Biometrika (2013), **100**, 1, *pp.* 17–73
doi: 10.1093/biomet/ass076

Printed in Great Britain

Biometrika highlights from volume 28 onwards

By D. M. TITTERINGTON

School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QQ, U.K.
michael.titterington@gla.ac.uk

Summary

Highlights, trends and influences are identified associated with the pages of *Biometrika* subsequent to the editorship of Karl Pearson.

Some key words: *Biometrika*; General statistical methodology; History of statistics.

1. Introduction

Volume 88 of *Biometrika* celebrated the centenary of the founding of the Journal by presenting a number of topic-based review articles containing a wealth of detail about the Journal's material. The papers were also reprinted in Titterington & Cox (2001) together with ten highlight papers dating from 1939 to 1971. As the Journal reaches another milestone, the publication of volume 100, this paper and Aldrich (2013) offer a further retrospective look at *Biometrika* and its influence.

To provide a different perspective from what was done in 2001, this paper takes a chronological rather than a topic-based approach, attempting to identify the most influential material as published through the decades, starting from the change of editorship from Karl Pearson to E. S. Pearson in 1936 and thereby complementing the period covered in Aldrich (2013). Even more so than in the 2001 papers, because of space constraints the coverage has to be selective. The choice of material is also somewhat personal, but the first level of selection has been mechanical in using Google Scholar to identify the Journal's 100 most highly cited papers published from volume 28 onwards, as listed on 14 September, 2012. Of these, 45 have been cited more than 1000 times and all have at least 487 citations. Altogether, over 650 *Biometrika* papers have been cited at least 100 times. In fact, a list of the overall top 100 papers would contain only four papers published earlier than 1936. The selection rule might also be criticized as favouring early papers that have had a longer period over which to earn citations. On the other hand, the increasing number of journals in existence provides more scope for the citation of recent papers, especially those that present methodology or ideas that influence fields outside statistics. In fact, the top ten papers all date from the period 1965–1995. Papers more recent than that arguably have not had a chance to make such a substantial impact. Of course, the papers in the top 100 fluctuate as time goes by but there is certainly short-term stability; between the preparation of two versions of this paper in April and September, 2012, there was no change in the identities of the papers included although minor changes in the ordering. In any case it may be of interest to readers in future years to observe the changes that will have occurred during the intervening period.

The overall period is considered decade by decade, all sections except for that corresponding to the most recent decade describing the relevant papers that appear in the top 100, followed by an account of other influential papers on the topics that were prominent in the Journal at that time;

although those papers have not reached the top 100, they have all been well cited. The chronology in § § 2–9 is followed by short sections intended to bring out the essence of the ten most highly cited papers.

The reference list first provides the top 100 papers, in chronological order but prefixed by their position in order of number of citations, followed by all other cited papers in the usual alphabetical order.

2. The nineteen thirties and forties

2·1. *Papers in the top* 100

In these years the flavour of *Biometrika* began to change with the transfer of the editorship from Karl to Egon Pearson. A good number of papers still appeared based on anthropometric measurements of items such as skulls and mandibles, many, such as Morant & Samson (1936), Von Bonin & Morant (1939) and Morant (1948), under the authorship of G. M. Morant, but their frequency declined during the period. Instead the Journal's pages became dominated by methodological advances of various sorts. The Second World War also had an effect, with publication becoming temporarily sporadic. However, the period saw three particularly influential papers in Hotelling (1936, placed 16 in terms of citations), Leslie (1945, 19) and Plackett & Burman (1946, 17).

Of these, Hotelling (1936, 16) established the methodology and underlying theory for canonical correlation analysis of the relationships between two sets of variables, including many results about matrices and determinants, the asymptotic standard errors, variances and covariances for the estimators of canonical correlations under certain conditions including multivariate normality of the set of variates, and much more besides. Subsequent work on canonical correlations was reported by Bartlett (1941), on tests of significance, and Hsu (1941), who extended the asymptotic distribution theory for the canonical correlation estimators. Perhaps less familiar in modern statistics is the type of material covered in Leslie (1945, 19), who examined matrices that arise from sets of equations representing the evolution of populations over time and which are instrumental in the calculation of stable age distributions. The material was followed up in Leslie (1948, 58), which included an extension to scenarios involving two populations, one of predators and one of prey. Plackett & Burman (1946, 17) used Hadamard matrices to develop unblocked designs for multifactorial experiments that are of minimum size subject to all main effects being estimable; more on multifactorial designs is available in Plackett (1946).

Kendall (1938, 36) introduced his rank correlation coefficient, usually denoted by τ, for use in comparing the ranking of items by two observers or the ranking by one observer with an established correct ranking, and there were several well-cited follow-up papers by Kendall himself and others: Kendall (1942) defined a partial τ that is 'a measure of the association of agreements [of two rankings] when compared in pairs with [a third ranking]'; Kendall (1948) and Sillito (1947) provided modifications to handle the case of tied ranks; Kendall (1949b) obtained expressions for expectations and variances of τ and Spearman's ρ; Daniels (1944) studied a version of τ based on a more general set of scores; and Moran (1948a) obtained the finite-sample mean and variance of Spearman's ρ when the data are normally distributed.

Advances in other areas also led to prominent publications. The Behrens–Fisher problem was investigated by B. L. Welch: Welch (1938, 59) examined the extent to which the usual two-sample t-test is valid when the two variances are unequal and evaluated the efficacy of using the t-distribution with a modified t-statistic, and Welch (1947, 38) considered a more general scenario in which the variance of the numerator of the test statistic of interest is a known linear combination of k variances. Johnson (1949, 45) proposed the construction of systems of frequency curves for a random variable Y based on the assumption that a linear function of a specified function of Y has a standard normal distribution. Box (1949, 63) provided detailed analysis of the properties of certain loglikelihood functions that lead to generalizations of Bartlett's test, including details of chi-squared and F approximations. Anscombe (1948, 64) studied transformations, such as the square-root transform for Poisson data, that improve the approximate normality of certain discrete distributions, and Patnaik (1949, 77) derived approximations to the cumulative distribution functions of noncentral χ^2 and F distributions, for use in power calculations in hypothesis-testing scenarios.

2·2. *Other influential contributions*

Further important work on design of experiments included Kempthorne (1947), who used group-theoretical methods to show the equivalence of fractional replication and confounding for certain factorial designs. There was also a debate, about the relative merits of randomized and systematic design, carried on by Welch (1937), Pearson (1937), Pearson (1938b), Pitman (1938), 'Student' (1938), Neyman & Pearson (1938), Jeffreys (1939) and Yates (1939); for a concise account of this debate see Atkinson & Bailey (2001, § 4).

Other prominent topics in this period were treatment of the 2×2 contingency table and foundational issues. Barnard (1947a) provided a detailed investigation of the 2×2 contingency table and the three models distinguished by whether both or just one or neither of the marginal totals is fixed. This was followed up immediately in the Journal by Pearson (1947), who illustrated the resulting variety of probabilistic structures, and by an associated technical note (Barnard, 1947b). Barnard (1947c) presented a related but more general discussion of the meaning of a significance level and the way in which it might be affected by accidental influences of the 'outside world'. Further work on the 2×2 table appeared in Haldane (1945b), concerning the null-hypothesis moments of the test statistic for the chi-squared test and of its square root, which corresponds to a product moment correlation coefficient, in Finney (1948), who provided a table of significance levels for the Fisher–Yates exact test when all cell counts are small, and in Patnaik (1948), concerning the power function when the table represents data from two binomial distributions being compared. Pearson & Merrington (1948) provided further power-related tables. Yates (1948) considered more general two-way tables with categorization based on variables that are numerical or at least ordinal; scores are assigned to the row and column categories and a regression approach is taken. Also for general two-way tables, Haldane (1940) evaluated the exact null expectation and variance of the χ^2 statistic. Lancaster (1949) showed that a k-cell multinomial probability can be expressed as the product of $(k-1)$ binomial probabilities and that the probability for an $r \times s$ contingency table can be written as the product of the probabilities for $(r-1)(s-1)$ 2×2 tables; in each case the result leads to an asymptotically exact partition of the relevant chi-squared statistic.

At a foundational level, a Kendall (1949a) essay, aimed at reconciling the attempts of frequentists and non-frequentists to establish a theory of probability, concluded that frequentists cannot avoid some prior position from which to start, but also that the probability calculations of non-frequentists must 'reflect, in some way, the behaviour of events. . . . Neither party can avoid using the ideas of the other in order to set up and justify a comprehensive theory'. He found it not possible to accept the fiducial approach. Similar disenchantment with the fiducial approach, in favour of confidence intervals, was admitted by Neyman (1941). However, some investigation of the fiducial approach did appear in the Journal, such as Garwood's (1936) estimation of a Poisson mean.

Throughout this period there were many papers giving detailed distribution theory for particular statistics, such as the sample range, sample moments and sample cumulants, usually for samples from normal distributions, and considerable space was devoted to the publication of statistical tables, by E. S. Pearson, H. O. Hartley and others. The range statistic in particular aroused much interest. Distributional work on the range of a normal sample was provided in a well-cited paper by Newman (1939), Pearson & Hartley (1942) investigated the range's cumulative distribution function, and the relationship of the range to the standard deviation was discussed by Lord (1947), in the context of a range-based modification of the one-sample t-test. Plackett (1947) showed that the ratio of the mean range to the population standard deviation for a continuous random variable can be arbitrarily close to zero; he also provided an upper bound for the ratio. Cox's (1948) contribution to the asymptotics constitutes this future long-term editor's first publication in the Journal. Geary (1936) obtained the moments of the ratio of the sample mean deviation to the sample standard deviation, with a possible test of normality in mind. Geary (1947) extended this by basing tests of normality on the ratios of the sample average of the cth power of deviations from the sample mean to the cth power of the sample deviation, for $c > 0$. Gayen (1949) studied the distribution of the t-statistic based on samples from nonnormal distributions represented by Edgeworth series. The comparison of the two variances in a bivariate normal distribution was investigated by Morgan (1939), in terms of their difference, and by Finney (1938), in terms of their ratio; power comparisons for certain cases in Morgan (1939) indicated that the two approaches performed very similarly. Finney (1938) method required the ad hoc use of an estimate for the correlation coefficient but Pitman (1939b) showed how to create an exact procedure when the correlation coefficient is unknown. Particular distributions under investigation included the noncentral t-distribution (Johnson & Welch, 1940). Pitman (1939a, c) discussed respectively estimation and testing for the parameters in a general location-scale model using invariance arguments.

Noteworthy papers appeared on a variety of other topics. David (1947) studied Neyman's smooth goodness-of-fit test; this involved a test statistic denoted by ψ^2, which for a simple null hypothesis had approximately a chi-squared distribution but which for so-called smooth alternatives provided more power than the usual chi-squared statistic. Haldane (1945a) obtained an unbiased estimator, and its variance, for the success probability in the negative binomial distribution. Finney (1947) fitted, and assessed the goodness of fit of, a probit regression model with two covariates to a dataset in which the binary response was the presence or absence of vaso-constriction. D. G. Kendall (1948) considered birth processes in which the generation time, which is the time taken for an individual to subdivide, follows a χ^2_{2k} distribution where k is an integer greater than unity, instead of $k = 1$, which corresponds to the simple version of the model. Kendall & Babington Smith (1940) formulated a method of paired comparisons based on individuals stating, for each pair of a set of items, which member of the pair they preferred; this led to assessment of an individual's consistency in ranking and of the level of concordance among a set of individuals. Yule (1939) used sentence length to characterize the style of authors and in particular to try to resolve two issues of disputed authorship, one involving Thomas à Kempis and Jean Charlier de Gerson and the other involving John Graunt and Sir William Petty; and Williams (1940) showed empirically that the distributions of word lengths from samples of the writings of Chesterton, Wells and Shaw are reasonably well approximated by discretized lognormal distributions. Rao (1948) presented a unified approach to the problem of tests of significance in multivariate analysis, involving for example Mahalanobis's D^2, and elucidated the nature of Bartlett's correction to the distribution of the generalized likelihood ratio statistic. Of great historical interest are E. S. Pearson's appreciation of K. Pearson's life and achievements (Pearson, 1936, 1938a) and M. Greenwood's papers about the development of medical statistics (Greenwood, 1941, 1942, 1943).

3. The nineteen fifties

3·1. *Papers in the top* 100

Page's (1954, 18) paper on control charts and a variety of CUSUM-type techniques for identifying changepoints in quality is the leading paper from this decade. The paper established rules for taking action. In follow-up work Page (1955a) proposed supplementing these action rules with warning rules which would lead to action if sufficiently frequently infringed, Page (1955b) considered the particular case of binary data and Page (1957) investigated more general scenarios using a likelihood approach. Other well-cited papers include Skellam's (1951, 26) work on the evolution of populations, especially of plants, by dispersal, population growth and inter-species competition, Simon's (1955, 25) study of what have come to be known as the Yule–Simon distributions, a class of skew discrete distributions that can be applied to data on word frequencies, species abundance and other contexts, and another angle on frequency modelling by Good (1953, 24), this time more nonparametric and involving smoothing, with follow-up in Good & Toulmin (1956), enabling in particular the estimation of the number of unrecorded species.

The year 1950 was particularly influential: Moran (1950a, 29) studied stochastic processes, with discrete one- and two-dimensional index sets, that are derived from processes with corresponding continuous index sets; Cochran (1950, 82) investigated the comparison of percentages in matched samples, the simplest version of which corresponds to a two-way table in which the same observational units contribute the row and column data and for which McNemar's test is the appropriate tool; and Anscombe (1950, 99) studied the sampling theory of the negative binomial and logarithmic series distributions, including maximum likelihood and other approaches to estimation. A fourth paper for this year was the first of the seminal papers by J. Durbin and G. S. Watson concerning serial correlation of errors in linear models and the development of their eponymous test statistic; the paper, Durbin & Watson (1950, 32), was followed up by Durbin & Watson (1951, 33), in which tables are provided along with details of a test statistic for the one-way and two-way classifications and polynomial trends.

Other classic papers include the following: Scheffé (1953, 28) presented his method for the simultaneous assessment of contrasts in one-way analysis of variance; Whittle (1954, 49) studied stationary processes in the plane, including two-dimensional autoregressions, and emphasized the substantial differences in what is required to handle spatial processes beyond what is needed for time series; Box (1953, 53) investigated in detail the non-robustness of normal-based tests for equality of variances, especially if there are more than two variances in question and if kurtosis levels are not compatible with normality; Quenouille (1956, 61) introduced the jackknife as a bias-reduction technique; and Lindley (1957, 83) presented his famous paradox. The paradox is that a significance test can lead to a rejection of a hypothesis at level, say, 5% and yet simultaneously the posterior probability of the truth of the hypothesis can be as high as 95%, even if the prior probability is quite small. A minor error identified by Bartlett (1957b) did not affect the underlying message.

In design of experiments the leading paper was Box & Lucas (1959, 86), which set out procedures for designing experiments when the regression function, such as in chemical kinetics models, is nonlinear in the parameters; the major complication is that for nonlinear problems optimal designs depend on the true value of the unknown parameter(s) and the paper assumes that a preliminary estimate is available in order that a locally optimal design can be constructed. Also related to design, Bradley & Terry (1952, 40) introduced their eponymous method for paired comparisons, based on a formulation involving a set of ratings, nonnegative and summing to unity, for the items being compared; a different way of modelling the paired-comparisons approach to ranking a set of items was taken by Mallows (1957).

Jonckheere (1954, 52) developed a distribution-free approach, based on a test statistic which can be regarded as a generalization of Kendall's τ, for testing equality of a set of continuous distributions that follow a specified ordering under the alternative hypothesis.

Finally, Bailey (1951, 88) considered the precision of estimators of population size, birth rates and death rates obtained from capture-recapture data from a closed population.

3·2. *Other influential contributions*

Time series was becoming a prominent topic: in what Tong (2001) calls a landmark paper, Bartlett (1950) established the importance of smoothing periodograms and developed a method based on the correlogram; Marriott & Pope (1954), with a technical addendum by Kendall (1954), considered the problem that the use of residuals, from stationary time series with the mean level and trend fitted, leads to biased estimators of autocorrelations, the bias resulting from a combination of, first, the dependence between the serial correlation and the estimated variance and, secondly, the need to estimate the mean; Whittle (1952) developed tests of fit based on minimum residual sums of squares corresponding to two time-series models, one nested within the other; and Durbin (1959) avoided the complications inherent in the calculation of maximum likelihood estimates of parameters in moving-average models by first fitting truncated versions of the equivalent infinite autoregressive process by least squares, which then makes available a set of linear equations for estimating the moving-average parameters. Further influential work on serial correlation was published: Watson (1955) studied the effect, on biases and efficiencies of estimators of parameters and on the significance levels of t and F tests, of analysing data from multiple regression models under the assumption of an incorrect but specified correlation matrix for the errors; building on Moran (1948b), Moran (1950b) evaluated the first two null sampling moments of the first-order cyclic correlation coefficient of residuals in a simple linear regression model as the basis of a test for serial independence of the residuals; and Daniels (1956) developed asymptotic theory based on saddlepoint approximations for the distributions of sample serial correlation coefficients in a number of autoregressive scenarios.

There were several contributions to ecology, especially concerning capture-recapture, and epidemics. Moran (1951) modelled an animal population, that is reduced over time, by a specified number of trappings. He used maximum likelihood to estimate the initial population size and the probability that a remaining animal is caught in the next trapping and he illustrated the method on data from rats in Sierra Leone. Craig (1953) discussed models appropriate for the marking and recapture of insects like butterflies, the data taking the form of the numbers of insects caught on particular numbers of occasions; the truncated Poisson distribution plays a key role in the modelling. Evans (1953) fitted the negative binomial, Pólya–Aeppli and Neyman Type A models to plant quadrat counts and insect counts; he observed empirically that the former were well fitted by the Neyman Type A whereas for the latter only the negative binomial was adequate. Leslie (1958) discussed simulation methods for discrete-time stochastic population models developed from deterministic models, first for a single species represented by a logistic process, then for two competing species and finally for a prey-predator scenario. Bartlett (1957a) provided further insight into stochastic models of these types. Darroch (1958) considered scenarios in which there is a fixed number of samplings and note is taken of on exactly which samplings a given individual is recaptured. Parameters, in particular the population size, are estimated by maximum likelihood based on the joint distribution of the numbers of individuals with the same capture history and on the conditional joint distribution of those variables given the numbers of individuals captured on the different samplings. The paper restricts attention to the case of a closed population, but

Darroch (1959) allows for immigration or death. Bailey (1950) set up the stochastic difference-differential equations underlying a simple, closed epidemic model, obtaining the time derivative of the mean epidemic size as well as characteristics of the distribution of the time until no susceptible individual remains. Bailey (1953) extended the model to allow for removal of infected individuals and obtained the probability distribution of the number of infected individuals. Whittle (1955) provided a simplified version of the required analysis based on the solution of a set of singly recurrent relationships rather than a set of doubly recurrent relationships.

There was still considerable activity in some areas with high profiles in the previous period. James (1951) constructed a test based on a weighted between-samples sum of squares for equality of k normal means or regression coefficients in k normal simple linear regressions when the corresponding variances are not assumed equal. If the sample sizes are large the critical region is expressed approximately in terms of a χ^2 percentage point but James also adapted the approach to deal with the case of small sample sizes. Welch (1951, 87) obtained a version of this in terms of an F reference distribution rather than χ^2 and James (1954) extended the approach to more general scenarios, including two-factor experiments and multivariate versions of the problem. Hartley (1950b) suggested using the ratio of the largest to the smallest of a set of sample variances as a test statistic for equality of the corresponding variances, as an alternative to Bartlett's test. Properties of the range of a normal sample were still of interest: Patnaik (1950) obtained and exploited an approximation to the distribution of the mean of the ranges of several normal subsamples, on the basis that this mean range leads to a more efficient estimator of the error variance than does the range of the whole sample; Hartley (1950a) used the mean range obtained from residuals from marginal means to estimate the error variance in randomized block experiments; and David et al. (1954) investigated and tabulated the distribution of the studentized range when the range and sample standard deviation are obtained from the same normal sample, rather than from two independent samples.

In contingency tables, Freeman & Halton (1951) sought exact methods for handling cases with small expected counts based on generalization of the probabilistic structure underlying Fisher's exact test for 2×2 tables; the cases of 2×3 and $2 \times 2 \times 2$ tables were described in detail. E. J. Williams (1952) investigated the possibility of assessing association in a two-way table on the basis of scores, which may have to be estimated, assigned to one or both of the row and column variables. Stuart (1955) constructed a test, based on the differences between the respective row and column marginal counts, for equality of the two marginal distributions in a two-way table. Cox (1958) noted the interpretation of the analysis of 2×2 contingency tables based on matched pairs as one of two further applications of logistic regression.

Well-cited work on design was reported in Box (1952), who constructed optimal designs for determining the slopes of a planar regression surface; since the optimality of the design does not depend on its orientation, the design can be rotated either, given prior knowledge of the response surface, in order to reduce bias or in order to eliminate certain systematic effects. Box & Hunter (1954) considered confidence regions for the solutions of simultaneous linear equations when the coefficients are subject to error, with application to finding the optimum of a second-degree response surface. Cox (1951) investigated designs for treatment allocations when an additive polynomial trend is present as well as treatment effects; a design is sought in which treatment differences are as close as possible to being orthogonal to the trend. R. M. Williams (1952) obtained designs for treatment comparisons when the errors are serially correlated, in particular following a first-order or second-order autoregression, with parameters requiring to be estimated; various systematic designs are compared for efficiency with randomized block designs.

Important contributions in multivariate analysis were provided by Deemer & Olkin (1951), who expounded relevant matrix theory transcribed from lecture notes of P. L. Hsu; this included methods for calculating the Jacobians of various transformations used in multivariate statistics. Lawley (1956a) obtained approximate tests of hypotheses concerning latent roots of sample covariance and correlation matrices, with application to principal components and factor analysis in mind. Similar calculations for more general likelihood ratio tests appeared in Lawley (1956b). In addition, Plackett (1954) showed how to express survivor functions of some multivariate normal density functions in terms of integrals of smaller order and Rao (1959) investigated normal linear models with correlated errors, with particular emphasis on the construction of simultaneous confidence intervals for a set of linear functions of the parameters; a growth-curve example is treated in detail. Quantal response models featured in Anscombe (1956), who discussed both the issue of solving the likelihood equations for linear logistic regression and the minimum logit χ^2 method of fitting that model, and in Aitchison & Silvey (1957), who were stimulated by a problem in entomology, in which an insect passes in time through a number of stages, to develop a multiple-response version of probit analysis.

Further well-cited papers include Kermack & Haldane (1950), who used a particular line called the line of organic correlation, regression-like but defined symmetrically in both variables, to investigate correlation in the context of regression models involving the type of power law used to relate body size and shape. Plackett (1950) discussed least-squares estimation for the linear model when the design matrix is singular and established formulae for updating estimators of parameters, their covariance matrix and the residual sum of squares when additional observations become available. Lloyd (1952) estimated the parameters in location-scale models using linear combinations of order statistics, fitted by least squares with the consequence that the estimators are unbiased and of minimum variance; further results were obtained for symmetric distributions. Bartlett (1953a, b) outlined the construction of confidence intervals and regions based on the asymptotic normality of the score function and vector, and incorporating skewness or higher-order moment corrections to the standard approximations; in the latter paper issues such as the problem of nuisance parameters were considered. Dunnett & Sobel (1954, 1955) studied respectively bivariate and multivariate versions of the t-distribution, including the construction of tables of percentage points in the bivariate case. Bechhofer et al. (1954) provided a method for ranking normal means when the variances are known multiples of an unknown parameter; a two-sample method is used, with the unknown parameter being estimated from the first sample. Cox & Smith (1954) investigated the process resulting from the pooling of several independent renewal processes, with particular emphasis on the long-term, equilibrium behaviour and with application to a problem in neurophysiology. Bliss & Owen (1958) described two ways of estimating the assumed common dispersion parameter in a set of negative binomial distributions. Bartholomew (1959a, b) developed tests for equality of a set of normal means when the alternative hypothesis specifies a particular ordering. The method is based on the likelihood ratio and is more generally applicable than the approach of Jonckheere (1954, 52), mentioned earlier, in § 3·1.

