DIRECTIONAL TESTS AND STATISTICAL FRAMES

Conditional inference has an ease of implementation that is generally unavailable with marginal inference. The main patterns for conditional inference are provided by the location and transformation families as initiated by Fisher, and by the exponential patterns as initiated by Neyman and Pearson; these are surveyed briefly together with some discussion as to how and why conditioning should be used in inference for them.

A more recent alternative pattern is provided by directional (or conical) tests and confidence methods; these lead to conditional inference for simple hypotheses with vector parameters, and can be extended to provide tests for treatment, for variance, and for treatment improvement, in the multivariate analysis of variance context.

This paper proposes the extraction of statistical frames by generalized conditioning procedures; these use versions of the conditional methods just mentioned, and then recombine the components by independence calculations. As an example, the partial likelihood of the proportional hazards model then becomes a standard likelihood for which third order asymptotics are available to give accurate tests and confidence intervals.

1. Introduction

Fisher (1934) examined the location and location-scale models and recommended inference conditional on an ancillary or configuration statistic, which describes the standardized spacings between the order statistics. Neyman and Pearson (1933) proposed similar tests and discussed their application in the context of exponential models. These two initiatives led to the two common patterns for conditional inference. This is discussed in Section 2 together with a brief survey of the sanctions for conditional approaches to inference.

Fraser and Massam (1985) proposed conical tests for a simple hypothesis concerning a vector parameter; the method also covered a vector interest parameter with nuisance parameters when a corresponding pivotal quantity was available. The approach examined an observed value conditionally in the cone radiating from the value expected under the hypothesis. Fraser (1987) extended this to more general contexts by analyzing in terms of the score function under the null parameter value. Skovgaard (1988) applied saddlepoint methods to this score function extension and used the term, directional tests, which we feel is more appropriate and adopt here. These developments are examined in Section 3.

Birnbaum (1962, 1972) as part of a detailed examination of inference principles used the term evidential meaning for the information in a model data combination; Evans, Fraser, Monette (1985, 1986) called a model-data combination an inference base and used the term content for the corresponding statistical information. In this paper a generalized model data combination is called a statistical frame, and extracting a statistical frame refers to the isolation of a generalized marginal or conditional model with its corresponding data.

This paper proposes a very general approach to inference: that the extraction of components can be done very generally and very freely, even in ways that violate global probabilities. The objective focusses on the quality of the retained statistical frame, largely ignoring the loss of perhaps contaminated information. This is consistent with the objective of scientific experimentation. One example of this extraction involves the construction of a one dimensional inference curve for a real parameter (Fraser and Reid, 1988). A device then for obtaining p-values and a significance function having third order accuracy is provided by the Lugannani and Rice (1980) formula and its extensions (Fraser, 1988; Fraser and Reid, 1990,1993). These results are surveyed in Section 4.

The proportion hazards model (Cox, 1972, 1975) is widely used particularly in the biomedical sciences; it leads to a partial likelihood which is generally viewed as not being a proper likelihood (Kalbfleisch and Prentice, 1973). The extraction principle and a related compounding principle present the partial likelihood as a true likelihood, and permit the use of the third order asymptotic theory for constructing tests and confidence regions. These results are surveyed in Section 5.

The approach is supportive and important to theory for meta-analysis.

2. Basic conditional methods and principles

2.1 Transformation model patterns

In a major paper entitled 'Two new properties of mathematical likelihood', Fisher (1934) discussed the analysis of a sample $(y_1, \ldots, y_n)$ from a location model $f(y - \theta)$ or from a location-scale model $\sigma^{-1} f\{\sigma^{-1}(y - \mu)\}$, where the functional form f is presumed known. His emphasis was on the use of ancillary information to recover information lost in considering only the maximum likelihood estimate; for some recent discussion see Kass (1989). The conditional method used has widespread implications in statistics.

The recovery of information, a technically defined information now called the Fisher information, was obtained by using the conditional model given an ancillary or configuration statistic rather than the marginal model of the maximum likelihood estimate itself. For the location model the configuration statistic was the sample complexion (a term he used)

$$a = (y_1 - \hat{\theta}, \ldots, y_n - \hat{\theta})' ,$$

with corresponding conditional distribution

$$f(\hat{\theta}|a;\theta) \propto \prod_1^n f(a_i + \hat{\theta} - \theta) ;$$

for the location-scale model the statistic was

$$a = \hat{\sigma}^{-1}(y_1 - \hat{\mu}, \ldots, y_n - \hat{\mu})' ,$$

with conditional distribution

$$f(\hat{\mu}, \hat{\sigma}|a;\mu,\sigma) \propto \prod_1^n f\{(\hat{\mu} - \mu + \hat{\sigma}a_i)/\sigma\} \left(\frac{\hat{\sigma}}{\sigma}\right)^n \hat{\sigma}^{-2} .$$

The generalization to the regression context was considered by Fraser (1957, 1967), Verhagen (1961).

Location and location scale models are simple examples of a general class of models, called transformation models (Peisakoff, 1951). The analysis of these models was partly developed in the context of fiducial analysis (Fraser, 1961a,b) and partly in the context of structural models (Fraser, 1966, 1968, 1979). Transformation models can generally be reexpressed as structural models, the addition being an explicit model component describing error in the system. We will describe the corresponding conditional analysis in terms of the structural models where the conditioning is automatic and conditioning principles are not needed.

A fairly general structural model has a response variable y that is generated by some transformation θ from an error variable e; the unknown transformation θ belongs to a transformation group G that is exact on the space of y and e. Exactness means that a transformation that carries some initial point into a new point is the only transformation that does so; transformations thus provide unique labels for image points of an initial point, and serve to describe how data are generated. For the continuous context the error has known density $f(e)$ with respect to say an invariant measure $dm(e)$; the corresponding model for y is

$$f(\theta^{-1}y)dm(y) .$$

With observed data y^0, the orbit of the error is observable

$$a = [e]^{-1}e = [y^0]^{-1}y^0 ,$$

where $[\cdot]$ is a transformation variable satisfying $[gy] = g[y]$ for g in G. The appropriate model given the observed function of the realized error is

$$cf(\theta^{-1}[y]a)d\mu([y]) \ ,$$

where μ is the left invariant measure on the group; the model describes $[y]|a = [y]^{-1}y$; for additional details, see Fraser (1979). We emphasis that the conditioning is a consequence of observing characteristics of the error, which in turn is a consequence of the transformation property in the model. Thus conditioning here is not the result of applying a principle of inference or desiring special structure such as similar regions, but is a consequence of how the data were generated; on the other hand, with a statistical model that presents response distributions rather than a single error distribution, a conditionality or ancillarity principal would typically be needed to obtain similar conditioning.

