On profile likelihood”’, *Journal of the American Statistical Association* **95**, 431–435.
- Robins, J. M. and van der Vaart, A. W. (2004), A unified approach to estimation in non-semiparametric models using higher order influence functions, Technical report, Department of Epidemiology, Harvard School of Public Health.
- Robins, J. M. and Wasserman, L. (2000), ‘Conditioning, likelihood, and coherence: A review of some foundational concepts’, *Journal of the American Statistical Association* **95**, 1340–1346.
- Robinson, G. K. (1991), ‘That BLUP is a good thing: The estimation of random effects (with discussion)’, *Statistical Science* **6**, 15–51.
- Robinson, P. (1988), ‘Root- N -consistent semiparametric regression’, *Econometrica* **56**, 931–954.
- Rodriguez-Iturbe, I. (2000), ‘Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamics’, *Water Resources Research* **36**, 3–9.
- Rodriguez-Iturbe, I., Cox, D. R. and Isham, V. S. (1987), ‘Some models for rainfall based on stochastic point processes’, *Proceedings of the Royal Society of London, series A* **410**, 269–288.
- Rodriguez-Iturbe, I., Cox, D. R. and Isham, V. S. (1988), ‘A point process model for rainfall: Further developments’, *Proceedings of the Royal Society of London, series A* **417**, 283–298.
- Rodriguez-Iturbe, I. and Porporato, A. (2005), *Ecohydrology of Water-Controlled Ecosystems*, Cambridge University Press, Cambridge.
- Rodriguez-Iturbe, I., Porporato, A., Ridolfi, L., Isham, V. and Cox, D. R. (1999), ‘Probabilistic modelling of water balance at a point: The role of climate, soil and vegetation’, *Proceedings of the Royal Society of London, series A* **455**, 3789–3805.
- Rosenbaum, P. R. (2002), *Observational Studies* (2nd edn.), New York: Springer.
- Rosenbaum, P. R. and Rubin, D. B. (1983), ‘The central role of the propensity score in observational studies for causal effects’, *Biometrika* **70**, 41–55.
- Rosenblatt, M. (1952), ‘Remarks on a multivariate transformation’, *Annals of Mathematical Statistics* **23**, 470–2.
- Ross, R. (1911), *The Prevention of Malaria*, 2nd edn, Murray, London.
- Ross, R. (1916), ‘An application of the theory of probabilities to the study of *a priori* pathometry, I’, *Proceedings of the Royal Society of London* **A92**, 204–30.
- Ross, R. and Hudson, H. P. (1917a), ‘An application of the theory of probabilities to the study of *a priori* pathometry, II’, *Proceedings of the Royal Society of London* **A93**, 212–25.
- Ross, R. and Hudson, H. P. (1917b), ‘An application of the theory of probabilities to the study of *a priori* pathometry, III’, *Proceedings of the Royal Society of London* **A93**, 225–40.
- Roth, K. F. (1955), ‘Rational approximations to algebraic numbers’, *Mathematika* **2**, 1–20, (corrigendum **2**, 168).
- Rothman, K. J. and Greenland, S. (1998), *Modern Epidemiology*, 2nd edn, Lippincott Williams and Wilkins, Philadelphia.
- Royal Statistical Society (1988), ‘Royal Statistical Society Meeting on AIDS’, *Journal of the Royal Statistical Society, series A* **151**, 1–136.
- Royall, R. (1997), *Statistical Evidence: A Likelihood Paradigm*, Chapman & Hall, London.

- Rubin, D. B. (1974), 'Estimating causal effects of treatments in randomized and non-randomized studies', *Journal of Educational Psychology* **66**, 688–701.
- Rubin, D. B. (1976), 'Inference and missing data', *Biometrika* **63**, 581–590.
- Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.
- Rubin, D. B. (1996), 'Multiple imputation after 18+ years', *Journal of the American Statistical Association* **91**, 473–489.
- Rubin, D. B. (2001), 'Estimating the causal effects of smoking', *Statistics in Medicine* **20**, 1395–1414.
- Rubin, D. B. (2002), 'The ethics of consulting for the tobacco industry', *Statistical Methods in Medical Research* **11**, 373–380.
- Ruczinski, I., Kooperberg, C. and LeBlanc, M. (to appear), 'Logic regression', *Journal of Computational and Graphical Statistics* **11**.
- Ruud, P. A. (1983), 'Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models', *Econometrica* **51**, 225–228.
- Ruud, P. A. (1986), 'Consistent estimation of limited dependent variable models despite misspecification of distribution', *Econometrica* **32**, 157–187.
