J. R. Statist. Soc. B (1982),
44, *No.* 2, *pp.* 234–243

Inference for the Linear Model using a Likelihood based on Ranks

By A. N. Pettitt,

University of Technology, Loughborough, Leics., U.K.

[Received April 1980. Final revision July 1981]

SUMMARY

An approximation to the marginal likelihood of the ranks of a set of observations is used to analyse data. The observations, after some arbitrary monotone transformation, are assumed to be independently normally distributed with means related by a linear model. An approximate posterior distribution for the parameters of the linear model is given and used to find predictive distributions for the ordering of future observations. A numerical example is given. Appendices give mathematical details of the approximation.

Keywords: DISTRIBUTION-FREE TESTS; KRUSKAL–WALLIS TEST; LINEAR MODEL; MULTIPLE COMPARISONS; NORMAL SCORES; PREDICTIVE PROBABILITY; RANK MARGINAL LIKELIHOOD

1. Introduction

MUCH has been written about the development of non-parametric and distribution-free statistical procedures and their use in the sciences, where the assumption of normal or quasi-normal data is dubious; see, for example, the introduction to the book of Bradley (1968).

Distribution-free procedures have been proposed for testing many statistical hypotheses. Most of these procedures are based directly on the ranks of the observations. Recent statistical theory and practice has tended to stress model building and estimation, rather than hypothesis testing. Models and estimation are naturally lacking from distribution-free or non-parameteric procedures, but estimation procedures have been developed which use only the ranks of observations, such as the Hodges-Lehmann estimate and robust R-estimates, see Huber (1977, Section 3.7) for example.

Recently Cox's regression model for life tables (Cox, 1972) has met with much success in the analysis of data. Kalbfleisch and Prentice (1973) show that Cox's likelihood is the marginal likelihood based on the ranks of observations, which are assumed to come, after some arbitrary increasing monotone transformation, from the extreme value distribution with density $\exp(y-e^y)$. The observations have different location parameters, but the same scale parameter. The extreme value distribution is skew and very appropriate for modelling life-times.

For experimental data, we usually assume that a symmetric error distribution is appropriate. In this paper we give an analysis, using an approximation to the rank marginal likelihood of a set of observations which, after some arbitrary monotone transformation, are assumed to come from the normal distribution, the observations having means which are related by a linear model. Our assumption about normality is therefore rather weak. Since the rank likelihood depends only on the ranks of the observations, we obviously do not need any genuinely quantitative observations, just ordinal observations, to compute the rank likelihood.

We use an approximation to compute a posterior distribution for the parameters of the linear model. We can use this posterior distribution to find Bayesian confidence intervals of predictive probabilities for future observations. This allows us to make inferences when some differences betweeen groups are inferred in the analysis, rather than just rejecting a hypothesis of no difference between groups.

The approximation can also be used to construct approximate likelihood ratio tests about subsets of parameters in the linear model. Under some circumstances, these likelihood ratio

 0035–9246/82/44234$2.00

tests have null distributions which are distribution-free. For the two-sample problem and the k-sample problem, the likelihood ratio tests are almost equivalent to the Terry–Hoeffding normal scores test for two-samples and a normal scores version of the Kruskal–Wallis statistic for the k-sample problem (see, for example, Bradley, 1968, Section 6, for tests based on normal scores).

In Section 2, we introduce the approximation; in Section 3 we consider inference, based on the approximation; Section 4 deals with numerical problems encountered in practice, including ties. A numerical example is given in Section 5; this illustrates the usefulness of the technique. The appendices include results deriving the approximation.

2. THE MARGINAL LIKELIHOOD FOR THE RANKS AND ITS APPROXIMATION

We assume that random variables $Y_1, \ldots, Y_N$, after some arbitrary monotone transformation, are independent normal with means $\mathbf{x}_i' \boldsymbol{\beta}$, $i = 1, \ldots, N$, and variances equal to one. Here $\mathbf{x}_i$ is the usual vector of known regressor variables, and $\boldsymbol{\beta}$ is the vector of k unknown parameters. We denote by X the usual $N \times k$ matrix of regressor vectors $\mathbf{x}_i'$.

Inference is to be made based only on the ranks of the Y_j's, so that the general location of the Y_j's cannot be estimated. We therefore assume that the linear model does not admit a constant term, μ say, and therefore no column of X is constant.

The marginal likelihood for the ranks $r_1, \ldots, r_N$ is given by

$$f(\mathbf{r} \mid \boldsymbol{\beta}) = \mathrm{pr}\,(Y_{\alpha_1} < Y_{\alpha_2} < \ldots < Y_{\alpha_N} \mid \boldsymbol{\beta}), \tag{2.1}$$

where α_i is the anti-rank of Y_j, that is $\alpha_i = j$ if, and only if, Y_j is the ith smallest of $Y_1, Y_2, \ldots, Y_N$. The probability in (2.1) is given by

$$\int \frac{1}{(2\pi)^{N/2}} \exp\{-\tfrac{1}{2}(\mathbf{y} - X\boldsymbol{\beta})'(\mathbf{y} - X\boldsymbol{\beta})\}\, d\mathbf{y}, \tag{2.2}$$

where the integration is over $y_{\alpha_1} < y_{\alpha_2} < \ldots < y_{\alpha_N}$.