The density factorization with the transformation model has the form

$$f(y;\theta) \ = \ f(y_1|a,\theta)f(a) \ ,$$

into conditional and marginal components where $y_1 = [y]$ and $a = [y]^{-1}y$. A generalization of this inference factorization would have

$$f(y;\theta) \ = \ f(y_1|y_2;\theta)f(y_2)$$

with information concerning θ entirely in the conditional component. More generally, with a separation $\theta = (\psi, \lambda)$ into an interest parameter ψ and nuisance parameter λ, the factorization becomes

$$f(y;\theta) \ = \ f(y_1|y_2;\psi,\lambda)f(y_2;\lambda)$$

where the conditional component contains *all* the ψ information. Some extended inference methods based on this factorization may be found in Fraser and Reid (1988, 1989). In this transformation type conditioning the nuisance parameter appears in the conditional distribution for the interest parameter, but can often be removed approximately; see for example, Fraser and Wong (1991).

2.2 Exponential model patterns

Neyman and Pearson (1933) in their development of hypothesis testing proposed similar size α tests for composite hypotheses. For a model with parameter $\theta = (\psi, \lambda)$ and composite hypothesis $\psi = \psi_0$ a similar size α test with critical region C satisfies

$$P(C;\psi_0,\lambda) \ \equiv \ \alpha \ .$$

For the case that the minimal sufficient statistic say $s(y)$ for (ψ_0, λ) is complete, a similar size α test satisfies the conditional property

$$P(C|s;\psi_0) = \alpha$$

which is λ free by the sufficiency of s. As a consequence in this context, a similar size α test can be viewed as a conditional test given the minimal sufficient statistic for the nuisance parameter under the hypothesis. For more general statistical inference this pattern suggests that inference concerning $\psi = \psi_0$ be based on the conditional model given the minimal sufficient statistic for (ψ_0, λ) or more generally given the maximum likelihood estimate of (ψ_0, λ).

The two sample nonparametric problem provides a familiar illustration. Let $y_1, \ldots, y_n$ be a sample from $F(y)$ and $y_{n+1}, \ldots, y_{n+m}$ from $F(y-\theta)$ where F is unknown and the hypothesis is $\theta = 0$. The conditional approach to this is to assess the observed rank vector given the order statistic $(y_{(1)}, \ldots, y_{(n+m)})$ vector for the composite sample.

In the parametric context the general exponential model

$$\exp\{\psi y_1 + \lambda y_2 - \kappa(\psi, \lambda) + h(y)\}$$

provides a major class of examples. The conditional distribution of y_1 given y_2 provides the largest-range conditional distribution that is λ-free, based on the full minimal sufficient statistic (y_1, y_2). Building on this, general statistical inference would argue that inference for ψ should be based on the conditional model for $y_1|y_2$.

The density factorization with the exponential model has the form

$$f(y_1, y_2; \theta) = f(y_1|y_2; \psi) f(y_2; \psi, \lambda) .$$

For more general models with this type of factorization, current statistical inference would recommend the use of the conditional model for $y_1|y_2$. Note that the conditional model contains *only* information concerning ψ. Some extended inference methods based on this factorization may be found in Fraser and Reid (1988, 1989). In this exponential type conditioning the nuisance parameter is absent from the conditional distribution for the interest parameter.

2.3 Why condition?

Much of the development associated with conditional methods has been closely tied to questions and imperatives concerning whether one should or should not condition. However, a dominating objective was to make the hypothetical long-run frequencies for interpretations as relevant to or consisent with the *unique* particular data under analysis.

Fisher's (1933) conditioning in the location context was motivated as a means of recovering information lost with the use of the maximum likelihood estimate alone. ¿From this grew however a general imperative that one should condition on a variable that has a fixed distribution, is *ancillary*. This evolved as a conditionality principle or ancillarity principle, which had some strong supporters but many who were indifferent or distrusting.

Birnbaum (1962, 1972) examined the three major principles, sufficiency, likelihood and conditionality, from an equivalence class viewpoint in terms of evidential meaning (or information content). He showed that sufficiency and conditionality implies likelihood. This was paradoxical, for many statisticians supported the first two but rejected the third. The likelihood principle mentioned in this is the *strong* likelihood principle: that inference should be based on *only* the observed likelihood function and thus be *independent* of any model information otherwise.

Evans, Fraser, Monette (1985) used cross-embedded Bernoulli models to show that conditionality alone implies likelihood. The method of proof was simple enough to show clearly the mechanism by which the conditionality principle, condition on a variable with a fixed distribution, led to the likelihood principle. And it indicated that the conditionality principle as commonly formulated was unacceptable as an imperative in statistics. Of course, a reexamination of Fisher showed that he had much more than a fixed distribution in mind, in particular his choice of words, configuration statistic, complexity of the sample, suggested physical characteristics of the context to which the model might be applied. Some aspects of these additional characteristics may be found in the formulation of the structural model.

There does remain a strong concern with a yes-no attitude towards the conditionality principle, particularly in the context of the Neyman-Pearson hypothesis testing framework; see Brown (1990) and the accompanying discussion.

Recently a more pragmatic approach to conditional inference is developing, perhaps a little in the pattern of that for nonparametric tests in its earlier period. Interestingly, many of the nonparametric procedures have a conditional structure and can provide a guide for the more general development of conditional methods. The concluding sections in this paper will promote a very general, exploratory approach to conditional inference. The idea is to seek model components that provide high quality information concerning the parameter of interest allowing the loss of possibly contaminated information as a price of getting the high quality information. This is some opposite of the usual concern over loss of information from the conditioning variable. We view it as a *positive* approach.

One exploratory approach to conditioning is that of conical or directional tests (Fraser and Massam, 1985) to be surveyed in Section 3. This was extended to the score function in Fraser

(1987); the application of saddlepoint methods was developed in Skovgaard (1988). These tests examine a vector parameter, typically without nuisance parameters although Section 3 provides an extension.

Another approach is to construct one dimensional curves to provide inference for a real parameter, typically in the presence of nuisance parameters. The emphasis is on the kind of inference that can be obtained conditionally, in contrast to the traditional focus on whether there is a loss of information when the marginal model is omitted. This is discussed in Section 4.5 as part of a strong emphasis on the freedom to examine any conditional model of interest.

3. Directional tests of significance

3.1 Origins of conical or directional tests

In preceding sections we have indicated how the development of conditional methods has been much dominated by imperatives connected with the conditionality principle. A more pragmatic approach may be found in recent developments of conical or directional tests.

A test of significance starts with some measure of the departure of data from what is expected under a hypothesis and then records a probability of being as far or farther from what is expected based on that measure of departure.

