THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY
OF NORTH CAROLINA

PARAMETRIC STATISTICAL INFERENCE

by

O. Barndorff-Nielsen
Department of Theoretical Statistics
University of Aarhus

Institute of Statistics Mimeograph Series #1513

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

PARAMETRIC STATISTICAL INFERENCE

O. Barndorff-Nielsen
Department of Theoretical Statistics
University of Aarhus

0. Introduction

In these lectures, the two main classes of statistical models discussed are transformation models and exponential models. It is known that very little overlap exists between these two classes in the one-dimensional case; however, as the dimension increases, so does the interplay between the two groups. Exponential transformation models, i.e., models which are both transformational and exponential, will be briefly considered. The likelihood approach will be discussed for the inferential aspects of the models.

The material to be presented is classified under the following headings.

1. Exponential models
2. Transformation models
3. Hyperbolic distributions
4. Exponential transformation models
5. A formula for the distribution of the maximum likelihood estimator
6. Modified profile likelihood

Let X be a sample space and P be a class of probability measures P_ω, $\omega\epsilon\Omega$, dominated by a σ-finite measure μ, i.e., $P = \{P_\omega: \omega\epsilon\Omega\} << \mu$. We call the corresponding Radon-Nikodym derivative $\frac{dP_\omega}{d\mu}(x) = p(x;\omega)$ a *model function* and we write $p(x;\omega)$ $<\mu>$ to emphasize the dominating measure μ.

1. Exponential models

Suppose a model function $p(x;\omega)$ can be expressed as follows:

$$p(x;\omega) = a(\omega)b(x)e^{\underset{\sim}{\theta}'(\omega)\cdot\underset{\sim}{t}(x)} \quad , \tag{1}$$

where $\underset{\sim}{\theta}$, $\underset{\sim}{t}$ are $k\times 1$ vectors. We call k the *order* of the exponential family $\mathcal{P}$, provided that k is the minimal dimension for which an exponential representation of the form (1) exists for $p(x;\omega)$.

Example. Inverse Gaussian Model, $N^{-}(\chi,\psi)$

For a single observation x the p.d.f. is given by $p(x;\chi,\psi) = \frac{\sqrt{\chi}}{\sqrt{2\pi}}x^{-3/2} e^{\sqrt{\chi\psi}}\, e^{-1/2(\frac{\chi}{x}+\psi x)}$, $x>0$, $\chi>0$, $\psi\geq 0$. This is an exponential model of order 2. If $y(t)$ is a Brownian motion(μ,σ^2) with $\mu\geq 0$ and if x is the first passage time to a level $c>0$, then x follows the inverse Gaussian distribution with $\chi = \frac{c^2}{\sigma^2}$, $\psi = \frac{\mu^2}{\sigma^2}$.

Let $x_1,\ldots,x_n$ be a random sample from $N^{-}(\chi,\psi)$ and $t = (\bar{u},\bar{x})$, where $\bar{u} = n^{-1}\sum_{i=1}^{n}x_i^{-1}$, and $\bar{x} = n^{-1}\sum_{i=1}^{n}x_i$. Then

$$n(\bar{u}-\bar{x}^{-1}) \sim \chi^{-1}\cdot\chi^2(n-1),$$

$$\bar{x} \sim N^{-}(n\chi,n\psi),$$

and $(\bar{u}-\bar{x}^{-1}) \perp \bar{x}$, where $\perp$ is used to denote statistical independence. Hence we can perform an analog of the ordinary t-test for the mean μ. Also hierarchical analysis of variance can be performed (Tweedie [1957]). □

2. Transformation models

We introduce some notation before discussing transformation models. Let f be a map such that $f: \mathcal{X}\to\mathcal{Y}$. Let ν be a measure defined on $\mathcal{X}$ and let $f\nu$ be the lifted (by f) measure on $\mathcal{Y}$. Let G be a group. Then we say that G *acts* on $\mathcal{X}$ if there is given a mapping γ, defined on G, such that $\forall\, g \in G$, $\gamma(g)$ is a one-to-one transformation of $\mathcal{X}$ onto itself and such that γ is a homomorphism, i.e., $\gamma(g'g) = \gamma(g') \circ \gamma(g)$ for all g and g' in G, where $\circ$ denotes composition of mappings. We say that γ is an *action*, and for simplicity we write gx for $\gamma(g)x$. Let the lifted probability measure $\gamma(g)p$ be denoted by gp.

Then a *transformation model* is defined to be a family $\mathcal{P}$ of probability measures on $\mathcal{X}$ such that $\mathcal{P} = \{gp: g \in G\}$. We shall furthermore assume that there exists a dominating measure μ (w.r.t. $p \in \mathcal{P}$), which is invariant under the action of G, i.e., $g\mu = \mu$ for $\forall\, g \in G$. Then, writing $p(x;g)$ for $\frac{dgp}{d\mu}$, we have

$$(2) \qquad p(x;g) = p(g^{-1}x) \; \langle\mu\rangle \; .$$

<u>Example</u>. <u>Location-scale models</u>

Let f be a known one-dimensional p.d.f. In this model a random sample $x_1, \dots, x_n$ has a joint p.d.f. $\Pi_{i=1}^n \frac{1}{\sigma} f(\frac{x_i - \xi}{\sigma})$. Define $G = \{[\xi,\sigma];\, \xi \in \mathbb{R}, \sigma > 0\}$ with an operation $[\xi,\sigma][\xi',\sigma'] = [\xi + \sigma\xi', \sigma\sigma']$. Then G acts on $\mathbb{R}^n$ if we define γ as $[\xi,\sigma](x_1,\dots,x_n) = (\xi + \sigma x_1, \dots, \xi + \sigma x_n)$. □

Very often the models we are dealing with are partly, but not fully transformational. We speak of the model as a *composite transformation model* if it is the disjoint union of transformation models all of which are transformational relative to one and the same action γ. That is, $\mathcal{P} = \cup\, \mathcal{P}_\kappa$ where $\mathcal{P}_\kappa$ is a transformation model relative to γ and indexed by some additional parameter κ which we shall refer to as the *index parameter*.