Let $Z_{\alpha_1} < \ldots < Z_{\alpha_N}$ be the order statistics of a sample of size N from the standard normal distribution, then the results of the Appendices show that (2.1) can be approximated by

$$f^*(\mathbf{r} \mid \boldsymbol{\beta}) = \text{const.} \exp\{-\tfrac{1}{2}\boldsymbol{\beta}' X' C X \boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{X}' a\}, \tag{2.3}$$

where $\mathbf{a} = E(\mathbf{Z})$ and $C = I - A$, where $A = \mathrm{var}\,(\mathbf{Z})$, with $\mathbf{Z}' = (Z_{\alpha_1}, \ldots, Z_{\alpha_N})$; thus, if ξ_i is the mean value of the ith order statistic in a random sample of size N from the standard normal distribution, and r_j is the rank of the jth observation, then $a_j = \xi_{r_j}$. The matrix A is the variance–covariance matrix for the normal order statistics $Z_{\alpha_1}, \ldots, Z_{\alpha_N}$, so that if (ξ_{ij}) is the variance–covariance matrix for the normal order statistics, then $(A)_{ij} = \xi_{r_i, r_j}$, where the ith and jth observations have ranks r_i and r_j respectively. For the present, we assume that the ranks can be uniquely given to the observations and leave until Section 4 the problem of ties.

We note (2.3) can be rewritten as

$$f^*(\mathbf{r} \mid \boldsymbol{\beta}) = \text{const.}\exp(\tfrac{1}{2}\mathbf{m}' M^{-1}\mathbf{m}) \exp\{-\tfrac{1}{2}(\boldsymbol{\beta} - \mathbf{m})' M^{-1}(\boldsymbol{\beta} - \mathbf{m})\}, \tag{2.4}$$

where

$$M^{-1} = X'CX, \quad \mathbf{m} = MX'\mathbf{a}.$$

The positive semi-definiteness of C and the non-singularity of M is considered in Appendix 2.

Equation (2.4) shows that the approximation $f^*(\mathbf{r} \mid \boldsymbol{\beta})$ resembles very much the analysis for the standard linear model, except that we have replaced the vector of observations by $\mathbf{a}$, the vector of normal scores, and modifed $X'X$ to become $X'CX$. Note that the approximation is *not* equivalent to a weighted least squares analysis, assuming the errors have covariance matrix A.

The approximation (2.3) is not adequate for large values of $\| \boldsymbol{\beta} \|$ for "extreme" rankings $\mathbf{r}$. By an "extreme" ranking, we mean a ranking for which the exact likelihood (2.1) is maximized when $\|\boldsymbol{\beta}\| \rightarrow \infty$ in some way. For example, if the linear model $\mathbf{x}_i'\boldsymbol{\beta}$ reduces to the simple regression $x_i \beta$, then the ranking $r_1, ..., r_N$ is extreme if $x_{\alpha_1} < ... < x_{\alpha_N}$ or $x_{\alpha_1} > ... > x_{\alpha_N}$, since the exact likelihood (2.1) is then maximized in the first case when $\beta \rightarrow \infty$ and in the second by $\beta \rightarrow -\infty$. The approximation (2.4) has a maximum at the finite value $\boldsymbol{\beta} = \mathbf{m}$. This point is further discussed in the next section.

3. Inference Based on the Rank-likelihood

3.1. *Introduction*

If we consider the approximation (2.4) as a true likelihood, then we can see that $(\mathbf{m}, M)$ are sufficient statistics for $\boldsymbol{\beta}$, the likelihood is maximized at $\hat{\boldsymbol{\beta}} = \mathbf{m}$, and the variance of $\hat{\boldsymbol{\beta}}$ is *estimated* to be M, this being the value of the inverse of the matrix $-\{(\partial^2/\partial\beta_i\,\partial\beta_j)\log f^*(\mathbf{r}\,|\,\boldsymbol{\beta})\}$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. This result is derived from the approximate identity $\Sigma f^*(\mathbf{r}\,|\,\boldsymbol{\beta}) = 1$, where the summation is over the $N!$ possible rankings $\mathbf{r}$. It can be postulated that much of the theory for the likelihood is applicable for $\boldsymbol{\beta}$ in a neighbourhood of $\mathbf{0}$, so that, for example, the quantity $(\boldsymbol{\beta}-\mathbf{m})' M^{-1}(\boldsymbol{\beta}-\mathbf{m})$ is a pivotal statistic having an aproximate χ^2 distribution under the sampling distribution of $\mathbf{r}$.

3.2. *Posterior Distribution for the Parameter* $\boldsymbol{\beta}$

A standard Bayesian analysis for the linear model can be applied to the approximate rank likelihood.

If $\boldsymbol{\beta}$ has a prior multivariate normal with mean $\mathbf{0}$ and covariance matrix $\sigma^2 I$, then the approximation (2.3) gives the corresponding approximate posterior distribution of $\boldsymbol{\beta}$ as being multivariate normal with mean $\mathbf{m}_1$ and covariance matrix M_1 where

$$M_1^{-1} = X'CX + \sigma^{-2} I, \tag{3.1}$$

$$\mathbf{m}_1 = M_1 X'\mathbf{a}. \tag{3.2}$$

From numerical work this approximation has been found to be very accurate for all rankings for a limited number of choices of X; see also Brooks (1978). The prior distribution is equivalent to some sample information supporting the value $\boldsymbol{\beta} = \mathbf{0}$ as most likely.

