THE ESTIMATION OF DISCONTINUITIES IN MULTIVARIATE DENSITIES, AND RELATED PROBLEMS IN STOCHASTIC PROCESSES

HERMAN RUBIN
MICHIGAN STATE UNIVERSITY

1. Summary

We consider estimation of the finite dimensional real parameter α from independent identically distributed observations on a generalized random variable with density discontinuous in the parameter on a set which depends on the parameter. We assume that certain regularity conditions are met; some of these will be explicitly stated, but many will not. These conditions are merely to insure that the approximations made are valid. Subject to these conditions, we show that hyper-efficient estimators, among them the maximum likelihood estimator, exist, and that asymptotically the estimation problem is equivalent to that for a nonstationary process with possibly multidimensional "time" and unknown center of nonstationarity.

The inference problem for the process is treated from the standpoint of the likelihood function, a method which has not been used as much as it should be.

2. Reduction to the asymptotic problems

Specifically, we consider the following problem: Let $x = (y, z)$, where y is k-dimensional, and let $R(\alpha)$ be a region in k-dimensional Euclidean space. We assume that the density is given by

$$f(x, \theta, \alpha) = g(x, \theta, \alpha), \qquad y \in R(\alpha), \tag{1}$$

$$f(x, \theta, \alpha) = h(x, \theta, \alpha), \qquad y \notin R(\alpha). \tag{2}$$

We do not assume that θ or z are finite dimensional. This density is with respect to $dy \times d\mu(z)$. From now on we shall assume all approcimations are valid.

One classic example is the case of the rectangular distribution with one or both endpoints unknown. Another is that of the endpoint of the exponential distribution. In both of these cases it is well known that the extreme order statistic or statistics are the maximum likelihood estimators and are hyperefficient.

Research partially supported by the Office of Naval Research.

The first nontrivial case considered was by Chernoff and Rubin [1]. In this paper y is one dimensional, α is one dimensional, and z is trivial. Also $R(\alpha) = (-\infty, \alpha)$. They showed that the maximum likelihood estimate and some related estimates are consistent and have an asymptotic variance $O(1/n^2)$. They also reduced the problem of computing the asymptotic distribution of the estimates obtained to the distribution of a corresponding function of a stochastic process. Breakwell and Chernoff, in some unpublished memoranda, obtained the asymptotic distribution of those estimates. It was observed that the mean is $O(1/n)$.

Since the mean is of the same order as the standard deviation, the question arises as to whether the maximum likelihood estimate is even asymptotically sufficient. This is not so, and in [2], Rubin considered the asymptotic estimation problem.

That there should be a close relationship between the inference problem for one-dimensional random observations and for Poisson processes is not as surprising as would seem. If we assume the number of observations has a Poisson distribution, the number of observations in an arbitrary set will be a Poisson variable, independent for disjoint sets, and hence will form a Poisson process. No great use has been made of this fact for inference purposes, and in most cases it will not be of great importance.

The special example used in [1] is very well suited to illustrate this point. In that example, $g(x) = \beta$ and $h(x) = \gamma$ on $(0, 1)$, and 0 otherwise. Then if the number of observations is Poisson with mean λ, and if $x' = \lambda(x - \alpha)$, the resulting process will have rate β on $(-\lambda\alpha, 0)$ and rate γ on $(0, \lambda(1 - \alpha))$. Intuitively the estimation of the actual 0 of such a process should not depend heavily on the tails, and so should be asymptotically independent of λ; also it should be asymptotically independent of the estimates of β and γ. Thus the asymptotic distribution of the estimate of α should have a scale factor of $1/\lambda$, and so should be hyperefficient, and also the large sample problem is approximately reduced to that of the corresponding stochastic process.