<u>Example</u>. <u>Inverse Gaussian distribution $N^-(\chi,\psi)$</u>

This is not a transformation model. But if we introduce $\kappa = \sqrt{\chi\psi}$, $\sigma = \sqrt{\chi/\psi}$, then it becomes a composite transformation model. □

Example. von Mises-Fisher distributions

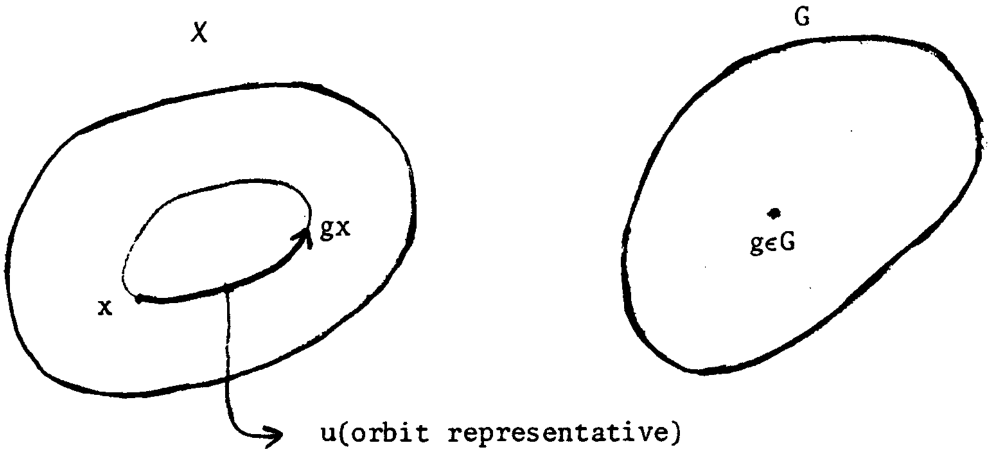
Let $X = S^{k-1}$, the unit sphere in $\mathbb{R}^k$ and let $P = \{p(x;\kappa,\xi): (\kappa,\xi) \in [0,\infty) \times S^{k-1}\}$ be the family of von Mises-Fisher distributions on S^{k-1} given by

$$p(x;\kappa,\xi) = a(\kappa)\, e^{\kappa\xi\cdot x},\ \xi \in S^{k-1},\ \kappa \in [0,\infty).$$

The norming constant $a(\kappa)$ depends on κ only and it may be expressed in terms of the modified Bessel function of the first kind and of order $\frac{k}{2}-1$.

Here P is a composite transformation model, with index parameter κ the acting group being the special orthogonal group $G = SO(k) = \{U: U^*U=I, |U|=1\}$ where U denotes a $k\times k$ matrix and where U^* is the transpose of U. □

For any $x \in X$ the set $Gx = \{gx: g\in G\}$ of points traversed by x under the action G is termed the *orbit* of x. The sample space X is partitioned into disjoint orbits.



If on each orbit we select a point u, to be called the *orbit representative,* then any point $x \in X$ can be determined by specifying the representative u of Gx and an element $z \in G$ such that $x = zu$. We speak of (z,u) as an *orbital decomposition* of x. The orbit representative u is a maximal invariant and hence ancillary statistic, and the inference proceeds by first conditioning on that statistic.

Now we present a key theorem for transformation models. Consider a statistic, say s, which is *equivariant*, i.e., $s(x) = s(x')$ implies $s(gx) = s(gx')$ for all $g \in G$. One can now see that if s is equivariant, then G acts on s by the prescription $gs=s(gx)$ when $s=s(x)$. The action of G on a space S is said to be *transitive* if S consists of a single orbit.

<u>Theorem</u> Let u be an invariant statistic with range space $\mathcal{U} = u(\mathcal{X})$, let s be an equivariant statistic with range space $\mathcal{S} = s(\mathcal{X})$, and assume that G acts transitively on $\mathcal{S}$. Furthermore, let μ be an invariant measure on $\mathcal{X}$. Then we have $(s,u)(\mathcal{X}) = \mathcal{S}\times\mathcal{U}$ and

(3) $$(s,u)\mu = \nu\times\rho,$$

where ν is an invariant measure on $\mathcal{S}$ and ρ is some measure on $\mathcal{U}$.

Suppose in addition that the transformation model has a model function $p(x;g) = p(g^{-1}x)$ relative to the invariant measure μ on $\mathcal{X}$ such that

(4) $$p(x) = q(u)r(s,w)$$

for some functions q and r and some invariant statistic w.

Then we have the following conclusions:

(i) The model function $p(x;g)$ is of the form

$$p(x;g) = q(u)r(g^{-1}s;w),$$

and hence the statistic (s,w) is sufficient.

(ii) s and u are conditionally independent given w.

(iii) The invariant statistic u has probability function

$$p(u) = q(u) \int r(s,w)d\nu(s) \quad \langle\rho\rangle$$

(iv) The conditional probability function of s given w is

$$p(s;g|w) = c(w)r(g^{-1}s,w) \quad \langle\nu\rangle,$$

where c(w) is a norming constant. □

For the requisite (extremely mild) regularity conditions and the proof, see Barndorff-Nielsen, Blaesild, Jensen and Jørgensen (1982).

By means of this theorem we can solve the distribution problems of transformation models. (The power of the theorem may be well illustrated by applying the theorem to the problem of multivariate two-way analysis of varaince, cf. Barndorff-Nielsen [1982b].)