For non-extreme rankings, we can let $\sigma^2 \rightarrow \infty$, to obtain a uniform vague prior for $\boldsymbol{\beta}$, and then the approximate posterior distribution $p^*(\boldsymbol{\beta}\,|\,\mathbf{r})$ is given by

$$p^*(\boldsymbol{\beta}\,|\,\mathbf{r}) \propto f^*(\mathbf{r}\,|\,\boldsymbol{\beta}),$$

that is an approximate multivariate normal distribution with mean $\mathbf{m}$ and covariance matrix M, see (2.4). For extreme rankings, $p^*(\boldsymbol{\beta}\,|\,\mathbf{r}) \propto f^*(\mathbf{r}\,|\,\boldsymbol{\beta})$ is a good approximation to the expression $f(\mathbf{r}\,|\,\boldsymbol{\beta})\, p_0(\boldsymbol{\beta})$, where $p_0(\boldsymbol{\beta})$ is a *locally uniform prior* distribution for $\boldsymbol{\beta}$ with $p_0(\boldsymbol{\beta}) \rightarrow 0$ as $\|\boldsymbol{\beta}\| \rightarrow \infty$. It could be argued that if we are considering ranked observations, then the information we would expect to gain from them is limited. Taking a prior $p_0(\boldsymbol{\beta})$, as above, is a reasonable prior to represent an "objective" analysis. The approximate posterior distribution for $\boldsymbol{\beta}$, with mean $\mathbf{m}$ and covariance matrix M (see (2.4)), has worked well in practice elsewhere, see Pettitt (1981).

3.3. *Predictive Probabilities*

The parameter $\boldsymbol{\beta}$ gives information about the orders of future observations. We might be interested in the predictive probability $\mathrm{pr}(Y^{(1)} < Y^{(2)}\,|\,\boldsymbol{\beta})$, where $Y^{(i)}$ is a future observation at the value $\mathbf{x}_{(i)}$ of the regressor variable (see Brooks, 1978, for a discussion of the two-sample problem). This would be of interest in both a standard multiple regression application and in a one- or two-way classified analysis of variance model. We note

$$\mathrm{pr}(Y^{(1)} < Y^{(2)}\,|\,\boldsymbol{\beta}) = \mathrm{pr}(Z < 0), \tag{3.3}$$

where Z is normally distributed with mean $(\mathbf{x}_{(1)} - \mathbf{x}_{(2)})'\boldsymbol{\beta}$ and variance equal to 2. The probability (3.3) is denoted by $g(\boldsymbol{\beta})$ and the posterior distribution of $\boldsymbol{\beta}$ can be used to make inferences about $g(\boldsymbol{\beta})$. A probability such as (3.3) is at the heart of some non-parametric analysis and is found in the definitions of the Mann–Whitney and Kendall's rank statistics. If the model is used in the analysis of variance, then we may wish to compute $\mathrm{pr}(Y^{(1)} < Y^{(2)} < Y^{(3)} \mid \boldsymbol{\beta})$, where $Y^{(i)}$ refers to a future observation in the ith group.

Comparing k future observations would involve the k multivariate normal distribution function, which can be computed using the algorithm of Milton (1970). It does not seem appropriate to consider linear combinations of the $Y^{(i)}$'s.

Returning to $g(\boldsymbol{\beta})$, (3.3), we note *a posteriori*, that

$$\mathrm{pr}(g(\boldsymbol{\beta}) \leqslant u \mid \mathbf{r}) \simeq \Phi\left\{\frac{2\Phi^{-1}(u)+b_1}{b_2}\right\}, \tag{3.4}$$

where $(\mathbf{x}_{(1)}-\mathbf{x}_{(2)})'\boldsymbol{\beta}$ has the posterior normal distribution with mean b_1 and variance b_2^2, and $\Phi(\cdot)$ is the standard normal distribution function. The result (3.4) can be used to find Bayesian confidence intervals for $g(\boldsymbol{\beta})$ over regions where the density of $g(\boldsymbol{\beta})$ is highest.

From Brooks (1978, equation (9)), we note that the posterior mean of $g(\boldsymbol{\beta})$, or the predictive probability $\mathrm{pr}(Y^{(1)} < Y^{(2)})$, can be approximated by

$$\Phi\left\{-\frac{b_1}{\sqrt{2}}\right\}+\frac{1}{8\sqrt{2}}b_2^2\,b_1\,\phi\left(\frac{b_1}{\sqrt{2}}\right), \tag{3.5}$$

where $\phi(\cdot)$ is the standard normal density function. Similar expansions can be found when comparing the order of k $Y^{(i)}$'s.

3.4. *Choosing between Models*

One approach, to investigate the plausibility of possible values of $\boldsymbol{\beta}$, is to follow Box and Tiao (1973, Section 2.7.1) and consider highest posterior density regions. The approximate posterior distribution of

$$Q(\boldsymbol{\beta}) = (\boldsymbol{\beta}-\mathbf{m})'M^{-1}(\boldsymbol{\beta}-\mathbf{m})$$

is χ^2 with k degrees of freedom. The value $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is included in a highest posterior density region or Bayesian confidence region of size $(1-\alpha)$ if $Q(\tilde{\boldsymbol{\beta}}) \leqslant \chi^2_{k,\alpha}$, where $\chi^2_{k,\alpha}$ is the upper α point of the χ^2 distribution with k degrees of freedom. In particular, $\boldsymbol{\beta} = \mathbf{0}$ is included in the region if

$$Q(\mathbf{0}) = \mathbf{m}'M^{-1}\mathbf{m} \leqslant \chi^2_{k,\alpha}. \tag{3.6}$$

If $\mathbf{m}'M^{-1}\mathbf{m} = \chi^2_{k,\alpha'}$, then $\mathbf{0}$ is just included in a $(1-\alpha')$ size confidence region for $\boldsymbol{\beta}$; the larger the value of α', the more the data support the value $\mathbf{0}$ of $\boldsymbol{\beta}$.