Fraser and Massam (1985) work with variables on some given or chosen vector space and use vector space properties to measure departure from expectation. They obtain probabilities by conditioning on the direction of departure. The focus is thus directly on the conditional model ignoring whether any information may be available in the conditioning variable. A simple advantage is the ease of computation; theoretical support is also available.

Consider the simple case of a location model $f(y_1 - \theta_1, y_2 - \theta_2)$ in the plane and the hypothesis $\theta = \theta_0$. For ease of discussion we assume that f has been centered so that the maximum is at the origin. Note that the location-parameter structure implicitly defines a vector space for (y_1, y_2) and thus will give uniqueness to the following procedure.

One motivational pattern for conical tests is obtained by means of a reparameterization relative to the hypothesis value $\theta = (\theta_{10}, \theta_{20})$. Let

$$\psi = |\theta - \theta_0| , \qquad \lambda = (\theta - \theta_0)/\psi$$

be new parameters recording the vector distance ψ of θ from the null θ_0 and the vector direction λ of θ from θ_0. This reparameterization explicitly uses the vector space implied by the location model.

The maximum likelihood estimate of the original parameter is $(\hat{\theta}_1, \hat{\theta}_2) = (y_1, y_2)$ and correspondingly the maximum likelihood estimates of ψ and λ are

$$\hat{\psi} = |y - \theta_0| \qquad \hat{\lambda} = (y - \theta_0)/\hat{\psi} \, .$$

The conical test uses $\hat{\psi}$ as the measure of the departure of the data y from the value θ_0 expected under the hypothesis. And the observed level of significance is obtained by conditioning on the maximum likelihood estimate $\hat{\lambda}$ of the nuisance parameter. The direction of departure $\hat{\lambda}$ is viewed as an irrelevant and inevitable data characteristic and the magnitude of departure is accordingly assessed conditionally.

The conditional structure of the test makes the computation almost trivial; the observed level of significance for θ_0 is

$$p(\theta_0) = \int_1^\infty f\{s(y^0 - \theta_0)\} s\, ds \Big/ \int_0^\infty f\{s(y^0 - \theta_0)\} s\, ds ,$$

where the underlying change of variable is from rectangular to polar coordinates with s as a radial distance in units of length $|y^0 - \theta_0|$. In practice the one dimensional integral is numerically trivial and easily managed on a hand calculator having an integration routine.

3.2 Formulation of directional tests

Conical tests as proposed in Fraser and Massam (1985) and Fraser (1987) use a vector space for the initial variable or for some derived pivotal variable. For the first, let $t = y - \hat{y}(\theta_0)$ be the departure of the variable y from the value, say maximum density, expected under the hypothesis. For the pivotal case let $t = t(y, \phi_0)$ where ϕ is the component parameter of interest and $t(y, \phi)$ is some pivotal measure of departure of the variable from its value expected under the parameter value ϕ.

Suppose that t is r dimensional with null density $f(t)$. The observed level of significance taken as the probability of as great or greater departure from that expected, and calculated conditionally given the direction of departure is

$$p(\phi_0) = \int_1^\infty f(st^0) s^{r-1} ds \Big/ \int_0^\infty f(st^0) s^{r-1} ds ,$$

based on an initial change of variable to radial and spherical coordinates; in this, t^0 is the observed value of the departure or pivotal measure.

As a first example consider the regression model $y = X\beta + e$, where e is a sample from a known distribution with density f which could be say normal $(0, \sigma_0^2)$; let $\beta = \beta_0$ be a hypothesis concerning the full set of regression coefficients.

For familiarity, let $b(y)$ be the least-squares regression coefficient vector and $d(y)$ be the residual vector. Then the distribution of $b(y)$ given the orbit variable $d(y)$ (transformation theory, Section 2.1) is

$$cf\big(X(b - \beta) + d\big) db$$

in r dimensions. Let $\hat{b}$ be the value that maximizes $f(Xb+d)$ for given d; then $\hat{\beta}^0 = b(y^0)-\hat{b}$. The observed level of significance for testing β_0 is then

$$p(\beta_0) = \int_1^\infty f[X(sb^0+\hat{b})+d]s^{r-1}ds \Big/ \int_0^\infty f[\cdots]s^{r-1}ds$$

where $b^0 = b(y^0) - \beta_0 - \hat{b}$. For the normal $(0, \sigma_0^2)$ case this becomes

$$p(\beta_0) = \int_{\chi_0^2}^\infty h_r(\chi^2)d\chi^2$$

where h_r is the chi-square density with r degrees of freedom and

$$\chi_0^2 = \big(b(y^0) - \beta_0\big)' X'X \big(b(y^0) - \beta_0\big)/\sigma_0^2$$

is the ordinary chi-square value.

As a second example consider the regression model $y = X\beta + \sigma e$, where e is a sample from a known distribution; let $\beta = \beta_0$ be a hypothesis concerning the full set of regression coefficients. ¿From transformation model theory the distribution of $\{b(y), s(y)\}$ where $b(y)$ is the least squares estimate and $s(y)$ the residual length is

$$f(b,s) = cf\big(X(b-\beta)+sd\big)s^{n-r-1}$$

on $R^r \times R^+$. The marginal distribution of $t(y) = \{b(y)-\beta\}/s(y)$ requires one-dimensional integration

$$f(t) = \int_0^\infty cf\big\{(Xt+d)s\big\}s^{n-1}ds\ ;$$

let $\hat{t}$ be the corresponding maximum density point. Then the observed level of significance for testing β_0 is

$$p(\beta_0) = \int_1^\infty f\{u(t^0-\hat{t})+\hat{t}\}u^{r-1}du \Big/ \int_0^\infty f\{\cdots\}u^{r-1}du$$

where $t^0 = t(y^0)$ and we have used u to designate the radial variable as s is a common designation for scaling in regression analysis.

Finding the maximizing point $\hat{t}$ for the density $f(t)$ on R^r could be numerically demanding; a reasonable compromise would be to use $\hat{t} = \hat{b}/\hat{s}$ where $(\hat{b},\hat{s})$ maximizes $f(b,s)$. The computation of $p(\beta_0)$ then requires just two dimensional integration which is manageable with routine programming.

For the case of a standard normal distribution for the coordinates of the error e, the expression above simplifies immediately to

$$p(\beta_0) = \int_{F^0}^\infty h(F)dF$$

where h is the F-density with p over $n-p$ degrees of freedom and F^0 is the observed value of the ordinary F ratio.