Topics beginning to make an impact included censored data, sequential analysis and directional data. In censored data Gupta (1952) applied maximum likelihood estimation and a best linear estimation approach to censored normal data and David & Johnson (1954) established formulae, such as expansions for the moments of order statistics, for use in analysing right-censored data with the number of censored values fixed in advance. In sequential analysis Armitage (1957) constructed a sequential procedure with a prescribed limit on the number of observations, with the possibility of stopping the experiment earlier and with probabilistic properties derived from a diffusion approximation. In directional data Watson & Williams (1956) investigated parameter estimation and hypothesis testing for von Mises–Fisher distributions, especially those on the circle and on the sphere.

The issues of *Biometrika* within this decade were particularly enriched by a large number of specialist and fascinating short notes in the *Miscellanea* section, small type being used so that much information is covered in very few pages, together with a large number of book reviews many of which were provided by F. N. David and D. E. Barton; the reviews provide intriguing reading in the present day since some of the books have become classics in the literature. These two sections had existed before the 1950s but the numbers of items included were typically much smaller. The 1951 volume contains a brief note by W. P. Elderton, marking the 50th anniversary of the founding of the Journal, as well as a eulogy for Karl Pearson by J. B. S. Haldane. Volume 42 in 1955 saw the first of the occasional series of 'Studies in the History of Probability and Statistics' in the form of an article on 'Dicing and Gaming', by F. N. David. In subsequent years many items in this series were contributed by M. G. Kendall.

4. The nineteen sixties

4·1. *Papers in the top* 100

The year 1965 stands out in this decade, with five of its papers contributing. Most prominent is the paper by Shapiro & Wilk (1965, 6), who presented the familiar Shapiro–Wilk test of normality described in more detail in § 10·6, followed by Gehan (1965a, 10), see § 10·10, who established another well-used test, this time for use in comparing two samples of singly censored observations; Gehan (1965b) provided the version for double censoring. The topic of population modelling reappeared, in the particular context of capture-recapture data: Jolly (1965, 23) established a general probability model for capture-recapture, he obtained and investigated simple estimators of parameters in the particular case of a homogeneous population, along with variances and covariances, and he showed that there is no essential difference in estimation formulae between single and multiple recapture scenarios; Seber (1965, 37) set up a somewhat different model from that of Darroch (1959) for handling a population with immigration and death, obtaining maximum likelihood estimators and associated variances for parameters. In addition, for the closed population considered by Darroch he constructed a test of the hypothesis that marked and unmarked individuals have the same probability of being caught. Finally for 1965, Rao (1965, 94) applied least squares to linear models in which the parameters are assumed to be random, with particular investigation of models for growth curves.

A further highly cited paper in this decade is by Gower (1966, 14), who obtained matrix-theory results for multivariate analysis, one goal being to find the coordinates of a set of multivariate observations, given all interobservation distances; of interest are both matrices representing measures of association between observations and matrices representing relationships between variables. Walker & Duncan (1967, 43) investigated logistic regression for binary and polytomous responses, based on an approximate linearized model and using a recursive least-squares procedure to estimate parameters. Cormack (1964, 60) provided a further contribution on population studies, for cases, such as the study of birds, in which ringed animals can be spotted in flight without actually having to be recaptured; although the total population size cannot be estimated, estimators with asymptotic variances and covariances can be obtained for survivor rates. The Cormack–Jolly–Seber model, which is the joint fruit of this paper together with the previously described papers by Jolly and Seber, is one of the basic models for capture-recapture data. Imhof (1961, 56) provided exact and approximate theory about the distribution of quadratic forms of normal variables, including those with nonzero means, essential for certain aspects of multivariate analysis, and Pothoff & Roy (1964, 65) developed seminal models for growth curves, based on a modification of the general framework of multivariate analysis of variance. Another highly influential paper is Day (1969, 76), which included an early application of maximum likelihood methodology to mixture distributions, in particular to mixtures of two normal distributions with equal but unknown covariance matrices; the algorithm he proposed was essentially an EM algorithm. Finally Wilk & Gnanadesikan (1968, 81) showed how to make use of valuable probability plotting techniques, starting with the empirical cumulative distribution function and moving on to Q-Q plots, P-P plots and extensions thereof.

4·2. *Other influential contributions*

In population models, Leslie & Gower (1960) investigated issues of stationary states and extinction probabilities in stochastic predator-prey models; the model considered has a stable stationary state unlike the stochastic version of the Lotka–Volterra model, the stationary state of which is unstable. Pollard (1966) considered stochastic versions of the population models, involving discrete time and discrete age scales, studied in papers such as Leslie (1945, 19). In particular he obtained the first two moments of the number of individuals in each age group at each time-point and he related the set-up to the multi-type Galton–Watson branching process. In capture-recapture methods, Darroch (1961) extended the discussion in Darroch (1958, 1959) to the case in which the population is sampled just twice and the data are stratified according to the different values of the probability of being caught in the second sample and according to different probability distributions for the second-sample stratum of those marked in the first sample; parameters are estimated mainly by maximum likelihood.

In time series, Whittle (1963) described the fitting of multivariate autoregressive models by fitting autoregressive schemes of increasing order, based on the univariate approach of Durbin (1960); a key achievement was to notice the time-irreversibility of stationary multivariate Gaussian time series. Box & Tiao (1965) discussed how to estimate a change in level of a time series that may be nonstationary, most of the paper being based on a simple version of the integrated moving-average model, and Jones & Brelsford (1967) treated time series in which nonstationarity takes the form of periodic structure, modelled by an autoregression. Hannan (1969) derived a necessary and sufficient condition for a vector mixed autoregressive moving-average process to be identifiable, mentioning also an analogous result for the case of continuous time, and Brillinger (1969) investigated the asymptotic theory of statistics such as the matrices of second-order periodograms of multivariate, strictly stationary time series. On the related topic of testing for serial correlation in regression analysis, Durbin (1969) used a Kolmogorov–Smirnov procedure based on the cumulated periodogram obtained from the least-squares residuals.

Influential papers on design of experiments include Box & Draper (1963) and Draper & Hunter (1966). Box & Draper (1963) constructed rotatable designs to minimize the mean-squared prediction error for a second-order response function, with the secondary aim of detecting the inadequacy of the second-order model when the true model is third-order. Draper & Hunter (1966) developed designs for multiple-response experiments that might involve nonlinear models. On the basis of an approximate Bayesian approach they considered how best to locate a specified number of extra observations after a number of observations have already been obtained.

In the area of sampling, leading contributions came from Sampford (1967), who extended the theory of sampling without replacement from a finite population to ensure that units appear in the sample with prescribed probabilities, and Hartley & Rao (1968), who developed a new sampling estimation theory for when to each experimental unit a label is attached with a known, finite set of possible values but when estimators of interest do not functionally depend on those labels.

Well-cited papers on circular data include Watson (1961), who proposed a modification of the Cramér–von Mises statistic for testing goodness of fit. The small-sample behaviour of the method was investigated in Watson (1962), as was a two-sample version. Ajne (1968) tested for a uniform distribution on the circle using, as test statistic, the maximum number of data points that can be covered by a semicircle.

In multivariate analysis, Kudo (1963) used a likelihood ratio approach to test the null hypothesis that a normal mean vector is zero against the alternative hypothesis that the mean vector lies in the positive orthant, in effect generalizing the work of Bartholomew (1959a, b); the underlying covariance matrix is assumed known. Beale et al. (1967) used multiple correlation coefficients as a tool for identifying unnecessary covariates in regression or for situations when variables can be discarded from joint distributions without losing much information.

Work on stochastic processes included Kendall (1960) on models for cellular development, and the observation thereof by the experimenter, that might mirror the mutations present in carcinogenesis, and Bartlett (1964) extended previous work on the spectral theory of a stationary point process in one dimension to the two-dimensional case. The paper included applications to the study of spatial patterns in natural plant distributions.

Among papers on foundational topics, Fraser (1961) examined a variety of issues related to the fiducial approach, including scenarios in which the method yields solutions that also have frequentist and Bayesian interpretations. Invariance ideas play a central role and the paper discusses the combination of fiducial and prior distributions and the combination of two fiducial distributions. Dempster (1967) provided an application of his method of upper and lower probabilities to a simple problem involving a finite population from which are drawn two samples, one after the other. The objective is to make inferences about the second sample having observed the first. The two types of probability, to be regarded as bounds on degrees of knowledge, gave rise to the concepts of belief and plausibility in the Dempster–Shafer theory of evidence. The method belongs to a general family that includes the Bayesian approach.

The topic of robustness became more prominent. Box & Watson (1962) showed the influence that the values taken by covariates can have on the validity of the normal-theory F-distribution as it arises in regression problems and Box & Tiao (1962) exhibited robustness in Bayesian inference for a mean based on a random sample assumed to come from a power distribution involving a kurtosis parameter that equals zero in the Gaussian case. Further contributions to the increasingly visible Bayesian approach include Box & Draper's (1965) treatment of multiple-response data; the errors are assumed to be normally distributed and inference is based on the posterior distribution with the error variance and covariance parameters integrated out. Tiao & Zellner (1964) dealt with the normal linear model when the data come from two experiments with different, unknown error variances; again the marginal posterior for the mean parameters is obtained together with an asymptotic normal approximation thereto. Box & Tiao (1968) modelled outliers as having a different distribution from that of non-outliers; in the case of the normal linear model the two error distributions differ in that the variance for outliers is a large, prescribed multiple of that for non-outliers and each observation is assumed to be an outlier with a fixed probability.

Finally, well-cited papers were published on a wide variety of particular topics. For the convex hull of a random set of points in two and three dimensions, Efron (1965) obtained expressions for the expectations of the area, perimeter, probability content and number of sides; particular versions are given for cases in which the underlying distribution is normal or uniform over a disc. Zelen & Feinleib (1969) considered models for aspects of the early detection of diseases, in particular the time between the detection by screening that a patient is in the preclinical state until the clinical state is entered. Ireland & Kullback (1968) studied parameter estimation in contingency tables with fixed marginal probabilities, using and establishing the convergence of the iterative proportional fitting algorithm. Hinkley (1969b) obtained the distribution of the ratio of correlated normal random variables, as used for instance in estimating the ratio of the intercept of the simple linear regression line with the covariate axis. This work also found application in the estimation of the point of intersection of two simple linear regression lines and consequently in the maximum likelihood estimation of the changepoint in dog-leg regression; see Hinkley (1969a). Hartley & Rao (1967) developed asymptotic theory for the maximum likelihood estimators of parameters in the general linear mixed model. Mosimann (1962) studied the compound multinomial model, the mixing distribution being the appropriate Dirichlet distribution. He investigated the distribution's moments and estimation of parameters, as well as properties of the Dirichlet distribution, which he calls the multivariate beta distribution. Harter (1961) used numerical integration to calculate expectations of normal order statistics and compared the results with approximate methods. In the context of goodness of fit, Durbin (1961) showed how manipulation of the ordered gaps between consecutive order statistics from the standard uniform distribution formed the basis of potentially more powerful modifications of the Kolmogorov–Smirnov test. Slater (1961) examined ways of counting the minimum number of inconsistencies present in orderings of items resulting from an individual's paired-comparisons exercise. In the context of what became known as the Farlie–Gumbel–Morgenstern distribution, Farlie (1960) compared the product moment correlation coefficient, Spearman's ρ, Kendall's τ and the maximum likelihood estimator of the distribution's parameter on the basis of asymptotic relative efficiency. Although not qualifying for the top 100, Turin's (1960) derivation of the characteristic function of Hermitian quadratic forms of complex normal variates is currently by some margin the most highly cited paper from 1960. Gart & Zweifel (1967) investigated various estimators of the logit of a binomial distribution parameter in terms of bias and variance estimation, with an application to linear logistic regression as used in bioassay. Watson & Leadbetter (1964) compared a number of nonparametric estimators of the hazard function, for example based on kernel density estimators. Tiku (1967) used the observation that the ratio of the density of the standard normal distribution to the cumulative distribution function is approximately linear to motivate approximations to maximum likelihood estimators of parameters from normal data that include fixed proportions of censored data in either tail.

M. G. Kendall (Kendall, 1963) contributed an obituary and bibliography of R. A. Fisher, and statistical tables continued to appear, constructed by D. E. Barton, E. S. Pearson and others. The pattern of many short communications and book reviews, up to about 15 of each per issue, continued from the 1950s, but by the end of the decade the book-review section was more-or-less reduced to a bare list of recent publications. In 1968 for the first time *Biometrika* was published in three issues per annum rather than the previous two and the mid-1960s saw the gradual inclusion of summaries at the beginning of papers as a regular feature. Also in the mid-1960s the editorship passed from E. S. Pearson to D. R. Cox, the last of the long-term editors of the Journal.

5. The nineteen seventies

5·1. *Papers in the top* 100

The decade started with a strong year, represented by five papers. The most highly cited is Hastings (1970, 5), at the time an unassuming publication but one that became central to the explosive development of Markov chain Monte Carlo methods, especially of the so-called Metropolis–Hastings type, that have had such a strong influence on Bayesian computational methods. Section 10·5 provides more detail, along with mention of Peskun's (1973) follow-up paper. Other important papers from that year are Mardia's (1970, 31) definition of measures of multivariate skewness and kurtosis, together with their use in constructing a test of multivariate normality and in assessing the robustness of Hotelling's T^2 test against nonnormality; Breslow's (1970, 48) version of the Kruskal–Wallis test for comparing a number of samples with different patterns of censorship, thereby generalizing Gehan (1965a, 10) two-sample work for right-censored data; Jöreskog's (1970, 70) general method for the maximum likelihood analysis of data from multivariate normal distributions with structured means and covariance matrices, as in factor analysis, growth-curve models and other special cases; and Hinkley's (1970, 95) approach to the maximum likelihood estimation of a changepoint, first assuming a general distribution and then assuming normality, in the latter case with various scenarios in which the initial and final means and the variance parameter are assumed known or unknown.

There are four particularly prominent papers from 1978. The leading contribution was the Ljung & Box (1978, 9) measure of lack of fit in time-series models; see § 10·9. Efron & Hinkley (1978, 93) provided a frequentist justification, based on a conditionality argument, for the use of the observed information instead of the Fisher information when calculating approximate variances of maximum likelihood estimators; the paper was discussed by O. E. Barndorff-Nielsen, A. T. James, G. K. Robinson and D. A. Sprott. Other new tools were Clayton's (1978, 44) frailty model for association in bivariate lifetables, motivated by the perception that the incidence times of onset of a chronic disease for two members of the same family are likely to be correlated, and Prentice's (1978, 100) work on linear rank tests, based on scores given to residuals, for regression models when the response is right-censored; the standard log-rank test can be regarded as a special case.

Prentice also contributed elsewhere, with Prentice & Pyke (1979, 69) exploiting the fact that, although a prospective logistic disease incidence regression model cannot be fully estimated from case-control data alone, odds ratios defined in the logistic regression model can be estimated. The paper builds on and clarifies Anderson (1972, 85), who described logistic regression in detail for different sampling schemes but required that the sample space of the regressors be finite; the work of Prentice & Pyke is not dependent on this pre-condition. For the comparison of two survival distributions under right censorship, Tarone & Ware (1977, 91) showed that the formulae for the statistics for the log-rank test and the modified Wilcoxon test are very similar. They also proposed a statistic that is a compromise between the two and avoids disadvantages of each.

This decade contains other subject-defining papers: Rubin (1976, 7), discussed in more detail in § 10·7, showed how crucial it is, from the point of view of inference methodology, to establish the ignorability or otherwise of the missingness process underlying incomplete data. Cox (1975, 13) established, in some generality, the approach of partial likelihood, highly valuable in certain problems with nuisance parameters of high dimension or complexity; the paper shows that, subject to predictable regularity conditions, the standard large-sample theory of maximum likelihood estimators and associated tests applies also to partial likelihood. Gabriel (1971, 30) introduced the biplot, which is based on the exact or approximate representation of an $n \times m$ data matrix as the product of an $n \times r$ matrix and an $r \times m$ matrix, each of rank r, where n and m are the numbers of observations and variables respectively; if $r = 2$ the method enables the observations and variables to be represented on the same two-dimensional plot. Miller (1974, 34) gave an overview of the jackknife which, in its simplest form, removes the $O(n^{-1})$ term in the bias of an estimator based on a sample of size n, and which contributes to robust interval estimation. Wedderburn (1974, 35) developed the concept of quasilikelihood, a least-squares-like criterion reflecting the relationship between the mean and variance of observations but usually not representing the exact loglikelihood. However, the maximum quasilikelihood estimator often has asymptotic properties similar to those of the maximum likelihood estimator itself and the paper considered, in particular, application to generalized linear models. Patterson & Thompson (1971, 11) described methods for recovering interblock information in designed experiments based on what became known as restricted maximum likelihood procedures, which have had a wide impact on areas such as the estimation of variance components; in this they built on Hartley & Rao (1967). Pocock (1977, 46) studied two-treatment trials in which a fixed number of groups of patients of specified size are sequentially entered into the trial, with termination possible at each stage and with prescribed overall significance level; the case of normally distributed responses with known variance is treated in detail and extensions to other scenarios are discussed. Goodman (1974, 54) developed latent structure analysis, investigating models, for data in multiway contingency tables, that include an underlying hidden categorical variable made up of latent classes; a variety of model structures are investigated, and maximum likelihood methods and test of fit are described.

Further listed papers include Shibata's (1976, 90) application of Akaike's information criterion to autoregressive models, showing that asymptotically the order selected by the method will not be less than the true value but may exceed it, and Davies's (1977, 50) careful account of a class of nonstandard hypothesis-testing problems; the test statistic used is the supremum, over the range of values for the nuisance parameter, of the process represented by the locally optimal test statistic that would be appropriate if the nuisance parameter were known. In this paper the test statistic is assumed to be normally distributed; in the follow-up paper, Davies (1987, 51), a chi-squared distribution is assumed. Marcus et al. (1976, 68) provided an approach to multiple testing of a set of hypotheses closed under intersection in which the experiment-wise error rate is the same as the level of the overall test. Buckley & James (1979, 74) showed how to handle simple linear regression problems when the response variable may be censored on the right and the error variance is unspecified; they used the usual normal equations but with the censored responses replaced by values that reflect the censoring. In an earlier attack on the same problem, Miller (1976) used a least-squares approach to estimate parameters, but the estimators are not consistent in general. Scott (1979, 72) investigated the histogram approach to nonparametric density estimation and in particular the selection of the histogram bin-width that minimizes asymptotically the integrated squared error of the density estimator; for practical application he used the version of the formula based on a normal distribution and the sample standard deviation. Finally, for CUSUM-type sampling inspection schemes with a discrete state space, Brook & Evans (1972, 97) provided a Markov-chain representation from which features such as average run length and run-length distribution can be calculated.

5·2. *Other influential contributions*

Strong papers on classical design of experiments included Box & Draper (1975), who list fourteen attributes of a good design and then concentrate on the issue of insensitivity to 'wild observations', identifying, as a measure of sensitivity for the design matrix X of the linear model, the sum of squares of the diagonal elements of $X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$, and Patterson & Williams (1976), on resolvable incomplete block designs for when the number of varieties is a multiple of the block size. The topic of optimal design also emerged, with contributions of which the most highly cited is Atkinson & Fedorov's (1975) paper on discrimination between two models, one of which is assumed to be true; the design is chosen so as to maximize the power of the test for departures from the incorrect model in the direction of the correct model. Silvey & Titterington (1973) used Lagrangian duality to establish a geometrical interpretation of optimal design. A 1974 conference on optimal design at Imperial College led to four papers in the Journal and optimal design was also the topic of what turned out to be *Biometrika*'s final book review, namely that by A. M. Herzberg and H. P. Wynn of V. V. Fedorov's *Theory of Optimal Experiments*.

There were many influential papers on time series. Akaike (1973) developed the maximum-likelihood approach to estimating the parameters in multi-dimensional zero-mean Gaussian ARMA models, using a Newton–Raphson-type algorithm in which the gradient and Hessian of the loglikelihood are handled by a frequency-domain approach. Bloomfield (1973) analysed a time series in terms of a parametric model for the spectral density function rather than modelling the time series directly; various estimators, including maximum likelihood, are obtained for the parameters and prediction procedures are developed. Newbold (1974), Ansley (1979) and Ljung & Box (1979) extended established methods for calculating the exact likelihood of moving-average processes to deal with the case of an ARMA process, Newbold (1974) and Ljung & Box (1979) extending a method of Box & Jenkins (1976, p. 271), and Ansley (1979) developing a method of Phadke & Kedem (1978) based on Cholesky decomposition of the covariance matrix. Box & Tiao (1977) proposed transformation of the state space of a multi-dimensional stationary autoregressive time series into independent, stationary and nearly nonstationary subspaces, the first two subspaces representing stable features and the third representing dynamic growth; they thereby brought canonical correlation analysis to time series, pre-dating the related concept of co-integration in econometrics. Davies et al. (1977) warned that asymptotic-theory significance levels of the Box–Pierce portmanteau test for assessing the fit of a time-series model are likely to be much larger than true values in practice, even for fairly large sample sizes; see Merikoski & Pukkila (1983) for some corrections. For selecting the order of an autoregressive model using Akaike's criterion Bhansali & Downham (1977) showed that increasing the penalty constant was beneficial in selecting the right model. Akaike (1979) built on their results, in particular creating a Bayesian extension of his criterion. Harvey & Phillips (1979) showed how application of Kalman filtering techniques facilitate the calculation of generalized least-squares estimates of the parameters in linear regression models with ARMA errors. Durbin & Watson (1971) completed their trilogy about serial correlation by investigating in detail the distribution of their test statistic, including the exploration of approximations of which the most successful were a beta distribution and the distribution of a moment-matching linear function of a certain upper-bounding statistic.

The area of censored data and survival analysis enjoyed accelerated growth, stimulated by Cox's (1972) proportional hazards model and the associated partial likelihood for estimating parameters. Some papers related to this have been mentioned already, but there were other prominent contributions: Kalbfleisch & Prentice (1973) demonstrated that Cox's partial likelihood can be regarded as a marginal likelihood for ranks, although only if there is no censoring and the covariates are time-independent, and they also described a way of handling tied survival times; Prentice (1973) considered what is in effect a proportional hazards model in which the baseline hazard and the covariates are constant in time; Tarone (1975) used the proportional hazards model as the basis of a test for monotone trend in the hazard function, for use in contexts such as dose-response in which it is felt that the hazard should increase with the dose level; Oakes (1977) obtained explicit formulae for the asymptotic variances of estimators of parameters in the proportional hazards model, including the maximizers of the full likelihood and of Cox's partial likelihood; and Prentice & Breslow (1978) adapted the proportional hazards model to cover case-control studies, with Cox's partial likelihood appearing as a valid tool for inference but with the risk-set at a particular time being the set of individuals recruited to the study at that time. Lagakos et al. (1978) modelled censored multistate data by semi-Markov models, estimating transition probabilities and sojourn time distributions nonparametrically; special cases include the Kaplan–Meier estimator of a survival function. Koziol & Green (1976) adapted the Cramér–von Mises statistic so as to handle data censored on the right, with the Kaplan–Meier product-limit estimator used instead of the empirical distribution function in the formula for the statistic and under the assumption that the survivor function of the censoring points is some power of the survivor function of the data themselves; asymptotic powers are calculated for various types of alternative.

Changepoint problems stimulated further well-cited contributions. In Hinkley (1971) cumulative sum tests provide an estimator of the point at which the mean of a sequence of normal variates with known variance changes downwards to some unknown value from a known initial value; the estimator is compared with the maximum likelihood estimator of the changepoint. Smith (1975) established Bayesian posterior distributions for a changepoint, first in a fairly general setting and then for the cases of binomial trials and normal distributions, with various scenarios involving known and unknown parameters. For the scenario in which the only unknown feature is the changepoint, Cobb (1978) approximated the conditional distribution of the maximum likelihood estimator of the changepoint conditional on the ancillary values of neighbouring observations. Bacon & Watts (1971) took a Bayesian approach to the related problem of dog-leg regression with normal errors, but allowing a transition parameter that accommodates a smooth change from one straight line to the other as well as an abrupt change; their method allows for more than one transition between intersecting straight lines.