If $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$ and $X = (X_0, X_1)$, then in model building we are interested frequently in whether the larger model $X\boldsymbol{\beta}$ can be reduced to the smaller model $X_0\boldsymbol{\beta}_0$ by putting $\boldsymbol{\beta}_1 = \mathbf{0}$. In this case, we are interested in the marginal distribution of $\boldsymbol{\beta}_1$ and whether $\mathbf{0}$ is a value for $\boldsymbol{\beta}_1$ supported by the data. If $\boldsymbol{\beta}_1$ has an approximate multivariate normal distribution with mean $\mathbf{m}_1$ and covariance M_1 then, following the ideas above, we consider the value of $\mathbf{m}_1' M^{-1}\mathbf{m}_1$ and refer it to the χ^2 distribution with k_1 degrees of freedom, k_1 being the dimension of $\boldsymbol{\beta}_1$. If $\mathbf{m}_1' M^{-1}\mathbf{m}_1$ is small, then this suggests that the model can be simplified, by putting $\boldsymbol{\beta}_1 = \mathbf{0}$.

These ideas have their sampling theory equivalents. If we consider the likelihood ratio test of H_0: $\boldsymbol{\beta} = \mathbf{0}$ against H_1: $\boldsymbol{\beta} \neq \mathbf{0}$, using the approximate likelihood $f^*(\mathbf{r} \mid \boldsymbol{\beta})$ (2.3), the corresponding likelihood ratio test is to reject H_0 if $\mathbf{m}'M^{-1}\mathbf{m} \geqslant c$, considering $-2\log(\text{likelihood ratio})$; for an α size test, we can postulate that $c = \chi^2_{k,\alpha}$ is an excellent approximation. The statistic $\mathbf{m}'M^{-1}\mathbf{m}$, which is equivalent to one considered by Prentice (1978, Sections 3.3 and 3.5), has a sampling distribution which is *distribution-free* when H_0 is true. For a test of H_0: $\boldsymbol{\beta}_1 = \mathbf{0}$, against H_1: $\boldsymbol{\beta}_1 \neq \mathbf{0}$, when $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$, the likelihood ratio test, based on (2.3), is to reject H_0 if

$$\mathbf{m}'M^{-1}\mathbf{m} - \mathbf{m}_0 M_0^{-1}\mathbf{m}_0 \geqslant c_1, \tag{3.7}$$

where $\mathbf{m}_0, M_0$ are defined by

$$\mathbf{m}_0 = M_0 X_0' \mathbf{a}, \quad M_0^{-1} = X_0' C X_0.$$

It can be postulated for an α size test, to a good approximation, that $c = \chi^2_{k_1,\alpha}$, where k_1 is the dimension of $\boldsymbol{\beta}_1$. It can be shown that the statistic on the left-hand side of (3.7) is equal to $\mathbf{m}_1' M_1^{-1} \mathbf{m}_1$, so making the Bayesian and likelihood ratio test statistics equivalent.

4. Practical Aspects

4.1. *Numerical Approximations*

To carry out the analysis of the previous sections we need to know $(\xi)_i$ and $(\xi)_{ij}$, the vector of the means and the matrix of covariances of the normal order statistics. These are tabulated in Pearson and Hartley (1972) and Tietjen *et al.* (1977), for example. For numerical work, good approximations to ξ_i are given by Blom's approximation

$$\xi_i^* = \Phi^{-1}\{(i-\tfrac{3}{8})/(N+\tfrac{1}{4})\},$$

where $\Phi(x)$ is the standard normal distribution function (see David, 1970, Section 4.4). The covariances can be approximated by a Taylor series (see David, 1970, Section 4.5). For any approximation, ξ_{ij}^*, the conditions $\Sigma_i \xi_{ij}^* = \Sigma_j \xi_{ij}^* = 1$ must hold so that the matrix C remains positive semi-definite.

In this work we used only the first term of the Taylor series approximation, suitably normalized; for the covariances, Davis and Stephens (1978) provide an algorithm which gives excellent approximations.

4.2. *Ties*

We deal with ties by using the "equiprobable" assumption discussed by Bradley (1968, Section 3.3.3). Suppose in a sample of N order statistics there are two groups of ties, say $Z_{(s+1)}, \ldots, Z_{(s+m)}$ and $Z_{(r+1)}, \ldots, Z_{(r+k)}$. Instead of believing the observations in each group to have the same value, suppose we believe that they really are distinct, but we have lost their labels denoting their order within the group. If we take one of the $Z_{(s+1)}, \ldots, Z_{(s+m)}$ at random and call it X, then X has marginal density given by

$$f_X(x) = \frac{1}{m} \sum_{j=s+1}^{s+m} f_j(x),$$

where $f_j(\cdot)$ is the density of the jth distinct order statistic when no ties are assumed. This is just a uniform mixture of the densities f_j. If we take two at random from the same group and denote them by X_1 and X_2, then we can find their joint density, which again is a mixture.

If we take one from the first group and one from the second group, and denote them respectively by X and Y, we can also find their joint density.