As a third example consider a statistical model $f(x;\theta)$ where x and θ are r-dimensional. Let $\theta = \theta_0$ be the hypothesis of interest. If the space of x does not have appropriate vector space properties, the use of an exponential tangent model at θ_0 may provide the appropriate linearity. Let $s = (\partial/\partial\theta) \ln f(x;\theta)|_{\theta_0}$ be the score function and $\bar{f}(s;\theta)$ be the corresponding density. Then the observed level of significance for testing θ_0 with data s^0 is

$$p(\theta_0) = \int_1^\infty \bar{f}(us^0;\theta_0)u^{r-1}du \Big/ \int_0^\infty \bar{f}(us^0;\theta_0)u^{r-1}du \,.$$

For further details see Fraser (1987) and for the use of saddlepoint asymptotics see Skovgaard (1988) and Cheah, Fraser, and Reid (1991).

3.3 Multivariate analysis of variance

Directional tests can provide simple easily interpreted tests for multivariate analysis of variance (Guttman, Fraser, Srivastava, 1991). In this the observed level of significance records how close the data value for treatment, for treatment improvement, or for error is to the edge of the corresponding null distribution; it explicitly calculates the amount of probability beyond the data value, not in terms of some macro-measure but physically at the actual data value itself; recall the discussion for the simple case in Section 3.1. For this, treatment improvement refers to a parameter that is assumed to be nonnegative and there is an interest in whether the value of the parameter is greater than zero.

Let $Y = BX + E$ with $p \times n$ response Y error E, $p \times r$ coefficient matrix B, $r \times n$ design matrix X, and with the columns of E independently distributed as $N_p(0;\Sigma)$.

The multivariate analysis of variance can separate t degrees of freedom for some treatment effect and $m = n - r$ degrees of freedom for error. The corresponding canonical version of the model has the form $(Y_1, Y_2) = (M, 0) + (E_1, E_2)$ with $p \times t$ response Y_1 with mean M and with observable error $Y_2 = E_2$; again the error column vectors are $N_p(0,\Sigma)$.

A treatment hypothesis $M = 0$ can be assessed in terms of a Student matrix. Let D be a $p \times m$ matrix of orthonormal vectors forming a basis for the *row space* of Y_2, with the same orientation as the rows but otherwise independent of Y_2, and let $Y_2 = CD$. The Student matrix $H = C^{-1}Y_1$ has the null density (Fraser, 1979, p. 293)

$$c|I + HH'|^{-(m+t)/2} \,.$$

For testing a value M_0 (say $=0$) for treatment matrix, we obtain an observed

$$H^0 = C^{-1}(Y_2^0)(Y_1^0 - M_0)$$

which will typically differ from the central value 0 under the null hypothesis. The observed level of significance is

$$p(M) = \int_1^\infty \{|I + s^2 H^0 H^{0\prime}|^{-(m+t)/2} s^{pt-1}\} ds / \int_0^\infty \{\cdots\} ds$$

by the argument earlier.

An error hypothesis, $\Sigma = \Sigma_0$ can be tested by examining the roots $\ell_1, \ldots, \ell_p$ of the Wishart matrix $S_2 = Y_2 Y_2'$. A directional test can then be constructed by examining the log-departures $d_i = \log \ell_i - \log m$.

A treatment improvement hypothesis arises with $t = 1$ and mean *vector* $M = \mu$. The hypothesis then is $\mu = 0$ in the context having $\mu \geq 0$. We simplify notation and write $(Y, E) = (\mu + F, E)$ where F and E are samples of 1 and m from the $N_p(0; \Sigma)$ distribution. Perlman (1969) presents essential multivariate distribution theory and Tang, Gnecco and Geller (1989) present an alternative to a test proposed by Perlman.

The conditional approach examines the maximum likelihood estimate of μ given $\mu \geq 0$, conditions on negative coordinates, and calculates the conical level of significance for the remaining coordinates; for details see Fraser, Guttman, Srivastava (1991).

Tang, Gnecco and Geller (1989) consider treatment improvement for a shift from a five drug therapy to a two drug therapy. For three response measures, the normalized treatment difference (y_1, y_2, y_3) is $(2.21, 1.92, -1.56)$ and the corresponding correlation matrix is

$$(1.00,\ 0.50,\ -0.15;\ 0.50,\ 1.00,\ -0.15;\ -0.15,\ -0.15,\ 1.00\ )\ .$$

If the constraint that the coordinates of the mean are $\mu \geq 0$ is introduced , the new maximum likelihood estimate in the standardized units is $(1.976, 1.686, 0)$; the change in the first two coordinates is attributable to the correlation structure. As the error degrees of freedom is $m = 145$ the distribution theory reduces from multivariate Student to multivariate normal with variances assumed known. The corresponding observed level of significance is

$$p = \int_1^\infty f(s\,1.976, s\,1.686) s\, ds / \int_0^\infty f(s\,1.976, s\,1.686) s\, ds = 10.644\%$$

where $f(y)$ is the bivariate normal density with means equal to zero, variances equal to 1, and corelations given above for the first two coordinates. This shows that the data value is not close (10%) to the edge of the distribution in contrast to two scalar function analyses (loc.cit) that indicated 5%.

4. Extracting statistical frames

4.1 Statistical frames

A central concern of statistical inference is the assessment of data together with possibilities for the origins of the data. Frequently the possibilities are organized in the form of a parametric or semi-parametric statistical model representing all or part of an application. ¿From this viewpoint statistical inference is concerned with implications and conclusions that can be drawn from the starting point of the model-data combination. In our generalization we will be using extracted model components not directly justifiable by marginal or conditional criteria in the initial model; accordingly we will call such a model data combination a *statistical frame*. Other terms have been used in restricted contexts; *instance of statistical evidence*, in Birnbaum (1962); *inference base* in Fraser (1979).

For the discussion here we take a model to be

$$\mathcal{M} = \{f(y;\theta) : \theta \in \Omega\} ,$$

a class of density functions; more generally, additional ingredients can be present, as for example the error structure of structural models. Correspondingly, the data are given by an observed value

$$\mathcal{D} = y^0$$

of the model variable y. A statistical frame is then $(\mathcal{M}, \mathcal{D})$. A simple example might be gamma sampling

$$\begin{aligned} \mathcal{M} &: (y_1, \ldots, y_n) \sim \text{gamma}(\mu, \beta) \\ \mathcal{D} &: (y_1^0, \ldots, y_n^0) \end{aligned}$$

where μ is the mean and β the shape for the gamma. Another example is the regression model,

$$\begin{aligned} \mathcal{M} &: y \sim X\beta + \sigma e;\ e_i \sim \text{independent, Student}(7) \\ \mathcal{D} &: (y_1^0, \ldots, y_n^0) , \end{aligned}$$

with a particular choice of nonnormal error.