Let us now examine our problem. We define

$$\lambda(x, \theta, \alpha) = \frac{g(x, \theta, \alpha)}{h(x, \theta, \alpha)}. \tag{3}$$

Let β and γ be two possible values of α, and ϕ, ψ two possible values of θ. If we consider the likelihood ratio of (ϕ, β) to (ψ, γ), from a sample of size N we obtain

$$\begin{aligned} l(\beta, \phi, \gamma, \psi) &= \frac{\prod f(x_i, \phi, \beta)}{\prod f(x_i, \psi, \gamma)} \\ &= \prod_{x_i \in R(\beta) \cap R(\gamma)} \frac{g(x_i, \phi, \beta)}{g(x_i, \psi, \gamma)} \prod_{x_i \notin R(\beta) \cup R(\gamma)} \frac{h(x_i, \phi, \beta)}{h(x_i, \psi, \gamma)} \\ &\quad \prod_{x_i \in R(\beta) \sim R(\gamma)} \frac{g(x_i, \phi, \beta)}{h(x_i, \psi, \gamma)} \prod_{x_i \in R(\gamma) \sim R(\beta)} \frac{h(x_i, \phi, \beta)}{g(x_i, \psi, \gamma)}. \end{aligned} \tag{4}$$

Let us assume that the likelihood function is well behaved for observations restricted to neither or both of $R(\beta)$ and $R(\gamma)$. That is, we shall assume that

$$l(\beta, \chi, \gamma, \chi) = l_1(\beta, \gamma, \chi) l_2(\beta, \gamma, \chi), \tag{5}$$

where l_1 is the contribution of the terms in (4) with x_i in neither or both of $R(\beta)$ and l_2 is the contribution of those x_i in exactly one.

Let M be the number of the x_i entering in l_2.

$$l(\beta, \chi, \gamma, \chi) = l_3(M, \beta, \gamma, \chi) l_1(\beta, \gamma, \chi | M) l_2(\beta, \gamma, \chi | M), \tag{6}$$

where l_3 is the likelihood ratio of M and the other terms are the conditional likelihood ratios. Let capital letters denote the logarithms of the corresponding quantities with small letters. Then if suitable regularity conditions are satisfied,

$$\begin{aligned} &L_1(\zeta, \alpha, \chi | M) \\ &\quad = (N - M)^{1/2} \sum V_i(\zeta_i - \alpha_i) - \frac{1}{2}(N - M) \sum W_{ij}(\zeta_i - \alpha_i)(\zeta_j - \alpha_j) \end{aligned} \tag{7}$$

for ζ near α, χ near θ, where the V_i have a limiting distribution and W_{ij} have probability limits as N becomes large. Now if we can show that

$$L_3(M, \zeta, \alpha, \chi) + L_2(\zeta, \alpha, \chi | M) < A - BN \sum |\zeta_i - \alpha_i|, \tag{8}$$

for ζ near α and χ near θ, where A has a limiting distribution and B is some positive constant, L_1 can be ignored. In fact, in [1] much weaker conditions than (7) were used. The proof of (8) will be given here only asymptotically, (that is, for the stochastic process approximation), but it can be carried out as in [1]. The remainder of this section will be devoted to the asymptotic reduction of the likelihood function to a stochastic process.

Let us take, as an example, the following generalization of the case of Chernoff and Rubin. Let y and α be one dimensional, $R(\alpha) = (-\infty, \alpha)$, and assume that g and h are continuous in y and α uniformly in z and θ. These conditions are unnecessarily restrictive, but will serve to clarify the situation.

Then,

$$g(x, \chi, \zeta) \sim \beta(z), \tag{9}$$

$$h(x, \chi, \zeta) \sim \gamma(z), \tag{10}$$

$$\Lambda(x, \chi, \zeta) \sim \log \beta(z) - \log \gamma(z) = \Lambda^*(z). \tag{11}$$