3. Hyperbolic distribution

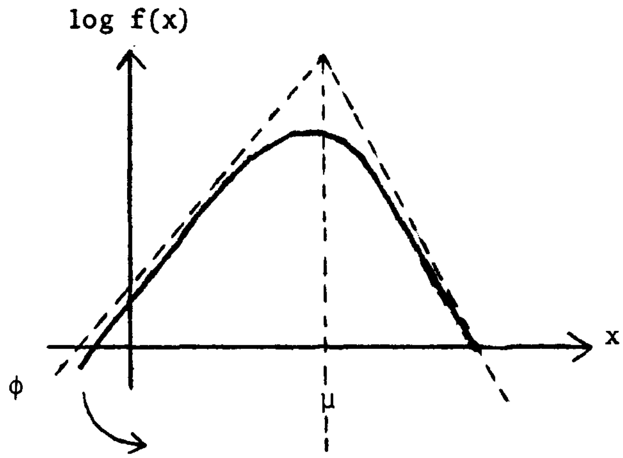
In the contexts of transformation models and robustness studies the hyperbolic distributions are of some particular interest as error laws in linear models, because they are heavy-tailed and because they yield log-concave likelihood functions (in contrast, for instance, to the student distributions). They are also of interest for various other reasons and we shall briefly describe some of their properties.

The hyperbolic distributions in r-dimensions ($r \geq 1$) are characterized by the fact that the graphs of their log probability density functions are hyperbolas or hyperboloids. The one-dimensional hyperbolic distribution has four parameters, α, β, μ and δ and it is denoted by $H(\alpha,\beta,\mu,\delta)$. It may be represented as a normal variance-mean mixture,

$$H(\alpha,\beta,\mu,\delta) = N(\mu+\beta\sigma^2,\sigma^2) \underset{\sigma^2}{\wedge} N_1^-(\delta^2,\alpha^2-\beta^2),$$

where $N(\xi,\sigma^2)$ denotes the normal distribution and $N_1^-(\chi,\psi)$ is one of the generalized inverse Gaussian distributions.

One can think of a hyperbolic distribution in the following way:



By pressing down the asymptote ϕ until it becomes a vertical line one can obtain a limit distribution on the half line, which may be called a positive hyperbolic distribution. Actually, the class of positive hyperbolic distributions coincides with the class of distributions N_1^- .

Using the above mixture representation it is possible to show that the hyperbolic distributions are self-decomposable (or belong to the Lévy class), i.e., there exists a sequence of independent random variables $x_1, x_2, \ldots,$ and norming constants a_n and b_n such that

$$\frac{(x_1+\ldots+x_n)-a_n}{b_n} \xrightarrow{\sim} \text{hyperbolic distribution}$$

as $n \to \infty$. There is at least one good statistical reason for being interested in the property of self-decomposability. Cox (1981) considered autoregressive stationary schemes

$$y_n = \theta\, y_{n-1} + u_n,$$

where the u_n's are i.i.d. random variables and he raised the problem of characterizing those one-dimensional distributions that can occur as the distribution of y_n in such a scheme.

The answer was given by L. Bondesson in the discussion to Cox's paper, and it is precisely that the distribution should be self-decomposable.

The hyperbolic distributions were originally motivated by some geological problems (Barndorff-Nielsen [1977]). Later, it was observed that the three-dimensional version of the hyperbolic distribution in the isotropic case has been known to statistical physicists since 1911.

According to Boltzmann and Gibbs the distribution of a momentum vector $p = (p_x, p_y, p_z)$ of a particle (x,y,z) in a physical system follows the Maxwell-Boltzmann law with a p.d.f.

(5) $$ae^{-\lambda K(p)} \quad \langle \text{Leb} \rangle ,$$

where a is a norming constant, λ is a simple function of Boltzmann's constant and the absolute temperature of the system, K(p) is the kinetic energy of a single particle in the system and 'Leb' stands for the Lebesgue measure. For the simplest possible case of an ideal gas the kinetic energy is

(6) $$K(p) = \frac{1}{2m}(p_x^2+p_y^2+p_z^2) = \frac{1}{2m}p\cdot p,$$

where m is the mass of a particle and $\cdot$ represents inner product and hence (p_x, p_y, p_z) follows an isotropic normal distribution. However, while (6) is the appropriate formula within the classical, Newtonian, framework of physics, it turns out that in Einstein's relativity theory K(p) has to be replaced by

(7) $$K(p) = c\sqrt{m^2c^2+p\cdot p} \quad ,$$

where c is the velocity of light. Note that Equation (7) yields precisely the formula for the density of the isotropic three-dimensional hyperbolic distribution. The general r-dimensional hyperbolic distribution has, except for a location-scale change, a p.d.f.

$$ae^{-\alpha\sqrt{1+x\cdot x}+\beta\cdot x}$$

where a is a norming constant.

This discussion is, in fact, not complete from the relativistic viewpoint, because it was assumed that the coordinate system of the observer of the ideal gas was stationary relative to the ideal gas. For a complete relativity treatment it is necessary to consider that an observer moves with a constant speed u relative to the ideal gas. This problem was solved by a Danish physicist, C. Møller, in 1968 by introducing a Lorentz transformation from one coordinate system to the other coordinate system which gives

$$K(p) = (c/\sqrt{1-u^2/c^2})(\sqrt{m^2c^2+p\cdot p} - (u/c,\ 0,0)\cdot p),$$

where c is the velocity of the light and u is the velocity of the observer. Møller also showed that the p.d.f. is a relativistic invariant. Together these results lead to the completely general form of hyperbolic distributions.

Additional information on the theory and applications of the hyperbolic and various related distributions may be found in Barndorff-Nielsen and Blaesild (1980, 1981), Blaesild and Jensen (1981) and the references given there.

4. Exponential transformation models

Statistical models that are exponential as well as transformational are called exponential transformation models. Such models are highly structured and we present here a theorem which shows some of this structure.

Theorem. Consider an exponential transformation model, i.e., the model function, relative to invariant measure μ on X, is of the form

$$p(x;g) = p(g^{-1}x) = a(g)b(x)e^{\theta(g)\cdot t(x)} \qquad \langle\mu\rangle . \tag{8}$$

Suppose that the exponential representation in (8) is minimal and of order k, and that p(x) and t(x) are continuous.