There were a number of important contributions concerning contingency tables: Gart (1970) investigated exact and approximate estimation of the assumed common odds ratio or relative risk of several 2×2 tables, assuming that marginal totals are fixed, and Zelen (1971) provided a test for whether or not several 2×2 tables have the same relative risk and also considered a linear model for the log relative risk given a putative explanatory variable. For multi-dimensional tables Goodman (1973) studied models in which some of the variables are modelled as dependent on some others and/or are explanatory variables for yet further variables; the paper shows how to obtain expected frequencies and chi-squared tests of model fit. Williams (1976) obtained scalar multipliers that corrected for $O(n^{-1})$ terms in the expected deviances associated with multiway tables with total sample size n and thereby improved the approximation to null distributions provided by the χ^2 distribution, along the lines of Lawley (1956b). Fienberg (1972) linked the topic of incomplete contingency tables, modelled by log-linear models, with that of multiple recapture in a closed population; incompleteness obtains since the number of individuals missing from all samplings is unknown but is estimated from a model fitted to all the other cells in the table. Other contributions on statistical ecology included Seber's (1970) treatment of data on the retrieval of bird-rings from dead birds, allowing retrieval rates and survival rates to vary from year to year, and Burnham & Overton's (1978) modelling of individual capture probabilities as a random sample from an arbitrary distribution on the interval [0, 1]. There is multiple recapture and a nonparametric estimator of the total population size is obtained.

Important contributions with a sequential flavour were provided by Efron (1971) and Siegmund (1978). Efron (1971) provided, in the form of biased-coin designs, a method for allocating individuals to treatments sequentially in order to maintain balance while retaining a strong element of randomization and being resistent to various types of bias. Siegmund (1978) showed how to obtain confidence intervals subsequent to sequential tests, an important observation underlying the construction of the intervals being that a sequential test will tend to stop earlier the further the true situation is from the null-hypothesis scenario.

There were a number of prominent articles about foundational issues: Barndorff-Nielsen (1973) defined M-ancillarity and Feigin & Reiser (1979) developed the concept of asymptotic ancillarity, applied also by Efron & Hinkley (1978, 93), and thereby could develop a conditional inference approach that is applicable to a class of regular, nonergodic processes that includes the Yule process. Sprott (1975) used likelihood as the basis for the development of marginal and conditional sufficiency; if there is an unknown nuisance parameter then no information is lost about the parameter of interest if inference is based on a marginal or conditional sufficient statistic. Godambe (1976) discussed conditional likelihood in the presence of nuisance parameters and showed that it leads to an estimating equation for the parameter of interest that is optimal in a certain sense. Kalbfleisch (1975) distinguished between ancillary statistics associated with the underlying experimental design and those resulting from the modelling, the latter not necessarily being unique, and then leading to different inferences about the parameter of interest; the paper was supplemented by discussion contributions from G. A. Barnard, O. E. Barndorff-Nielsen, A. D. McLaren and A. Birnbaum.

In finite-population sampling, Royall (1970) applied his superpopulation approach based on linear regression models in scenarios with a variable of interest and an auxiliary variable. As § 5.1 of Smith (2001) emphasizes, the paper has been highly influential in arguing the case for using model-based rather than randomization inferences, on the grounds of efficiency. Cassel et al. (1976) further extended the superpopulation approach in similar contexts, seeking estimators that have good properties under both the traditional and the model-based approaches.

An emerging area of interest was that of smoothing and nonparametric density estimation. Good & Gaskins (1971) maximized the loglikelihood with an additive penalty in the form of a linear combination of the integrated squared first and second derivatives of the square root of the underlying density. The coefficients of the linear combination act as smoothing parameters, values being chosen on a fairly ad hoc basis in this paper. Aitchison & Aitken (1976) applied the kernel method to multivariate binary data, with the smoothing parameter chosen by maximizing the leave-one-out likelihood, which is formally a version of crossvalidation; they used the method in the context of discriminant analysis. For choosing a smoothing parameter in univariate kernel-based density estimation Silverman (1978) proposed the use of 'test graphs' of the density estimate's second derivative; the degree of smoothing can be set so as to realize asymptotic characteristics of the optimal density estimate.

Other influential papers included Kettenring (1971) on generalizations to more than two sets of variables of the standard theory of canonical correlation analysis; several approaches were developed, all of which reduced to the familiar method in the case of two sets of variables. Hawkes (1971) developed the theoretical properties of a class of point processes for which possible applications include epidemics and particle emission. In the context of cluster analysis Fisher & Van Ness (1971) listed various notions of admissibility in clustering and identified to what extent a number of cluster analysis methods achieve the types of admissibility, and Binder (1978) developed a Bayesian approach, incorporating a prior distribution on the number of clusters and a vector of cluster indicators for the individual observations, and specifying a loss function that relates the true and estimated clusterings; the case of multivariate normal cluster-specific distributions is treated in detail. Efron & Morris (1972) constructed estimators, of a set of k normal mean vectors of dimension p, that dominate the usual least-squares estimators; the univariate version, for $p = 1$, corresponds to the James–Stein estimator. Geisser (1974) compared a number of methods, including two based on predictive sample reuse, for estimating means in the random-effect model. Clifford & Sudbury (1973) developed what they refer to as a fair stochastic model for conflict between two species in competition for territory, based on a lattice representation of the territory and rules governing invasion and other moves. Harville (1974) provided a Bayesian follow-up to the work of Patterson & Thompson (1971, 11) on estimating variance components, showing that use of the restricted likelihood within the Bayesian paradigm is equivalent to ignoring prior information on the linear-model fixed effects and using all the data. Strauss (1975) proposed a model for the spatial clustering of points when the degree of closeness between two points is a binary variable; it turns out that the probability density of the points depends on a single clustering parameter. Certain aspects of the paper were tidied up by Kelly & Ripley (1976). Aitchison (1975) used a measure of closeness based on the Kullback–Leibler divergence to assess the performance of a data-based distribution as an estimator of a true parametric distribution. He showed that in this respect the Bayesian posterior predictive distribution was the optimal estimator and in particular was superior to distributions, based on plug-in efficient estimators of the parameters, which typically underestimated uncertainty. Efron & Thisted (1976) used a model based on Poisson processes to estimate the number of different words in an author's total vocabulary, given the author's body of work; the application of a number of methods, including those of Fisher et al. (1943) and Good & Toulmin (1956), provided remarkably consistent lower bounds for Shakespeare's vocabulary. Besag (1977) investigated the efficiency of his pseudo-likelihood method in the context of certain Gaussian fields and Stone (1977) investigated the possibility of lack of asymptotic consistency in crossvalidatory assessment. Wedderburn (1976) investigated the properties of maximum likelihood estimators of parameters in certain generalized linear models. Finally, Bowman & Shenton (1975) presented an influential method for testing for normality and Cox & Small (1978) approached the testing for multivariate normality through coordinate-dependent and invariant procedures for testing for the linearity of regression functions.

The 1972 issue saw the initiation of the inclusion of keywords at the beginning of papers and during the decade the *Miscellanea* section provided over 200 short communications, including a historical piece by I. J. Good about Turing's wartime statistical work (Good, 1979).

Finally, the paper by Pearson et al. (1977) on testing departures from normality represents a milestone in being the first author's final paper in *Biometrika*.

6. The nineteen eighties

6·1. *Papers in the top* 100

The 1970s and 1980s represent a particularly influential period in *Biometrika*'s history, at least in terms of citations, in that almost half of the papers in the top 100, and six of the top ten, come from those years. Indeed, the 1980s saw the publication of the three most highly cited articles.

Top of the list, with over 9400 citations, is Liang & Zeger (1986, 1), see § 10·1, which established the concept of generalized estimating equations within the context of repeated measures as prevalent in longitudinal data. The estimating equations are based on score functions reflecting simplified working assumptions about intrasubject correlations and the asymptotic properties of the resulting estimators are obtained. Rosenbaum & Rubin (1983, 2) boasts about 8300 citations for its introduction, in the context of estimating a treatment effect, of the propensity score, which is a function of covariates that is the coarsest so-called 'balancing score', use of which can lead to unbiased estimation of the treatment effect; see § 10·2. The third paper is Phillips & Perron (1988, 3), cited over 7300 times, and a seminal article about the oft-investigated time-series issue of testing for a unit root, that is, in its simplest form, of the hypothesis that the parameter of a first-order autoregression is equal to 1; see § 10·3. A further contribution on the topic of testing for unit roots is provided by Said & Dickey (1984, 20).

Further influential time-series papers in this decade include Hosking (1981, 12), on fractional differencing, who provided models that explain both short-term and long-term structure of a time series and in particular exhibit long-term persistence, a feature especially relevant to applications in economics and hydrology. Hurvich & Tsai (1989, 21) developed a small-sample bias correction to Akaike's AIC, aimed at combatting the tendency of AIC to overfit, and applied it to model selection in regression and time-series, and Luukkonen et al. (1988, 62) examined a set of procedures for testing for the linear autoregressive model within the wider class of smooth-transition autoregressive, STAR, models. Note that, when these models were introduced by Chan & Tong (1986), the second letter in the acronym STAR was intended to stand for threshold.

The year of Liang & Zeger (1986, 1) also included three other very strong papers: Simes (1986, 47) described a less conservative version of the basic Bonferroni procedure for multiple testing, based on the ordered values of the p-values of the individual tests; Prentice (1986, 71) began the year's volume with case-cohort designs in which covariate histories are assembled only for a randomly selected subcohort and for the cases, rather than for the whole cohort; and Goldstein (1986,

79) set out a general multilevel mixed effects model and showed how estimation of parameters is facilitated by iterative generalized least-squares algorithms.

Also in this decade, Hochberg (1988, 15) provided further refinement of the Bonferroni approach to multiple testing, extending the recent method of Simes (1986, 47) to allow for inferences about individual hypotheses, Owen (1988, 42) constitutes the fundamental paper that initiated the huge literature on empirical likelihood, deriving confidence intervals, for features of the underlying distribution, based on the empirical likelihood ratio, which is shown to be a nonparametric analogue of parametric likelihood ratios, and Efron (1981, 73) described how to estimate nonparametrically the standard error of point estimates using a variety of techniques such as the jackknife and the bootstrap; the paper is focussed on the estimation of the correlation coefficient in a bivariate normal distribution. Prentice et al. (1981, 78) discussed the regression analysis of failure-time data when the subjects may experience multiple failures, such as a sequence of asthmatic attacks; different models were considered for the baseline hazard and regression parameters were estimated from partial likelihoods. Harrington & Fleming (1982, 96) developed k-sample generalizations of the log-rank test for censored data, establishing asymptotic normality of the test statistics under the appropriate null hypotheses and investigating asymptotic relative efficiencies. Schoenfeld (1982, 55) defined partial residuals, expressed in terms of the covariate values, intended to detect deviation from the proportional hazards model as well as highlighting possibly outlying covariate values, and Lan & DeMets (1983, 39) provided a flexible way of defining discrete boundaries for clinical trials based on the α-spending rate; the boundaries are sequential in that they are determined at a given time on the basis of previous and current decisions only. Cook & Weisberg (1983, 92) used the score statistic as the basis for a diagnostic test for heteroscedasticity in regression, given particular parametric models for the nature of the heteroscedasticity. Finally, Bowman (1984, 98) introduced L_2 crossvalidation for choosing a kernel bandwidth in density estimation as an alternative to the Kullback–Leibler–based maximum leave-one-out likelihood approach.

6·2. *Other influential contributions*

A major feature of this decade was the large number of well-cited contributions on survival analysis. Schoenfeld (1980) developed omnibus goodness-of-fit tests for the proportional hazards model; there are various versions of the tests based on different partitionings of the space of covariate values and failure times, corresponding to which observed and expected frequencies can be obtained. Rather than produce an omnibus procedure, Gill & Schumacher (1987) provided a test of the proportional hazards assumption for the two-sample case, designed to detect monotone departures from the hypothesis that the hazard ratio is constant. Hall & Wellner (1980) derived simultaneous confidence bands for a survival function based on right-censored data, to go with the Kaplan–Meier product-limit estimator and to provide an analogue of the Kolmogorov bands for uncensored data. Breslow (1981) studied the properties of certain estimators of the common odds ratio of 2×2 tables when the number of tables increases but the possible marginal configurations remain fixed; the most attractive estimator of those examined, in terms of consistency and efficiency even for fairly large odds ratios, turns out to be the Mantel–Haenszel estimator. Schoenfeld (1981) investigated tests for comparing survival functions given censored data; he examined the asymptotic properties of the tests of Tarone & Ware (1977, 91) and showed that other well-known tests are asymptotically equivalent to those of Tarone & Ware. Gail et al. (1981) presented a recursion for calculating a conditional likelihood function associated with a proportional hazards model applicable to matched case-control studies. Prentice (1982) developed a partial likelihood function through which to estimate a proportional hazards model when the covariates are subject to error; a particular activity covered was testing equality of survival functions when the covariate is group membership and individuals are susceptible to misclassification. Hougaard (1984) studied frailty models for accommodating heterogeneous populations, going beyond the already well-investigated gamma-distributed frailty and examining in particular the inverse-Gaussian case; and, in further investigation of potential frailty distributions for heterogeneous populations, Hougaard (1986a) developed a three-parameter family of distributions that include those studied in Hougaard (1984) as well as, generally, the stable distributions on the positive numbers. Hougaard (1986b) proposed multivariate failure-time distributions that allowed for dependence among individuals within a group, such as a family, as well as accommodating covariates. Struthers & Kalbfleisch (1986) considered the impact of model misspecification on estimation of the proportional hazards model, for example if the true model were based on accelerated failure time or if the model proposed did not include all relevant covariates. Tsai et al. (1987) established the asymptotic properties, such as weak convergence to a Gaussian process, of the Kaplan–Meier estimator as modified to cope with left-truncation as well as right-censoring. Lagakos et al. (1988) adapted survival analysis to cope with right-truncated data, necessary for application to data concerning the onset of AIDS; the analysis is facilitated by transforming to reverse survival time within which the data are left-truncated. Finally Heckman & Honoré (1989) investigated the effect of the introduction of covariates on the identifiability of competing-risks models, including models based on proportional hazards or on accelerated hazards.

In the area of foundations and general theory a group of papers appeared on non-Bayesian inference, following on from Efron & Hinkley (1978, 93): for certain one-parameter scenarios in which no exact ancillary statistic exists, Cox (1980) showed how to obtain a locally ancillary statistic, using Edgeworth expansion to approximate the appropriate conditional distribution and establishing that it yields significance tests and confidence intervals with required probability levels to $O(n^{-1})$; for a one-parameter curved exponential family model, Hinkley (1980) showed that, to the same order of approximation, the normalized likelihood function is approximately pivotal, providing the conditional density of the maximum likelihood estimator, given a natural approximate ancillary; Barndorff-Nielsen (1980) showed that density to be, in particular in exponential family models, approximately proportional to the product of the square root of the determinant of the observed information matrix, evaluated at the maximum likelihood estimator, and the ratio of the likelihoods corresponding to the parameter and to its maximum likelihood estimator; Barndorff-Nielsen (1983) showed that this approximation is accurate to $O(n^{-1})$, sometimes to $O(n^{-3/2})$ and sometimes exact, as well as using it as basis for a modified profile likelihood; Durbin (1980) discussed a similar approximation formula for the distribution of a sufficient statistic, but involving the expected, rather than the observed, information matrix; for one-parameter problems, McCullagh (1984) established the existence of a statistic, to be called second-order locally sufficient, that is independent of any corresponding second-order locally ancillary statistic and in fact is a linear function of a signed loglikelihood ratio statistic; and, in contexts with two sets of parameters, Barndorff-Nielsen (1986) studied the signed loglikelihood ratio for one set of parameters, showing how to establish its asymptotic normality and revealing its role as an ancillary if inference about the other set of parameters is of interest. On a different tack, Nelder & Pregibon (1987) extended the definition of quasilikelihood for generalized linear models to permit the comparison of variance functions.

There was much activity in logistic regression. Tsiatis (1980) constructed a goodness-of-fit test of the linear logistic model based on partitioning the space of the covariates, reminiscent of Schoenfeld (1980); the partitioning is represented by indicator variables, with parameters that are zero if the linear logistic model holds, and a version of the efficient scores test is developed.

Aranda-Ordaz (1981) proposed symmetric and asymmetric families of transformations of a binomial probability, special cases of which correspond to the logistic transformation or, approximately, to the probit or arc-sine transformation. Albert & Anderson (1984) studied the impact of properties such as degrees of separation of data clouds on the existence of maximum likelihood estimators in multinomial logistic regression models. For a multi-category response with a baseline category, Begg & Gray (1984) considered the use of binary-response logistic models for comparing individual response categories with the baseline category rather than the corresponding, more complex, polytomous analysis; empirical results indicated that the individualized approach is often very efficient. For the binary-response case Carroll et el. (1984) allowed for errors in some of the covariates in binary regression, be it logistic or probit; they took a structural-modelling approach with normally distributed errors in the variables. Finally, Breslow & Cain (1988) showed how to modify the usual logistic regression model as applied to case-control data to incorporate information from a preliminary random sample that provides supplementary information about joint incidence of the disease of interest and an important risk factor.

In the somewhat related area of clinical trials, DeMets & Ware (1980) described three modifications of Pocock's (1977, 46) work on group-sequential methods designed to handle one-sided alternative hypotheses, an important development given that many trials have particular interest in the possible superiority of a new treatment; Gail et al. (1984) investigated the biases in estimating treatment effects resulting from model nonlinearity and omitted covariates, for models in which the expected response is a known nonlinear function of a linear combination of treatment effects and covariates; and Kim & DeMets (1987) investigated issues related to Lan & DeMets's (1983, 39) approach to group-sequential methods, including extension to cover the case of asymmetric two-sided boundaries and rules for specifying sample sizes. Wei & Johnson (1985) investigated treatment-versus-control scenarios with repeated measurements and some missing values, establishing overall test statistics that combine the evidence from the various stages.

Well-cited advances in multivariate analysis include Muirhead & Waternaux (1980), who investigated canonical correlation analysis especially for data from elliptical populations, for which the asymptotic distributions of the sample canonical correlation coefficients and of certain test statistics take quite simple forms; for example some likelihood ratio statistics are asymptotically chi-squared except for a scale factor. Tyler (1983) established conditions under which such an adjustment is possible and developed robust chi-squared tests based on affine-invariant M-estimators of scatter. Dawid (1981) introduced an advantageous notation for certain matrix-variate distributions, justifying it by characteristics such as its consistency under marginalization. Bartholomew (1984) developed a firm theoretical basis, in which Bayesian sufficiency plays a part, for latent structure methods including factor analysis, while revealing that much current practice relied on untestable distributional assumptions. Darroch & Mosimann (1985) developed multivariate methods for scale-free shape variables constructed from a set of original variables representing the same physical quantity and on the same scale, such as the four length measurements making up the iris data; work concentrated particularly on the logarithms of shape variables. Critchley (1985) provided ways of identifying influential observations in principal component analysis, proposing influence curves and empirical versions thereof for the eigenvalues and eigenvectors of the underlying covariance matrix. Little & Schluchter (1985) applied the EM algorithm to multivariate data involving both continuous and categorical variables, modelled by the the general location model of Olkin & Tate (1961), a model that includes a number of important special cases, and subject to missing values. Genest (1987) investigated a class of bivariate distributions on the unit square that are Archimedean copulas and can be used to construct bivariate families of distributions with arbitrary marginals.

There were a number of important contributions about graphical models and contingency tables. Wermuth & Lauritzen (1983) described graphical and recursive models that correspond to assumptions of conditional independence of pairs of variables. Edwards & Kreiner (1983) concentrated mainly on graphical models, expressing a preference for them rather than nongraphical hierarchical models on the grounds of interpretability. Asmussen & Edwards (1983) discussed necessary and sufficient conditions for collapsibility of hierarchical log-linear models. Finally, Edwards & Havranek (1985) developed a strategy for selecting the simplest acceptable models for a multi-dimensional contingency table, based on the rules that if a model is found acceptable then so also are models that include it and if a model is rejected then so also are models that it includes, thereby avoiding the need to fit all models; they provided versions for both graphical models and hierarchical log-linear models.

Leading contributions in sampling theory were by Särndal (1980) and Chambers & Dunstan (1986). For estimating the mean of a variable in a finite population in the presence of covariates, Särndal considered two weighting schemes for probability sampling, one of the so-called π-inverse type and one model-based; he found that both estimators were asymptotically unbiased and had similar efficiencies. Chambers & Dunstan used a model-based method, in which auxiliary information can be incorporated through a regression relationship, for estimating the population distribution function and quantiles for a finite-population variable. Phillips (1987) discussed asymptotic theory for autoregressions.

In time series, Tiao & Grupe (1980) investigated the implications, for example in terms of loss of forecasting efficiency, of fitting an ARIMA model when the true model is of a periodic ARMA type; Haggan & Ozaki (1981) obtained a discrete-time time-series model with similar properties to those of nonlinear deterministic vibrations, usually modelled in continuous time; and Hannan & Rissanen (1982) developed a recursive method that allows economical calculation of a criterion used in estimating the order of an ARMA process. Keenan (1985) adapted the Tukey nonadditivity test to a time-series context, creating a test of the linearity of a time-series model as expressed as a Volterra expansion; this was followed up by Tsay (1986), who provided a similar but more powerful version. Li & McLeod (1986) investigated maximum likelihood estimation in fractionally differenced ARMA models, presenting an algorithm and asymptotic theory, and Davies & Harte (1987) examined the power of tests for distinguishing between fractional Gaussian noise and white noise of an AR(1) process, aiming for optimal tests for detecting long-term dependence.

There continued to be influential work about changepoints: Worsley (1986) used maximum likelihood methods to test for a change and to give point and interval estimates of a changepoint in the expected value of a sequence of exponential family variables, with particular emphasis on the exponential distribution; in Raftery & Akman's (1986) Bayesian approach the parameters of interest were the rates of a Poisson process before and after a changepoint and the changepoint itself; James et al. (1987) compared tests, based for example on the likelihood ratio or on recursive residuals, for a change in the mean in the Gaussian case, but without the aim of estimating the changepoint itself; and Kim & Siegmund (1989) covered tests in the context of simple regression when the change might be either in the intercept alone, in which case they also obtained confidence regions for the changepoint, the two intercepts and the common slope, or in both parameters.

A number of prominent papers came under the general heading of regression. In the context of AIC-like criteria for model choice in linear models, Atkinson (1980) showed that the optimum value for the penalty constant depends on the parameters of the true model and on what the model is meant for, such as prediction. Shibata (1981) proposed a method for selecting covariates that asymptotically optimizes a prediction criterion and is asymptotically equivalent to methods including C_p and AIC. Atkinson (1981) developed graphical methods, in the form of half-normal plots based on jackknife residuals and modified Cook's distances, for identifying observations with outlying values in, respectively, the response variable or the covariates. Siegel (1982) described the repeated-medians method, a generalization of the standard univariate median that has a breakdown value of 50%. Miller & Halpern (1982) compared four existing approaches, including those of Cox's (1972), Miller (1976) and Buckley & James (1979, 74), to the case of censoring on the response variable when no parametric model is used for the underlying distribution; the Cox and Buckley & James methods provide the best analysis of the Stanford heart transplant data. Mardia & Marshall (1984) developed a maximum likelihood treatment of linear models with normal noise and covariance structure defined by a parametric model, with particular emphasis on spatial data; they established conditions under which estimators of parameters are asymptotically normal and weakly consistent. Stefanski (1985) obtained a general formulation of the errors-in-variables problem with the parameters of interest estimated by an M-estimator that would be unbiased if there were no measurement error but in general is biased; a less biased estimator is proposed and is illustrated in the context of generalized linear models. For nonlinear regression models Gasser et al. (1986) based a nonparametric estimator of residual variance on pseudo-residuals generated by local linear fitting. Mallet (1986) estimated by nonparametric maximum likelihood the distribution of parameters in random-coefficient regression models and linked the problem to optimal design theory, one consequence being that the optimal distribution has finite support. Longford (1987) developed a scoring algorithm for handling certain linear mixed models, Stefanski & Carroll (1987) derived unbiased scores for generalized linear models with normally distributed measurement errors on the covariates, covering both functional and structural models, and Zeger (1988) considered a log-linear regression model for counts with temporal correlation, with parameter estimation carried out by an estimating-equation approach analogous to quasilikelihood.