As in [1], we shall assume that β and γ are both nonzero. It is clear that, by continuity, the general nature of the results must remain valid. However, technical difficulties occur which have not yet been adequately treated. Suppose $\alpha < \zeta$. From (8), we may take $\zeta = \alpha + \tau/N$. Then under α, the joint distribution of the numbers w_i of the x with y between α and ζ and $z \in S_i$ is multinomial with means

$$m_i(\alpha) = \int_\alpha^\zeta \int_{S_i} h(x, \chi, \alpha)\, d\mu(z)\, dx \tag{12}$$

and, under ζ, h is replaced by g. The likelihood ratio of these distributions is

$$l^*(\zeta, \alpha, \chi) = \left[\frac{1 - \sum m_i(\zeta)}{1 - \sum m_i(\alpha)}\right]^{N - \Sigma w_i} \prod \left(\frac{m_i(\zeta)}{m_i(\alpha)}\right)^{w_i}. \tag{13}$$

As $N \to \infty$ with τ fixed and the S_i shrink to sets on which Λ is nearly constant, we obtain

$$L^*(\zeta, \alpha, \chi) \sim \sum_{\alpha < y_i < \zeta} \Lambda^*(z_i) - \tau \int [\beta(z) - \gamma(z)] \, d\mu(z). \tag{14}$$

A similar argument holds if $\zeta < \alpha$, where the sign of the sum is reversed, as in [1], equation (41).

Now suppose we consider a random measure m on $(-\infty, \infty) \times Z$, with $m(E)$ Poisson with mean $\mu(E)$ and independent for disjoint E and $\mu(E) = \int \beta(z) \, d\mu(z) \, dy$ for $E \subseteq (-\infty, \rho) \times Z$ and β is replaced by γ for $E \subseteq (\rho, \infty) \times Z$. Then as in [2], the likelihood ratio of $\rho + \tau$ with respect to ρ is given by

$$\Lambda^*(\rho + \tau, \rho) = \sum_{\rho < y_i < \rho + \tau} \Lambda^*(z_i) - \tau \int [\beta(z) - \gamma(z)] \, d\mu(z), \tag{15}$$

which is approximately the same as (14), and the appropriate remarks apply. Furthermore, the distribution of L is asymptotically the same as that of L^*.

This can still be somewhat simplified. If we set

$$e^q = \frac{\beta(z)}{\gamma(z)}, \tag{16}$$

then (15) becomes

$$L^*(\rho + \tau, \rho) = \sum_{\rho < y_i < \rho + \tau} q_i - \tau \int (e^q - 1) \, d\nu(q). \tag{17}$$

This corresponds to a separable process with independent increments with characteristic function per unit interval

$$\begin{aligned} &\int [(e^{i\eta q} - 1)e^q - i\eta(e^q - 1)] \, d\nu(q), && y < \rho \\ &\int [(e^{i\eta q} - 1) - i\eta(e^q - 1)] \, d\nu(q), && y > \rho. \end{aligned} \tag{18}$$

The condition that we started with a finite sample size implies that

$$\int (e^q + 1) \, d\nu(q) < \infty. \tag{19}$$

However, the sample size might be infinite, but the conditional sample size given z is finite. In this case the condition on ν can be written in many ways; one is

$$\int (e^{q/2} - 1)^2 \, d\nu(q) < \infty. \tag{20}$$

Now let us consider the multivariate case. First let us look at an example. Let the range of y be the unit disc, z be trivial, α be the product of the unit circle with the open interval (0, 1), and θ the open interval (0, 1). Let $R(\alpha)$ be the ellipse with major axis 2 and minor axis $2\alpha_2$, with center at 0 and major axis in the direction $\alpha_1/2$.

Let

$$g(x, \theta, \alpha) = f_1(\theta, \alpha) f_2(y_1^2 + y_2^2, \theta), \tag{21}$$

$$h(x, \theta, \alpha) = f_1(\theta, \alpha) f_3(y_1^2 + y_2^2, \theta), \tag{22}$$

and let us assume that f_2 and f_3 are uniformly integrable and uniformly continuous. Then the regularity conditions are satisfied.