From these results we find, for the Z's from the standard normal, that

$$E(X) = \frac{1}{m} \sum_{i=s+1}^{s+m} \xi_i, \tag{4.1}$$

$$\operatorname{var}(X) = \frac{1}{m} \sum_{s+1}^{s+m} \xi_{ii} + \frac{1}{m} \sum_{s+1}^{s+m} \xi_i^2 - \{E(X)\}^2 \tag{4.2}$$

$$\operatorname{cov}(X_1, X_2) = \frac{1}{m(m-1)} \sum_{i \neq j, s+1}^{s+m} \xi_{ij} + \frac{1}{m-1}\{E(X)\}^2 - \left(\sum_{s+1}^{s+m} \xi_i^2\right) \bigg/ m(m-1), \tag{4.3}$$

$$\operatorname{cov}(X, Y) = \frac{1}{km} \sum_{i=s+1}^{s+m} \sum_{j=r+1}^{r+k} \xi_{ij}. \tag{4.4}$$

We note $E(Y)$ is obtained from (4.1) but the summation is over $(r+1)$ to $(r+k)$.

We now use $E(X)$ for the mean of $Z_{(s+j)}, j = 1, \ldots, m$, $E(Y)$ for any $Z_{(r+j)}, j = 1, \ldots, k$, and similarly for the variances and covariances, using var(X), etc. The elements of **a** and A are now adjusted in line with these formulae, and the tied observations in the data. Of course, the technique deals with more than two groups of ties. We note that the vector of expected values, (ξ_i), sums to zero as does the vector **a**, adjusted for ties. Similarly, all the rows and columns of the matrix A, adjusted for ties, sum to one.

5. An Example

5.1. *Introduction*

In this section we consider an example taken from Siegel (1956)

In the analyses, we use a locally uniform prior, the results of Section 3.2 and the numerical approximations of Section 4..

5.2. *One-way Analysis of Variance*

Consider some data given by Siegel (1956, p. 187) as an example of the Kruskal–Wallis test for the one-way analysis of variance. The ranks of the three groups are: group I: 7, 13, 14, 12; group II: 2, 8, 10, 11, 6; group III: 4, 9, 3, 1, 5, reversing the order of the groups in Siegel.

For some arbitrary monotone function $h(\cdot)$, we propose the model

$$E\{h(Y_{ij})\} = \beta_i, \quad j = 1, \ldots, n_i, \quad i = 1, 2, 3$$

for the jth observation in the ith group of size n_i. Since the rank analysis does not involve a location parameter, we assume $\beta_3 = 0$, without loss of generality. Then β_1 and β_2 represent the differences between the first and third groups, and the second and third groups, respectively. Using a locally uniform prior for $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, we can first compute $Q(\mathbf{0})$ (3.6) to see whether $\boldsymbol{\beta} = (0, 0)'$ is a plausible value. We find $Q(\mathbf{0}) = 7{\cdot}75$, which when referred to the χ^2 distribution with two degrees of freedom, corresponds to about the upper 2 per cent point of the distribution. Taking a Bayesian stance, $\boldsymbol{\beta} = 0$ would then be just included in the 98 per cent region for $\boldsymbol{\beta}$. Alternatively, if we calculate minus twice the log rank likelihood ratio, for testing H_0: $\boldsymbol{\beta} = \mathbf{0}$ against H_1: $\boldsymbol{\beta} \neq \mathbf{0}$, this is also given by $Q(\mathbf{0})$, and so the observed value 7·75 has observed significance probability equal to about 2 per cent. This should be compared with the value of the Kruskal–Wallis statistic, 6·4, which has a significance probability less than 0·049, but greater than 0·01 (see Siegel, 1956, p. 187).

The null hypothesis H_0: $\boldsymbol{\beta} = \mathbf{0}$ is, of course,, equivalent to the hypothesis that there is no difference between groups, the null hypothesis tested by the Kruskal–Wallis test.

If we investigate the parameters of the model, we find, *a posteriori*, that $\boldsymbol{\beta}'$ has mean (2.14, 0·82) and $\text{var}(\beta_1) = 0{\cdot}59$, $\text{var}(\beta_2) = 0{\cdot}43$, $\text{cov}(\beta_1, \beta_2) = 0{\cdot}25$. We can use the posterior distribution of $\boldsymbol{\beta}$ to find $\text{pr}(Y^{(1)} < Y^{(2)} \mid \boldsymbol{\beta})$, $\text{pr}(Y^{(1)} < Y^{(3)} \mid \boldsymbol{\beta})$, $\text{pr}(Y^{(2)} < Y^{(3)} \mid \boldsymbol{\beta})$ where $Y^{(i)}$, $i = 1, 2, 3$, represents a future observation from the ith group.