Important contributions to multiple testing came from Schweder & Spjøtvoll (1982), who based simultaneous evaluation of a large number of tests on a plot of cumulative p-values, a linear plot indicating that the null hypotheses are all true and the actual number of true number of hypotheses able to be estimated from deviations from linearity, and Hommel (1988), who presented further improvements on the Bonferroni approach, creating a method that permits statements about individual hypotheses as well as about the whole set. Worsley (1982) provided an improved upper bound of the inequality on which the Bonferroni method relies and applied it in contexts such as outlier detection.

In the developing topic of smoothing, Azzalini (1981) applied the kernel method, previously introduced in density estimation, to the estimation of cumulative distribution functions and quantiles, showing that the asymptotically optimal bandwidth is $O(n^{-1/3})$, and Azzalini et al. (1989) used nonparametric regression to check the validity of a parametric model, by developing a pseudo likelihood ratio test and a graphical approach that compare a nonparametric estimate with confidence bands simulated for the parametric model.

Bootstrap and related methods represented an exciting new direction: Efron (1985) showed how confidence intervals based on the bias-corrected percentile bootstrap method can in small-sample nonlinear problems improve substantially upon those based on standard approximations; Davison et al. (1986) demonstrated how balancing, that is, ensuring that all data points are included equally often in the totality of all bootstrap samples, together with the use of approximations with known moments that do not involve simulation, can improve bootstrap methods; Silverman & Young (1987) investigated how to judge whether or not it pays to incorporate smoothing in the bootstrap, using simulation from an either so-called shrunk or unshrunk kernel density estimator; Davison & Hinkley (1988) showed how bootstrap applications involving estimators based on sample moments or monotonic estimating equations can greatly benefit from the use of accurate saddlepoint approximations; Beran (1987) showed that the accuracy of bootstrap confidence sets can be improved by one or more iterations of prepivoting, that is, transforming the boundary of a confidence set by its estimated bootstrap cumulative distribution function; and Hall & Martin (1988) provided a unifying approach to various inference procedures including bootstrap resampling that leads on in a natural way to iteration. Burman (1989) showed how v-fold crossvalidation and the repeated learning-testing method can be effective alternatives to the computationally more expensive ordinary crossvalidation, which corresponds to $v = n$, and Besag & Clifford (1989) developed methods for constructing Monte Carlo tests when it is hard to simulate from the null distribution of the test statistic; their approach used Markov chains for which the distribution of interest was the equilibrium distribution so as to generate exchangeable realizations against which the observed value is assessed.

A wide variety of other topics were covered in influential contributions. Aitchison & Shen (1980) and Aitchison (1983) discussed the logistic-normal distribution crucial for facilitating the application of multivariate analysis techniques in a meaningful way to compositional data. Jupp & Mardia (1980) developed a measure of correlation that applies to bivariate directional data but also much more generally, when the items being correlated belong to Euclidean vector spaces or to compact Riemannian manifolds. In optimal design Ford & Silvey (1980) raised the issue of inference and adaptive design and Atkinson (1982) introduced the ideas of optimal design into the balancing of sequential clinical trials. Wieand et al. (1989) developed methods that include the receiver operating characteristic curve as a special case, for use in comparing diagnostic markers. Clarke (1980) considered bivariate regression situations in which it is not clear which variable should be treated as the response and which as the covariate, in which case the so-called reduced major axis is often used as a linear estimator of the relationship between the variables. Kent (1982) studied the asymptotic distribution of the likelihood ratio statistic when the data do not come from the parametric model under test but where the member of the parametric family nearest to the true distribution does satisfy the null hypothesis under test. Kent (1983) used the concept of information gain to define a measure of correlation, rather in the spirit of Nagelkerke (1991, 22). Harrell & Davis (1982) proposed the estimation of quantiles by a linear combination of order statistics that corresponds to an estimator of the expectation of the exact sample quantile. Cox (1983) discussed the effect that overdispersion has on parameter estimation, identifying circumstances in which the effect of overdispersion is detectable and deriving a test for overdispersion. Smith (1985) examined the asymptotic properties of maximum likelihood estimators in a class of models in which regularity conditions fail; for some of the models the asymptotic properties still all hold, for some they hold partially and for some the estimators may not even exist. Application is made to extreme value theory, a topic also considered by Tawn (1988), who provided an extensive treatment of bivariate extreme value distributions, including the development of parametric models for the dependence function, given that the marginal distributions are, without loss of generality, unit exponential. Godambe (1985) obtained a version of the Gauss–Markov Theorem for stochastic processes, based on the identification of optimal unbiased estimating functions. Chaloner & Brant (1988) identified outliers in a linear model by comparing the associated residuals with the posterior distribution of disturbances. Huggins (1989) considered the estimation of the size of a closed population assuming that capture probabilities differed between individuals according to a linear logistic regression model involving covariates; for small-sample inference, parameter estimation is carried out using a conditional bootstrap method. Mardia & Dryden (1989) initiated the modelling strategy for shape by looking for distributions in shape space, in particular by proposing a normal distribution for landmarks. Finally, Tibshirani (1989) imposed the requirement that marginal posterior intervals have accurate marginal coverage in constructing a prior distribution that is noninformative for a single parameter in the presence of nuisance parameters and Tierney et al. (1989) discussed Laplace-type approximations for marginal densities of nonlinear functions of parameters under conditions on the joint distribution of the parameters that lead to asymptotic properties similar to those of the corresponding saddlepoint approximation.

The 1981 volume began with an obituary for E. S. Pearson by Bartlett & Tippett (1981).

In 1987 the number of issues per annum was increased to four and there were about 30 *Miscellanea* items per volume, including a brief item, stimulated by one of Julian Besag's examination students (Besag, 1989), that itself has accumulated 60 citations!

7. The nineteen-nineties

7·1. *Papers in the top* 100

We are now reaching recent times and one would expect the number of relevant papers in the top 100 to fall away, with insufficient time having elapsed for papers to have generated sufficient interest to be rated highly. While this is true, in that only 11 papers qualify for discussion in this subsection, two of them have made major impacts.

Donoho & Johnstone (1994, 4), with nearly 5800 citations, represents a major work in the wavelet literature for estimating a function nonparametrically from noisy data. Cited over 3200 times, Green (1995, 8) introduced the highly influential reversible jump Markov chain Monte Carlo method which facilitates switching among parameter spaces of different dimensions. More detail about these papers is given in § § 10·4 and 10·8.

Perhaps a surprising inclusion is the very short paper by Nagelkerke (1991, 22) about the generalization, from the classical case of linear models to general regression models, of the coefficient of determination; in the latter case the coefficient is defined as a function of the difference between the fitted and null loglikelihoods in such a way that it fits in with the classical definition. Continuing work on survival data is represented by Grambsch & Therneau (1994, 27) and Therneau et al. (1990, 84). In the former paper, a score test for the proportional hazards assumption is shown to be equivalent to a generalized least-squares test on the residuals of Schoenfeld (1982, 55) and to include many well-known tests as special cases. The other paper uses different types of residual for assessing characteristics of survival models, such as the adequacy of the proportional hazards assumption and the functional form of the influence of a new covariate. More on Markov chain Monte Carlo was provided by Carter & Kohn (1994, 41), in the context of linear state-space models with nonconstant coefficients and errors, in the state and observation equations, that are Gaussian mixtures; sampling has to be carried out for all parameters, the hidden states and the indicator variables for the mixture components. The EM algorithm literature was enhanced by Meng & Rubin's (1993, 57) variation, the ECM, or Expectation/Conditional Maximization algorithm, in which a complicated M-step is replaced by a set of simple conditional-maximization steps.

For estimating generalized linear models with random effects, Schall (1991, 67) provided algorithms based on linearization of the link function and related to algorithms for maximum likelihood estimation and restricted maximum likelihood estimation in the linear-model case. Pearl's (1995, 66) paper on the use of diagrams and the associated graphical models in causal inference was supplemented by a discussion involving 13 people, Azzalini & Dalla Valle (1996, 75) investigated the properties of a multivariate version of the skew-normal distribution, with application to body-dimension data from Australian athletes, and Genest et al. (1995, 89) developed semiparametric estimation in copula-based multivariate distributions with particular emphasis on the dependence structure, based on maximizing a pseudo-loglikelihood function.

7·2. *Other influential contributions*

Topics with an increasingly high profile during this decade included aspects of Bayesian analysis, wavelets and nonparametric methods.

In Bayesian analysis Berger & Bernardo (1992) applied the reference-prior approach to scenarios, such as the multinomial case, in which parameters are gathered into groups of different levels of importance for inference with the prior constructed through a series of conditioning steps. Liu et al. (1994) investigated the covariance structure of a Markov chain generated by the Gibbs sampler in missing-data contexts, indicating the advantages offered by Rao–Blackwellization. This postprocessing technique, which typically provides estimators of higher precision than that of standard estimators, was explored also by Casella & Robert (1996) in the context of the accept-reject and Metropolis algorithms. Gelfand et al. (1995) showed how to reparameterize normal linear mixed models by hierarchical centring so as to improve the convergence of computational procedures and Raftery (1996) developed, in the context of generalized linear models, a modification of Laplace's method for obtaining approximate Bayes factors; the method uses maximum likelihood estimates of parameters, the deviance and the observed or expected Fisher information matrix, all of which are accessible from the output from standard software. Bush & MacEachern (1996) suggested, for randomized block experiments, a nonparametric modification of the traditional Bayesian hierarchical model, incorporating a Dirichlet process model for the distribution of the block effects, and Roberts & Tweedie (1996) obtained conditions under which Metropolis–Hastings algorithms do or do not converge at a geometric rate, a property that underpins stability of the estimation procedures and leads to important central limit theorems. Müller et al. (1996) and Denison et al. (1998) developed Bayesian approaches to curve or surface fitting and classification trees respectively; Müller et al. based their approach on multivariate Dirichlet process mixtures of normal distributions and Denison et al. used reversible jump Markov chain Monte Carlo to obtain a posterior distribution on the space of possible trees. Reversible jump was also used by Giudici & Green (1999) and Dellaportas & Forster (1999) in the determination of different types of graphical model; Giudici & Green considered decomposable Gaussian models, represented by their concentration matrices, whereas Dellaportas & Forster modelled high-dimensional contingency tables. Finally, Gelfand & Ghosh (1998) were motivated by drawbacks, in terms of interpretability and computation, underlying the use of Bayes factors to provide a prediction-based approach to model choice founded on formal utility maximization.

Wavelet methods were the subject of well-cited papers by Wang (1995), who developed theory, tests and estimation procedures for the use of the wavelet transform of the data in detecting jumps or cusps in a function, and Percival (1995), who investigated two estimators of the wavelet variance, one based on the discrete wavelet transform and the other on an interpretation of wavelets as a filter. Bruce & Gao (1996) studied aspects of WaveShrink estimators, such as the calculation of ideal thresholds and investigation of finite-sample behaviour, as well as deriving computationally efficient formulae for the bias, variance and L_2 risk of WaveShrink estimators, Clyde et al. (1998) modelled wavelet shrinkage using Bayesian hierarchical models, providing implementation through analytical approximations and computational options, and Abramovich & Silverman (1998) applied two wavelet-based decomposition methods to noise-corrupted inverse problems; the wavelet-vaguelette method is based on the wavelet series for the target function and the vaguelette-wavelet method on the wavelet series for the data.

The estimation of error variance in nonparametric regression was the subject of Hall & Marron (1990), who used a simple scaled sum of squares of residuals, and Hall et al. (1990), who presented difference-based estimators that are optimal in minimizing asymptotic mean-squared error. Also, Fan & Yao (1998) estimated the variance function of a nonparametric regression model by applying local linear regression to the squared residuals, having used local linear regression also to estimate the regression function itself; it is shown that the variance function is thereby estimated asymptotically as well as if the true regression function were known. Hall et al. (1991) used higher-order kernels together with a more accurate than usual asymptotic representation of the optimal bandwidth to obtain an improved bandwidth estimator in kernel-based density estimation, Müller (1991) provided modifications to kernel estimators that are needed in order to provide satisfactory estimates near endpoints of the support, Koenker et al. (1994) designed smoothing splines so as to be robust estimators of quantile functions and Linton & Nielsen (1995) produced a kernel approach, to nonparametric regression involving more than one covariate, that can be used to distinguish among different model structures.

There was again strong activity in logistic regression and related topics. A number of papers considered scenarios involving correlated binary responses: Zhao & Prentice (1990) used a quadratic exponential model parameterized in terms of marginal means and pairwise correlations estimated from a set of score estimating equations; Lipsitz et al. (1991) modelled association between binary variables through the parameterized odds ratios for binary responses at pairs of times and used two sets of estimating equations, one for the regression coefficients and one associated with the set of odds ratios; Fitzmaurice & Laird (1993) modelled the correlation in terms of conditional log odds ratios given all other responses and implemented maximum likelihood estimation; Carey et al. (1993) iterated between estimating equations for estimating logistic regression models for the binary responses given covariates and logistic regression models for each response given other responses in the same cluster to update odds-ratio parameters; finally, a multivariate probit model was developed by Chib & Greenberg (1998), who provided a Bayesian procedure in the form of Markov chain Monte Carlo simulation of the posterior distribution and a maximum likelihood analysis using a Monte Carlo EM algorithm. Goldstein (1991) established a multilevel-modelling approach for discrete responses, as a particular application of a linearization of a nonlinear multilevel model. Kuk & Chen (1992) created a model for censored data by combining a logistic regression model for the probability of the event with a proportional hazards model for the time till the event; parameters are estimated from a Monte Carlo approximation to the marginal likelihood corresponding to elimination of the conditional baseline hazard function which is then estimated using the EM algorithm. Neuhaus et al. (1992) considered logistic regressions with cluster-specific random intercepts whose distribution may be misspecified, this leads to asymptotic biases in the estimators of the cluster-independent regression coefficients, but approximate analysis, backed up by empirical evidence, indicates that the biases will be small. Qin & Zhang (1997) showed that, under case-control sampling, the logistic regression model is equivalent to a two-sample semiparametric model in which the log ratio of the two densities is linear in the data; the validity of the logistic regression model can then be checked with a Kolmogorov–Smirnov-type test. In a similar context Scott & Wild (1997) extended Breslow & Cain's (1988) finding that for certain logistic models the prospective fitting of a pseudo-model provides maximum likelihood estimates; the new developments were that an iteration based on this finding provides maximum likelihood estimates for any model and that the methodology can be extended to cover stratified case-control studies.

In general regression and longitudinal data analysis influential contributions came from Rotnitzsky & Jewell (1990), who developed tests of hypotheses, including versions of the Wald and score tests incorporating working correlation structures, in the context of semiparametric marginal generalized linear models for cluster-correlated data, and Davidian & Gallant (1993), who took a maximum likelihood approach to the nonlinear mixed effects model, popular in areas such as pharmacokinetics, in which the only assumption made about the random effects distribution is that of smoothness. Breslow & Lin (1995), followed up in Lin (1997), used first- and second-order Laplace approximations to the integrated likelihood to investigate asymptotic biases of approximate estimators for a class of generalized linear models. Reilly & Pepe (1995) studied regression scenarios in which some observations are complete but for others some covariates are missing although there may be auxiliary variables that are not in the model but are informative about the missing covariates and are known for all observations; the regression parameters are estimated through a mean score vector resulting from nonparametric incorporation of the auxiliary data. Crowder (1995) showed that in some cases of the Liang & Zeger (1986, 1) working-correlations approach asymptotic properties can break down, especially for any parameters in the working correlation matrix, Molenberghs et al. (1997) considered longitudinal ordinal data with nonrandom drop-out, modelled by logistic regression, using the EM algorithm to maximize the likelihood; as usual in contexts with nonrandom missingness, inference involves uncheckable assumptions. For a varying-coefficient regression model with repeated measurements, within-subject correlation and mean function linear in a set of smooth nonparametric functions, Hoover et al. (1998) estimated the latter using smoothing splines and locally weighted polynomials, of which kernel estimators form a special case; the necessary smoothing parameters were selected by crossvalidation and asymptotic properties were established for the kernel case. Wolfinger (1993) applied an approximation based on Laplace's method to the marginal distributions of data from nonlinear mixed models, with links to parameter-estimation methods such as those of Schall (1991, 67) mentioned in § 7·1. Pourahmadi (1999) used covariates to model a covariance matrix as well as the mean, through a parameterization of the covariance matrix that automatically enforces positive-definiteness.

Leading contributions to design of experiments were in classical design by Wu (1993), who provided a way of constructing supersaturated designs involving factors with two levels by supplementing a saturated design with columns defined by its partially aliased interactions, and in optimal design by Mentré et al. (1997), in the context of finding, given a cost constraint, D-optimal designs for the parameters of random effects models, namely the fixed effects and the variance of the random effects distribution.

There were influential advances in survival analysis: Wei et al. (1990) provided simple methods for making inference about a subset of the parameters in additive failure-time models, with the remaining parameters treated as nuisance parameters, Prentice & Cai (1992) developed estimators of survivor functions of bivariate and, more generally, multivariate failure times, and Lin et al. (1993) proposed checking the proportional hazards model using cumulative sums of martingale residuals, these processes being approximately zero-mean Gaussian processes under the model; this allows the observed process to be compared with realizations simulated from the null model. Cheng et al. (1995) considered models in which an unknown transform of the survival time is linearly related to the covariates and estimating equations are developed for examining the regression parameters. For the additive risk model Lin & Ying (1994) provided procedures for making inferences about the regression parameters given that the baseline hazard function is unspecified and McKeague & Sasieni (1994) described a partly parametric version in which the influence of some covariates is time-varying, modelled nonparametrically, and that of the rest remains constant, represented by regression parameters that are of prime interest.

Two particularly well-cited papers in general theory are Barndorff-Nielsen (1991) and Firth (1993). The former paper studied a modified form of the signed loglikelihood ratio for a scalar parameter of interest that asymptotically has a standard normal distribution with error of order $O(n^{-3/2})$; the expression for the new form contains a term which can itself be interpreted as a test statistic. The second paper showed how to remove the first-order term from the asymptotic bias of maximum likelihood estimators of parameters in regular scenarios by modifying the score function in a way that corresponds, in the case of exponential family distributions expressed in terms of canonical parameters, to adding the Jeffreys prior to the likelihood as a penalty. Contributions on the calculation of tail probabilities were provided by DiCiccio & Martin (1991) and Fraser et al. (1999).

In time series, Auestad & Tjøstheim (1990) was the first paper to address the order-determination problem in nonlinear autoregressive modelling without assuming the functional form of the autoregression, that is, in a nonparametric context. Shephard & Pitt (1997) used computational techniques involving Taylor expansions and blocked Markov chain Monte Carlo samplers to develop likelihood and Bayesian analysis for non-Gaussian state space scenarios. This paper follows on from a number of other contributions to the use of simulation techniques to handle state-space models: Shephard (1994) used simulation to extend standard filtering and smoothing procedures to a class of non-Gaussian models, such as models involving 'Student's' t or Gaussian mixture disturbances and Markov switching models; de Jong & Shephard (1995) proposed multistate samplers for the disturbances rather than for the states on the grounds of simplicity and economy; Carter & Kohn (1996) developed a Markov chain Monte Carlo procedure for Bayesian analysis of a class of state-space models that includes changepoint models; and Durbin & Koopman (1997) obtained accurate approximations to loglikelihoods for linear state-space models with non-Gaussian distributions by using simulation methods that were computationally more economical than Markov chain Monte Carlo techniques. Away from simulation techniques, Koopman (1993) examined a smoothed estimator of the disturbance vector and subsequently of the state vector in a linear state-space model. In the context of nonlinear dynamical systems, Fan et al. (1996) used locally polynomial regression to estimate conditional densities and their square roots, and their partial derivatives; methods were also developed to assess the sensitivity of a stochastic dynamical system to its initial value. In spatial statistics, Besag & Kooperberg (1995) discussed conditional autoregressions and their often advantageous limiting forms given by intrinsic autoregressions, as well as describing ways of combatting problems that arise with applications involving small arrays or nonlattice domains. Finally, in spatio-temporal modelling, Wikle & Cressie (1999) provided a method for handling large space-time datasets using the Kalman filter; a dimension-reduction procedure is developed to overcome the computational difficulties inevitable with large spatial domains and parameters are estimated by the method of moments.

There were leading contributions in a number of other established areas: in multiple testing Rom (1990) improved upon Hochberg's (1988, 15) procedure by enlarging the rejection region so that the Type-I error equals the nominal value; in sample surveys Rao et al. (1990) obtained design-based ratio and difference estimators of a population distribution function using auxiliary information in the form of values of a covariate known for each member of the population; also in finite-population inference, Chen & Qin (1993) showed how population characteristics of an auxiliary variable can be exploited in empirical likelihood methods for making inferences about the variable of interest; and for surveys with missing values Rao & Shao (1992) developed a jackknife variance estimator for stratified multi-stage surveys based on a completed dataset involving just single imputations, which in practice would have to be flagged, generated by a particular hot-deck approach. In the context of errors-in-variables models Nakamura (1990) introduced so-called corrected score functions, whose expectations under the measurement error distribution equal the usual score functions based on the true independent variables; the corrected score functions lead to corrected estimators that may be consistent but, asymptotically, cannot have smaller variance than those of the maximum likelihood estimators in the absence of measurement error. For work on semiparametric models for which the p-dimensional parameter of interest is estimated consistently through estimating functions, Parzen et al. (1994) developed resampling methods for obtaining interval estimators based on pivotal estimating functions; that is, the functions' distribution can be generated by a p-dimensional random vector whose distribution is completely known.