- Rvachev, L. A. and Longini, I. M. (1985), 'A mathematical model for the global spread of influenza', *Mathematical Biosciences* **75**, 3–23.
- Sasieni, P. (2003), 'Evaluation of the UK breast screening programmes', *Annals of Oncology* **14**, 1206–1208.
- Sato, K. (1999), *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge.
- Scaccia, L. and Green, P. J. (2003), 'Bayesian growth curves using normal mixtures with nonparametric weights', *Journal of Computational and Graphical Statistics* **12**, 308–331.
- Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, London: Chapman & Hall/CRC.
- Schaffer, W. M. (1985), 'Can nonlinear dynamics elucidate mechanisms in ecology and epidemiology?', *IMA Journal of Mathematics Applied in Medicine and Biology* **2**, 221–252.
- Scharfstein, D. O. and Robins, J. M. (2002), 'Estimation of the failure time distribution in the presence of informative censoring', *Biometrika* **89**, 617–634.
- Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999), 'Adjusting for non-ignorable drop-out using semi-parametric non-response models', *Journal of the American Statistical Association* **94**, 1096–1146.
- Schenzle, D. (1984), 'An age-structured model of the pre- and post-vaccination measles transmission', *IMA Journal of Mathematics Applied in Medicine and Biology* **1**, 169–191.
- Schervish, M. J. (1995), *Theory of Statistics*, Springer, New York.
- Schwert, G. W. (1989), 'Why does stock market volatility change over time?', *Journal of Finance* **44**, 1115–1153.
- Schwert, G. W. (1990), 'Stock market volatility', *Financial Analysts Journal* **46**, 23–34.
- Sellek, T. (1983), 'On the asymptotic distribution of the size of a stochastic epidemic', *Journal of Applied Probability* **20**, 390–394.
- Sen, A. and Srivastava, M. (1990), *Regression Analysis: Theory, Methods, and Applications*, Springer, New York.
- Severini, T. A. (2000a), *Likelihood Methods in Statistics*, Clarendon Press, Oxford.

- Severini, T. A. (2000b), 'The likelihood ratio approximation to the conditional distribution of the maximum likelihood estimator in the discrete case', *Biometrika* **87**, 939–945.
- Shephard, N. (1994), 'Partial non-Gaussian state space', *Biometrika* **81**, 115–131.
- Shephard, N., (ed.) (2004), *Stochastic Volatility: Selected Readings*, Oxford University Press, Oxford.
- Siegel, C. L. (1942), 'Iterations of analytic functions', *Annals of Mathematics* **43**, 607–612.
- Silverman, B. W. (1981), 'Using kernel density estimates to investigate multimodality', *Journal of the Royal Statistical Society, series B* **43**, 97–99.
- Simon, J. (1969), *Basic Research Methods in Social Science: The Art of Experimentation*, Random House, New York.
- Simon, J. (1993), *Resampling: The New Statistics*, preliminary edn, Duxbury, Belmont, CA.
- Simpson, J. A. and Weiner, E. S. C., (eds) (1989), *Oxford English Dictionary*, Oxford University Press, London.
- Singpurwalla, N. D. (2002), 'On causality and causal mechanisms', *International Statistical Review* **70**, 198–206.
- Skovgaard, I. M. (1996), 'An explicit large-deviation approximation to one-parameter tests', *Bernoulli* **2**, 145–166.
- Skovgaard, I. M. (2001), 'Likelihood asymptotics', *Scandinavian Journal of Statistics* **28**, 3–32.
- Small, C. G. and McLeish, B. G. (1994), *Hilbert Space Methods in Probability and Statistical Inference*, Wiley, New York.
- Smith, J. Q. (1985), 'Diagnostic checks of non-standard time series models', *Journal of Forecasting* **4**, 283–91.
- Snijders, T. A. B. (2001), 'The statistical evaluation of social network dynamics', *Sociological Methodology* **31**, 361–395.
- Snijders, T. A. B. (2002), 'Markov chain Monte Carlo estimation of exponential random graph models', *Journal of Social Structure*, **3**, part 2.
- Snijders, T. A. B. and Bosker, R. J. (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*, London: Sage.
- Sobel, M. (2000), 'Causal inference in the social sciences', *Journal of the American Statistical Association* **95**, 647–651.
- Soper, H. E. (1929), 'Interpretation of periodicity in disease prevalence', *Journal of the Royal Statistical Society* **92**, 34–73.
- Spence, R. (2001), *Information Visualization*, Addison-Wesley, Harlow.
- Spiegelhalter, D. J. (2002), 'Funnel plots for institutional comparison', *Quality and Safety in Health Care* **11**, 390–391.