Defining $g_{12}(\boldsymbol{\beta}) = \text{pr}(Y^{(1)} < Y^{(2)} \mid \boldsymbol{\beta})$,. we find that, using (3.4), *a posteriori*, $\text{pr}(0{\cdot}00 < g_{12}(\boldsymbol{\beta}) < 0{\cdot}46) = 0{\cdot}90$ and the posterior mean of $g_{12}(\boldsymbol{\beta})$ is 0·19, using (3.5). Defining $g_{13}(\boldsymbol{\beta})$ and $g_{23}(\boldsymbol{\beta})$ similarly, we find that: $\text{pr}(0{\cdot}00 < g_{13}(\boldsymbol{\beta}) < 0{\cdot}27) = 0{\cdot}90$, $E\{g_{13}(\boldsymbol{\beta})\} \simeq 0{\cdot}08$; $\text{pr}(0{\cdot}09 < g_{23}(\boldsymbol{\beta}) < 0{\cdot}58) = 0{\cdot}90$, $E\{g_{23}(\boldsymbol{\beta})\} \simeq 0{\cdot}29$. The Bayesian confidence intervals for the various $g(\boldsymbol{\beta})$ are *not* highest posterior density regions. We can estimate $\text{pr}(Y^{(i)} < Y^{(j)})$ using the Mann–Whitney statistic computed for the ith and jth groups. These estimates are found to be the following: for $\text{pr}(Y^{(1)} < Y^{(2)})$, 0·15; for $\text{pr}(Y^{(1)} < Y^{(3)})$, 0·05; for $\text{pr}(Y^{(2)} < Y^{(2)})$, 0·24. It would be difficult to give confidence intervals for these estimates, unless we were to use some procedure like the jack-knife. To compare all three groups, we can consider $g_{321}(\boldsymbol{\beta}) = \text{pr}(Y^{(3)} < Y^{(2)} < Y^{(1)} \mid \boldsymbol{\beta})$, or any other ordering of the Y's. Using the algorithm of Milton (1970) to compute multivariate normal integrals, we find that the mean of $g_{321}(\boldsymbol{\beta})$ is 0·540, the posterior mode is 0·556 and the median is 0·547. A 90 per cent Bayesian confidence interval is (0·31, 0·73), and a 95 per cent interval is (0·28, 0·76). A reference value for $g_{321}(\boldsymbol{\beta})$ might be $(3!)^{-1}$ or 1/6 which is equal to $\text{pr}(Y^{(3)} < Y^{(2)} < Y^{(1)} \mid \boldsymbol{\beta} = \mathbf{0})$, that is the value when there is no difference between the groups. We see that 1/6 is not included in the 95 per cent interval for $g_{321}(\boldsymbol{\beta})$ and not well supported by the data. A direct sample estimate of $\mathrm{pr}(Y^{(3)} < Y^{(2)} < Y^{(1)})$ is given by the proportion of triples (y_1, y_2, y_3) of observations satisfying $(y_3 < y_2 < y_1)$. This is found to be 62 out of 100. Again, finding confidence intervals for this quantity would be involved.

5.3. *Further Possible Analyses*

The generality of the techniques of this paper allow us to investigate more complex models than just the previous one of Section 5.2. For example, suppose the data referred to some experiment and the observations were taken in a known chronological order, $t_1, t_2, \ldots, t_{15}$,, say, t_i being the time (absolute or relative) of the ith observation. We might suspect a time trend in the observations which could be modelled by adding the component $\beta_4 t_i$ to the model. If a serial time effect were suspected, then we might add the term $\beta_5(-1)^{t_i}$ to the model, where we now assume $t_1, \ldots, t_{15}$ is a permutation of $1, \ldots, 15$, giving the relative time order of the observations. Other possible covariates can be added to the model similarly,, and their importance investigated, using the techniques of Section 3.4.

6. FURTHER POINTS

We have found the approximation of the marginal rank likelihood useful for the analysis of many types of data. It appears that the technique is useful in deciding whether quadratic terms in polynomial regression or interaction terms in factorial models can be reduced by transformation of the response variable. The rank analysis gives some information about the intrinsic non-linearity or non-monotonicity of the data. The technique has also proved useful in the analysis of biological assay data and the problem of rankings of objects by groups of judges. These applications are the subjects of future papers.

Computer programs, in FORTRAN,, are available from the author to carry out the analysis of this paper.

It is possible to incorporate the analysis in a GENSTAT program and would be extremely useful if it could be incorporated in GLIM.

The Appendix includes details of the approximation for densities other than the normal density. Explicit results are given for the logistic, extreme-value and double-exponential. The logistic density gives rise to Wilcoxon or uniform scores, the extreme-value density gives rise to exponential scores and the double-exponential density to scores giving less weight to observations with extreme ranks.

AKNOWLEDGEMENTS

Part of this work was undertaken while the author was supported by a University of Queensland Postdoctoral Research Fellowship. The author is extremely grateful for the interest shown by Dr Ray Meddis and for the comments of referees and an editor on an earlier draft of this paper.

REFERENCES

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families*. Chichester: Wiley.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, Mass.: Addison-Wesley.

BRADLEY, J. V. (1968). *Distribution Free Statistical Tests*. New Jersey: Prentice-Hall.

BROOKS, R. J. (1978). Bayesian analysis of a two-sample problem based on the rank order statistics. *J. R. Statist. Soc.* B, **40**, 50–57.

COX, D. R. (1972). Regression models and life tables. *J. R. Statist. Soc.* B, **34**, 187–220.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

DAVID, H. A. (1970). *Order Statistics*. New York: Wiley.

DAVIS, C. S. and STEPHENS,, M. A. (1978), AS 128. Approximating the covariance matrix of normal order statistics. *Appl. Statist.*, **27**, 206–212.

HAJEK, J. and SIDAK, Z. S. (1967). *Theory of Rank Tests*. New York: Academic Press.

HUBER, P. J. (1977). Robust statistical procedures. *Regional Conf. Series in Applied Maths.*, No. 27. Philadelphia: S.I.A.M.
KALBFLEISCH, J. D. and PRENTICE, R. L. (1973). Marginal likelihoods based on Cox's regression and life models. *Biometrika*, **60**, 267–278.
LINDLEY, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint. Part 2, Inference.* London: Cambridge University Press.
MILTON, R. C. (1970). *Rank Order Probabilities: Two-sample Normal Shift Alternatives.* New York: Wiley.
PEARSON, E. S. and HARTLEY, H. O. (1972). *Biometrika Tables for Statisticians*, Vol. 2. London: Cambridge University Press.
PETTIT, A. N. (1981). Posterior probabilities for a change-point using ranks. *Biometrika*, **68**, 443–450.
PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, **65**, 167–179.
SIEGEL, S. (1956). *Nonparametric Statistics for the Behavioural Sciences.* New York: McGraw-Hill.
TIETJEN, G. L., KAHANER, D. K. and BECKMAN, R. J. (1977). Variances and covariances of normal order statistics for sample sizes 2 to 50. In *Selected Tables in Mathematics Statistics*, V, 1–74. Providence: American Mathematical Society.