A wide variety of other influential articles include Besag & Clifford (1991), concerning sequential Monte Carlo tests in which an observed value of a test statistic is compared against a sequence of values simulated according to the null hypothesis; for a number of strategies for stopping the sequence to save computational effort the rules for stating the p-value are described along with the corresponding null distribution; the paper also contains a method for when the null hypothesis provides only the equilibrium distribution of a Markov chain, which is what is simulated. Meng & Rubin (1992) proposed a complete-data loglikelihood procedure for obtaining significance levels from multiply-imputed data, Chao & Yang (1993) developed stopping rules for use in testing for bugs in software and for estimating the number of bugs still undetected; the model allows for different failure rates among the bugs and for recapture-debugging, in which, although a bug is detected and corrected, any occasion on which the bug would have recurred is flagged. Liu & Rubin (1994) and Liu et al. (1998) described further advances on the EM algorithm; Liu & Rubin's Expectation/conditional maximization Either, ECME, algorithm allows some of the conditional maximization steps in the ECM algorithm, which maximize constrained expected complete-data loglikelihoods, to be replaced by steps that maximize the correspondingly constrained actual loglikelihood, thereby providing convergence that is still monotonic but generally faster; in the PX-EM algorithm Liu et al. also aim to accelerate EM, this time by improving the efficiency of the maximization step thanks to a covariance adjustment provided by a parameter expansion of the complete-data model. Little (1994) applied pattern-mixture models, under which the joint distribution of a set of variables depends on the missingness pattern, to bivariate normal data with one variable missing from some of the observations; except under special circumstances the missingness process is nonignorable and unverifiable assumptions have to be imposed if all parameters are to be identifiable. Konishi & Kitagawa (1996) proposed AIC-like information criteria for evaluating models when the specified family of distributions does not contain the true model; different versions are presented corresponding to different model-estimation paradigms, such as ordinary, robust and penalized maximum likelihood and Bayesian inference. In the context of extreme-value threshold exceedance models, Ledford & Tawn (1996) developed an improved multivariate model for joint tail estimation when the underlying variables are almost independent and Smith et al. (1997) considered extremes of univariate time-series data, assuming that the time series is Markovian and modelling the corresponding transition distributions using models for bivariate extremes. Gallant & Long (1997) investigated the properties of a minimum chi-squared estimator for the parameters of an ergodic system of stochastic differential equations with state that is observed at discrete time-points. Catchpole & Morgan (1997) obtained necessary and sufficient conditions for the parameter redundancy of a wide class of nonlinear exponential family models, the indicator of redundancy being whether or not the matrix of derivatives of the means with respect to the parameters is of symbolic full rank, a property that can be checked with a symbolic algebra package. Foster & Vohra (1998) attacked the problem of forecasting the probability of an event so as to match the event's empirical probability, showing that approximate calibration is possible if the forecaster is allowed to randomize; the work was stimulated by Oakes's (1985) finding that no deterministic forecasting sequences can be calibrated for all sequences. For the analysis of spatially dependent count data Wolpert & Ickstadt (1998) developed Bayesian hierarchical models that are doubly stochastic Poisson processes the intensities of which are mixtures of inhomogeneous, infinitely divisible random fields; posterior distributions are explored using a Metropolis/Gibbs approach with data augmentation. Finally, Basu et al. (1998) proposed a robust parameter estimation method based on a class of density-based divergence measures, to be minimized and parameterized by a single parameter one value of which corresponds to maximum likelihood; Baggerly (1998) showed that the empirical likelihood method as applied to an n-sample can be regarded as estimating cell probabilities in an n-cell contingency table so as to optimize a goodness-of-fit criterion such as those of the Cressie–Read class; Barnard & Rubin (1999) calculated small-sample values of the degrees of freedom for the t-distribution used with multiply-imputed data when a normal distribution would be used with complete data; Liu & Pierce (1994) proposed improvements to Gauss–Hermite numerical integration; and Frangakis & Rubin (1999) considered the impact of all-or-none compliance and subsequent missing outcomes on the estimation of the intention-to-treat effect of assignment in randomized studies.

8. The period 2000–2008

8·1. *The leading papers*

For a paper to appear among the top 100 it has to have gathered at least 487 citations and it is therefore not very surprising that only one of the papers published in the year 2000 or later has achieved this level of impact. In that paper Lo et al. (2001, 80) attacked the thorny problem of testing the number of components in a normal mixture. The usual likelihood ratio test runs into difficulties because of a lack of identifiability under the null hypothesis of a more parsimonious model, and instead Lo et al. developed a likelihood ratio statistic based on Kullback–Leibler divergence for which the null distribution is a weighted sum of independent χ_1^2 distributions. Although no other paper of this period appears in the top 100, some of the more recent papers appear nevertheless to be well on the way to making a significant impact; in the rest of this subsection we discuss the leaders in terms of current citations.

George & Foster (2000) performed variable selection in the normal linear model, based on penalized least-squares criteria that have a hierarchical Bayes interpretation, and hyperparameters are estimated by empirical Bayes. For data from an exponential family distribution, DiMatteo et al. (2001) investigated Bayesian curve-fitting using splines with the number and locations of knots treated as free parameters; reversible jump Markov chain Monte Carlo is used to handle the variable number of knots. Croux & Haesbroeck (2000) based robust principal components analysis on robust estimation of the covariance or correlation matrix, deriving influence functions and asymptotic variances of the corresponding eigenvalues and eigenvectors. Continuing related work in Carter & Kohn (1994, 41), de Jong & Shephard (1995) and Durbin & Koopman (1997), Durbin & Koopman (2002) presented an approach to simulating data-conditional distributions of parameters and disturbances in linear Gaussian state-space models with and without an additive linear regression term in the observation equation.

The next two papers, already showing their influence only a few years after publication, represent areas of very high current interest, namely sparsity and multiple testing: Yuan & Lin (2007) developed a sparse, shrinkage-based, positive-definite estimator of the concentration matrix for the Gaussian graphical model, with the degree of shrinkage dictated by a BIC-like penalty; and, for improving the linear step-up procedure for controlling the false discovery rate in multiple testing first proposed in Benjamini & Hochberg (1995), Benjamini et al. (2006) developed adaptive procedures based on estimating the number of true null hypotheses.

8·2. *Other influential contributions*

There was considerable activity in Bayesian methods. To facilitate the analysis of Dirichlet process mixture models by so-called conditional methods, Ishwaran & Zarepour (2000) developed random probability measures that are easy to construct and are good approximations to the Dirichlet process; Papaspiliopoulos & Roberts (2008) then devised the new technique of retrospective sampling through which the approximations can be avoided; and the conditional approach together with retrospective sampling was also adopted by Dunson & Park (2008) in developing kernel stick-breaking processes for uncountable collections of dependent random probability measures. Roberts & Stramer (2001) introduced a new Markov chain Monte Carlo approach for analysing discretely observed diffusion processes, with the paths between any two data points regarded as missing data; because of the dependence between the missing paths and the volatility of the diffusion, convergence of algorithms can be slow, and to combat this a transformation of the diffusion is proposed that removes the problematic dependence. Chopin (2002) presented a particle-filter approach, normally employed in dynamic scenarios, that can provide efficient estimation of a posterior distribution in a nonsequential setting. In the context of Metropolis–Hastings algorithms, Møller et al. (2006) described an auxiliary-variable approach for sampling from a distribution with an intractable normalizing constant; the proposal distribution of the auxiliary variable is so defined that the problematic normalizing constant cancels out in the overall Metropolis–Hastings ratio. Kennedy & O'Hagan (2000) and Oakley & O'Hagan (2002) developed methods for analysing complex computer codes: based on the representation of prior beliefs about the codes by Gaussian processes, Kennedy & O'Hagan proposed combining expensive runs of complex versions of the codes with cheap runs of simpler approximate versions, in order to improve efficiency; Oakley & O'Hagan examined aspects of the uncertainty distribution induced for the output of a computer model as a result of uncertainty about one or more inputs.

In the perennially active area of censored data and survival analysis, Bang & Tsiatis (2000) showed how to use a simple weighted complete-case estimator and refinements thereof to estimate mean total medical costs when there may be right-censoring, and Tsiatis & Davidian (2001) developed a method, based on the conditional score approach of Stefanski & Carroll (1987), for estimating the parameters of the proportional hazards model in the presence of some time-independent covariates along with time-dependent covariates modelled by a linear mixed effects model; no distributional assumption about the random effects is required. Jin et al. (2003) developed a class of rank-based monotone estimating functions for the semiparametric accelerated failure-time model, with estimators that are obtained by linear programming, and Abbring & van den Berg (2007) provided a formal justification of the common practice of using a gamma distribution in hazard models with proportional unobserved heterogeneity, in the analysis of duration time. Of particular note is Zhang & Lu (2007), who combined the highly topical lasso technique for variable selection with one of the Journal's most frequently investigated models, that of proportional hazards; the source of estimators is a penalized log partial likelihood with adaptively weighted L_1 penalties on the regression coefficients.

In the area of repeated measures, Huang et al. (2002) used function approximation through basis expansions for estimating the parameters of a varying-coefficient model, with inference procedures based on subject-level bootstrap resampling. He et al. (2002) developed M-estimators for models for longitudinal data that are semiparametric in including, additively, a nonparametric smooth function for which the paper uses a regression spline.

A number of influential papers concerned information criteria: Konishi et al. (2004) developed a version of BIC for handling models estimated by methods such as maximum penalized likelihood and applied the methods to radial basis function network generalized linear models; in the context of model selection for linear mixed effects models with clustering, Vaida & Blanchard (2005) proposed the conditional Akaike information for use when focussing on particular clusters, prediction at that level being conditional on the clusters with the random effects acting as parameters; Yang (2005) attempted, without success, to formulate an approach to regression modelling that shared the strengths of AIC, which is minimax-rate optimal for estimating the correct regression function, and BIC, which consistently selects the right model; Ando (2007) proposed an information criterion for assessing the goodness of certain predictive distributions when the specified family of distributions does not contain the true model; and Chen & Chen (2008) provided extensions of BIC for large model spaces, based on a certain type of prior distribution on the model space.

In nonparametric methods there was considerable emphasis on penalized splines: Crainiceanu et al. (2005) used likelihood ratio tests and restricted likelihood ratio tests to test for parametric regression versus a general model based on penalized splines, and Breidt et al. (2005) used semiparametric penalized spline regression on an auxiliary variable in constructing a modification of the Horvitz–Thompson estimator of a finite population total. Theoretical properties of the approach were examined by Hall & Opsomer (2005) and Li & Ruppert (2008): the former paper obtained consistency results for the penalized spline regression estimator and showed that the estimator converged at the optimal nonparametric rate; for a univariate version based on B-splines, the second paper showed that penalized splines behave similarly to Nadaraya–Watson kernel estimators, and also established the asymptotic distributional properties of the estimator.

The topic of composite likelihood stimulated influential contributions, from Cox & Reid (2004), who explored the use and asymptotic properties of approximate likelihoods based on univariate and bivariate marginal distributions for making inferences about high-dimensional distributions, and Varin & Vidoni (2005), who provided an integrated approach to inference and model selection, the latter based on an information criterion derived from a Kullback–Leibler divergence appropriate for composite likelihood.

There were innovative contributions with biomedical applications. In handling case-control studies of possible gene-environment associations with disease when genetic and environmental factors can be assumed not to be independent in the general population, Chatterjee & Carroll (2005) investigated semiparametric estimation of logistic regression parameters that exploit gene-environment independence. Green & Mardia (2006) devised procedures for matching configurations of points, using a hierarchical model based on a Poisson process for the locations of the hidden true points; the method was applied to two contexts in bioinformatics, matching protein gels in two dimensions and aligning active sites of proteins in three dimensions.

In multivariate analysis Pourahmadi (2000), Wu & Pourahmadi (2003) and Huang et al. (2006) are representative of a number of papers around this time on the estimation of covariance matrices following on from Pourahmadi (1999) and concentrating on the Cholesky decomposition of the inverse matrix: for example, Huang et al. carried out penalized maximum likelihood estimation with L_1 and L_2 penalties imposed on the loglikelihood function leading to stable, and in the case of the L_1 penalty parsimonious, shrinkage-based estimators. Wong et al. (2003) took a Bayesian approach to the estimation of an inverse covariance matrix, with parsimonious parameterization encouraged by a prior that allows off-diagonal elements of the concentration matrix to be zero.

A miscellany of further influential work includes James et al. (2000), who used a reduced rank model involving a spline basis to obtain principal component functions for a set of curves when the set of points at which the curves are measured may be sparse and may not be the same from curve to curve, and Lee & Nelder (2001), who generalized the method of restricted maximum likelihood so as to be applicable to their hierarchical generalized linear models. Genovese et al. (2006) developed a method for multiple testing that controls the false discovery rate while incorporating prior information about the hypotheses in the form of p-value weights related to the perceived relative likelihoods that the individual hypotheses are false. Robins & Wang (2000) obtained asymptotic variances of estimators based on simple and multiple imputation; the variance estimator is consistent even if the models for imputation and analysis are misspecified and mutually incompatible. Fuentes (2002) presented various parametric approaches to the estimation of the spectral density of a nonstationary spatial process, establishing the properties of the estimators using shrinking asymptotics, as well as using a convolution of locally stationary processes to create a new class of nonstationary processes. Tsiatis & Mehta (2003) showed that, for monitoring clinical trials, standard group-sequential tests based on the sequentially computed likelihood ratio statistic are uniformly more efficient than adaptive designs in which the sample size is modified on the basis of sequentially computed observed treatment differences. In shape analysis, Kent & Mardia (2001) established the definitive coordinate representation required for studying bilateral symmetry. In causal inference, Robins et al. (2003) established the nonexistence of uniformly consistent causal inference procedures even though there are examples in which pointwise consistency obtains; in other words there may be 'tests guaranteed to yield correct answers with an infinite sample size [but no test that] can make such guarantees in finite samples, even approximately'. In survey analysis with auxiliary variables Deville & Tillé (2004) introduced the cube method for choosing approximately balanced samples, with the aim of thereby reducing the variances of estimators of population totals of variables of interest. In the context of Gaussian graphical models, Drton & Perlman (2004) obtained conservative simultaneous confidence intervals for the pairwise partial correlations making up the concentration matrix, resulting in a method of model selection that controls the overall rate of wrongly including an edge.

9. CURRENT DIRECTIONS

In this section we pick out, from each of the three most recent volumes, covering the years 2009–2011, a few papers dealing with prominent current topics, which are already accumulating citations and which may well turn out to be particularly influential.

The continuing substantial activity in Bayesian statistics includes Beaumont et al. (2009), on the approximate Bayesian computation method used when there is no closed form for the likelihood; a sequential version is provided that avoids a bias in the resulting approximation to the target posterior distribution. Chopin & Robert (2010) studied the asymptotic properties and applicability of Skilling's (2006) nested sampling method. Fearnhead et al. (2010) described a new particle filter to be used as a smoother, in that it aims to estimate previous states from a batch of noisy time-series data; the computational cost of the method is less by an order of magnitude than that of most other smoothers. In further work on particle filters, Poyiadjis et al. (2011) developed methods for computing the score vector and observed information matrix recursively in the context of nonlinear, non-Gaussian state-space models. Carvalho & Scott (2009) presented a method for model selection in decomposable Gaussian graphical models that involves priors that handle the issue of multiple testing as well as a version of the hyper-inverse Wishart prior that is suitable for restricted covariance matrices. In Bayesian nonparametrics, Rodriguez et al. (2009) developed a hierarchical model for functional estimation based on independent Dirichlet process mixtures of Gaussian distributions for the joint distribution of predictors and outcomes, and Dunson (2009) sought a prior for an unknown random effects distribution within a hierarchical model; a local partition process prior was adopted that induces dependent local clustering.

Prominent papers in design include Bingham et al. (2009), who developed a class of designs, ranging from Latin hypercube designs to two-level fractional factorial designs as special cases, with computer experiments in mind.

Another area of increasing interest is that of sparsity, with Huang et al. (2009) basing variable selection in the linear model on regression with a group-bridge penalty consisting of terms that group together parameters in a way consistent with known groupings of the predictors, and Carvalho et al. (2010) estimating sparse systems using the robust so-called horseshoe estimator that arises from a prior based on multivariate normal scale mixtures. Bhattacharya & Dunson (2011) carried out sparse modelling of high-dimensional covariance matrices with Bayesian latent factor models; a multivariate gamma process shrinkage prior is assumed for the factor loadings and an adaptive Gibbs sampler is used to truncate automatically the infinite loading matrix that the prior allows. In jointly analysing several Gaussian graphical models that share some of the variables and some of the dependence structure, Guo et al. (2011) took an approach involving a hierarchical penalty designed to remove common zeros in the concentration matrices. A number of topical tools for analysing data of very high dimension were examined, including the smoothly clipped absolute deviation method (Wang et al., 2007), the lasso (Hans, 2009; Belloni et al., 2011; Bien & Tibshirani, 2011) and the Dantzig selector (James & Radchenko, 2009). In particular, Bien & Tibshirani (2011) attacked the estimation of high-dimensional covariance matrices.

False discovery rates continue to inspire important work, including Efron & Zhang (2011), Schwartzman & Lin (2011) and Siegmund et al. (2011): Efron & Zhang investigated so-called copy number data from 5000 chromosome marker positions on 150 subjects, estimating false discovery rates for each position and subject, as well as estimating the number of subjects carrying a variant at a given position; with the assumption of a negative binomial distribution for the number of false discoveries, Schwartzman & Lin investigated the distribution of the standard false discovery rate estimator in the presence of correlation; and Siegmund et al. adapted the method to scanning statistics, grouping neighbouring rejections together and regarding them as a single discovery in view of the likely local dependence. The topic of composite likelihood was addressed by Mardia et al. (2009), who showed that, in the context of closed exponential families, the maximizers of the composite likelihood and the full likelihood are the same. There was also continuing interest in penalized spline regression, with Kauermann & Opsomer (2011) describing likelihood-based criteria for choosing the number of basis functions and Claeskens et al. (2009) adding to the underlying theory; they showed that, depending on factors such as the number of knots and the penalty function, the theoretical properties are similar to those of either regression splines or smoothing splines.

Although sparsity and false discovery rates are currently among the most topical areas of statistical research, they both owe their origins to papers from past decades (Tibshirani, 1996; Benjamini & Hochberg, 1995), and the same is true of the final two recent papers to be mentioned. Kosmidis & Firth (2009) extended Firth's (1993) work on reducing the bias of maximum likelihood estimators to a wide range of generalized nonlinear models, and Crump et al. (2009), by some distance the currently most highly cited paper from 2009, includes, as a key reference in its investigation of the problems that lack of overlap in the covariate distribution cause in the estimation of average treatment effects, Rosenbaum & Rubin's (1983, 2) seminal paper on the propensity score, which ranks second on the main list.

10. The ten most cited papers

10·1. *Liang & Zeger (1986, 1)*

Let (y_{it}, x_{it}) be a response and a vector of covariates for subject i at time t, where $t = 1, \ldots, n_i$ and $i = 1, \ldots, K$. The repeated measures on a given subject inevitably create correlation, and implementation of likelihood methods is difficult except in special cases. Let $Y_i = (y_{i1}, \ldots, y_{in_i})^{\mathrm{T}}$ and let $X_i = (x_{i1}, \ldots, x_{in_i})^{\mathrm{T}}$ be for subject i respectively the $n_i \times 1$ vector of responses and the $n_i \times p$ matrix of covariate values. An exponential family marginal distribution is assumed for y_{it}:

$$f(y_{it}) = \exp\,[\{y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})\}\phi],$$

where $\theta_{it} = h(\eta_{it})$ and $\eta_{it} = x_{it}^{\mathrm{T}}\beta$, so that

$$E(y_{it}) = a'(\theta_{it}), \quad \mathrm{var}(y_{it}) = a''(\theta_{it})/\phi.$$

The parameter of main interest is β.

To avoid the complications caused by intrasubject correlation, estimating equations are developed, based on score functions reflecting the ‘independence working assumption’ that there is no such correlation. These equations are

$$U_I(\beta) = \sum_{i=1}^{K} X_i^{\mathrm{T}} \Delta_i S_i = 0, \tag{1}$$

where $\Delta_i = \mathrm{diag}(d\theta_{it}/d\eta_{it})$ and $S_i = Y_i - a_i'(\theta)$, in which $a_i'(\theta)$ is the vector with tth element $a'(\theta_{it})$. The estimator $\hat{\beta}_I$ is the solution of (1). It is shown that $\hat{\beta}_I$ is typically consistent for β and has asymptotically a normal distribution with a stated covariance matrix for which there is an explicit estimator of a so-called sandwich form.

More general versions of the estimating equations take some account of intrasubject correlations; for example, in the so-called exchangeable case the correlations are all assumed to be equal, the common value being a parameter to be estimated. Asymptotic properties are established for the case of each working correlation structure and the paper includes numerical illustration of efficiencies for different combinations of true and working correlation structures.

10·2. *Rosenbaum & Rubin (1983, 2)*

Consider two treatment groups, made up of treated units ($z = 1$) and control units ($z = 0$). Let x denote covariates and let (r_1, r_0) be the pair of responses that would be observed given either treatment; in practice of course only one of these responses is observed for a given unit. A balancing score is a function, $b(x)$, of x such that x is independent of z, given $b(x)$. Let $e(x)$ denote $\mathrm{pr}(z = 1 \mid x)$. Then $e(x)$ is called the propensity score. It is the coarsest possible balancing score, whereas x itself is the finest. At any value of a balancing score, in particular of the propensity score, the difference between the treatment and control means is an unbiased estimator of the average treatment effect at that value, provided that the treatment assignment process is strongly ignorable, in that, for any x for which the propensity score is strictly between zero and one, the treatment assignment is conditionally independent of the possible responses, (r_1, r_0), given x. In these circumstances, unbiased estimators of the population treatment effect, $E(r_1 - r_0)$, are available via various routes, such as pair-matching of treatments and controls on the balancing score, subclassification on the balancing score or covariance adjustment on the balancing score.

Implementation of these properties is described in detail for the case of the propensity score and the resulting benefits are identified. For example, adjustment for the propensity score is important in analysing observational studies and, for the scenario of using the propensity score to construct matched samples, numerical evidence is provided of the reduction that can be available in the bias of the estimator of the treatment effect.

10·3. *Phillips & Perron (1988, 3)*

As the number of citations for this paper suggests, the problem of testing for a unit root is of great interest in time series, especially to econometricians. In the simplest scenario, the time series $\{y_t\}$ is generated by

$$y_t = \alpha y_{t-1} + u_t,$$

for $t = 1, \ldots, T$, with $\alpha = 1$ and y_0 is arbitrary, where the $\{u_t\}$ are innovations satisfying certain technical conditions. It is of interest to know whether or not it is plausible that $\alpha - 1 = 0$. For stationarity of the series it is required that $\alpha < 1$. The paper considers two fitted regressions:

$$y_t = \hat{\mu} + \hat{\alpha} y_{t-1} + \hat{u}_t, \tag{2}$$

$$y_t = \tilde{\mu} + \tilde{\beta}(t - T/2) + \tilde{\alpha} y_{t-1} + \tilde{u}_t, \tag{3}$$

in which $(\hat{\mu}, \hat{\alpha})$ and $(\tilde{\mu}, \tilde{\beta}, \tilde{\alpha})$ are the corresponding ordinary least-squares estimators. The limiting distributions of $T(\hat{\alpha} - 1)$, $T(\tilde{\alpha} - 1)$ and the regression t statistics of $\hat{\alpha}$ and $\tilde{\alpha}$, together with $\hat{\mu}$, $\tilde{\mu}$ and $\tilde{\beta}$, are obtained under the hypothesis that $\alpha = 1$, $\mu = 0$ and $\beta = 0$. The unknown variance parameters in the models lead to dependencies that are removed, asymptotically, by simple transformations to so-called Z statistics that are used to test the unit-root hypothesis. Asymptotic power functions are obtained for local alternatives to $\alpha = 1$ and empirical comparisons are made with the competing test of Said & Dickey (1984, 23). Between the two models in (2) and (3) the approach allows one to test for a unit root against both stationarity and time-trend alternatives. For more unit-root references see Tong (2001, § 11).

10·4. *Donoho & Johnstone (1994, 4)*

The paper concerns nonparametric estimation of a function f, based on data

$$y_i = f(t_i) + e_i \quad (i = 1, \ldots, n),$$

in which the $t_i = i/n$ are equally spaced on the interval [0, 1], and the e_i are independent and distributed as $N(0, \sigma^2)$. The performance of an estimator $\hat{f}$ is measured by the risk

$$R(\hat{f}, f) = n^{-1} \sum_{i=1}^{n} E\{\hat{f}(t_i) - f(t_i)\}^2.$$

The estimator to be chosen is of the form

$$\hat{f}(\cdot) = T\{y, d(y)\}(\cdot),$$

in which $d(y)$ is a data-adaptive version of a spatial smoothing parameter δ, which must be chosen. This structure includes methods such as piecewise-polynomial fitting, variable-knot spline fitting and variable-bandwidth kernel estimators.

The strategy is to target an oracle that will identify the best δ for the true underlying f; such a δ would provide an ideal $\hat{f}$ but this strategy is of course not available in practice. The paper adopts a selective wavelet approach, which has the advantage that one can evaluate precisely the degree to which one can approach ideal performance when no oracle is available and choice of δ, which for wavelets defines the optimum degree of shrinkage of wavelet coefficients, has to be based on the data alone. A data-based method, RiskShrink, is developed that is shown theoretically to mimick the performance of an oracle 'as well as it is possible to do', with explicit upper bounds on risk as a multiple of the ideal risk. The method is consequently superior to data-based versions of the other approaches mentioned above; given the appropriate oracle the ideal versions are not much better than the wavelet method, also with an oracle, and there is no proof that the performance of the data-based versions can approach that of the ideal versions. That claim is justified by extensive theoretical analysis of the methodology. This is supplemented by practical application to noisy versions of a number of known functions, which allows a comparison between the ideal estimators, which benefit from the assistance of an oracle, and the data-based alternatives.