- Spielman, R. S., McGinnis, R. E. and Ewens, W. J. (1993), 'Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)', *American Journal of Human Genetics* **52**, 506–516.
- Stark, J., Iannelli, P. and Baigent, S. (2001), 'A nonlinear dynamics perspective of moment closure for stochastic processes', *Nonlinear Analysis* **47**, 753–764.
- Stein, C. (1956), Efficient nonparametric testing and estimation, in 'Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability', Vol. 1, University of California Press, Berkeley, pp. 187–195.
- Stein, M. L. (1999), *Interpolation of Spatial Data*, Springer, New York.

- Stewart, L. A. and Parmar, M. K. B. (1993), 'Meta-analysis of the literature or of individual patient data: Is there a difference?', *Lancet* **341**, 418–220.
- Stone, C. (1975), 'Adaptive maximum likelihood estimation of a location parameter', *Annals of Statistics* **3**, 267–284.
- Storvik, G. (2002), 'Particle filters in state space models with the presence of unknown static parameters', *IEEE Transactions on Signal Processing*, **50**, 281–289.
- Tardella, L. (2002), 'A new Bayesian method for nonparametric capture-recapture models in presence of heterogeneity', *Biometrika* **89**, 807–817.
- Taylor, S. J. and Xu, X. (1997), 'The incremental volatility information in one million foreign exchange quotations', *Journal of Empirical Finance* **4**, 317–340.
- Thorin, O. (1977), 'On infinite divisibility of the lognormal distribution', *Scandinavian Actuarial Journal* pp. 141–149.
- Thorin, O. (1978), 'An extension of the notion of a generalized gamma convolution', *Scandinavian Actuarial Journal* pp. 121–148.
- Tibshirani, R. J. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society, series B* **58**, 267–288.
- Tikhonov, A. N. (1963), On the solution of incorrectly put problems and the regularisation method, in 'Outlines of the Joint Symposium on Partial Differential Equations (Novosibirsk, 1963)', Academy of Science, USSR, Siberian Branch, Moscow, pp. 261–265.
- Titterton, D. M. and Cox, D. R., (eds) (2001), *Biometrika: 100 Years*, Oxford University Press, Oxford.
- Titterton, D. M., Smith, A. F. M. and Makov, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York.
- Troynikov, V. S. (1999), 'Probability density functions useful for parametrization of heterogeneity in growth and allometry data', *Bulletin of Mathematical Biology* **60**, 1099–1122.
- Tukey, J. W. (1974), 'Named and faceless values: An initial exploration in memory of Prasanta C. Mahalanobis', *Sankhyā, series A* **36**, 125–176.
- Tukey, J. W. (1977), *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.
- van de Vijver, M. J., He, Y. D., van 't Veer, L. J., Dai, H., Hart A. A. M., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H. and Bernards, R. (2002), 'A gene-expression signature as a predictor of survival in breast cancer', *New England Journal of Medicine* **347**, 1999–2009.
- van der Geer, S. (2000), *Empirical Processes in M-Estimation*, Cambridge University Press, Cambridge.
- van der Laan, M. J. and Robins, J. M. (2003), *Unified Methods for Censored Longitudinal Data and Causality*, Springer, New York.
- van der Vaart, A. W. (1988), *Statistical Estimation in Large Parameter Spaces*, CWI Tracts 44, Centrum voor Wiskunde en Informatica, Amsterdam.
- van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press.
- van der Vaart, A. W. (2000), Semiparametric statistics, in P. Bernard, (ed.), *Lectures on Probability Theory and Statistics*, Springer, Berlin.
- Vardi, Y. (1985), 'Empirical distributions in selection bias models', *Annals of Statistics* **13**, 178–203.

- Venables, W. N. and Ripley, B. D. (1994), *Modern Applied Statistics with S-Plus*, Springer-Verlag, New York.
- Venables, W. N. and Ripley, B. D. (2002), *Modern Applied Statistics with S*, 4th edn, Springer-Verlag, New York.
- Versluis, M., Schmitz, B., von der Heydt, A. and Lohse, D. (2000), 'How snapping shrimp snap: Through cavitating bubbles', *Science* **289**, 2114–2117.
- Voiculescu, D. (2000), Lectures on free probability, in *Lectures on Probability Theory and Statistics*, Springer, Heidelberg, pp. 283–349. Ecole d'Été de Saint Flour XXVIII.
- von Mering, C., Zdobnov, E. M. and Tsoka, S. (2003), 'Genome evolution reveals biochemical networks and functional modules', *Proceedings of the National Academy of Sciences of the USA* **100**, 15428–15433.