Our major interest is the extraction of statistical frames, for particular purposes of interest, from some initial statistical frame. The extraction may be to focus on a component parameter of interest, to computationally achieve inference where little or none was obviously available initially, to increase precision or accuracy from that ostensibly available initially, or to obtain reliable information not compromised by components confounded with nuisance parameters.

A partition $\mathcal{P}$ on the sample space *splits* an initial model $\mathcal{M}$ into a marginal model $\mathcal{M}_*$ and a conditional model $\mathcal{M}^*$, which of course has an input parameter from the variable of the marginal model. Let $\mathcal{D}_*$ and $\mathcal{D}^*$ designate the corresponding data arrays. Then the partition $\mathcal{P}$ splits the frame $(\mathcal{M}, \mathcal{D})$ into the marginal $(\mathcal{M}_*, \mathcal{D}_*)$ and the conditional $(\mathcal{M}^*, \mathcal{D}^*)$.

For example with the model $y_1, y_2 \sim$ uniform $(\psi \pm \delta/2)$, the partition generated by $y_2 - y_1$ produces

$$\mathcal{M}_* \ : \ d = y_2 - y_1 \ \sim \ \text{triangular}(-\delta, +\delta)$$

$$\mathcal{M}^* \ : \ \bar{y} \ \sim \ \text{uniform}\{\psi \pm (\delta - |d|)/2\} \ .$$

Location model theory would suggest the conditional model $\mathcal{M}^*$ for inference concerning ψ.

We will not examine them here but more complex splittings are possible based on pivotal structure. For example the partition $\mathcal{P}$ could depend on an interest parameter, as with certain nonparametric tests of location.

We will speak of *extracting* a statistical frame when an initial frame is replaced by the marginal $(\mathcal{M}_*, \mathcal{D}_*)$ or the conditional $(\mathcal{M}^*, \mathcal{D}^*)$. Either can be of interest but our main focus will be on the conditional which typically admits greater ease of computation and analysis. Our exploratory inference approach is to consider all extracted frames from some initial statistical frame. This admits both marginal and conditional extraction for arbitrary partitions $\mathcal{P}$. As indicated earlier the objective may be: (i) to eliminate nuisance parameters exactly or approximately, (ii) to facilitate computation and analysis, (iii) to get better or appropriate inference results for an interest parameter.

We suggest an *extraction principle*, that an extracted frame provides valid inferences without prejudice: (i) Clearly information is omitted from the complementary frame, but in some sense this is a price paid to achieve whatever inference is possible from the extracted frame. (ii) No effect or prejudice exists in the evaluation of the extracted frame itself, other than the deliberate omission of whatever is in the complementary frame. (iii) The choice of $\mathcal{P}$ should not depend on the data, although for computation only certain aspects relevant to the data may need to be considered. The second may seem to be covered by the first, but is included as part of personal deep respect for the subtleties and complexities of conditional probability.

Why deliberately omit the complementary frame? By itself it may not produce useful information easily: this is pragmatic. It may not contain information of interest: this is theoretical and indeed may be difficult to measure. It may be computationally difficult: this is pragmatic. Thus choose an extracted frame to achieve useful, convenient, appropriate, accurate inferences.

The choice of a frame would be based on the inferences possible within the model component of the frame and of cause would not be influenced by the data value in the frame. This would of course be comparable to not choosing a test statistic based on the p value it gives.

4.2 One dimensional conditional models, an example of extraction

The directional tests discussed in Section 3 provide a central example of statistical extraction: directly focussing on how far data is from expectation, and ignoring the conditioning variable as having inaccessible or confounder information. In this section we survey briefly the construction of one dimensional curves for inference concerning a real parameter ψ in the presence of nuisance parameters λ; for details see Fraser and Reid (1988).

The one dimensional inference curve, say C, through the data point is computed iteratively, progressing in both directions from the observed data point y^0: $\cdots, y^{-2}, y^{-1}, y^0, y^1, y^2, \cdots$. The likelihood function $\ell(\theta; y) = \ln f(y;\theta)$ is assumed available and it describes key model characteristics at each examined point, in particular at a point y^i the likelihood $\ell(\theta; y^i)$ is examined primarily for θ values near the corresponding maximum likelihood value $\hat{\theta}(y^i)$. The iterative procedure produces a new point $y^{i+1} = y^i + \delta v(y^i)$ by going in a vector direction $v(y^i)$ a distance δ from the old point; typically δ would be small, as in the iterative solution of differential equations. The curve through the observed data point is determined by the unit vector function $v(y)$. The vector function $v(y)$ not only determines succeeding points, but also determines the Jacobian effect $J(y)$ along the sample space curves:

$$J(y^{i+1}) \;=\; J(y^i)[1 + \delta\,\mathrm{div}\{v(y^i)\}]$$

where $\mathrm{div}\, v(y) = \Sigma \partial v_j(y)/\partial y_j$ is the divergence function of analysis. Normed likelihood methods (Fraser, 1988; Fraser & Reid, 1993a) provide an alternative to the direct calculation of the divergence function.

The conditional model then at points distance $s = i\delta$ along the curve is

$$f(s;\theta)ds \;=\; c(\theta) f(y^i;\theta) J(y^i) ds\ .$$

For any θ value of interest the relative density $f(y^i;\theta)J(y^i)$ can be calculated. Probability left of the data point provides an observed significance $p(\theta)$ and is available as a partial sum divided by a complete sum (numerical integration with norming).

The extraction to be accomplished or the partition to be chosen is determined by the vector direction function $v(y)$. First order elimination of nuisance parameters is obtained by having $v(y)$ satisfy equations based on the nuisance parameter scores; higher order inference properties concerning the interest parameter are obtained by having $v(y)$ satisfy equations based on higher order scores with respect to the interest parameters. For some details see Fraser and Reid (1988, 1989, 1990).

As a simple example consider a sample of n pairs (x_{1i}, x_{2i}) where x_{1i} is exponential λ and x_{2i} is exponential with rate $\psi\lambda$; the minimal sufficient statistic is $(y_1, y_2) = (\Sigma x_1, \Sigma x_2)$. For the interest parameter ψ marginal analysis is indicated by location model theory: $y_2/y_1\psi$ has an F-distribution with $2n$ over $2n$ degrees of freedom. To illustrate the present theory we consider the construction of a curve for conditional inference in the (y_1, y_2) space. Reasonably straightforward computation leads to the distribution of y_2/y_1 conditional on $y_1 y_2$. This conditional result can be shown to approximate closely the marginal analysis described above. For details see Fraser and Reid (1988).

If a model is a location model marginal analysis is indicated for a component parameter; if it is exponential a conditional analysis is indicated. For general models a computationally feasible approach is that discussed in this section; it borrows on the patterns of the exponential model analysis and gives the appropriate exact result for such models, and gives a feasible approximation more generally.