Then there exists, uniquely, a k-dimensional representation A(g) of the group G and k-dimensional vectors B(g) and $\tilde{B}(g)$ such that

$$t(gx) = A(g)t(x) + B(g)$$

$$\theta(g) = A(g^{-1})^*\theta(e) + \tilde{B}(g).$$

Moreover, for $g',g \in G$ we have

$$B(gg') = A(g)B(g') + B(g)$$

$$\tilde{B}(gg') = A(g^{-1})^*\tilde{B}(g') + \tilde{B}(g) .$$

Let δ be the function given by

$$\delta(g) = \frac{a(e)}{a(g)} e^{-\theta(g)\cdot B(g)}, \quad g \in G .$$

We then have

$$\delta(gg') = \delta(g)\delta(g')e^{\tilde{B}(g^{-1})\cdot B(g')}, \quad g',g \in G$$

and

$$b(gx) = \delta(g)b(x)e^{\tilde{B}(g^{-1})\cdot t(x)}.$$

If B(g) is constant, its value is necessarily 0, and similarly for $\tilde{B}(g)$. □

For further discussion of exponential transformation models, including the proof of the above theorem, the reader is referred to Barndorff-Nielsen, Blaesild, Jensen and Jørgensen (1982) and Barndorff-Nielsen (1982b).

5. A formula for the (conditional) distribution of the maximum likelihood estimator

We proceed to discuss a general formula for the distribution of the maximum likelihood estimator, conditional on an ancillary statistic a.

Let $\hat{\omega}$ be the maximum likelihood estimator of ω. Typically, the dominating measure μ is either Lebesgue measure (or, more generally, geometric measure on manifolds) or counting measure. Let t be a minimal sufficient statistic for the model and suppose there exists an, exact or approximate, ancillary statistic a such that $(\hat{\omega},a)$ is a one-to-one transformation of t. The dimensions of t and ω will be denoted by m and d, respectively, and we call the correspoonding model an *(m,d) model*.

Consider the conditional probability function $p(\hat{\omega};\omega|a)$ of $\hat{\omega}$ given a and define another probability function for $\hat{\omega}$ given ω and a by

$$p^*(\hat{\omega};\omega|a) = c|\hat{j}|^{\frac{1}{2}}\bar{L} . \tag{9}$$

Herewith $L = L(\omega) = L(\omega;x) \propto p(x;\omega)$ being an arbitrary version of the likelihood function, $\bar{L} = L(\omega)/L(\hat{\omega})$ is the normed likelihood function, $|\hat{j}|$ is the determinant of $\hat{j} = j(\hat{\omega})$ where $j = j(\omega) = (-\frac{\partial^2 \log L}{\partial\omega\, \partial\omega^*})_{d\times d}$ denotes the observed information matrix, and $c = c(a;\omega)$ is a norming constant. The probability function $p^*(\hat{\omega};\omega|a)$ is to be viewed as an approximation to $p(\hat{\omega};\omega|a)$, i.e.

$$p(\hat{\omega};\omega|a) \simeq c|\hat{j}|^{\frac{1}{2}}\bar{L} = p^*(\hat{\omega};\omega|a), \qquad \langle\mu\rangle \tag{A}$$

which we call formula (A) (cf. Barndorff-Nielsen [1980,1982]). Formula (A) expresses the fact that the conditional probability function $p(\hat{\omega};\omega|a)$ can very widely be calculated approximately, and in many important cases even exactly, by the probability function $p^*(\hat{\omega};\omega|a)$.

Example. Gamma distribution

Let $x_1,\ldots,x_n$ be a random sample from the p.d.f.

$$p(x;\lambda,\alpha) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1}e^{-\alpha x}, \quad \alpha>0,\ \lambda>0,\ x>0,$$

with λ known. Disregarding a factor depending on the observations only, the likelihood function is

$$L(\alpha) = \alpha^{n\lambda} e^{-\alpha x_{\bullet}}, \qquad \text{where } x_{\bullet} = \sum_{i=1}^{n} x_i .$$

Thus $\hat{\alpha} = \lambda \bar{x}^{-1}$ and $j(\alpha) = \lambda\alpha^{-2}$. Note that m=1 implies that no ancillary statistic a comes into the model. Hence

$$p^*(\hat{\alpha};\alpha) = c\,\bar{x}\,\bar{x}^{-\lambda}\alpha^{\lambda}e^{-n\alpha\bar{x}} = p(\hat{\alpha}:\alpha)$$

Hence, in this case, formula (A) is exact. □

<u>Example</u>. <u>Location-scale models (continued)</u>

Let $x_1,\ldots,x_n$ be a random sample from $\frac{1}{\sigma} f(\frac{x-\xi}{\sigma})$. Assume $(\hat{\xi},\hat{\sigma})$ exists uniquely (which is the case if log f(x) is concave or strongly unimodal). In general, $t = (x_{(1)},\ldots,x_{(n)})$, hence m=n, where $x_{(1)},\ldots,x_{(n)}$ are the order statistics. The ancillary statistic a is given by

$$a = u = \left(\frac{x_1-\bar{x}}{s},\ldots,\frac{x_n-\bar{x}}{s}\right),$$

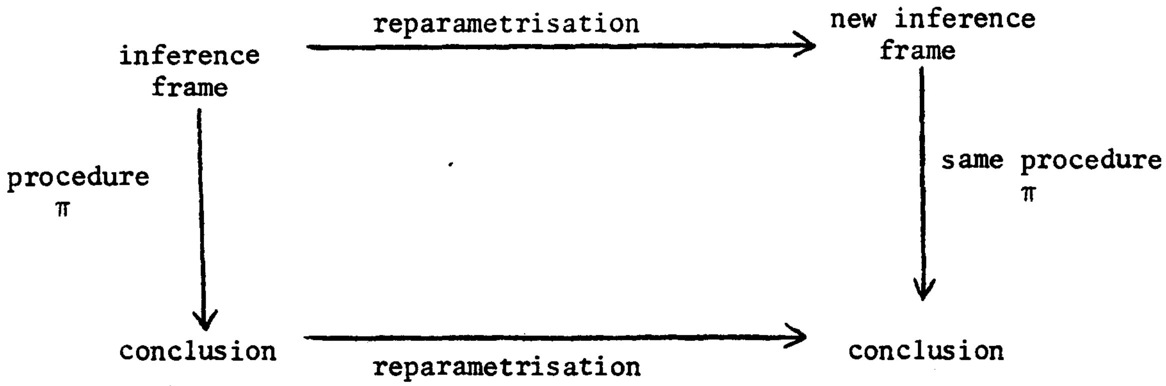
where u is an orbit representative, which is a maximal invariant statistic. Now