- Wahba, G. (1985), 'A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem', *Annals of Statistics* **13**, 1378–1402.
- Wahba, G. (1990), *Spline Models for Observational Data*, SIAM, Philadelphia, PA.
- Wakefield, J. (2004), 'Ecological inference for 2×2 tables (with discussion)', *Journal of the Royal Statistical Society, series A* **167**, 385–425.
- Waterman, R. P. and Lindsay, R. G. (1996), 'Projected score methods for approximating conditional scores', *Biometrika* **83**, 1–14.
- Watson, J. D. and Crick, F. H. C. (1953), 'Molecular structure of nucleic acids', *Nature* **171**, 737–738.
- Watts, D. J. and Strogatz, S. H. (1998), 'Collective dynamics of "small-world" networks', *Nature* **393**, 440–442.
- Welch, B. L. and Peers, H. W. (1963), 'On formulae for confidence points based on integrals of weighted likelihoods', *Journal of the Royal Statistical Society, series B* **25**, 318–329.
- West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, Springer, New York.
- Whittle, P. (1955), 'The outcome of a stochastic epidemic—a note on Bailey's paper', *Biometrika* **42**, 116–122.
- Whittle, P. (1957), 'On the use of the normal approximation in the treatment of stochastic processes', *Journal of the Royal Statistical Society, series B* **19**, 268–281.
- Wilkinson, L. (1999), *The Grammar of Graphics*, Springer-Verlag, New York.
- Wilson, E. B. and Burke, M. H. (1942), 'The epidemic curve', *Proceedings of the National Academy of Science* **28**, 361–367.
- Wilson, E. B. and Burke, M. H. (1943), 'The epidemic curve, II', *Proceedings of the National Academy of Science* **29**, 43–48.
- Winkel, M. (2001), The recovery problem for time-changed Lévy processes. Research Report 2001-37, MaPhySto, University of Aarhus.
- Witten, I. H. and Franke, E. (2000), *Data Mining. Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, San Francisco.
- Wjst, M., Fischer, G., Immervoll, T., Jung, M., Saar, K., Rueschendorf, F., Reis, A., Ulbrecht, M., Gomolka, M., Weiss, E. H., Jaeger, L., Nickel, R., Richter, K., Kjellman, N. I. M., Griese, M., von Berg, A., Gappa, M., Riedel, F., Boehle, M., van Koningsbruggen, S., Schoberth, P., Szczepanski, R., Dorsch, W., Silbermann, M., Loesgen, S., Scholz, M., Bickeboller, H. and Wichmann, H. E. (1999), 'A genome-wide search for linkage to asthma. German Asthma Genetics Group', *Genomics* **58**, 1–8.

- Wood, S. N., Kohn, R., Shively, T. and Jiang, W. (2002), 'Model selection in spline non-parametric regression', *Journal of the Royal Statistical Society, series B* **64**, 119–139.
- Woodroffe, M. (1970), 'On choosing a delta-sequence', *Annals of Mathematical Statistics* **41**, 1665–1671.
- Wulfsohn, M. S. and Tsiatis, A. A. (1997), 'A joint model for survival and longitudinal data measured with error', *Biometrics* **53**, 330–339.
- Xie, Y. (1992), 'The log-multiplicative layer effect model for comparing mobility tables', *American Sociological Review* **57**, 380–95.
- Yang, Y. H. and Speed, T. (2002), 'Design issues for cDNA microarray experiments', *Nature Reviews: Genetics* **3**, 579–588.
- Yau, P., Kohn, R. and Wood, S. N. (2003), 'Bayesian variable selection and model averaging in high-dimensional multinomial nonparametric regression', *Journal of Computational and Graphical Statistics* **12**, 23–54.
- Yu, J. (2004), 'On leverage in a stochastic volatility model', *Journal of Econometrics*. Forthcoming.
- Zeger, S. L., Liang, K.-Y. and Albert, P. S. (1988), 'Models for longitudinal data: a generalized estimating equation approach', *Biometrics* **44**, 1049–1060.
- Zerhouni, E. (2003), 'The U.S. National Institutes of Health Roadmap', *Science* **302** (5642), 63–72.
- Zhang, L., Mykland, P. and Aït-Sahalia, Y. (2003), A tale of two time scales: determining integrated volatility with noisy high-frequency data. Unpublished paper presented at the Realized Volatility Conference, Montreal, 8th November 2003.
- Zhuang, J., Ogata, Y. and Vere-Jones, D. (2002), 'Stochastic declustering of space-time earthquake occurrences', *Journal of the American Statistical Association* **97**, 369–380.