4.3 Approximate levels of significance

In this Section 4 we have been promoting the extraction of quite arbitrary conditional models with the objective of finding a satisfactory model for the type of inference desired. Our examples so far, the directional conditional model in Section 3 and the iteratively defined conditional model in Section 4.2, have one dimensional conditional distributions. In this section we describe briefly some asymptotic results that can give quite accurate significance levels $p(\psi)$ starting initially with only a likelihood function $\ell(\psi)$ and a reparameterization $\phi(\psi)$ where ψ is real. At the end of the section we give some indication how a likelihood $\ell(\psi)$ for a component parameter can be developed or approximated starting with a likelihood $\ell(\psi,\lambda)$ for the full parameter $\theta(\psi,\lambda)$.

The origins of the approximation lie with the saddlepoint methods of applied mathematics and introduced to statistics by Daniels (1954, 1987), Barndorff-Nielsen and Cox (1979). These methods focussed on a third-order accurate conversion of a cumulant generating function to the corresponding density. For exponential models this becomes a conversion from likelihood function to corresponding density.

For many statistical problems a distribution function is of more interest than a density function. Lugannani and Rice (1980) produced a third order formula for going from cumulant generating function to distribution function, or from likelihood to distribution function with exponential models. Further developments from this provide major approximations for statistical inference.

For the exponential model

$$f(y;\psi) \;=\; \exp\{\psi t - \kappa(\psi) + h(y)\}$$

the left tail probability for the sufficient statistic $t(y)$ can be approximated,

$$p(\psi) \;=\; P(t \le t^0) \;=\; \Phi(r) + \phi(r)\left\{\frac{1}{r} - \frac{1}{q_1}\right\} + O(n^{-\frac{3}{2}}) \;,$$

from an observed likelihood $\ell(\psi; y^0)$, where ϕ and Φ are the standard normal density and distribution functions and

$$q_1 = (\hat{\psi}^0 - \psi)\hat{\jmath}^{\frac{1}{2}} ,$$
$$r = \mathrm{sgn}(\hat{\psi}^0 - \psi) \cdot [2\{\ell(\hat{\psi}^0) - \ell(\psi)\}]^{\frac{1}{2}} ,$$

are respectively the standardized maximum likelihood statistic and signed square root of the likelihood ratio statistic, $\hat{\psi}^0 = \hat{\psi}(y^0)$, and $\hat{\jmath} = -\partial^2\ell(\psi)/\partial\psi^2|_{\hat{\psi}}$ is the observed information. The order $O(n^{-\frac{3}{2}})$ corresponds to the sampling case with $t = \Sigma y_i$ being the minimal sufficient statistic. This Lugannani and Rice (ibid) formula is expressed here in terms of likelihood for an exponential model but can be given equivalently in terms of the conversion from cumulant generating function to distribution function.

The standardized maximum likelihood estimate in the formulas above require ψ to be the canonical parameter of the model. The formula can be made parameterization invariant (Fraser, 1988) for an exponential model $f(y;\psi)$ by using the constructed parameter

$$\phi = \frac{d}{dy} \ln f(y;\psi)|_{y^0} = \frac{d}{dy}\ell(\psi;y)|_{y^0} = \dot{\ell}(\psi;y^0)$$

as the parameter in the preceding formula. In effect this extracts the canonical parameter from the likelihood expressed in terms of some other parameter; the choice of directional derivative has only a scale effect with no final effect on the $p(\psi)$ approximation. Having a parameter depond on an observed data point seems very contrived indeed. In fact, it is purely a computation device so that exponential model theory can be used more generally: the approximating exponential model at a data point provides tail probabilities at that point which turn out to be third order accurate (Cakmak, Cheah, Fraser, Reid, Tapia, 1993; Fraser & Reid, 1993).

For a general model $f(y;\psi)$ satisfying simple asymptotic properties with y real, the parameterization invariant version (Fraser and Reid, 1990) remains accurate to $O(n^{-\frac{3}{2}})$. In effect an observed likelihood function $\ell(\psi)$ and a local reparameterization $\phi = \phi(\psi)$ produce an $O(n^{-\frac{3}{2}})$ approximation to the significance function $p(\psi)$ describing left tail probability at the observed data.

Extensions of these results to the nuisance parameter context may be found in DiCiccio, Field, Fraser (1990), Fraser, Lee, Reid (1990) and Fraser and Reid (1990).

For comparison, the common first order methods of going from observed likelihood $\ell(y)$ to left tail probability $p(\psi) = P(\hat{\psi} \le \hat{\psi}^0)$ are based on the standardized maximum likelihood estimate q_1 , the standardized score function

$$q_2 = S(\psi)\hat{\jmath}^{-\frac{1}{2}}$$

and the signed square root of the likelihood ratio r ; the approximations are obtained by inserting q_1, q_2 , and r into the standard normal distribution function $\Phi(\cdot)$; thus $p(\psi) \approx \Phi(q_1), \Phi(q_2), \Phi(r)$. These approximations are accurate to $O(n^{-\frac{1}{2}})$.

As an illustration of accuracy we consider an example from Fisher (1956, p. 72) where an observed $y^0 = 3$ is examined for the binomial $(14, p)$: 14 trials and 3 successes with a probability p. The four approximations, (mle, score, likelihood ratio, and third order) to the left tail probability $Pr(\hat{p} \leq \hat{p}^0) = P(y \leq y^0; p)$ are plotted against $\log\{p/(1-p)\}$ in Fig. 1. The exact value was taken to be the mid-p value calculated from the incomplete beta; it exactly equals the third order approximate to the accuracy shown in the diagram.

The determination of likelihood for component parameters is discussed in Cox and Reid (1987) and Fraser and Reid (1989). The application of these to obtain a significance function is discussed in Fraser and Reid (1990).

5. Compounding frames, and the proportional hazards model

5.1 Compounding statistical frames

In Section 4.1 we discussed the splitting of a statistical frame into a marginal and conditional frame and the extraction of one of these for special purposes of statistical inference. We argued that an extracted frame could be used on its own merits, unprejudiced by the extraction or the omission of the complementary frame. In this Section 5 we will primarily be concerned with extracting conditional statistical frames.

Suppose the statistical frames

$$(\mathcal{M}_1, \mathcal{D}_1), \ (\mathcal{M}_2, \mathcal{D}_2), \ \ldots$$

have been successively extracted in each case with an omitted complementary frame. ¿From the origins of the frames we argue that the respective probability models are valid and by a compounding principle we put the valid probability structures together to obtain a compound statistical frame. In most interesting cases the compounded model will not be available by marginalization or conditioning from the initial model. This theory leads to a theoretical basis for meta-analysis.