Again the likelihood ratio depends mainly on the observations falling in one of the regions, and not in the other. Let us assume, as in [1], that we employ a consistent estimate of θ. Now if $N(\zeta - \alpha) = \tau$, the number of observations in $R(\zeta)$ and not in $R(\alpha)$ with $y_1^2 + y_2^2 \in S$ will be approximately Poisson with mean

$$N \int_{[R(\zeta) \sim R(\alpha)] \cap S} h(x, \theta, \alpha)\, d\alpha \tag{23}$$

under α and g instead of h under ζ. Also the numbers in disjoint regions are asymptotically independent. Intuitively, the point of the boundary corresponding to a point in $R(\zeta) \sim R(\alpha)$, which in this case can be taken to be the nearest point on the boundary at the same distance from the origin, now enters the argument.

We can now consider the general case. As remarked previously, certain difficulties occur in the passage to the limit. We shall ignore this problem. To avoid difficulties with infinities, let us define

$$\xi(x, \theta, \alpha) = \frac{g(x, \theta, \alpha)}{g(x, \theta, \alpha) + h(x, \theta, \alpha)}. \tag{24}$$

Now consider $E_S = \{x_i: \xi(x_i, \theta, \alpha) \in S \text{ and } y_i \in R(\zeta) \sim R(\alpha)\}$, and $F_S = \{x_i: \xi(x_i, \theta, \alpha) \in S \text{ and } y_i \in R(\alpha) \sim R(\zeta)\}$. The number of observations in E_S is approximately Poisson with mean

$$\begin{aligned} &N \int_{\xi^{-1}S \cap [R(\zeta) \sim R(\alpha)] \times Z} h(x, A, \alpha)\, dy\, d\mu(z) \qquad \text{under } \alpha, \\ &N \int_{\xi^{-1}S \cap [R(\zeta) \sim R(\alpha)] \times Z} g(x, A, \zeta)\, dy\, d\mu(z) \qquad \text{under } \zeta, \end{aligned} \tag{25}$$

with corresponding expressions for the case of F. With $\tau = N(\zeta - \alpha)$ fixed, these expressions approach a limit which is positively homogeneous in τ. Thus the distribution function of L^* corresponds to a separable process with multidimensional time.

Suppose now that the boundary $\beta R(\alpha)$ of $R(\alpha)$ is such that a point in the neighborhood of it can be expressed in the form (u, v), $u \in \beta R(\alpha)$, and suppose that $R(\zeta)$ is such that

$$\begin{aligned} &(u, v) \in R(\zeta) \sim R(\alpha) \quad \text{if} \quad u \in Q_1(\zeta - \alpha) \quad \text{and} \quad 0 < v \leqq \phi_1(u, \zeta - \alpha), \\ &(u, v) \in R(\alpha) \sim R(\zeta) \quad \text{if} \quad u \in Q_2(\zeta - \alpha) \quad \text{and} \quad 0 > v \geqq \phi_2(u, \zeta - \alpha). \end{aligned} \tag{26}$$

Suppose further that $N\phi_i(u, \eta/N)$ approach a limit $\psi_i(u, \eta)$ and also $Q_i(\theta, \eta/N)$ converge to $V_i(\eta)$, and $\psi_1(u_1 - \eta) = \psi_2(u, \eta)$, $V_1(-\eta) = V_2(\eta)$.

Let $dm(u)\, dv = dy$ be the representation of Lebesgue measure on the neighborhood of $\beta R(\alpha)$ in the (u, v) coordinate system. Let $\tau^i = N(\zeta^i - \alpha)$. Then the number of observations in $E_S^1 \cap \cdots \cap E_S^n$ is asymptotically Poisson with mean

$$\int_U h\{[(u, 0), z], \theta, \alpha\} \min_i \psi_1(u, \tau^i)\, dm(u)\, d\mu(z), \tag{27}$$

where

$$U = \bigcap_{i=1}^{n} V_1(\tau^i) \cap \xi^{-1}(S) \cap [\beta R(\alpha) \times Z], \tag{28}$$

if (27) is continuous at S. The reason for this latter restriction is that, loosely speaking, a "large" number of points can move in and out of S by a slight change of parameters if it is not met. A similar expression holds for F. The asymptotic independence of disjoint sets completes the description of the asymptotic reduction to the process, which otherwise proceeds as in the univariate case. The details are left to the reader.