$$p^*(\hat{\xi},\hat{\sigma};\xi,\sigma|u) = c(u)\,\sigma^{-n}\hat{\sigma}^{n-2}\prod_{i=1}^{n} f(\hat{\xi}-\xi+\hat{\sigma}u_i)$$

$$= p(\hat{\xi},\hat{\sigma};\xi,\sigma|u).$$

Again, formula (A) is exact in this (n,2) model. □

Here, we give a few remarks on invariance of statistical procedures, in particular, *parametrisation invariance*. If we think of an *inference frame* as consisting of the data in conjunction with the model and a particular parametrisation of the model, and of a *statistical procedure* π (e.g., least squares estimation, maximum likelihood estimation, some particular construction of confidence regions, etc.) as a method which leads from the inference frame to a conclusion formulated in terms of the parametrisation of the inference frame, then parametrisation invariance may be formally specified as commutativity of the diagram



Procedures such as confidence interval procedures, likelihood ratio tests, maximum likelihood estimation and score tests (using expected information) are all parametrisation invariant. However, Wald's test is not parametrisation invariant. Procedures that are not parametrisation invariant require some care in their use. For the Wald test this is discussed in Vaeth (1981).

Concerning formula (A) we note the following properties:

a) (A) is parametrisation invariant.

b) (A) is also invariant under 1-1 transformations of the data x.

c) The expression $p^* = p^*(\hat{\omega};\omega|a)$ is the same whatever version L of the likelihood function is selected for the construction of p^*.

d) Suppose a is exactly distribution constant, and let L_0 denote the particular version of the likelihood function given by $L_0(\omega) = p(x;\omega)$ where the probability density function $p(x;\omega)$ is relative to the relevant Lebesgue measure. Then the formula (A) is exact if and only if $|\hat{j}|^{\frac{1}{2}} \hat{L}_0$ depends on a only (where $\hat{L}_0 = L_0(\hat{\omega})$).

e) For exponential models the formula (A) is the renormalized saddle point approximation. Application and theory of the approximation to the exponential models where m=d was discussed in Barndorff-Nielsen and Cox (1979). These saddle point approximations are known to be highly accurate. Normally, approximations in statistics are accurate to order $O(n^{-\frac{1}{2}})$, whereas renormalized saddle point approximations are always accurate to order $O(n^{-1})$ and often $O(n^{-3/2})$, where n denotes the sample size (Durbin [1980]).

f) Formula (A) is exact in a wide class of cases. A complete solution to the problem of when formula (A) is exact is not available. However, we have

Theorem. (Barndorff-Nielsen [1980, 1982]) *Formula (A) is exact for all transformation models.* □

(The starting point for the proof of this result is conclusion (iv) of the key theorem for transformation models presented earlier.) It should be noted that there are models other than those of transformational type for which formula (A) is exact.

Example. Inverse Gaussian model, $N^{-}(\chi_0,\psi)$

The model function of the family of inverse Gaussian distributions is

$$P(x;\chi,\psi) = \frac{\sqrt{\chi}}{\sqrt{\pi}}\, e^{\sqrt{\chi\psi}}\, x^{-3/2}\, e^{-1/2(\chi x^{-1}+\psi x)},\quad x>0,\ \chi>0,\ \psi\geq 0.$$

The submodel obtained by fixing χ is exponential of order 1 and has no ancillary statistic. Also it is not of transformational type. Nevertheless, formula (A) with a degenerate yields the exact expression for the distribution of $\hat{\psi} = \chi/x^2$. □

Daniels (1980) showed that the only linear exponential models of order 1 for which formula (A) is exact for all sample sizes are the normal distribution with known variance, the gamma distribution with known shape parameter, and the inverse Gaussian distribution with known χ.

Now we give two further examples of exponential families, with (m,d) equal to (2,2) and (3,3) respectively, which are not transformation models but do have formula (A) exact.

Example. Inverse Gaussian model $N^-(\chi,\psi)$ (continued)

Let $x_1,\ldots,x_n$ be a sample of size $n>1$ from the inverse Gaussian distribution $N^-(\chi,\psi)$. When χ and ψ are both unknown we have $(m,d)=(2,2)$ and $t = (\sum x_i^{-1}, \sum x_i)$. As noted previously, $n^{-1}\sum x_i$ has again an inverse Gaussian distribution $N^-(n\chi,n\psi)$, while $w = \sum(x_i^{-1}-\bar{x}^{-1})$ is independent of $\sum x_i$ and follows the gamma distribution $\Gamma(\frac{n-1}{2}, \frac{\chi}{2})$. From these facts it is easy to see that formula (A) is exact. □

Example. Hyperboloid model, $H_0(\xi,\kappa)$

This model pertains to observations x on the unit hyperboloid $H^{k-1} = \{x\in \mathbb{R}^k;\ x*x=1,\ x_0>0\}$, where $x = (x_0,x_1,\ldots,x_{k-1})$ and $x*y=x_0y_0-x_1y_1-\ldots-x_{k-1}y_{k-1}$. The hyperboloid model function relative to a certain invariant measure μ on H^{k-1} is

$$(10)\qquad p(x;\xi,\kappa) = a_k(\kappa)\ e^{-\kappa\xi*x} \qquad \langle\mu\rangle,$$

where $x \in H^{k-1}$, $\xi \in H^{k-1}$, and $\kappa > 0$, and where

$$(11)\qquad a_k(\kappa) = \kappa^{(k/2)-1}/\{(2\pi)^{(k/2)-1}2K_{(k/2)-1}(\kappa)\}$$

with $K_{(k/2)-1}$ a Bessel function. The parameters ξ and κ are called, respectively, the mean direction and the precision.