5.2 Proportional hazards model

The proportional hazards model (Cox, 1972, 1975) is widely used in biological, medical and reliability sciences.

As a simple illustration of this suppose that three units $i = 1, 2, 3$ are under test and that the instantaneous probability of failure in $(t, t + dt)$ given survival to time t is proportional to $\lambda(t)\exp\{X_i\beta\}$ where $\lambda(t)$ is a base hazard rate, X_i is a row vector of r characteristics of unit i, and β is a column vector of differential coefficients such that $X_i\beta$ is the logarithmic adjustment to the base rate to specialize it to individual i. We allow that individuals can withdraw or be withdrawn at random or by external forces.

Consider the following data pattern: the three units have uneventful exposure E_1 for a period of time; unit 1 withdraws W; units 2, 3 have uneventful exposure E_2 for a period of time; unit 3 fails F; unit 2 has exposure E_3 for the remaining time in the investigation. We describe this in terms of a sequence of marginal and conditional components. The only component clearly focussed on the failure rate comparisons is the failure F with Bernoulli model and data,

Unit	Relative probability	Data
2	$\exp\{X_2\beta\}$	0
3	$\exp\{X_3\beta\}$	1

The indicator 1 is used for failure.

We propose for this case the direct suppression of all model components other than the Bernoulli where two were on test and one failed; the data corresponds to an indicator that the third unit failed.

In a more general context we extract the multivariate Bernoulli frames that record which unit failed among a group of units under exposure at the failure time. By doing this we are directly ignoring or suppressing the connecting models describing continuing exposure or withdrawal. We argue that the probabilities in each Bernoulli model are valid probabilities describing an event at the corresponding time. It is of no avail that we might have been in some other situation indicated by linking models that are omitted; we do have probabilities describing events at particular times. For inference we directly compound the models to get the overall frame for inference. The compounded model is not in general a conditional model from the original model (Kalbfleisch & Prentice, 1973), although a conditional viewpoint seems to have directly influenced the construction of Cox's (1972) model.

The familiar partial likelihood can then be analyzed as an ordinary likelihood for a compound Bernoulli or exponential model.

5.3 An example

Kalbfleisch and Prentice (1980, p.1) examine data from Pike (1966) involving 40 rats, under one or other of two pretreatment regimes, with time from carcinogenic insult to death from cancer recorded; four of the rats were censored with only time to withdrawal recorded. The proportional hazard for cancer was taken to be $\lambda(t)\exp\{\beta x\}$ where x is an indicator for preregime 2 as opposed to 1; β is the log hazard effect of a shift from regime 1 to 2.

The first order likelihood analysis (ibid, p. 81f) records the data ordered in time. We record here the first portion of the table with changes in notation for the present analysis

Failure time t	Pretreatment indicator x	Bernoulli for individual failure $L(\beta)$
142	1	$e^{\beta}/(19+21e^{\beta})$
143	0	$1/(19+20e^{\beta})$
146	1	$e^{\beta}(18+20e^{\beta})$
163	1	$e^{\beta}(18+19e^{\beta})$
...		
188	0,0	$1/(17+18e^{\beta})(16+18e^{\beta})$
...		

The data contains withdrawals that are ignored. They also contain some ties due to time grouping. In each case the ties have the same pretreatment indicator so that a tie breaking approximation is not needed; for example at $t = 188$.

The compounded Bernoulli model is exponential and β is the canonical parameter. Thus the third-order accurate significance function can be calculated directly from the Lugannani and Rice formula in Section 4.3 giving

$$p(\beta) \;=\; P(\hat{\beta} \leq \hat{\beta}^0; \beta)$$

as a function of β . The right end of this significance function is plotted in Figure 2 together with the first order approximations, the maximum likelihood $\Phi(q_1)$ and the likelihood ratio $\Phi(r)$. These third order probabilities are not availablve without an extension such as that here to statistical frames.

We note that the likelihood ratio approximation $\Phi(r)$ is fairly close to the third order $p(\beta)$. This conforms to patterns in many other examples.

The 95% confidence interval for β by the three procedures are

First order maximum likelihood	$(-1.306, 0.141)$
First order likelihood ratio	$(-1.254, 0.121)$
Third order	$(-1.253, 0.117)$

with maximum likelihood $\hat{\beta}^0 = -0.57$. The value $\hat{\beta} = -0.60$ in Kalbfleisch and Prentice corresponds to the use of an approximate rather than exact calculation of the likelihood.

6. Concluding remarks

In Section 2 we examined two standard patterns of conditional inference and briefly discussed grounds that have been given for conditioning. At this time the exponential pattern seems the most promising for new developments.

The theme in succeeding sections is that conditioning should be investigated very freely allowing arbitrary choice of conditioning to obtain component conditional models that address the inference issues of interest. On the other hand, the notion that sufficiently simple problems have clear unequivocal solutions is important; we feel, however, that this is consistent with the objective of the theme. In Section 3 the directional tests allow simple easily implemented tests for a simple hypothesis with many parameters. In Section 4 one dimensional inference for a real parameter of interest is presented as a direct method of computer analysis of a model with data. Section 5 promotes the arbitrary extraction and recombination of model components; typically the combined model is not a conditional model in the context of standard theory. As a result we can demonstrate partial likelihood as an ordinary likelihood for the proportional hazards model.

7. Postscript

Several readers from different viewpoints have commented on material in this paper.

The paper recommends a very free and general extraction of component models using marginal and conditional methods, and the subsequent direct recomposition of selected components. The choice of components would be based on the quality of the inference procedure, not of course on the inference itself. Some criteria for choice are discussed briefly but this is a quite general issue of statistical design. The present focus is on the free examination of conditional components, some opposite of the long time concern for uniqueness and properties of ancillary variables.

One reader was concerned about limitations for directional tests, specifically the limited ability to handle nuisance parameters. One objective of the general approach to component model extraction is precisely to escape such limitations, to have free options to seek good inference that is approximately or exactly free of nuisance parameters.

Another reader felt that the pragmatic approach to conditioning would be welcomed by those who believe in conditioning but some discussion of uniqueness of conditioning statistics would be helpful. The proposed approach is some opposite of the concern for uniqueness; it does seem that the development of conditional methods has been seriously delayed by the long standing preoccupation with uniqueness.

The reader also wondered whether the Lugannani and Rice formula applied to the compounded Bernoulli's gave the usual third order accuracy. The probabilities are valid third order approximations: they are in the context of repetitions of the component frames or models, not repetitions of the total model; for inference, we feel this is the appropriate reference. A similar issue arises in meta-analysis.