Another way of looking at the reduction is as follows. In the neighborhood of any particular point on the boundary, the movement of points in and out of $R(\zeta)$ as ζ varies from α depends essentially on a linear function of τ. Thus the resulting likelihood process is a continuous convolution of one-dimensional processes.

3. Description of the stochastic process

Let us first describe a general process with independent increments and n-dimensional time. For convenience the notation used in this and the following section is independent of that of the preceding part. Since the process is to have independent increments along each line,

$$\log E\{e^{i\lambda[X(t)-X(u)]}\} = i\alpha(t, u)\lambda - \frac{1}{2}\sigma^2(t, u)\lambda^2 + \int [e^{i\lambda x} - 1 - i\lambda f(x)]\mu(dx, t, u). \tag{29}$$

This can be looked at, in the usual manner, as a convolution of a Gaussian process with a continuous convolution of Poisson processes. Let us for the moment consider only the two Poisson processes with means 1 and -1. In this case

$$\log E\{e^{i\lambda[X(t)-X(u)]}\} = a(t, u)(e^{i\lambda} - 1) + b(t, u)(e^{-i\lambda} - 1). \tag{30}$$

Some inequalities can be obtained, but the joint distribution cannot be obtained from these data. Let us consider, however, a set of the processes introduced by consideration of (27). Let a function ϕ be given on $E_n \times W$, linear on E_n, and let ν be a measure on W. We can obtain a stationary generalized Poisson process as follows. Let F be a finite subset of E_n. For each nonempty subset G of F, define

$$\gamma(G) = \int [\min_{t\in G}{}^{+} \phi(t, w) - \max_{t\notin G}{}^{+} \phi(t, w)]^{+} \, d\nu(w),$$

(31)

$$\delta(G) = \int [\min_{t\in G}{}^{+} \{-\phi(t, w)\} - \max_{t\notin G}{}^{+} \{-\phi(t, w)\}]^{+} \, d\nu(w),$$

where the superscript + makes the expression 0 if it is negative or undefined. Let $g(G)$, $h(G)$ be independent Poisson variables with means $\gamma(G)$, $\delta(G)$ respectively. Then set

$$X(t) = \sum_{t\in G} g(G) - h(G). \tag{32}$$

In the one-dimensional case, this gives the general difference of two Poisson processes, and, in fact, W can be taken to consist of two points, one for each process. We shall consider stationary processes which are infinite convolutions of these processes, with possibly a Gaussian component.

The processes we have given have another description which is rather interesting. That is, we can look at the process as a continuous convolution of a one-dimensional process, the convolution being over the space of linear functionals. As two proportional linear functionals combine into one, this integration is over proportionality classes of linear functionals.

Can we now get an interpretation of the Poisson component for this process? If we attempt this, we obtain precisely the previous formulation. In fact, each jump corresponds to one linear functional. We thus obtain a Poisson distribution of the number of jumps in a given size range lying in a given set.

The location of a given jump can be used to supply a precise value to the arbitrary constant of proportionality by having the linear functional take the value 1 at the jump, if desired. We shall not use this fact.

Another consequence of this approach is that we can consider a nonstationary process as a convolution of nonstationary processes. If the process is to be locally stationary except at 0, the only difference is that a constant of proportionality must be inserted into the right side of one, or both, lines of (31). This type of process, with no Gaussian component, corresponds precisely to the problem at hand.

4. Inference about the point of nonstationarity

As we can consider the process as the convolution of one-dimensional processes plus a Gaussian term, all of which are independent, we can obtain the likelihood ratio by an investigation of the one-dimensional case together with a consideration of the Gaussian part.