For any fixed κ, the hyperboloid distributions $p(x;\xi,\kappa)$ constitute a transformation model under the action of the pseudo-orthogonal group of type (1,k-1)

$$SO^{\uparrow}(1,k-1) = \{ \underset{k\times k}{U} \; ; \; U^*\tilde{I}\, U = \tilde{I},\ |U| = 1, \quad u_{00}>0\},$$

where $$\tilde{I} = \begin{bmatrix} 1 & & & & 0 \\ & -1 & & & \\ & & -1 & & \\ & & & \ddots & \\ 0 & & & & -1 \end{bmatrix}$$, and u_{00} is the upper-left element of U.

The invariance of the dominating measure μ, referred to above, is relative to this action.

Now suppose we have a sample $x_1,\ldots,x_n$ from the distribution (10) and let $r = \sqrt{x_\cdot^* x_\cdot}$, where $x_\cdot = \sum_{i=1}^n x_i$. The statistic r is an analog of the resultant length of a sample from the von Mises-Fisher distribution. It can be shown that $p(\hat{\xi};\kappa,\xi|r)$, where $\hat{\xi} = x_\cdot/r$ is the MLE of ξ, is again a hyperboloid model function. In fact, $\hat{\xi}|r \sim H_0(\xi, r\kappa)$. This is in complete analog with the von Mises-Fisher situation. Moreover, for k=3 one has $r-n \sim \kappa\chi^2(n-1)$ (Jensen [1981]). The full model (10) with k=3, which is a (3,3) model, is not transformational. However, using $\hat{\xi}|r \sim H_0(\xi, r\kappa)$ and $r-n \sim \kappa\chi^2(n-1)$ it is simple to show that formula (A) is exact for this model. □

We conclude the discussion of formula (A) by constructing some (4,4) exponential models for which formula (A) is exact.

The construction of higher order models for which the formula (A) is exact may proceed as follows. Suppose $p(u;\gamma)$ is a model function for which (A) is exact, and for each fixed u, let $p(v;\delta|u)$ be a model function, also with formula (A) exact. The joined model

$$p(u,v;\gamma,\delta) = p(u;\gamma)\cdot p(v;\delta|u)$$

will then, under three elementary conditions, have formula (A) exact (Barndorff-Nielsen [1982]).

Example. Combination of two inverse Gaussian distirbutions, $[N^-, N^-]$

Let $u \sim N^-(\chi,\psi)$ and $v|u \sim N^-(d(u)\kappa,\lambda)$, where $d(u)$ is an arbitrary positive function of u. Then formula (A) is exact for the joint distribution $p(u,v;\chi,\psi,\kappa,\lambda)$, provided χ and κ are presumed known.

If we take a sample $(u_1,v_1),\ldots,(u_n,v_n)$ from $p(u,v;\chi,\psi,\kappa,\lambda)$ then we can ask whether formula (A) provides the exact distribution for the maximum likelihood estimator of the four-dimensional parameter $(\chi,\psi,\kappa,\lambda)$. This is not the case in general; however, if $d(u)=u^2$ then we can show that formula (A) is exact for any sample size $n>1$. Setting

$$\alpha = \psi - 2\sqrt{\kappa\lambda}$$

we may write the joint density $p(u,v;\chi,\psi,\kappa,\lambda)$ of (u,v) as

$$p(u,v;\chi,\psi,\kappa,\lambda) = \sqrt{\chi\kappa}\, e^{\sqrt{\chi}\sqrt{\alpha+2\sqrt{\kappa\lambda}}}\, b(u,v)\, e^{-\frac{1}{2}\{\chi u^{-1}+\alpha u+\kappa u^2 v^{-1}+\lambda v\}}, \tag{12}$$

where $b(u,v) = (2\pi)^{-1} u^{-1/2} v^{-3/2}$.

The above model (12) with all four parameters unknown, will be denoted by $[N^-,N^-](\chi,\psi,\kappa,\lambda)$.

It follows immediately from the form of the norming constant $\sqrt{\chi\kappa}$ $\exp\{\sqrt{\chi}\sqrt{\alpha+2\sqrt{\kappa\lambda}}\}$ that

$$(\bar{u},\bar{v}) \sim [N^-,N^-](n\chi,n\psi,n\kappa,n\lambda), \tag{13}$$

where $\bar{u} = n^{-1}u_{\bullet} = n^{-1}\sum u_i$, $\bar{v} = n^{-1}v_{\bullet} = n^{-1}\sum v_i$. In particular, if $\psi=\lambda=0$, then $(n^{-2}u_{\bullet},\ n^{-4}v_{\bullet})$ has a distribution which does not depend on n. We call this distribution the *bivariate stable law of index* $(\frac{1}{2}, \frac{1}{4})$.

More can be said about this model concerning maximum likelihood estimators. Suppose $n > 1$ and let

$$w = \sum(u_i^{-1}-\bar{u}^{-1}),$$

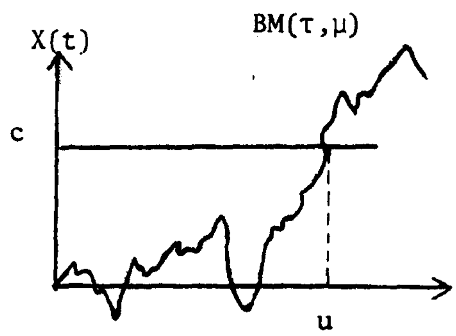
and

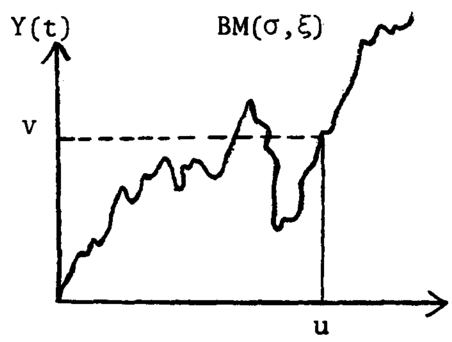
$$z = \sum(u_i^2 v_i^{-1}-\bar{u}^2\bar{v}^{-1}).$$