Acknowledgments

A conference, Data Analysis and Statistical Foundations, was held at the University of Toronto, May 31 – June 1, 1990; the organizing committee was N. Reid with D.F. Andrews, D. Brenner, M. Evans, J. Gillies. This paper was presented on June 1 using a Mac IIci computer and NEC graphics projector. Special appreciation and thanks go to N. Reid and the organizing committee, to Augustine Wong, and to the University of Toronto Media Centre.

The author would like to express his very deep appreciation to one referee who provided very helpful and insightful comments and suggestions.

References

Barndorff-Nielsen, O.E., Cox, D.R. (1979). Edgeworth and saddlepoint approximations, *J. Royal Statist. Soc. B* **41,** 279–312.

Birnbaum, A. (1962). On the foundations of statistical inference, *J. Amer. Statist. Assoc.* **57,** 269–332.

Birnbaum, A. (1972). More on concepts of statistical evidence, *J. Amer. Statist. Assoc.* **67,** 858–886.

Brown, L.J. (1990). An ancillarity paradox which appears in multiple linear regression, *Ann. Math. Statist.,* **18,** 471–538.

Cakmak, S., Cheah, P.K., Fraser, D.A.S., Reid, N., Tapia. (1993). Third order asymptotic model: Exponential types approximation, in revision.

Cheah, P.K., Fraser, D.A.S., Reid, N. (1991). Multiparameter testing in exponential models: third order approximations from likelihood, in revision for *Biometrika*.

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

Cox, D.R. (1975). Partial likelihood, *Biometrika* **62,** 269–276.

Cox, D.R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference, *J.R. Statist. Soc. B* **49,** 1–39.

Daniels, H.E. (1954). Saddlepoint approximations in statistics, *Ann. Math. Statist.* **25,** 631–650.

Daniels, H.E. (1987). Tail probability approximation, *Int. Statist. Rev.* **55,** 37–48

DiCiccio, T.J., Field, C.A., Fraser, D.A.S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters, *Biometrika* **77,** 77–96.

Evans, M.J., Fraser, D.A.S., Monette, G. (1985). Mixtures, embedding and ancillarity, *Canadian J. Statist.* **13,** 1–6.

Evans, M.J., Fraser, D.A.S., Monette, G. (1986). On principles and arguments to likelihood, *Canadian J. Statist.* **14,** 181–199.

Fisher, R.A. (1934). Two new properties of mathematical likelihood, *Proc. Royal Soc.* **A144,** 385–407.

Fisher, R.A. (1956). *Statistical Methods and Scientific Inference,* London: Oliver and Boyd.

Fraser, D.A.S. (1957). A regression analysis using the invariance method, *Ann. Math. Statist.* **28,** 517–520.

Fraser, D.A.S. (1961a). On fiducial inference, *Ann. Math. Statist.* **32,** 661–676.

Fraser, D.A.S. (1961b). On fiducial methods and invariance, *Biometrika* **48,** 261–280.

Fraser, D.A.S. (1966). Structural probability and a generalization, *Biometrika* **53,** 1–9.

Fraser, D.A.S. (1967). Data transformations and the linear model, *Ann. Math. Statist.* **38,** 1456–1465.

Fraser, D.A.S. (1968). *The Structure of Inference,* New York: Wiley.

Fraser, D.A.S. (1979). *Inference and Linear Models,* New York: McGraw Hill.

Fraser, D.A.S. (1987). Fibre analysis and tangent models, *Statistical Papers* **28,** 163–181.

Fraser, D.A.S. (1988). Normed likelihood as saddlepoint approximation, *J. Mult. Anal.* **27,** 181–193.

Fraser, D.A.S. (1990). Tail probabilities from observed likelihoods, *Biometrika* **77,** 65–76.

Fraser, D.A.S., Guttman, I., Srivastava, M. (1991). Conditional inference for treatment and error in multivariate analysis, *Biometrika* **78,** 565–72.

Fraser, D.A.S., Massam, H. (1985). Conical tests: observed levels of significance and confidence regions, *Statistical Papers* **26,** 1–18.

Fraser, D.A.S, Lee, H.S., Reid, N. (1990). Nonnormal linear regression; an example of significance levels in high dimensions, *Biometrika* **77,** to appear.

Fraser, D.A.S., Reid, N. (1988). Fibre analysis and conditional inference, Proc. Second Pacific Area Statistical Conference, *Statistical Theory and Data Analysis II* (ed: K. Matusita) North Holland, 241–248.

Fraser, D.A.S., Reid, N. (1988). On conditional inference for a real parameter: a differential approach on the sample space, *Biometrika* **75,** 251–264.

Fraser, D.A.S., Reid, N. (1989). Adjusting profile likelihood, *Biometrika* **76,** 477–488.

Fraser, D.A.S., Reid, N. (1990). ¿From multiparameter likelihood to tail probability for a scalar parameter, Technical Report: University of Toronto. Revised in Fraser & Reid (1993)

Fraser, D.A.S., Reid, N. (1993). Simple asymptotic connections between density and cumulant functions leading to accurate approximations for distribution functions, *Statistica Sinica*, to appear.

Fraser, D.A.S., Wong, A.(1991). Approximate Studentization for marginal and conditional inference. Technical Report, University of Toronto.

Kalbfleisch, J.D., Prentice, R.L. (1973). Marginal likelihoods based on Cox's regression on life model, *Biometrika* **60,** 267–278.

Kalbfleisch, J.D., Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data,* New York: Wiley.

Kass, R.E. (1989). The geometry of asymptotic inference. *Statistical Science* **4,** 188–234.

Lugannani, R., Rice, S. (1980). Saddlepoint approximation for the distribution function of the sum of independent variables, *Adv. Appl. Prob.* **12,** 475–490.

Neyman, J., Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses, *Phil. Trans. Royal Soc.* **A231,** 289–337.

Peisakoff, M. (1951). Tranformation parameters, Ph.D. Thesis, Princeton University.

Perlman, M.D. (1969). One sided testing problems in multivariate analysis, *Ann. Math. Statist.* **40,** 549–567.

Pike, M.C. (1966). A method of analysis of a certain class of experiments in carcinogenics, *Biometrika* **22,** 142–161.

Skovgaard, I. (1988). Saddlepoint expansions for directional test probabilities, *J.R. Statist. Soc. B* **50,** 269–280.

Tang, D.I., Gnecco, C., Geller, N.L. (1989). Likelihood ratio tests with nonnegative components, *Biometrika* **76,** 577–583.

Verhagen, A.M.W. (1961). The estimation of regression and error-scale parameters when the joint distribution of the errors is any continuous form and known apart from a scale parameter, *Biometrika* **48,** 125–132.

D.A.S. Fraser
Department of Mathematic and Statistics
York University
Toronto, Canada, M3J 1P3