In the one-dimensional case, let ξ be the point of nonstationarity, and let

$$(33)\quad \begin{aligned} &\frac{1}{t-\xi}\log E\{e^{i\lambda[X(t)-X(\xi)]}\} \\ &\qquad = i\alpha\lambda - \frac{1}{2}\sigma^2\lambda^2 + \int [e^{i\lambda x} - 1 - i\lambda f(x)]\, d\mu(x), \qquad t > \xi \\ &\frac{1}{\xi - t}\log E\{e^{i\lambda[X(\xi)-X(t)]}\} \\ &\qquad = i\beta\lambda - \frac{1}{2}\tau^2\lambda^2 + \int [e^{i\lambda x} - 1 - i\lambda g(x)]\, d\nu(x), \qquad t < \xi. \end{aligned}$$

Let us compute the likelihood ratio of the point η of nonstationarity to the point ξ. We may assume $\xi < \eta$. Then the increment process for $t < \xi$ and $t > \eta$ has the same distribution under ξ and η, so only the process between ξ and η is relevant. Under each of ξ, η the process is stationary there. An easy consideration (see, for example, [3]), shows that a sufficient statistic for a stationary process over a finite interval is the sequence of its jumps (which can be ordered, for example, by the magnitude of the jump, the locations of the jumps being unimportant), the variance (which can be estimated exactly with probability one) of its Gaussian component, and its increment over the interval.

The inference problem can be trivial, that is, ξ can be computed with probability one. While manipulating the process, we shall impose any condition we find necessary to prevent this.

Thus the first consideration for the problem to be nontrivial is that $\sigma = \tau$. Next, the number of jumps whose sizes are in S is Poisson with mean $(\eta - \xi)\mu(S)$ under ξ and mean $(\eta - \xi)\nu(S)$ under η. Let

$$(34)\qquad \phi = \frac{d\nu}{d(\mu + \nu)}.$$

First consider $\mu\phi^{-1}\{0\}$. This is the probability per unit time that a ξ-admissible jump will occur under ξ which cannot occur under η. Obviously the likelihood ratio will be 0 if such a jump occurs. Now if $\mu\phi^{-1}\{0\} = \infty$, there will be such a jump in every interval to the right of ξ with probability one, and in no interval to the left of ξ with probability one, if ξ is the true value. Note that any test of ξ against η has constant power on $(-\infty, \xi]$ and $[\eta, \infty)$. Thus, we may assume that $\mu\phi^{-1}\{0\}$, and $\nu\phi^{-1}\{1\}$, are finite.

We shall next show that $(\mu + \nu)\phi^{-1}(A) < \infty$ if $1/2$ is not a limit point of A for the inference problem to exist. If this measure is infinite, say $(\mu + \nu)\phi^{-1}(B) = \infty$, $B \subseteq [\rho, 1)$, $\rho > 1/2$. Now $\phi^{-1}(B) = \bigcup S_n$, with $\mu(S_n) < \infty$ for all n, $S_n \subseteq S_{n+1}$. For the case in which B has one point, see [2].

Let Y_n be the number of jumps whose sizes are in S_n. Now under ξ,

$$(35)\qquad E(Y_n) = \operatorname{Var}(Y_n) = (\eta - \xi)\mu(S_n) \geqq \rho(\eta - \xi)(\mu + \nu)(S_n)$$

so that for $r < \rho$,

(36)

$$P\{Y_n \geqq r(\eta - \xi)(\mu + \nu)(S_n)\} \geqq \frac{(\eta - \xi)[\mu(S_n) - r(\mu + \nu)(S_n)]^2}{(\eta - \xi)[\mu(S_n) - r(\mu + \nu)(S_n)]^2 + \mu(S_n)}$$

$$\geqq \frac{(\eta - \xi)(\rho - r)^2(\mu + \nu)(S_n)}{(\eta - \xi)(\rho - r)^2(\mu + \nu)(S_n) + 1}.$$

Thus for large n, we have $Y_n \geqq (\eta - \xi)(\rho - \epsilon)(\mu + \nu)(S_n)$ with probability close to one under ξ, and similarly $Y_n \leqq (1 - \rho + \epsilon)(\eta - \xi)(\mu + \nu)(S_n)$ with probability close to one under η, so that ξ and η are distinguishable with probability one. It is not difficult to extend the argument to show that the value of ξ can be inferred with probability one.