Then $\hat{\chi} = n/w$, $\hat{\psi} = n/(w\bar{u}^2)$, $\hat{\kappa} = n/z$, $\hat{\lambda} = n\bar{u}^2/(z\bar{v}^2)$. Furthermore, using $(\bar{u},\bar{v}) \sim [N^-,N^-](n\chi,n\psi,n\kappa,n\lambda)$ we can prove that w,z and $(\bar{u},\bar{v})$ are independent, that $w \sim \Gamma(\frac{n-1}{2}, \frac{\chi}{2})$ and $z \sim \Gamma(\frac{n-1}{2}, \frac{\kappa}{2})$, and that (A) is exact. Here we have a bivariate extension of the inverse Gaussian distribution with essentially all the nice properties of the inverse Gaussian distribution. (Construction of a bivariate generalization of the inverse Gaussian distribution for which both marginals are inverse Gaussians does not seem to be possible.) □

Example. Combination of inverse Gaussian and normal, $[N^-,N]$

Let $BM(\alpha,\beta)$ be a Brownian motion with diffusion coefficient α and drift coefficient β. Consider two Brownian motions $X(t)$ and $Y(t)$, where $X(t) \sim BM(\tau,\mu)$ and $Y(t) \sim BM(\sigma,\xi)$. Let u be the first passage time of $X(t)$ to a level $c>0$ and v be the value of $Y(t)$ when $t=u$.





Then we have

$$u \sim N^{-}(\chi,\psi),$$

where $\chi = c^2/\tau^2$, $\psi = \mu^2/\tau^2$, and

$$v|u \sim N(u\xi,\ u\sigma^2).$$

We denote the distribution of (u,v) by $[N^{-},N](\theta)$, i.e.

$$(u,v) \sim [N^{-},N](\theta),$$

where θ, which is the canonical parameter of the model, is

$$\theta = (c^2/\tau^2,\ 1/\sigma^2,\ \mu^2/\tau^2 + \xi^2/\sigma^2,\ 2\xi/\sigma^2).$$

Now, if $(u_1,v_1),\ldots,(u_n,v_n)$ is a sample from $[N^{-},N]$, then

$$(\bar{u},\bar{v}) \sim [N^{-},N](n\theta).$$

Furthermore, the maximum likelihood estimators of χ and κ follow

$$\hat{\chi}^{-1} = n^{-1}\sum u_i^{-1} - \bar{u}^{-1} \sim \Gamma(\tfrac{n-1}{2},\ n\chi/2),$$

$$\hat{\kappa}^{-1} = n^{-1}\sum v_i^2 u_i^{-1} - \bar{v}^2\bar{u}^{-1} \sim \Gamma(\tfrac{n-1}{2},\ n\kappa/2),$$

and $\hat{\chi}$, $\hat{\kappa}$ and $(\bar{u},\bar{v})$ are independent. From these results it is again simple to show exactness of (A). Moreover, we can perform an exact F test for the hypothesis H: $\xi=\mu$ and if $\xi=\mu$, then we can make exact inference on ξ. For a detailed discussion, see Barndorff-Nielsen and Blaesild (1982). □

One can build similar models of any multi-dimensions by essentially the same construction method.

An important tool for proving the above results is the following general and useful formula for calculating the marginal model function for a statistic $u = u(x)$. We may find a model function $p(u;\omega)$ of u in the following way:

(14) $$p(u;\omega) = E_{\omega_0} \{\frac{p(x;\omega)}{p(x;\omega_0)}|u\} \, p(u,\omega_0),$$

where ω_0 is an arbitrarily chosen value of ω. Thus, calculating a model function $p(u;\omega)$ in general is reduced to computing a conditional expectation provided we can choose a ω, which makes calculation of $p(u,\omega_0)$ particularly tractable.

Also, using this formula, we can often simplify calculation of the marginal likelihood function of ω based on u, since (14) implies that

(15) $$L(\omega;u) = E_{\omega_0} \{\frac{L(\omega)}{L(\omega_0)}|u\}.$$

(An interesting instance of formula (14) occurs in a recent issue of the *Journal of Mathematical Physics*, see Lavenda and Santamato [1981].)

In view of the nice and useful properties of the above mentioned examples it would seem worthwhile to make a systematic study for characterizations of the class of models for which formula (A) is exact.

6. Modified profile likelihood

As a derivative of formula (A) we now introduce a *modified profile likelihood* and discuss its applications.

Consider a model with parameter (ω,κ), where κ is the parameter of interest. The profile (or partially maximized) likelihood $\tilde{L}(\kappa)$ for κ is given by

(16) $$\tilde{L}(\kappa) = \sup_{\omega|\kappa} L(\kappa,\omega).$$

The study of formula (A) suggests (for the detailed reasoning see Barndorff-Nielsen [1982b]) introducing a *modified profile likelihood* $\overset{+}{L}(\kappa)$, which we define as follws:

(17) $$\overset{\dagger}{L}(\kappa) = \left|\frac{\partial\hat{\omega}}{\partial\hat{\omega}_\kappa}\right| |\hat{j}_\kappa|^{-\frac{1}{2}} \tilde{L}(\kappa),$$

where $\left|\frac{\partial\hat{\omega}}{\partial\hat{\omega}_\kappa}\right|$ is the Jacobian determinant for the transformation from $\hat{\omega}$ to $\hat{\omega}_\kappa$, the maximum likelihood estimate of ω for given known value of κ, and where $\hat{j}_\kappa$ is the observed information on ω given κ. It seems that the modified profile likelihood function will often exhibit better inferential properties than the profile likelihood function itself.

We may note that the formula for $\overset{\dagger}{L}(\kappa)$ is parametrisation invariant.

We have a particularly simple situation if our model function is of the following form:

(18) $$p(x;\kappa,\omega) = b(x,\kappa)\, e^{\alpha(\kappa)\phi(x,\omega)}$$

which generalizes the generalized linear models of Nelder and Wedderburn.