APPENDIX 1

The Approximation

Suppose $Y_1, ..., Y_N$ are independent random variables such that Y_j has the distribution function $F(y-\theta_j)$ and density $f(y-\theta_j)$. We assume that $F(\cdot)$ is continuous and $f(\cdot)$ satisfies regularity conditions so that the asymptotic theory for maximum likelihood estimation of θ holds when a random sample is taken from a distribution with density $f(y-\theta)$ and $-\infty<\theta<\infty$ (see, for example, Cox and Hinkley, 1974, Section 9.1). Define $\alpha_j, j=1,...,N$, to be the anti-ranks of the Y's so that $\alpha_j = i$ if, and only if, Y_i is the jth smallest in the sample. We want to find approximations to the probability of the rank order $Y_{\alpha_1}<Y_{\alpha_2}<...<Y_{\alpha_N}$. Now

$$\operatorname{pr}(Y_{\alpha_1}<...<Y_{\alpha_N}) = \int \left\{ \prod_{j=1}^{N} f(y_j-\theta_j) \right\} dy_1, ..., dy_N, \tag{A1}$$

where the integration is over the region $y_{\alpha_1}<...<y_{\alpha_N}$. Approximations to (A1) have been given in the literature. Milton (1970) suggests a Taylor series expansion for $f(y-\theta)$ about $f(y)$ involving θ and θ^2.

Exact numerical results for (A1) are available for the two-sample problem, see Milton (1970). Brooks (1978) suggests, for the two-sample problem again, an approximation of the form

$$\exp(\theta\gamma_1 - \tfrac{1}{2}\theta^2\gamma_2), \tag{A2}$$

where $\theta_1 = ... = \theta_n = 0$, and $\theta_{n+1} = ... = \theta_N = \theta$ and γ_1, γ_2 are rank statistics. He does this for $F(\cdot)$ corresponding to normal and logistic distribution functions. We generalize Brooks' result for any distribution satisfying the regularity conditions mentioned above and for the multisample problem, rather than the two-sample problem.

Assuming $f(\cdot)$ satisfies the regularity conditions mentioned above, we can expand $\ln f(y-\theta)$ by Taylor series about $\ln f(y)$.

Defining $g(y) = -f'(y)/f(y)$, we have

$$\ln f(y-\theta) \simeq \ln f(y) + \theta g(y) - \frac{\theta^2}{2} g'(y)$$

or

$$f(y-\theta) \simeq f(y) \exp\left\{ \theta g(y) - \frac{\theta^2}{2} g'(y) \right\} \tag{A3}$$

We now replace $f(y_j-\theta_j)$ by the right-hand side of (A3) in (A1) to obtain

$$\operatorname{pr}(Y_{\alpha_1}<...<Y_{\alpha_N}) \simeq \int \exp\{\sum \theta_j g(y_j) - \tfrac{1}{2}\sum \theta_j^2 g'(y_j)\} \left\{ \prod_{j=1}^{N} f(y_j) \right\} dy_1 ... dy_N, \tag{A4}$$

where the integration is over $y_{\alpha_1}<...<Y_{\alpha_N}$.

We can rewrite (A4) as

$$(N!)^{-1}\,E[\exp\{\sum\theta_j g(V_{r_j})-\tfrac{1}{2}\sum\theta_j^2\,g'(V_{r_j})\}], \tag{A5}$$

where $V_1<\ldots<V_N$ are the order statistics of a sample of size N from the population with distribution function $F(y)$ and $r_1,\ldots,r_N$ are the ranks of $Y_1,\ldots,Y_N$.

Consider first the case where $f(y)$ is the standard normal density, then (A3) is exact with $g'(y)=1$, and, in this case, $\Sigma\,\theta_j^2\,g'(V_{r_j})$ becomes $\Sigma\,\theta_j^2$. We now consider an approximation to $E\exp[\Sigma\{\theta_j\,g(V_{r_j})\}]$. We assume that $\Sigma\,\theta_j\,g(V_{r_j})$ can be well approximated by a normal random variable. Write $\mathbf{W}=(W_1,\ldots,W_N)'=(g(V_{r_1}),\ldots,g(V_{r_N}))'$, then we assume that $\mathbf{W}$ has a multivariate distribution with mean $\mathbf{a}=(\xi_{r_1},\ldots,\xi_{r_N})'$ and covariance matrix A, so that $(A)_{ij}=\xi_{r_i,r_j}$. Here we assume ξ_i is the mean of $g(V_i)$ *and* ξ_{ij} is the covariance of $g(V_i)$ and $g(V_j)$.