We are now in a position to compute the contribution to the likelihood function of the jumps. If x_i is the ith jump in some order, a restatement of (17) yields

$$L = \sum \log \frac{\phi(x_i)}{1 - \phi(x_i)} - (\eta - \xi) \int_0^1 (2w - 1)\, d(\mu + \nu)\phi^{-1}(w). \tag{37}$$

The sum should be taken negatively for $\eta < \xi$.

Two remarks are appropriate in connection with (37). First, it takes into account correctly the possibility that $\phi = 0$ or $\phi = 1$. Observe that if $\phi = 0$ or 1, then $L = -\infty$ or ∞, respectively. However, our previous discussion about this possibility shows that $\phi(x) = 0$ can only occur for $x > \xi$ and $\phi(x) = 1$ for $x < \xi$, where ξ is the true value, which only restricts the range of η. Second, the separate terms on the right side need not exist. Let us write

$$L_J = \sum L_{Jn}, \tag{38}$$

where

$$L_{Jn} = \sum_{\phi(x_i) \in S_n} \log \frac{\phi(x_i)}{1 - \phi(x_i)} - (\eta - \xi) \int_{S_n} (2w - 1)\, d(\mu + \nu)\phi^{-1}(w), \tag{39}$$

and $1/2$ is not a limit point of S_n for any n.

By our preceding observations, each L_{Jn} exists with probability one. Also the L_{Jn} are independent random variables. We may select the S_n to suit our convenience. Consequently, set $S_1 = [0, 1/3) \cup (2/3, 1]$. Then for $n > 1$, if ξ is the true value,

$$E(L_{Jn}) = (\eta - \xi) \int_{S_n} \log \frac{w}{1 - w}\, d\mu\, \phi^{-1}(w) - (2w - 1)\, d(\mu + \nu)\phi^{-1}(w), \tag{40}$$

$$\operatorname{Var}(L_{Jn}) = (\eta - \xi) \int_{S_n} \left(\log \frac{w}{1 - w}\right)^2 d\mu\, \phi^{-1}(w). \tag{41}$$

Both integrands are comparable with $(2w - 1)^2\, d(\mu + \nu)\phi^{-1}$ on $[1/3, 2/3]$ and the integrand in (40) is negative. A modification of the argument used to show that $(\mu + \nu)\phi^{-1}$ is finite away from $1/2$ shows that

$$\int_{1/3}^{2/3} (2w - 1)^2 \, d(\mu + \nu)\phi^{-1}(w) < \infty, \tag{42}$$

and hence

$$\int_0^1 (2w - 1)^2 \, d(\mu + \nu)\phi^{-1}(w) < \infty. \tag{43}$$

If (43) holds, all our previous conditions are satisfied. Since (42) then implies that $\sum_{n>1} E(L_{Jn})$ and $\sum_{n>1} \mathrm{Var}\,(L_{Jn})$ are finite, the sum in (38) exists with probability one.

We have deliberately left f and g vague. We shall now show that we may take $f\, d\mu = g\, d\nu$, and that after that is done, the inference problem is trivial if $\alpha \neq \beta$ unless $\sigma > 0$.

The usual procedure is to take f to be $f_0 = x/(1 + x^2)$. Furthermore the condition usually given on μ is that $\int x^2/(1 + x^2)\, d\mu(x) < \infty$. Since this also holds for ν,

$$\int \frac{x^2}{1 + x^2}\, d(\mu + \nu) < \infty. \tag{44}$$

Now clearly f can be taken to be any function with $f - f_0$ integrable $[\mu]$, and hence it is sufficient that $f - f_0$ be integrable $[\mu + \nu]$. Let $f = 2f_0\phi$. Then

$$f(x) - f_0(x) = \frac{x}{1 + x^2}\,[2\phi(x) - 1]. \tag{45}$$

It easily follows from (43) and (44) that $f - f_0$ is integrable $[\mu + \nu]$. Similarly we may take $g = 2(1 - \phi)f_0$. Then since $d\mu = (1 - \phi)\, d(\mu + \nu)$, $d\nu = \phi\, d(\mu + \nu)$, we see that