In the model (18) we have

$$\hat{\omega} = \hat{\omega}_\kappa \quad \text{for all } \kappa,$$

and

(19) $$\overset{\dagger}{L}(\kappa) = |\alpha(\kappa)|^{-d/2}\, b(x,\kappa)\, e^{\alpha(\kappa)\phi(x,\hat{\omega})},$$

where $d = \dim(\omega)$.

<u>Example.</u> <u>Pairs of observations from normal distributions</u>

Suppose $(x_{i1},x_{i2}) \sim N(\xi_i,\sigma^2)$, $i=1,2,\ldots,n$. Then, by maximizing $\tilde{L}(\sigma^2)$ we may obtain the full MLE $\hat{\sigma}$ which is known to be inconsistent but to converge to $\sigma^2/2$ as $n \to \infty$. However, by maximizing $\overset{\dagger}{L}(\sigma^2)$ we obtain the usual estimator, say $'\sigma^2$, which satisfies

$$'\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2 \longrightarrow \sigma^2 .$$ □

Example. Hyperboloid distribution

The model function for a single observation was

$$a_k(\kappa)\ e^{-\kappa\xi * x}$$

Note that $r = \sqrt{x_\cdot * x_\cdot}$ is the maximal invariant statistic for the model with fixed κ. Hence the distribution $p(r;\kappa)$ depending on κ alone is naturally used for inference on κ.

For $k = 3$ we have

(20) $$2(r-n) \sim \kappa^{-1}\chi^2(2(n-1)),$$

and for $k = 2$

(21) $$p(r;\kappa) = \{[a_k(\kappa)]^n/a_k(r\kappa)\}\ \pi^{n+1} r\ \mathrm{Re}\{i^{n+1}\int_0^\infty [H_0^{(1)}(x)]^n\ J_0(rx)x\,dx,$$

where $H_0^{(1)}$ is a Hankel function, - a rather redoubtable expression. The modified profile likelihood for κ is found to be

(22) $$\overset{\dagger}{L}(\kappa) = [a_k(\kappa)]^n\ \kappa^{(k-1)/2-r\kappa}$$

Using (11) and the asymptotic formula for the Bessel function $K_\nu(x)$ as $x \to \infty$, i.e.

$$K_\nu(x) \sim \sqrt{\pi/2}\ x^{-\frac{1}{2}}\ e^{-x},$$

one can obtain the following asymptotic expression for the modified profile likelihood:

(23) $$\overset{\dagger}{L}(\kappa) \sim \kappa^{(n-1)(k-1)/2}\ e^{-\kappa(r-n)}$$

as $\kappa \to \infty$.

Now, we may substitute (23) into formula (A), thereby obtaining an expression for the distribution of $\hat{\kappa}$. If the modified profile likelihood function gives an approximately correct answer, the derived expression should be an approximation to the distribution of $\hat{\kappa}$, at least for κ large. We note that the approximating formula in (23) equals the likelihood function of a gamma distribution so that without any calculation we find from formula (A)

$$2(r-n) \simeq \kappa^{-1}\chi^2((n-1)(k-1)). \tag{24}$$

As we can see from (20), (24) is, in fact, exact for $k = 3$. For general k formula (24) is asymptotically valid for $\kappa \to \infty$, as has been established via a traditional approach by Jensen (1981). Clearly, for $k = 2$, the approximation is a considerable simplification over the exact result (21). □

For further discussion and exemplification of the modified profile likelihood see Barndorff-Nielsen (1982b).

References

Barndorff-Nielsen, O. (1977) Exponentially decreasing distributions for the logarithm of particle size, *Proc. R. Soc. London, A 353,* 401-419.

Barndorff-Nielsen, O. (1982a) The hyperbolic distribution in statistical physics, *Scand. J. Statist., 9,* 43-46.

Barndorff-Nielsen, O. (1982b) Parametric statistical models and inference: some aspects. Research Report 81, Dept. Theoretical Statistics, Aarhus University.

Barndorff-Nielsen, O. and Blaesild, P. (1980) Hyperbolic distributions. (To appear in *Encyclopedia of Statistical Sciences,* Wiley, New York.)

Barndorff-Nielsen, O. and Blaesild, P. (1981) Hyperbolic distributions and ramifications: contributions to theory and application. In C. Taillie, G. P. Patil and B. A. Baldessari (Eds.) *Statistical Distributions in Scientific Work, 4,* 19-44, D. Reidel Publ. Co., Dordrecht.

Barndorff-Nielsen, O. and Blaesild, P. (1982) Reproductive exponential families. Research Report 87, Dept. Theoretical Statistics, Aarhus University.

Barndorff-Nielsen, O., Blaesild, P., Jensen, J. L. and Jørgensen, B. (1982) Exponential transformation models, *Proc. R. Soc. London, A 379,* 41-65.

Barndorff-Nielsen, O. and Cox, D. R. (1979) Edgeworth and saddle-point approximations with statistical applications. (With discussion) *J. R. Statist. Soc., B 41,* 279-312.

Blaesild, P. and Jensen, J. L. (1981) Multivariate distributions of hyperbolic type. In C. Taillie, G. P. Patil and B. A. Baldessari (Eds.) *Statistical Distributions in Scientific Work, 4,* 45-66, D. Reidel Publ. Co., Dordrecht.

Cox, D. R. (1981) Statistical analysis of time series: some recent developments. (With discussion) *Scand. J. Statist., 8,* 93-115.

Daniels, H. E. (1980) Exact saddlepoint approximations, *Biometrika, 67,* 53-58.

Durbin, J. (1980) Approximations for densities of sufficient estimators, *Biometrika, 67,* 311-333.

Jensen, J. L. (1981) On the hyperboloid distribution, *Scand. J. Statist., 8,* 193-206.

Lavenda, B . H. and Santamato, E. (1981) The kinetic analog of Boltzmann's principle, *J. Math. Phys., 22,* 2926-2933.

Tweedie, M. C. K. (1957) Statistical properties of inverse Gaussian distributions, *Ann. Math. Statist., 28,* 362-377.

Vaeth, M. (1981) On the use of Wald's test in exponential families. Research Report 70, Dept. Theoretical Statistics, Aarhus University.