10·5. *Hastings (1970, 5)*

One way of creating realizations from a probability distribution $\pi = \{\pi_i\}$ on the finite state space $\{0, 1, \ldots, S\}$ is to generate realizations of an irreducible Markov chain with transition matrix $P = \{p_{ij}\}$ for which the equilibrium distribution is π. A sufficient condition on P for this to work is the reversibility property of detailed balance:

$$\pi_i p_{ij} = \pi_j p_{ji},$$

for each i and j. The version of this known as the Metropolis algorithm (Metropolis et al., 1953) first uses a symmetric transition matrix $Q = \{q_{ij}\}$ to propose a move from state i to state j and then accepts the move with probability α_{ij}, with

$$\alpha_{ij} = \min\left(1, \frac{\pi_j}{\pi_i}\right).$$

The corresponding p_{ij} is just $q_{ij}\alpha_{ij}$, for $j \neq i$. It is easy to show that this P satisfies detailed balance.

The Hastings paper showed how to extend the method dramatically, to more general forms of the proposal and acceptance probabilities, and to multi-dimensional, infinite and continuous state spaces. The resulting wide class of Metropolis–Hastings algorithms have greatly facilitated computational Bayesian methods, especially with target distributions that have complicated normalizing constants. The paper describes application to the generation of random orthogonal and unitary matrices. Peskun (1973), also well cited, discussed optimization of the acceptance probability for given proposal distribution so as to maximize the precision of estimators obtained from the resulting chain. The optimal formula, mentioned by Hastings but proved optimal by Peskun, is

$$\alpha_{ij} = s_{ij}/(1 + t_{ij}),$$

where $t_{ij} = \pi_i q_{ij}/(\pi_j q_{ji})$ and $s_{ij} = 1 + \min(t_{ij}, t_{ji})$.

10·6. *Shapiro & Wilk* (1965, 6)

Shapiro & Wilk's W test statistic for normality is calculated as follows from an ordered version of a random sample: $x_{(1)}, \ldots, x_{(n)}$. Let S^2 be the usual corrected sum of squares.

(i) If $n = 2k$ is even, compute

$$b = \sum_{i=1}^{k} a_{n-i+1}(x_{(n-i+1)} - x_{(i)}),$$

for certain constants a_{n-i+1}, which are tabulated in the paper.

(ii) If n is odd, equal to $2k + 1$, the same formula applies, since $a_{k+1} = 0$ when $n = 2k + 1$.

Finally, calculate $W = b^2/S^2$.

The formula for b can also be written as $b = \sum_{i=1}^{n} a_i x_{(i)}$, and the n-vector $a = (a_i)$ is given by

$$a^{\mathrm{T}} = m^{\mathrm{T}} V^{-1} / (m^{\mathrm{T}} V^{-1} V^{-1} m)^{-1/2},$$

in which m denotes the vector of expected values of n standard normal order statistics and V denotes the corresponding covariance matrix. Clearly, $a_i = -a_{n-i+1}$ and the statistic W is location-scale invariant, thereby being available for a test of the composite null hypothesis of normality; there is no need to estimate parameters, in contrast to other procedures such as the Kolmogorov–Smirnov and chi-squared tests.

The paper contains some theoretical results, for instance that $na_1^2/(n-1) \leqslant W \leqslant 1$, approximations are proposed for the elements of a, some percentage points for W under the null hypothesis are tabulated, as are values of elements of a for a range of n, and the method is illustrated on examples as well as being compared empirically with competing tests.

10·7. *Rubin (1976, 7)*

The analysis of incomplete data is typically made easier if the process by which the incompleteness occurs can be ignored. The paper establishes conditions under which different types of inference can proceed straightforwardly when data might be missing. Let $u = (u_1, \ldots, u_n)$ be a realization of random variables with probability density function f_θ and let $m = (m_1, \ldots, m_n)$ denote a realization of the set of binary missing-data indicators, with a distribution $g_\phi(m \mid u)$. Let $u_{(0)}$ be the part of u that is missing and let $u_{(1)}$ be the part that is observed.

The missing data are called missing at random, MAR, if, for each value of ϕ, $g_\phi(m \mid u)$ is the same for all possible values of $u_{(0)}$ and the observed data are called observed at random, OAR, if for each ϕ and $u_{(0)}$ $g_\phi(m|u)$ is the same for all possible values of $u_{(1)}$. If MAR and OAR obtain then the missing data are called missing completely at random, MCAR. The parameters θ and ϕ are called distinct if the joint parameter space and any Bayesian prior density, factorize.

The paper establishes, with careful argument, essentially that, given MAR and parameter-distinctiveness, the missingness process can be ignored in likelihood and Bayesian inference about θ. However, for valid frequentist inference about θ, for the missingness process to be ignored MCAR and parameter-distinctiveness must be true. If the missingness process is non-ignorable then it has to be modelled and this creates considerable practical difficulties; without knowledge of the missing data themselves, validation of any such model would be an act of faith.

Other issues are also covered and the paper concludes with a discussion with R. J. A. Little about the role and meaning of ancillarity in this context.

10·8. *Green (1995, 8)*

Markov chain Monte Carlo methods are used to explore a model space in such a way that the equilibrium situation corresponds to a target distribution of interest. The Metropolis–Hastings methods discussed in § 10·5 provide one general way of proceeding, in which a possible move is proposed and then either accepted or turned down. Difficulties arise if the overall model space consists of subspaces of different dimensions. In the Bayesian context this means that parameter spaces of different dimensions are involved, and moving from one subspace to another is not straightforward. The paper provides a general, constructive procedure, that of reversible jump Markov chain Monte Carlo, for handling this issue, creating Markov chains that satisfy the required detailed balance conditions. The method is actually Metropolis–Hastings, adapted to a more general setting. Suppose that a move is being considered from subspace $k = 1$, parameterized by the n_1-dimensional parameter $\theta^{(1)}$, to subspace $k = 2$, parameterized by the n_2-dimensional parameter $\theta^{(2)}$. Unless $n_1 = n_2$ some dimension matching has to be achieved.

For this a random vector $u^{(1)}$ of dimension m_1 is generated and $\theta^{(2)}$ is set at some deterministic function of $\theta^{(1)}$ and $u^{(1)}$. For the reverse move, a random vector $u^{(2)}$ of dimension m_2 is generated and $\theta^{(1)}$ is set at some deterministic function of $\theta^{(2)}$ and $u^{(2)}$. It is essential that $n_1 + m_1 = n_2 + m_2$ and that the transformations be mutually consistent and differentiable. The appropriate acceptance probability of the proposed move is

$$\min\left\{1, \frac{p(2,\theta^{(2)}|y)j(2,\theta^{(2)})q_2(u^{(2)})}{p(1,\theta^{(1)}|y)j(1,\theta^{(1)})q_1(u^{(1)})}\left|\frac{\partial(\theta^{(2)},u^{(2)})}{\partial(\theta^{(1)},u^{(1)})}\right|\right\},$$

where y denotes the data and the ps, js and qs denote respectively the joint posterior densities, the probabilities of choosing the move type and the proposal densities for the us. In many applications m_1 or m_2 will be zero.

The paper covers other general issues and illustrations, with particular emphasis on problems concerning one-dimensional multiple changepoint problems, in which the number of changepoints is unknown along with their heights and locations, and noisy realizations of images with piecewise-constant intensity functions.

10·9. *Ljung & Box (1978, 9)*

The paper proposes an improvement to the test of Box & Pierce (1970) for testing the fit of autoregressive moving-average models. Suppose that $a_1, \ldots, a_n$ are the independently and normally distributed disturbances associated with a series from an ARMA(p, q) model and that $\hat{a}_1, \ldots, \hat{a}_n$ are the residuals from the fitted model, with autocorrelations

$$\hat{r}_k = \sum_{l=k+1}^{n} \hat{a}_l \hat{a}_{l-k} \Big/ \sum_{l=1}^{n} \hat{a}_l^2 \quad (k = 1, 2, \ldots).$$

The Box–Pierce test statistic,

$$Q(\hat{r}) = n \sum_{k=1}^{m} \hat{r}_k^2,$$

is, for large n and if the model is correct, distributed approximately as χ^2_{m-p-q}, but has a tendency to produce unsatisfactorily low values. The Ljung–Box modification is to use

$$\bar{Q}(\hat{r}) = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} \hat{r}_k^2.$$

It turns out that, if r denotes the set of true residuals, then asymptotically $E\{Q(r)\} = m$ but, for finite n,

$$E\{Q(r)\} = \frac{mn}{n+2}\left(1 - \frac{m+1}{2n}\right).$$

This can be noticeably less than m unless n is large relative to m whereas $E\{\bar{Q}(r)\} = m$, even for finite n. This discrepancy carries over to $Q(\hat{r})$ and $\bar{Q}(\hat{r})$, rendering the latter as a more reliable statistic in tests that assume an approximate null distribution of χ^2_{m-p-q}. The method is also applicable to more complicated transfer-function models.

10·10. *Gehan (1965a, 10)*

The paper introduces a rank-based two-sample test for data that might be censored on the right. In the terminology of a clinical trial, n_1 and n_2 individuals are allocated randomly to treatments A and B respectively and their survival times are denoted by $x_1^*, \ldots, x_{r_1}^*, x_{r_1+1}, \ldots, x_{n_1}$ and $y_1^*, \ldots, y_{r_2}^*, y_{r_2+1}, \ldots, y_{n_2}$, in which $*$ denotes censoring on the right. All values are in the range $[0, T]$. The distributions of failure times are $F_1(t)$ and $F_2(t)$. The null hypothesis of interest is $H_0 : F_1(t) = F_2(t) (0 \leqslant t \leqslant T)$, a one-sided alternative hypothesis is $H_1 : F_1(t) < F_2(t)$ $(0 \leqslant t \leqslant T)$ and a two-sided hypothesis is $H_2 : F_1(t) < F_2(t)$ $(0 \leqslant t \leqslant T)$ or $F_1(t) > F_2(t)$ $(0 \leqslant t \leqslant T)$. Define

$$U_{ij} = \begin{cases} -1, & \text{if } x_i < y_j \text{ or } x_i \leqslant y_j^*, \\ \ \ 0, & \text{if } x_i = y_j \text{ or } x_i^* < y_j \text{ or } x_i > y_j^* \text{ or } (x_i^*, y_j^*), \\ +1, & \text{if } x_i > y_j \text{ or } x_i^* \geqslant y_j, \end{cases}$$

and calculate the test statistic

$$W = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} U_{ij}.$$

If there is no censoring then $W = n_2(n_1 + n_2 + 1) - R_2$, where R_2 is the rank sum of the second sample in the combined order statistic, revealing the link to the Wilcoxon–Mann–Whitney test. The paper provides the mean, namely zero, and variance of W under the null hypothesis, establishes asymptotic normality of W and consistency of the test, compares the test's efficiency against a parametric test based on exponentially distributed survival times and works through an illustrative example.

11. Discussion

The papers at the top of the highly cited list are generally still very relevant. The reference facility JSTOR provides, for any journal, lists of the top twenty most highly cited and electronically most accessed papers over the most recent periods. The top 8 of each of those lists include 6 of our top 8, at the time of writing.

Two trends are worthy of comment. The first is the change in the *Miscellanea* section, which now tends to include only a small number of articles and those are longer than the short notes that typified the section in earlier times. The second is the decrease in the proportion of papers that have single authors: this proportion has declined from 91% in 1950 to 60% in 1980 and 15% in 2010, perhaps because of a combination of the change in the nature of statistical research, the increased involvement of doctoral and postdoctoral researchers and the increase of collaboration through modern communication facilities such as the Internet.

What of the future? For the period covered by this review *Biometrika* has created and maintained a position as one of a small group of leading general journals in statistical theory and methodology, aiming neither towards work of mathematical interest only nor towards very applied work, not specializing in a particular branch of methodology but attempting to publish influential material across a wide spectrum of topical areas. That this policy is still in force is exemplified by the cutting-edge nature of the material referred to in § 9. Valuable insights and aspirations for particular areas were provided in the centenary papers by Davison (2001, § 12), Atkinson & Bailey (2001, § 14), Oakes (2001, § 10·3) and Tong (2001, § 16). Some developments have already taken place, such as the increasing emphasis on computational issues (Davison, 2001, § 12; Tong, 2001, § 16) and models for computer experiments (Atkinson & Bailey, 2001, § 14); see § 8·2 in the present paper. The hope is that *Biometrika* will continue to publish material that is as influential in the decades to come as has been the case in the decades reviewed in this paper.

ACKNOWLEDGEMENT

I am very grateful for much helpful feedback, from Anthony Atkinson, Sir David Cox, Peter Green, Peter Hall, Byron Morgan, David Oakes, Christian Robert, Howell Tong, two referees and the editor, that has influenced this version of the paper.

REFERENCES

The top-100 papers since 1936 in chronological order, 14th September, 2012

[16] HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28**, 321–77 (2122 citations).
[36] KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika* **29**, 81–93 (1180).
[59] WELCH, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika* **29**, 350–62 (824).
[19] LESLIE, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika* **33**, 183–212 (1924).
[17] PLACKETT, R. L. & BURMAN, J. (1946). The design of optimum multifactorial experiments. *Biometrika* **33**, 305–25 (2100).
[38] WELCH, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika* **34**, 28–35 (1155).
[58] LESLIE, P. H. (1948). Some further notes on the use of matrices in population mathematics. *Biometrika* **35**, 213–45 (826).
[64] ANSCOMBE, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika* **35**, 246–54 (779).
[45] JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* **36**, 149–76 (1038).
[77] PATNAIK, P. B. (1949). The non-central χ^2- and $F-$distribution and their applications. *Biometrika* **36**, 202–32 (612).
[63] BOX, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36**, 317–46 (779).
[29] MORAN, P. A. P. (1950a). Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1556).
[82] COCHRAN, W. G. (1950). The comparison of percentages in matched samples. *Biometrika* **37**, 256–66 (584).
[99] ANSCOMBE, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika* **37**, 358–82 (488).
[32] DURBIN, J. & WATSON, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika* **37**, 409–28 (1353).
[33] DURBIN, J. & WATSON, G. S. (1951). Testing for serial correlation in least squares regression: II. *Biometrika* **38**, 159–78 (1336).
[26] SKELLAM, J. G. (1951). Random dispersal in theoretical populations. *Biometrika* **38**, 196–218 (1640).
[88] BAILEY, N. T. J. (1951). On estimating the size of mobile populations from recapture data. *Biometrika* **38**, 293–306 (549).
[87] WELCH, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika* **38**, 330–6 (551).
[40] BRADLEY, R. A. & TERRY, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **39**, 324–45 (1126).
[28] SCHEFFÉ, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* **40**, 87–110 (1556).
[24] GOOD, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–64 (1716).
[53] BOX, G. E. P. (1953). Non-normality and tests on variances. *Biometrika* **40**, 318–55 (894).
[18] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–15 (1946).
[52] JONCKHEERE, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika* **41**, 133–45 (896).
[49] WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434–49 (934).
[25] SIMON, H. A. (1955). On a class of skew distribution functions. *Biometrika* **42**, 425–40 (1683).
[61] QUENOUILLE, M. H. (1956). Notes on bias in estimation. *Biometrika* **43**, 353–60 (811).
[83] LINDLEY, D. V. (1957). A statistical paradox. *Biometrika* **44**, 187–92 (580).
[86] BOX, G. E. P. & LUCAS, H. L. (1959). Design of experiments in non-linear situations. *Biometrika* **46**, 77–90 (551).

[56] IMHOF, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419–26 (851).
[60] CORMACK, R. M. (1964). Estimation of survival from the sighting of marked animals. *Biometrika* **51**, 429–38 (815).
[65] POTHOFF, R. F. & ROY, S. N. (1964). A general multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313–26 (773).
[10] GEHAN, E. A. (1965a). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52**, 203–23 (2782).
[23] JOLLY, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration – stochastic model. *Biometrika* **52**, 225–47 (1794).
[37] SEBER, G. A. F. (1965). A note on the multiple-recapture census. *Biometrika* **52**, 249–59 (1159).
[94] RAO, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52**, 447–58 (510).
[6] SHAPIRO, S. S. & WILK, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611 (5269).
[14] GOWER, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–38 (2194).
[43] WALKER, S. H. & DUNCAN, D. B. (1967). Estimation of the probability of an event as a union of several variables. *Biometrika* **54**, 167–79 (1111).
[81] WILK, M. B. & GNANADESIKAN, R. (1968). Probability plotting methods for the analysis of data. *Biometrika* **55**, 1–17 (587).
[76] DAY, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika* **56**, 463–74 (613).
[95] HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1–17 (498).
[5] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (5694).
[70] JÖRESKOG, K. G. (1970). A general method for analysis of covariance structures. *Biometrika* **57**, 239–51 (697).
[31] MARDIA, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519–30 (1409).
[48] BRESLOW, N. E. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* **57**, 579–94 (942).
[30] GABRIEL, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **58**, 453–67 (1447).
[11] PATTERSON, H. D. & THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–54 (2371).
[85] ANDERSON, J. A. (1972). Separate sample logistic discrimination. *Biometrika* **59**, 19–35 (564).
[97] BROOK, D. & EVANS, D. A. (1972). An approach to the probability distribution of CUSUM run length. *Biometrika* **59**, 539–49 (493).
[34] MILLER, R. G. (1974). The jackknife – a review. *Biometrika* **61**, 1–15 (1309).
[54] GOODMAN, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215–31 (882).
[35] WEDDERBURN, R. W. M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61**, 439–47 (1223).
[13] COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76 (2278).
[90] SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike’s information criterion. *Biometrika* **63**, 117–26 (535).
[7] RUBIN, D. B. (1976). Inference and missing data (with discussion). *Biometrika* **63**, 581–92 (3576).
[68] MARCUS, R., PERITZ, E. & GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63**, 655–60 (722).
[91] TARONE, R. E. & WARE, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika* **64**, 156–60 (530).
[46] POCOCK, S, J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64**, 191–9 (990).
[50] DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247–54 (929).
[44] CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–51 (1063).
[100] PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–79 (487).
[9] LJUNG, G. M. & BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303 (3123).
[93] EFRON, B. & HINKLEY, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika* **65**, 457–83 (512).

[69] Prentice, R. L. & Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66**, 403–11 (702).
[74] Buckley, J. & James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–36 (658).
[72] Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika* **66**, 605–10 (665).
[12] Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165–76 (2347).
[78] Prentice, R. L., Williams, B. J. & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373–9 (609).
[73] Efron, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika* **68**, 589–99 (660).
[55] Schoenfeld, D. (1982). Partial residuals for the proportional hazards model. *Biometrika* **69**, 239–41 (867).
[96] Harrington, D. P. & Fleming, T. R. (1982). A class of rank test procedures for censored data. *Biometrika* **69**, 553–66 (496).
[92] Cook, R. D. & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70**, 1–10 (528).
[2] Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55 (8319).
[39] Lan, K. K. G. & DeMets, D. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659–63 (1147).
[98] Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **71**, 353–60 (488).
[20] Said, S, E. & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**, 599–607 (1906).
[71] Prentice R. L. (1986). A case-cohort design for epidemiologic cohort sudies and disease prevention trials. *Biometrika* **73**, 1–11 (671).
[1] Liang, K.-Y & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22 (9424).
[79] Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* **73**, 43–56 (608).
[47] Simes, R. G. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 751–4 (966).
[51] Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**, 33–43 (916).
[42] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–49 (1123).
[3] Phillips, P. C. B. & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* **75**, 335–46 (7348).
[62] Luukkonen, R., Saikkonen, P. & Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika* **75**, 491–9 (807).
[15] Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**, 800–2 (2164).
[21] Hurvich, C. M. & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297–307 (1880).
[84] Therneau, T. M., Grambsch, P. B. & Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147–60 (567).
[22] Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–2 (1812).
[67] Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78**, 719–27 (695).
[57] Meng, X.-L. & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**, 267–78 (838).
[4] Donoho, D. & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–55 (5790).
[27] Grambsch, P. M. & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515–26 (1561).
[41] Carter, C. K. & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541–53 (1125).
[89] Genest, C., Ghoudi, K. & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–52 (538).
[66] Pearl, J. (1995). Causal diagrams for empirical research (with discussion). *Biometrika* **82**, 669–88 (766).
[8] Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–32 (3287).
[75] Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–26 (615).

[80] Lo, Y., Mendell, N. R. & Rubin, D. B. (2001). Testing the number of components in a mixture. *Biometrika* **88**, 767–78 (587).

Other references

Abbring, J. H. & van den Berg, G. J. (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika* **94**, 87–99.

Abramovich & Silverman, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85**, 115–29.

Aitchison, J. (1975). Goodness of prediction fit. *Biometrika* **62**, 547–54.

Aitchison, J. (1983). Principal component analysis of compositional data. *Biometrika* **70**, 57–65.

Aitchison, J. & Aitken, C. G. G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika* **63**, 413–20.

Aitchison, J. & Shen, S. M. (1980). Logistic-normal distributions: Some properties and uses. *Biometrika* **67**, 261–72.

Aitchison, J. & Silvey, S. D. (1957). The generalization of probit analysis to the case of multiple responses. *Biometrika* **44**, 131–40.

Ajne, B. (1968). A simple test for uniformity of a circular distribution. *Biometrika* **55**, 343–54.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* **60**, 255–65.

Akaike, H. (1979). A Bayesian extension of the minimum aic procedure of autoregressive model fitting. *Biometrika* **66**, 237–42.

Albert, A. & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71**, 1–10.

Aldrich, J. (2013). Karl Pearson's *Biometrika*: 1901–36. *Biometrika* **100**, 3–15.

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika* **94**, 443–58.

Anscombe, F. J. (1956). On estimating binomial response relations. *Biometrika* **43**, 461–4.

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika* **66**, 59–65.

Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data. *Biometrika* **68**, 357–63.

Armitage, P. (1957). Restricted sequential procedures. *Biometrika* **44**, 9–26.

Asmussen, S. & Edwards, D. (1983). Collapsibility and response variables in contingency tables. *Biometrika* **70**, 567–78.

Atkinson, A. C. (1980). A note on the generalized information criterion for choice of a model. *Biometrika* **67**, 413–8.

Atkinson, A. C. (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika* **68**, 13–20.

Atkinson, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69**, 61–7.

Atkinson, A. C. & Bailey, R. A. (2001). One hundred years of the design of experiments on and off the pages of *Biometrika*. *Biometrika* **88**, 53–97.

Atkinson, A. C. & Fedorov, V. V. (1975). The design of experiments for discriminating between two rival models. *Biometrika* **62**, 57–70.

Auestad, B. & Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika* **77**, 669–87.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* **68**, 326–8.

Azzalini, A., Bowman, A. W. & Härdle, W. (1989). On the use of nonparametric regression for model checking. *Biometrika* **76**, 1–11.

Bacon, D. W. & Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika* **58**, 525–34.

Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika* **85**, 535–47.

Bailey, N. T. J. (1950). A simple stochastic epidemic. *Biometrika* **37**, 193–202.

Bailey, N. T. J. (1953). The total size of a general stochastic epidemic. *Biometrika* **40**, 177–85.

Bang, H. & Tsiatis, A. A. (2000). Estimating medical costs with censored data. *Biometrika* **87**, 329–43.

Barnard, G. (1947a). Significance tests for 2×2 tables. *Biometrika* **34**, 123–38.

Barnard, G. (1947b). 2×2 tables. A note on E. S. Pearson's paper. *Biometrika* **34**, 168–9.

Barnard, G. (1947c). The meaning of a significance level. *Biometrika* **34**, 179–82.

Barnard, J. & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika* **86**, 948–55.

Barndorff-Nielsen, O. (1973). On M-ancillarity. *Biometrika* **60**, 447–55.

Barndorff-Nielsen, O. (1980). Conditionality resolutions. *Biometrika* **67**, 293–310.

Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–65.
Barndorff-Nielsen, O. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307–22.
Barndorff-Nielsen, O. E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**, 557–63.
Bartholomew, D. J. (1959a). A test of homogeneity for ordered alternatives. *Biometrika* **46**, 36–48.
Bartholomew, D. J. (1959b). A test of homogeneity for ordered alternatives. II. *Biometrika* **46**, 328–35.
Bartholomew, D. J. (1984). The foundations of factor analysis. *Biometrika* **71**, 221–32.
Bartlett, M. S. (1941). The statistical significance of canonical correlations. *Biometrika* **32**, 29–37.
Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika* **37**, 1–16.
Bartlett, M. S. (1953a). Approximate confidence intervals. *Biometrika* **40**, 12–9.
Bartlett, M. S. (1953b). Approximate confidence intervals. II. More than one unknown parameter. *Biometrika* **40**, 306–17.
Bartlett, M. S. (1957a). On theoretical models for competitive and predatory biological systems. *Biometrika* **44**, 27–42.
Bartlett, M. S. (1957b). A comment on D. V. Lindley's statistical paradox. *Biometrika* **44**, 533–4.
Bartlett, M .S. (1964). The spectral analysis of two-dimensional point processes. *Biometrika* **51**, 299–311.
Bartlett, M .S. & Tippett, L. H. C. (1981). Egon Sharpe Pearson, 1895–1980. *Biometrika* **68**, 1–11.
Basu, A., Harris, I. R., Hjort, N. L. & Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85**, 549–59.
Beale, E. M. L., Kendall, M. G. & Mann, D. W. (1967). The discarding of variables in multivariate analysis. *Biometrika* **54**, 357–66.
Beaumont, M. A., Cornuet, J.-M., Marin, J.-M. & Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika* **97**, 983–90.
Bechhofer, R. E., Dunnett, C. W. & Sobel, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika* **41**, 170–6.
Begg, C. B. & Gray, R. (1984), Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika* **71**, 11–8.
Belloni, A., Chernozhukov, V. & Wang, L. (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika* **98**, 791–806.
Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* B **57**, 289–300.
Benjamini, Y., Krieger, A. M. & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507.
Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74**, 457–68.
Berger, J. O. & Bernardo, J. M. (1992). Ordered group reference priors with application to the multinomial problem. *Biometrika* **79**, 25–37.
Besag, J. E. (1977). Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika* **64**, 616–8.
Besag, J. (1989). A candidate's formula: a curious result in Bayesian prediction. *Biometrika* **76**, 183.
Besag, J. & Clifford, P. (1989). Generalized Monte Carlo significance tests. *Biometrika* **76**, 633–42.
Besag, J. & Clifford, P. (1991). Sequential Monte Carlo p-values. *Biometrika* **78**, 301–4.
Besag, J. & Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–46.
Bhansali, R. J. & Downham, D. Y. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's fpe criterion. *Biometrika* **64**, 547–51.
Bhattacharya, A. & Dunson, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika* **98**, 291–306.
Bien, J. & Tibshirani, R. J. (2011). Sparse estimation of a covariance matrix. *Biometrika* **98**, 807–20.
Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika* **65**, 31–8.
Bingham, D., Sitter, R. R. & Tang, B. (2009). Orthogonal and nearly orthogonal designs for computer experiments. *Biometrika* **96**, 51–65.
Bliss, C. I. & Owen, A. R. G. (1958). Negative binomial distributions with a common k. *Biometrika* **45**, 37–58.
Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60**, 217–26.
Bowman, K. O. & Shenton, L. R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika* **62**, 243–50.
Box, G. E. P. (1952). Multi-factor designs of first order. *Biometrika* **39**, 49–57.
Box, G. E. P. & Draper, N. R. (1963). The choice of a second order rotatable design. *Biometrika* **50**, 335–52.
Box, G. E. P. & Draper, N. R. (1965). The Bayesian estimation of common parameters from several responses. *Biometrika* **52**, 355–65.
Box, G. E. P. & Draper, N. R. (1975). Robust designs. *Biometrika* **62**, 347–52.
Box, G. E. P. & Hunter, J. S. (1954). A confidence region for the solution of a set of simultaneous equations with an application to experimental design. *Biometrika* **41**, 190–9.
Box, G. E. P. & Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control,* 2nd ed. San Francisco: Holden Day.

Box, G. E. P. & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Statist. Assoc.* **65**, 1509–26.
Box, G. E. P. & Tiao, G. C. (1962). A further look at robustness via Bayes's theorem. *Biometrika* **49**, 419–32.
Box, G. E. P. & Tiao, G. C. (1965). A change in level of a non-stationary time series. *Biometrika* **52**, 181–92.
Box, G. E. P. & Tiao, G. C. (1968). A Bayesian approach to some outlier problems. *Biometrika* **55**, 119–29.
Box, G. E. P. & Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrika* **64**, 355–65.
Box, G. E. P. & Watson, G. S. (1962). Robustness to non-normality of regression tests. *Biometrika* **49**, 93–106.
Breidt, F. J., Claeskens, G. & Opsomer, J. D. (2005). Model-assisted estimation for complex surveys using penalised splines. *Biometrika* **92**, 831–46.
Breslow, N. E. (1981). Odds ratio estimators when the data are sparse. *Biometrika* **68**, 73–84.
Breslow, N. E. & Cain, K. C. (1988). Logistic regression for two-stage case-control data. *Biometrika* **75**, 11–20.
Breslow, N. E. & Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* **82**, 81–91.
Brillinger, D. R. (1969). Asymptotic properties of spectral estimates of second order. *Biometrika* **56**, 375–90.
Bruce, A. G. & Gao, H.-Y. (1996). Understanding WaveShrink: Variance and bias estimation. *Biometrika* **83**, 727–45.
Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika* **76**, 503–14.
Burnham, K. P. & Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 625–33.
Bush, C. A. & MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83**, 275–85.
Carey, V., Zeger, S. L. & Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika* **80**, 517–26.
Carroll, R. J., Spiegelman, C. H. & Lam, K. K. G. (1984). On errors-in-variables for binary regression models. *Biometrika* **71**, 19–25.
Carter, C. K. & Kohn, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika* **83**, 589–601.
Carvalho, C. M. & Scott, J. G. (2009). Objective Bayesian model selection in Gaussian graphical models. *Biometrika* **96**, 497–512.
Carvalho, C. M., Polson, N. G & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97**, 465–80.
Casella, G. & Robert, C. P. (1996). Rao-Blackwellisation of sampling schemes. *Biometrika* **83**, 81–94.
Cassel, C. M., Särndal, C. E. & Wretman, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika* **63**, 615–20.
Catchpole, E. A. & Morgan, B. J. T. (1997). Detecting parameter redundancy. *Biometrika*, **84**, 187–96.
Chaloner, K. & Brant, R. (1988). A Bayesian approach to outlier detection and residual analysis. *Biometrika* **75**, 651–9.
Chambers, R. L. & Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika* **73**, 597–604.
Chan, K. S. & Tong, H. (1986). On estimating thresholds in autoregressive models. *J. Time Ser. Anal.* **7**, 179–90.
Chao, A. & Yang, M. C. K. (1993). Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika* **80**, 193–201.
Chatterjee, N. & Carroll, R. J. (2005). Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika* **92**, 399–418.
Chen, J. & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95**, 759–71.
Chen, J. & Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **80**, 107–16.
Cheng, S, C., Wei, L. J. & Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835–45.
Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika* **85**, 347–61.
Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika* **89**, 539–51.
Chopin, N. & Robert, C. P. (2010). Properties of nested sampling. *Biometrika* **97**, 742–55.
Claeskens, G., Krivobokova, T. & Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* **96**, 529–44.
Clarke, M. R. B. (1980). The reduced major axis of a bivariate sample. *Biometrika* **67**, 441–6.
Clifford, P. & Sudbury, A. (1973). A model for spatial conflict. *Biometrika* **60**, 581–8.
Clyde, M., Parmigiani, G. & Vidakovic, B. (1998). Multiple shrinkage and subset selection in wavelets. *Biometrika* **85**, 391–401.
Cobb, G. W. (1978). The problem of the Nile: conditional solution to a changepoint problem. *Biometrika* **65**, 243–51.
Cox, D. R. (1948). A note on the asymptotic distribution of range. *Biometrika* **35**, 310–5.
Cox, D. R. (1951). Some systematic experimental designs. *Biometrika* **38**, 312–23.
Cox, D. R. (1958). Two further applications of a model for binary regression. *Biometrika* **45**, 562–5.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B **34**, 187–220.
Cox, D. R. (1980). Local ancillarity. *Biometrika* **67**, 279–86.
Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika* **70**, 269–74.
Cox, D. R. & Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91**, 729–37.
Cox, D. R. & Small, N. J. H. (1978). Testing multivariate normality. *Biometrika* **65**, 263–72.
Cox, D. R. & Smith, W. L. (1954). On the superposition of renewal processes. *Biometrika* **41**, 91–9.
Craig, C. C. (1953). On the utilization of marked specimens in estimating populations of flying insects. *Biometrika* **40**, 170–6.
Crainiceanu, C., Ruppert, D., Claeskens, G. & Wand, M. P. (2005). Exact likelihood ratio tests for penalised splines. *Biometrika* **92**, 91–103.
Critchley, F. (1985). Influence in principal components analysis. *Biometrika* **72**, 627–36.
Croux, C. & Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika* **87**, 603–18.
Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika* **82**, 407–10.
Crump, R. K., Hotz, V. J., Imbens, G. W. & Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96**, 187–99.
Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika* **33**, 129–35.
Daniels, H. E. (1956). The approximate distribution of serial correlation coefficients. *Biometrika* **43**, 169–85.
Darroch, J. N. (1958). The multiple-recapture census: I. Estimation of a closed population. *Biometrika* **45**, 343–59.
Darroch, J. N. (1959). The multiple-recapture census: II. Estimation when there is immigration or death. *Biometrika* **46**, 336–51.
Darroch, J. N. (1961). The two-sample capture-recapture census when tagging and sampling are stratified. *Biometrika* **48**, 241–60.
Darroch, J. N. & Mosimann, J. E. (1985). Canonical and principal components of shape. *Biometrika* **72**, 241–52.
David, F. N. (1947). A χ^2 ‘smooth’ test for goodness of fit. *Biometrika* **34**, 299–310.
David, F. N. & Johnson, N. L. (1954). Statistical treatment of censored data part I. Fundamental formulae. *Biometrika* **41**, 228–40.
David, H. A., Hartley, H. O. & Pearson, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika* **41**, 482–93.
Davidian, M. & Gallant, A. R. (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika* **80**, 475–88.
Davies, N., Triggs, C. M. & Newbold, P. (1977). Significance levels of the Box-Pierce portmanteau statistic in finite samples. *Biometrika* **64**, 517–22.
Davies, R. B. & Harte, D. S. (1987). Tests for Hurst effect. *Biometrika* **74**, 95–101.
Davison, A. C. (2001). *Biometrika* Centenary: Theory and general methodology. *Biometrika* **88**, 13–52.
Davison, A. C. & Hinkley, D.V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75**, 417–31.
Davison, A. C., Hinkley, D.V. & Schechtman, E. (1986). Efficient bootstrap simulation. *Biometrika* **73**, 555–66.
Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika* **68**, 265–74.
Deemer, W. L. & Olkin, I. (1951). The Jacobians of certain matrix transformations useful in multivariate analysis: Based on lectures of P. L. Hsu at the University of North Carolina, 1947. *Biometrika* **38**, 345–67.
de Jong, P. & Shephard, N. (1995). The simulation smoother for time series models. *Biometrika* **82**, 339–50.
Dellaportas, P. & Forster, J. J. (1999). Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika* **86**, 615–33.
DeMets, D. L. & Ware, J. H. (1980). Group sequential methods for clinical trials with a one-sided hypothesis. *Biometrika* **67**, 651–60.
Dempster, A. P. (1967). Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika* **54**, 515–28.
Denison, D. G. T., Mallick, B. K. & Smith, A. F. M. (1998). A Bayesian cart algorithm. *Biometrika* **85**, 363–77.
Deville, J. C. & Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika* **91**, 893–912.
DiCiccio, T. J. & Martin, M. A. (1991). Approximations of marginal tail probabilities for a class of smooth functions with application to Bayesian and conditional inference. *Biometrika* **78**, 891–902.
DiMatteo, I., Genovese, C. R. & Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88**, 1055–71.
Draper, N. R. & Hunter, W. G. (1966). Design of experiments for parameter estimation in multiresponse situations. *Biometrika* **53**, 525–33.
Drton, M & Perlman, M. D. (2004). Model selection for Gaussian concentration graphs. *Biometrika* **91**, 591–602.
Dunnett, C. W. & Sobel, M. (1954). A bivariate generalization of Student’s t-distribution, with tables for certain special cases. *Biometrika* **41**, 153–69.

Dunnett, C. W. & Sobel, M. (1955). Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution. *Biometrika* **42**, 258–60.

Dunson, D. B. (2009). Nonparametric Bayes local partition models. *Biometrika* **96**, 249–62.

Dunson, D. B. & Park, J.-H. (2008). Kernel stick-breaking processes. *Biometrika* **95**, 307–23.

Durbin, J. (1959). Efficient estimation of parameters in moving-average models. *Biometrika* **46**, 306–16.

Durbin, J. (1960). The fitting of time-series models. *Rev. Int. Inst. Statist.* **28**, 233–44.

Durbin, J. (1961). Some methods of constructing exact tests. *Biometrika* **48**, 41–55.

Durbin, J. (1969). Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals. *Biometrika* **54**, 1–15.

Durbin, J. (1980). Approximations for densities of sufficient estimators. *Biometrika* **67**, 311–33.

Durbin, J. & Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* **84**, 669–84.

Durbin, J. & Koopman, S. J. (2002). A simple and efficient smoother for state space time series analysis. *Biometrika* **89**, 603–16.

Durbin, J. & Watson, G. S. (1971). Testing for serial correlation in least squares regression. III. *Biometrika* **58**, 1–19.

Edwards, D. & Havranek, T. (1985). A fast procedure for model search in multidimensional contingency tables. *Biometrika* **72**, 339–51.

Edwards, D. & Kreiner, S. (1983). The analysis of contingency tables by graphical models. *Biometrika* **70**, 553–65.

Efron, B. (1965). The convex hull of a random set of points. *Biometrika* **52**, 331–43.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58**, 403–17.

Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72**, 45–58.

Efron, B. & Morris, C. (1972). Empirical Bayes on vector observations: an extension of Stein's method. *Biometrika* **59**, 335–47.

Efron, B. & Thisted, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **93**, 435–47.

Efron, B. & Zhang, N. R. (2011). False discovery rates and copy number variation. *Biometrika* **98**, 251–71.

Evans, D. A. (1953). Experimental evidence concerning contagious distributions in ecology. *Biometrika* **40**, 186–211.

Fan, F. & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645–60.

Fan, J., Yao, Q. & Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* **83**, 189–206.

Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47**, 307–323.

Fearnhead, P., Wyncoll, D. & Tawn, J. A. (2010). A sequential smoothing algorithm with linear computational cost. *Biometrika* **97**, 447–64.

Feigin, P. D. & Reiser, B. (1979). On asymptotic ancillarity and inference for Yule and regular nonergodic processes. *Biometrika* **66**, 279–83.

Fienberg, S. E. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika* **59**, 591–603.

Finney, D. J. (1938). The distribution of the ratio of estimates of the two variances in a sample from a normal bi-variate population. *Biometrika* **30**, 190–2.

Finney, D. J. (1947). The estimation from individual records of the relationship between dose and quantal response. *Biometrika* **34**, 320–34.

Finney, D. J. (1948). The Fisher-Yates test of significance in 2×2 contingency tables. *Biometrika* **35**, 157–75.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38.

Fisher, L. & Van Ness, J. W. (1971). Admissible clustering procedures. *Biometrika* **58**, 91–104.

Fisher, R. A., Corbet, A. S. & Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58.

Fitzmaurice, G. M. & Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika* **80**, 141–51.

Ford, I. & Silvey, S. D. (1980). A squentially constructed design for estimating a nonlinear parametric function. *Biometrika* **67**, 381–8.

Foster, D. P. & Vohra, R. V. (1998). Asymptotic calibration. *Biometrika* **85**, 379–80.

Frangakis, C. E. & Rubin, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* **86**, 365–79.

Fraser, D. A. S. (1961). The fiducial methods and invariance. *Biometrika* **48**, 261–80.

Fraser, D. A. S., Reid, N. & Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* **86**, 249–64.

Freeman, G. H. & Halton, J. H. (1951). Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika* **38**, 141–9.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika* **89**, 197–210.

Gail, M. H., Lubin, J. H. & Rubinstein, L. V. (1981). Likelihood calculations for matched case-control studies and survival studies with tied death times. *Biometrika* **68**, 703–7.
Gail, M. H., Wieand, S. & Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* **71**, 431–44.
Gallant, A. R. & Long, J. R. (1997). Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika* **84**, 125–41.
Gart, J. J. (1970). Point and interval estimation of the common odds ratio in the combination of 2×2 tables with fixed marginals. *Biometrika* **57**, 471–5.
Gart, J. J. & Zweifel, J. R. (1967). On the bias of various estimators of the logit and its variance with application to quantal bioassay. *Biometrika* **54**, 181–7.
Garwood, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika* **28**, 437–42.
Gayen, A. K. (1949). The distribution of 'Student's' t in random samples of any size drawn from non-normal universes. *Biometrika* **36**, 353–69.
Gasser, T., Sroka, L. & Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73**, 625–33.
Geary, R. C. (1936). Moments of the ratio of the mean deviation to the standard deviation for normal samples. *Biometrika* **28**, 295–307.
Geary, R. C. (1947). Testing for normality. *Biometrika* **34**, 209–42.
Gehan, E. A. (1965b). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* **52**, 650–3.
Geisser, S. (1974). A predictive approach to the random effect model. *Biometrika* **61**, 101–7.
Gelfand, A. E. & Ghosh, S. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika* **85**, 1–11.
Gelfand, A. E., Sahu, S. K. & Carlin, B. P. (1995). Efficient parameterisations for normal linear mixed models. *Biometrika* **82**, 479–88.
Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika* **74**, 549–55.
Genovese, C. R., Roeder, K. & Wasserman, L. (2006). False discovery control with p-value weighting. *Biometrika* **93**, 509–24.
George, E. I. & Foster, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87**, 731–47.
Gill, R. & Schumacher, M. (1987). A simple test of the proportional hazards assumption. *Biometrika* **74**, 289–300.
Giudici, P. & Green, P. J. (1999). Decomposable graphical Gaussian model determination. *Biometrika* **86**, 785–801.
Godambe, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika* **63**, 277–84.
Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika* **72**, 419–28.
Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika* **78**, 45–51.
Good, I. J. (1979). Studies in the history of probability and statistics. XXXVII. A. M. Turing's statistical work in World War II. *Biometrika* **66**, 393–6.
Good, I. J. & Gaskins, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika* **58**, 255–77.
Good, I. J. & Toulmin, G. H. (1956). The number of new species, and the increase in population coverage, when the sample size is increased. *Biometrika* **43**, 45–63.
Goodman, L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika* **60**, 179–92.
Green, P. J. & Mardia, K. V. (2006). Bayesian alignment using hierarchical models, with applications in protein bioinformatics. *Biometrika* **93**, 235–54.
Greenwood, M. (1941). Medical statistics from Graunt to Farr. *Biometrika* **32**, 101–27.
Greenwood, M. (1942). Medical statistics from Graunt to Farr (continued). *Biometrika* **32**, 203–25.
Greenwood, M. (1943). Medical statistics from Graunt to Farr (concluded). *Biometrika* **32**, 1–24.
Guo, J., Levina, E., Michailidis, G. & Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika* **98**, 1–15.
Gupta, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika* **39**, 260–73.
Haggan, V. & Ozaki, T. (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* **68**, 189–96.
Haldane, J. B. S. (1940). The mean and variance of χ^2, when used as a test of homogeneity, when expectations are small. *Biometrika* **31**, 346–55.
Haldane, J. B. S. (1945a). On a method of estimating frequencies. *Biometrika* **33**, 222–5.
Haldane, J. B. S. (1945b). Moments of r and χ^2 for a fourfold table in the absence of association. *Biometrika* **33**, 231–3.
Hall, P. & Marron, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* **77**, 415–9.
Hall, P. & Martin, M. A. (1988). On bootstrap resampling and iteration. *Biometrika* **75**, 661–71.
Hall, P. & Opsomer, J. D. (2005). Theory for penalised spline regression. *Biometrika* **92**, 105–18.

HALL, P., KAY, J. W. & TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77**, 521–8.
HALL, P., SHEATHER, S. J., JONES, M. C. & MARRON, J. S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* **78**, 263–9.
HALL, W. J. & WELLNER, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67**, 133–43.
HANNAN. E. J. (1969). The identification of vector mixed autoregressive-moving average systems. *Biometrika* **56**, 223–5.
HANNAN. E. J. & RISSANEN, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika* **82**, 81–94.
HANS, C. (2009). Bayesian lasso regression. *Biometrika* **96**, 835–45.
HARRELL, F. E. & DAVIS, C. E. (1982). A new distribution-free quantile estimator. *Biometrika* **69**, 635–40.
HARTER, H. L. (1961). Expected values of normal order statistics. *Biometrika* **48**, 151–65.
HARTLEY, H. O. (1950a). The use of range in analysis of variance. *Biometrika* **37**, 271–80.
HARTLEY, H. O. (1950b). The maximum F-ratio as a short-cut test for heterogeneity of variance. *Biometrika* **37**, 308–12.
HARTLEY, H. O. & RAO, J. N. K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93–108.
HARTLEY, H. O. & RAO, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika* **55**, 547–57.
HARVEY, A. C. & PHILLIPS, G. D. A. (1979). Maximum likelihood estimation of regressive models with autoregressive-moving average disturbances. *Biometrika* **66**, 49–58.
HARVILLE, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383–5.
HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90.
HE, X., ZHU, Z.-Y. & FUNG, W.-K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89**, 579–90.
HECKMAN, J. J. & HONORÉ, B. E. (1989). The identifiability of the competing risks model. *Biometrika* **76**, 325–30.
HINKLEY, D. V. (1969a). Inference about the intersection in two-phase regression. *Biometrika* **56**,495–504.
HINKLEY, D. V. (1969b). On the ratio of two correlated normal random variables. *Biometrika* **56**, 635–9.
HINKLEY, D. V. (1971). Inference about the change-point from cumulative sum tests. *Biometrika* **58**, 509–23.
HINKLEY, D. V. (1980). Likelihood as approximate pivotal distribution. *Biometrika* **67**, 287–92.
HOMMEL, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75**, 383–6.
HOOVER, D. R., RICE, J. A., WU, C. O. & YANG, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809–22.
HOUGAARD, P. (1984). Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika* **71**, 75–83.
HOUGAARD, P. (1986a). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387–96.
HOUGAARD, P. (1986b). A class of multivariate failure time distributions. *Biometrika* **73**, 671–8.
HSU, P. L. (1941). On the limiting distribution of the canonical correlations. *Biometrika* **32**, 38–45.
HUANG, J., MA, S., XIE, H. & ZHANG, C. H. (2009). A group bridge approach for variable selection. *Biometrika* **96**, 339–55.
HUANG, J. Z., WU, C. O. & ZHOU, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111–28.
HUANG, J. Z., LIU, N., POURAHMADI, M. & LIU. L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* **93**, 85–98.
HUGGINS, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **76**, 133–40.
IRELAND, C. T. & KULLBACK, S. (1968). Contingency tables with given marginals. *Biometrika* **55**, 179–88.
ISHWARAN, H. & ZAREPOUR, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* **87**, 371–90.
JAMES, B., JAMES, K. L. & SIEGMUND, D. (1987). Tests for a change-point. *Biometrika* **74**, 71–83.
JAMES, G. M. & RADCHENKO, P. (2009). A generalized Dantzig selector with shrinkage tuning. *Biometrika* **96**, 323–37.
JAMES, G. M., HASTIE, T. & SUGAR, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587–602.
JAMES, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika* **38**, 324–9.
JAMES, G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika* **41**, 19–43.
JEFFREYS, H. (1939). Random and systematic arrangements. *Biometrika* **31**, 1–8.
JIN, Z., LIN, D. Y., WEI, L. J. & YING, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–53.
JOHNSON, N. L. & WELCH, B. L. (1940). Applications of the non-central t-distribution. *Biometrika* **31**, 362–89.