With $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_N)'$, we have the approximation

$$E[\exp\{\sum\theta_j g(V_{r_j})\}]\simeq\exp(\boldsymbol{\theta}'.\mathbf{a}+\tfrac{1}{2}\theta'.A\boldsymbol{\theta}),$$

so that

$$\begin{aligned}\mathrm{pr}(Y_{\alpha_1}<\ldots<Y_{\alpha_N})&=\mathrm{pr}(Y_j\text{ has rank }r_j,\,j=1,\ldots,N)\\&\simeq(N!)^{-1}\exp\{-\tfrac{1}{2}\boldsymbol{\theta}'.\boldsymbol{\theta}+\tfrac{1}{2}\boldsymbol{\theta}'A\boldsymbol{\theta}+\boldsymbol{\theta}'.\mathrm{a}\}.\end{aligned}$$

For other densities, consider approximations to $E[\exp\{\Sigma\,\theta_j g(V_{r_j})-\tfrac{1}{2}\Sigma\,\phi_j g'(V_{r_j})]$, where $\phi_j=\theta_j^2$. Up to quadratic terms in the θ_j, we have

$$E[\exp\{\sum\theta_j g(V_{r_j})-\tfrac{1}{2}\sum\phi_j g'(V_{r_j})\}]\simeq\exp\{\boldsymbol{\theta}'.\mathbf{a}+\tfrac{1}{2}\boldsymbol{\theta}'.A.\boldsymbol{\theta}-\tfrac{1}{2}\boldsymbol{\phi}'.\mathbf{b}\},$$

by considering the joint moment generating function of $\mathbf{W}$ and $(g'(V_{r_1}),\ldots,g'(V_{r_N}))'$. Here $\boldsymbol{\phi}=(\phi_1,\ldots,\phi_N)'$ and $\mathbf{b}=(b^1,\ldots,b_N)$ with $b_i=E\{g'(V_{r_i})\}$. If B is a diagonal matrix and $B_{ii}=b_i$, then the approximation becomes

$$\exp\{\boldsymbol{\theta}'.\mathbf{a}+\tfrac{1}{2}\boldsymbol{\theta}'.\mathbf{A}.\boldsymbol{\theta}-\tfrac{1}{2}\boldsymbol{\theta}'.B.\boldsymbol{\theta}\},$$

giving finally the result

$$\mathrm{pr}(Y_{\alpha_1}<\ldots<Y_{\alpha_N})\simeq(N!)^{-1}\exp\{\boldsymbol{\theta}'.\mathbf{a}-\tfrac{1}{2}\boldsymbol{\theta}'.C.\boldsymbol{\theta}\}, \tag{A6}$$

where $C=B-A$.

Suppose now the θ_j's are connected by the linear model $\theta_j=\mathbf{x}_j'.\boldsymbol{\beta}$ and we replace $\boldsymbol{\theta}$ by $X\boldsymbol{\beta}$, then (A6) becomes

$$\mathrm{pr}(Y_{\alpha_1}<\ldots<Y_{\alpha_N})\simeq(N!)^{-1}\exp\{-\tfrac{1}{2}\boldsymbol{\beta}'X'CX\boldsymbol{\beta}+\boldsymbol{\beta}'X'\mathbf{a}\}. \tag{A7}$$

APPENDIX 2

The Positive Semi-definiteness of C

If it is assumed that the density $f(y)$ is strongly unimodal (Hajek and Sidak, 1967) then the probability of rank order (A1) is strongly unimodal with respect to $\theta_1,\ldots,\theta_N$, when any one of the θ's is set equal to zero. This follows from Barndorff-Nielsen (1978, Theorem 6.4). It therefore follows that C is positive semi-definite of rank $N-1$ with all its columns (or rows) adding to zero. Thus for non-trivial choices of X, which do not include a column or linear combinations of columns of X proportional to the vector of unities, $X'CX$ is positive definite and equation (2.4) holds.

APPENDIX 3

Results for Particular Densities

(a) *Normal distribution*

For $f(.)$ chosen to be standard normal, we obtain the results of Section 2.

(b) *Logistic distribution*

For $f(y) = e^{-y}/(1+e^{-y})^2$, the logistic density, we find

$$a_j = 2r_j/(N+1)-1$$
$$(A)_{ij} = 4r_j(N+1-r_i)/\{(N+1)^2(N+2)\}, \quad r_j \leqslant r_i$$
$$(B)_{jj} = \tfrac{1}{2}(N+1)\,A_{jj}.$$

(c) *An extreme value distribution*

For the extreme value distribution with density $f(y) = \exp(y-e^y)$, we find

$$a_j = \frac{1}{N}+\frac{1}{N-1}+\ldots+\frac{1}{N-r_j+1}-1,$$
$$(A)_{ij} = \frac{1}{N^2}+\frac{1}{(N-1)^2}+\ldots+\frac{1}{(N-r_j+1)^2}, \quad r_j \leqslant r_i,$$
$$(B)_{jj} = a_j+1.$$

These results can also be deduced from the explicit result for (A1), which Kalbfleisch and Prentice (1973) give.

(d) *Double exponential distribution*

For the double exponential distribution with $f(y) = \frac{1}{2}\exp\{-|y|\}$, after a limiting argument overcoming the discontinuity of $g(y)$ at $y = 0$, we find

$$a_j = \xi_{r_j},$$
$$(A)_{ij} = 1-\xi_{r_i}+\xi_{r_j}-\xi_{r_i}\xi_{r_j}, \quad r_j \leqslant r_i,$$
$$(B)_{jj} = 2^{-(N-1)}r_j\binom{N}{r_j}$$

where

$$\xi_j = 2^{-N}\sum_{i=j}^{N}\binom{N}{i}.$$

We can apply the techniques of Sections 2–5 with $\mathbf{a}$ and $C = B-A$ replaced by the appropriate form given by any of the choices above. A slight modification has to be made to the correction for ties given in Section 4.