Jones, R. H. & Brelsford, W. M. (1967). Time series with periodic structure. *Biometrika* **54**, 403–8.
Jupp, P. E. & Mardia, K. V. (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika* **67**, 163–73.
Kalbfleisch, J. D. (1975). Sufficiency and conditionality (with discussion). *Biometrika* **62**, 251–68.
Kalbfleisch, J. D. & Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60**, 267–78.
Kauermann, G. & Opsomer, J. D. (2011). Data-driven selection of the spline dimension in penalized spline regression. *Biometrika* **98**, 225–30.
Keenan, D. M. (1985). A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika* **72**, 39–44.
Kelly, F. P. & Ripley, B. D. (1976). A note on Strauss's model for clustering. *Biometrika* **63**, 357–60.
Kempthorne, O. (1947). A simple approach to confounding and fractional replication in factorial experiments. *Biometrika* **34**, 255–72.
Kendall, D. G. (1948). On the role of variable generation time in the development of a stochastic birth process. *Biometrika* **35**, 316–30.
Kendall, D. G. (1960). Birth-and-death processes and the theory of carcinogenesis. *Biometrika* **47**, 13–21.
Kendall, M. G. (1942). Partial rank correlation. *Biometrika* **32**, 277–83.
Kendall, M. G. (1945). The treatment of ties in ranking problems. *Biometrika* **33**, 239–51.
Kendall, M. G. (1949a). On the reconciliation of theories of probability. *Biometrika* **36**, 101–16.
Kendall, M. G. (1949b). Rank and product-moment correlation. *Biometrika* **36**, 177–93.
Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika* **41**, 403–4.
Kendall. M. G. (1963). Ronald Aylmer Fisher, 1890–1962. *Biometrika* **50**, 1–15.
Kendall, M. G. & Babington Smith, B. (1940). On the method of paired comparisons. *Biometrika* **31**, 324–45.
Kennedy, M. C. & O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87**, 1–13.
Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika* **69**, 19–37.
Kent, J. T. (1983). Information gain and a general measure of correlation. *Biometrika* **70**, 163–73.
Kent, J. T. & Mardia, K. V. (2001). Procrustes tangent projections and bilateral symmetry. *Biometrika* **88**, 469–85.
Kermack, K. A. & Haldane, J. B. S. (1950). Organic correlation and allometry. *Biometrika* **37**, 30–41.
Kettenring, J. R. (1971). Canonical analysis of several sets of variables *Biometrika* **58**, 433–51.
Kim, H. J. & Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika* **76**, 409–23.
Kim, K. & DeMets, D. L. (1987). Design and analysis of group sequential tests based on the type I error spending rate function. *Biometrika* **74**, 149–54.
Koenker, R., Ng, P. & Portnoy, S. (1994). Quantile smoothing splines. *Biometrika* **81**, 673–80.
Konishi, S. & Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika* **83**, 875–90.
Konishi, S., Ando, T. & Imoto, S. (2004). Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika* **91**, 27–43.
Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* **80**, 117–28.
Kosmidis, I. & Firth, D. (2009). Bias reduction in exponential family nonlinear models. *Biometrika* **96**, 793–804.
Koziol, J. A. & Green, S. B. (1976). A Cramér-von Mises statistic for randomly censored data. *Biometrika* **63**, 465–74.
Kudo, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50**, 403–18.
Kuk, A. Y. C. & Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79**, 531–41.
Lagakos, S. W., Barraj, L. M. & De Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with application to aids. *Biometrika* **75**, 515–23.
Lagakos, S. W., Sommer, C. J. & Zelen, M. (1978). Semi-Markov models for partially censored models. *Biometrika* **65**, 311–7.
Lancaster, H. O. (1949). The derivation and partition of χ^2 in certain discrete distributions. *Biometrika* **36**, 117–29.
Lawley, D. N. (1956a). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika* **43**, 128–36.
Lawley, D. N. (1956b). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303.
Ledford, A. W. & Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–87.
Lee, Y. & Nelder, J, A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models random-effect models and structured dispersions. *Biometrika* **88**, 987–1006.
Leslie, P. H. (1958). A stochastic model for studying the properties of certain biological systems by numerical methods. *Biometrika* **45**, 16–31.
Leslie, P. H. & Gower, J. C. (1960). The properties of a stochastic model for the predator-prey type of interaction between two species. *Biometrika* **47**, 219–234.
Li, W. K. & McLeod, A. I. (1986). Fractional time series modelling. *Biometrika* **73**, 217–21.

Li, Y. & Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95**, 415–36.

Lin, D. Y. & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

Lin, D. Y., Wei, L. J. & Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–72.

Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84**, 309–26.

Linton, O. & Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93–100.

Lipsitz, S. R., Laird, N. M. & Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika* **78**, 153–60.

Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–83.

Little, R. J. A. & Schluchter, M. D. (1985). Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika* **72**, 497–512.

Liu, C. & Rubin, D. B. (1994). The ecme algorithm: a simple extension of em and ecm with faster monotone convergence. *Biometrika* **81**, 633–48.

Liu, C., Rubin, D. B. & Wu, Y. N. (1998). Parameter expansion to accelerate em: the px-em algorithm. *Biometrika* **85**, 755–70.

Liu, J. S., Wong, W. H. & Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81**, 27–40.

Liu, Q. & Pierce, D. A. (1994). A note on Gauss-Hermite quadrature. *Biometrika* **81**, 624–9.

Ljung, G. M. & Box, G. E. P. (1979). The likelihood function of stationary autoregressive-moving average models. *Biometrika* **66**, 265–70.

Lloyd, E. H. (1952). Least-squares estimation of location and scale parameters using order statistics. *Biometrika* **39**, 88–95.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika* **74**, 817–27.

Lord, E. (1947). The use of range instead of standard deviation in the t-test. *Biometrika* **34**, 41–67.

Mallet, A. (1986). A maximum likelihood method for random coefficient regression models. *Biometrika* **73**, 645–56.

Mallows, C. L. (1957). Non-null ranking models. I. *Biometrika* **44**, 114–30.

Mardia, K. V. & Dryden, I. L. (1989). The statistical analysis of shape data. *Biometrika* **76**, 271–81.

Mardia, K. V. & Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135–46.

Mardia, K. V., Kent, J. T., Hughes, G. & Taylor, C. C. (2009). Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika* **96**, 975–82.

Marriott, F. H. C. & Pope, J. A. (1954). Bias in the estimation of autocorrelations. *Biometrika* **41**, 390–402.

McCullagh, P. (1984). Local sufficiency. *Biometrika* **71**, 233–44.

McKeague, I. W. & Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501–14.

Meng, X. L. & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika* **79**, 103–11.

Mentré, F., Mallet, A. & Baccar, D. (1997). Optimal design in random-effects regression models. *Biometrika* **84**, 429–42.

Merikoski, J. K. & Pukkila, T. M. (1983). A note on the expectations of products of autocorrelations. *Biometrika* **70**, 528–9. Correction (1984), **71**, 655.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–92.

Miller, R. & Halpern, J. (1982). Regression with censored data. *Biometrika* **69**. 521–31.

Miller, R. G. (1976). Least squares regression with censored data. *Biometrika* **63**, 449–64.

Molenberghs, G., Kenward, M. G. & Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika* **84**, 33–44.

Møller, J., Pettitt, A. N., Reeves, R. & Berthelsen, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika* **93**, 451–8.

Moran, P. A. P. (1948a). Rank correlation and product-moment correlation. *Biometrika* **35**, 203–6.

Moran, P. A. P. (1948b). Some theorems on time series II. The significance of the serial correlation coefficient. *Biometrika* **35**, 255–60.

Moran, P. A. P. (1950b). A test for the serial independence of residuals. *Biometrika* **37**, 178–81.

Moran, P. A. P. (1951). A mathematical theory of animal trapping. *Biometrika* **38**, 307–11.

Morant, G. M. (1948). Comparisons of heights and weights of German civilians recorded in 1946-7 and Royal Air Force and other British series. *Biometrika* **35**, 368–96.

Morant, G. M. & Samson, O. (1936). An examination of investigations by Dr. Maurice Fishberg and Professor Franz Boas dealing with measurements of Jews in New York. *Biometrika* **28**, 1–31.

Morgan, W. A. (1939). A test for the significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika* **31**, 13–9.

Mosimann, J. E. (1962). On the compound multinomial distribution, the multivariate β distribution, and correlations among proportions. *Biometrika* **49**, 65–82.
Muirhead, R. J. & Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis. *Biometrika* **67**, 31–43.
Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* **78**, 521–30.
Müller, P., Erkanli, A. & West, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* **83**, 67–79.
Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika* **77**, 127–37.
Nelder, J. A. & Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika* **74**, 221–32.
Neuhaus, J. M., Hauck, W. W. & Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika* **79**, 755–62.
Newbold, P. (1974). The exact likelihood function for a mixed autoregressive-moving average process. *Biometrika* **61**, 423–6.
Newman, D. (1939). The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika* **31**, 20–30.
Neyman, J. (1941). Fiducial argument and the theory of confidence intervals. *Biometrika* **32**, 128–50.
Neyman, J. & Pearson, E. S. (1938). Note on some points in "Student's" paper on 'Comparison between balanced and random arrangements of field plots'. *Biometrika* **29**, 379–88.
Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika* **64**, 441–8.
Oakes, D. (1985). Self-calibrating priors do not exist. *J. Am. Statist. Soc.* **80**, 339.
Oakes, D. (2001). *Biometrika* Centenary: Survival analysis. *Biometrika* **88**, 99–142.
Oakley, J. & O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika* **89**, 769–84.
Olkin, I. & Tate, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Ann. Math. Statist.* **32**, 448–65.
Page, E. S. (1955a). Control charts with warning lines. *Biometrika* **42**, 243–57.
Page, E. S. (1955b). A test for a change in a parameter occurring an at unknown point. *Biometrika* **42**, 523–7.
Page, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika* **44**, 248–52.
Papaspiliopoulos, O. & Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95**, 169–86.
Parzen, M. I., Wei, L. J. & Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–50.
Patnaik, P. B. (1948). The power function of the test for the difference between two proportions in a 2 × 2 table. *Biometrika* **35**, 157–75.
Patnaik, P. B. (1950). The use of mean range as an estimator of variance in statistical tests. *Biometrika* **37**, 78–87.
Patterson, H. D. & Williams, E. R. (1976). A new class of resolvable incomplete block designs. *Biometrika* **63**, 83–92.
Pearson, E. S. (1936). Karl Pearson: An appreciation of some aspects of his life and work, Part I. *Biometrika* **28**, 193–257.
Pearson, E. S. (1937). Some aspects of the problem of randomization. *Biometrika* **29**, 53–64.
Pearson, E. S. (1938a). Karl Pearson: An appreciation of some aspects of his life and work, Part II. *Biometrika* **29**, 161–248.
Pearson, E. S. (1938b). Some aspects of the problem of randomization: II. An illustration of "Student's" inquiry into the effect of 'balancing' in agricultural experiment. *Biometrika* **30**, 159–79.
Pearson, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2 × 2 table. *Biometrika* **34**, 139–67.
Pearson, E. S., D'Agostino, R. B. & Bowman, K. O. (1977). Tests for departure from normality: comparison of powers. *Biometrika* **64**, 231–46.
Pearson, E. S. & Hartley, H. O. (1942). The probability integral of the range in samples of n observations from a normal population. *Biometrika* **32**, 301–10.
Pearson, E. S. & Merrington, M. (1948). 2 × 2 tables: the power function of the test on a randomized experiment. *Biometrika* **35**, 331–45.
Percival, D. P. (1995). On estimation of the wavelet variance. *Biometrika* **82**, 619–31.
Peskun, P. H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60**, 607–12.
Phadke, M. S. & Kedem, G. (1978). Computation of the exact likelihood function of multivariate moving average models. *Biometrika* **65**, 511–9.
Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* **74**, 535–47.
Pitman, E. J. G. (1938). Significance tests which may be applied to samples from any populations: III. The analysis of variance tests. *Biometrika* **29**, 322–35.
Pitman, E. J. G. (1939a). The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika* **30**, 391–421.

PITMAN, E. J. G. (1939b). A note on normal correlation. *Biometrika* **31**, 9–12.
PITMAN, E. J. G. (1939c). Tests of hypotheses concerning location and scale parameters. *Biometrika* **31**, 200–15.
PLACKETT, R. L. (1946). Some generalizations in the multifactorial design. *Biometrika* **33**, 328–32.
PLACKETT, R. L. (1947). Limits of the ratio of mean range to standard deviation. *Biometrika* **34**, 120–2.
PLACKETT, R. L. (1950). Some theorems in least squares. *Biometrika* **37**, 149–57.
PLACKETT, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika* **41**, 351–60.
POLLARD, J. H. (1966). On the use of the direct matrix product in analysing certain stochastic population models. *Biometrika* **53**, 397–415.
POYIADJIS, G., DOUCET, A. & SINGH, S. S. (2011). Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika* **98**, 65–80.
POURAHMADI, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86**, 677–90.
POURAHMADI, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **87**, 425–35.
PRENTICE, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69**, 331–42.
PRENTICE, R. L. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika* **60**, 279–88.
PRENTICE, R. L. & BRESLOW, N. E. (1978). Retrospective studies and failure time models. *Biometrika* **65**, 153–8.
PRENTICE, R. L. & CAI, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79**, 495–512. Correction (1993), **80**, 711–2.
QIN, J. & ZHANG, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika* **84**, 609–18.
RAFTERY, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika* **83**, 251–66.
RAFTERY, A. E. & AKMAN, V. E. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika* **73**, 85–9.
RAO, C. R. (1948). Tests of significance in multivariate analysis. *Biometrika* **35**, 58–79.
RAO, C. R. (1959). Some problems involving linear hypotheses in multivariate analysis. *Biometrika* **46**, 49–58.
RAO, J. N. K. & SHAO, J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79**, 811–22.
RAO, J. N. K., KOVAR, J. G. & MANTEL, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika* **77**, 365–75.
REILLY, M. & PEPE, M. S. (1995). A mean score method for missing and auxiliary covariate data in regression models. *Biometrika* **82**, 299-314.
ROBERTS, G. O. & STRAMER, O. (2001). On inference for partially observed diffusion models using the Metropolis-Hastings algorithm. *Biometrika* **88**, 603–21.
ROBERTS, G. O. & TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83**, 95–110.
ROBINS, J. M. & WANG, N. (2000). Inference for imputation estimators. *Biometrika* **87**, 113–24.
ROBINS, J. M., SCHEINES, R., SPIRTES, P. & WASSERMAN, L. (2003). Uniform consistency in causal inference. *Biometrika* **90**, 491–515.
RODRIGUEZ, A., DUNSON, D. B. & GELFAND, A. E. (2009). Bayesian nonparametric functional data analysis through density estimation. *Biometrika* **96**, 149–62.
ROM, D. M. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika* **77**, 663–5.
ROTNITZSKY, A. & JEWELL, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* **77**, 485–97.
ROYALL, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika* **57**, 377–87.
SAMPFORD, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika* **54**, 499–513.
SÄRNDAL, C. E. (1980). On π-inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika* **67**, 639–50.
SCHOENFELD, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika* **67**, 145–53.
SCHOENFELD, D. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* **68**, 316–9.
SCHWARTZMAN, A. & LIN, X. (2011). The effect of correlation in false discovery rate estimation. *Biometrika* **98**, 199–214.
SCHWEDER, T. & SPJØTVOLL, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* **69**, 493–502.

Scott, A. J. & Wild, C. J. (1997). Fitting regression models to case-control data by maximum likelihood. *Biometrika* **84**, 57–71.
Seber, G. A. F. (1970). Estimating time-specific survival and reporting rates for adult birds from band returns. *Biometrika* **57**, 313–8.
Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika* **81**, 115–31.
Shephard, N. & Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653–67.
Shibata, R. (1981). An optimal selection of regression variables. *Biometrika* **68**, 45–54.
Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika* **69**, 242–4.
Siegmund, D. (1978). Estimation following sequential tests. *Biometrika* **65**, 341–9.
Siegmund, D. O., Zhang, N. R. & Yakir, B. (2011). False discovery rate for scanning statistics. *Biometrika* **98**, 979–85.
Sillito, G. P. (1947). The distribution of Kendall's τ coefficient of rank correlation in rankings containing ties. *Biometrika* **34**, 36–40.
Silverman, B. W. (1978). Choosing the window width when estimating a density. *Biometrika* **65**, 1–11.
Silverman, B. W. & Young, G. A. (1987). The bootstrap: to smooth or not to smooth? *Biometrika* **74**, 469–79.
Silvey, S. D. & Titterington, D. M. (1973). A geometric approach to optimal design theory. *Biometrika* **60**, 21–32.
Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Anal.* **1**, 833–60.
Slater, P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika* **48**, 303–12.
Smith, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika* **62**, 407–16.
Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67–90.
Smith, R. L., Tawn, J. A. & Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika* **84**, 249–68.
Smith, T. M. F. (2001). *Biometrika* Centenary: Sample surveys. *Biometrika* **88**, 167–94.
Sprott, D. A. (1975). Marginal and conditional sufficiency. *Biometrika* **62**, 599–606.
Stefanski, L. A. (1985). The effects of measurement error on parameter estimation. *Biometrika* **72**, 583–92.
Stefanski, L. A. & Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703–16.
Stone, M. (1977). Asymptotics for and against cross-validation. *Biometrika* **64**, 29–35.
Strauss, D. J. (1975). A model for clustering. *Biometrika* **62**, 467–75.
Struthers, C. A. & Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73**, 363–9.
Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* **42**, 412–6.
'Student' (1938). Comparison between balanced and random arrangements of field plots. *Biometrika* **29**, 363–78.
Tarone, R. E. (1975). Tests for trend in life table analysis. *Biometrika* **62**, 679–82.
Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika* **75**, 397–415.
Tiao, G. C. & Grupe, M. R. (1980). Hidden periodic autoregressive-moving average models in time series. *Biometrika* **67**, 365–73.
Tiao, G. C. & Zellner, A. (1964). Bayes's theorem and the use of prior knowledge in regression analysis. *Biometrika* **51**, 219–30.
Tibshirani, R. (1989). Noninformative priors for one parameter too many. *Biometrika* **76**, 604–8.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Statist. Soc.* B **58**, 267–88.
Tierney, L., Kass, R. E. & Kadane, J. B. (1989). Approximate marginal densities of nonlinear functions. *Biometrika* **76**, 425–33.
Tiku, M. L. (1967). Estimating the mean and standard deviation from a censored normal sample. *Biometrika* **54**, 155–65.
Titterington, D. M. & Cox, D. R., Ed. (2001). *Biometrika: One Hundred Years*. Oxford: Oxford University Press.
Tong, H. (2001). A personal journey through time series in *Biometrika*. *Biometrika* **88**, 195–218.
Tsai, W. Y., Jewell, N. P. & Wang, M.-C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74**, 883–6.
Tsay, R. (1986). Nonlinearity tests for time series. *Biometrika* **73**, 461–6.
Tsiatis, A. A. (1980). A note on a goodness-of-fit test for the logistic regression model. *Biometrika* **67**, 250–1.
Tsiatis, A. A. & Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447–58.
Tsiatis, A. A. & Mehta, C. (2003). On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika* **90**, 367-78.
Turin, G. L. (1960). The characteristic function of Hermitian quadratic forms in complex normal variables. *Biometrika* **47**, 199–201.
Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika* **70**, 411–20.
Vaida, F. & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* **92**, 351–70.
Varin, C. & Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika* **92**, 519–28.

VON BONIN, G. & MORANT, G. M. (1939). Indian races in the United States. A survey of previously published cranial measurements. *Biometrika* **30**, 94–129.
WANG, H., LI, R. & TSAI, C.-L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* **94**, 553–68.
WANG, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82**, 385–97.
WATSON, G. S. (1955). Serial correlation in regression analysis. I. *Biometrika* **42**, 327–41.
WATSON, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika* **48**, 109–14.
WATSON, G. S. (1962). Goodness-of-fit tests on a circle. II. *Biometrika* **49**, 57–63.
WATSON, G. S. & LEADBETTER, M. R. (1964). Hazard analysis. I. *Biometrika* **51**, 175–84.
WATSON, G. S. & WILLIAMS, E. J. (1956). On the construction of significance tests on the circle and the sphere. *Biometrika* **43**, 344–52.
WEDDERBURN, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* **63**, 27–32.
WEI, L. J. & JOHNSON, W. E. (1985). Combining dependent tests with incomplete repeated measurements. *Biometrika* **72**, 359–64.
WEI, L. J., YING, Z. & LIN, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests *Biometrika* **77**, 845–51.
WELCH, B. L. (1937). On the z-test in randomized blocks and Latin squares. *Biometrika* **29**, 21–52.
WERMUTH, N. & LAURITZEN, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika* **70**, 537–52.
WHITTLE, P. (1952). Tests of fit in time series. *Biometrika* **39**, 309–18.
WHITTLE, P. (1955). The outcome of a stochastic epidemic - a note on Bailey's paper. *Biometrika* **42**, 116–22.
WHITTLE, P. (1963). On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix. *Biometrika* **50**, 129–34.
WIEAND, S., GAIL, M. H., JAMES, B. R. & JAMES, K. L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika* **76**, 585–92.
WIKLE, C. K. & CRESSIE, N. (1999). A dimension-reduction approach to space-time Kalman filtering. *Biometrika* **86**, 815–29.
WILLIAMS, C. B. (1940). A note on the statistical analysis of sentence-length as a criterion of literary style. *Biometrika* **31**, 356–61.
WILLIAMS, D. A. (1976). Improved likelihood ratio tests for complete contingency tables. *Biometrika* **63**, 33–7.
WILLIAMS, E. J. (1952). Use of scores for the analysis of association in contingency tables. *Biometrika* **39**, 274–89.
WILLIAMS, R. M. (1952). Experimental designs for serially correlated observations. *Biometrika* **39**, 151–67.
WOLFINGER, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80**, 791–5.
WOLPERT, R. L. & ICKSTADT, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–67.
WONG, F., CARTER, C. K. & KOHN, R. (2003). Efficient estimation of covariance selection models. *Biometrika* **90**, 809–30.
WORSLEY, K. J. (1982). An improved Bonferroni inequality and applications. *Biometrika* **69**, 297–302.
WORSLEY, K. J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika* **73**, 91–104.
WU, C. F. J. (1993). Construction of supersaturated designs through partially aliased interactions. *Biometrika* **80**, 661–9.
WU, W. B. & POURAHMADI, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90**, 831–44.
YANG, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika* **92**, 937–50.
YATES, F. (1939). The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments. *Biometrika* **30**, 440–66.
YATES, F. (1948). The analysis of contingency tables with groupings based on quantitative characters. *Biometrika* **35**, 176–81.
YUAN, M. & LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94**, 19–35.
YULE, G. U. (1939). On sentence-length as a statistical characteristic of style in prose: With application to two cases of disputed authorship. *Biometrika* **30**, 363–90.
ZEGER, S. L. (1988). A regression model for time series of counts. *Biometrika* **75**, 621–9.
ZELEN, M. (1971). The analysis of several 2×2 contingency tables. *Biometrika* **58**, 129–37.
ZELEN, M. & FEINLEIB, M. (1969). On the theory of screening for chronic diseases. *Biometrika* **56**, 601–14.
ZHANG, H. H. & LU, W. (2007). Adaptive lasso for Cox's proportional hazards model. *Biometrika* **94**, 691–703.
ZHAO, L. P. & PRENTICE, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* **77**, 642–8.

[*Received April* 2012. *Revised October* 2012]