$$f\, d\mu = g\, d\nu. \tag{46}$$

Now let $[f_\epsilon(x), g_\epsilon(x)]$ be $[f(x), g(x)]$ if $|x| \leqq \epsilon$, and $(0, 0)$ if $|x| > \epsilon$. Since $f - f_\epsilon$ and $g - g_\epsilon$ are integrable $[\mu + \nu]$, and also $\int (f - f_\epsilon)\, d\mu = \int (g - g_\epsilon)\, d\nu$, we could have used f_ϵ for f and g_ϵ for g without altering the equality or inequality of α and β. If we break the process into the two independent parts, corresponding to the jumps of size greater than ϵ in absolute value and the other part Y_ϵ, we find that $E_\xi(Y_\epsilon) - E_\eta(Y_\epsilon)$ approaches $(\eta - \xi)(\alpha - \beta)$ and the variance of Y_ϵ approaches $\sigma^2(\eta - \xi)$ under both ξ and η.

Consequently, if $\sigma = 0$ and $\alpha \neq \beta$, we may distinguish ξ from η with probability one. If $\alpha \neq \beta$ and $\sigma > 0$, we may obtain the logarithm of the likelihood ratio of the Gaussian component in the usual manner as

$$L_G = \frac{\beta - \alpha}{\sigma^2}\left[X_G(\eta) - X_G(\xi) - \frac{(\xi - \eta)(\alpha + \beta)}{2}\right]. \tag{47}$$

We may combine (37) and (47) to obtain the logarithm of the likelihood ratio, and we see that this is a process with independent increments and characteristic function

$$(48)\quad \begin{aligned} &\frac{1}{t-\xi}\log E\{e^{i\lambda L(t,\xi)}\} = -\frac{(\beta-\alpha)^2}{2\sigma^2}(\lambda^2 + i\lambda) \\ &+ \int \left\{\left[\left(\frac{\phi}{1-\phi}\right)^{i\lambda} - 1\right](1-\phi) - (2\phi - 1)i\lambda\right\} d(\mu+\nu), \qquad t-\xi > 0, \\ &\frac{1}{\xi - t}\log E\{e^{-i\lambda L(t,\xi)}\} = -\frac{(\beta-\alpha)^2}{2\sigma^2}(\lambda^2 - i\lambda) \\ &+ \int \left\{\left[\left(\frac{\phi}{1-\phi}\right)^{i\lambda} - 1\right]\phi - (2\phi - 1)i\lambda\right\} d(\mu+\nu), \qquad t-\xi < 0. \end{aligned}$$

These expressions need the appropriate interpretation if $(\mu + \nu)\{0, 1\} > 0$. The likelihood ratio process of the likelihood ratio process does not yield anything new.

Suppose $(\mu + \nu)\{0, 1\} = 0$. (We can obtain even better results if this is not so.) Then under ξ, we have $E[\exp L(t, \xi)] = 1$, but

$$(49)\quad E[L(t,\xi)] = \begin{cases} -(t-\xi)\left[\dfrac{(\beta-\alpha)^2}{2\sigma^2} - \displaystyle\int (1-\phi)\log\frac{\phi}{1-\phi} - (2\phi - 1)\, d(\mu+\nu)\right], & t-\xi > 0, \\ (t-\xi)\left[\dfrac{(\beta-\alpha)^2}{2\sigma^2} + \displaystyle\int \phi\log\frac{\phi}{1-\phi} - (2\phi - 1)\, d(\mu+\nu)\right], & t-\xi < 0. \end{cases}$$

These integrals might not exist; however, a failure of the integral to exist would make the expectation $-\infty$ in the usual sense. In any case, $E_\xi[L(t, \xi)] < 0$. Then by the strong law of large numbers, $L(t, \xi) < A - B|t - \xi|$ for all t, where A is a finite number with probability one.

Similar results hold for the case of n-dimensional time. A rather interesting result is that the likelihood ratio statistic only depends on the one-dimensional marginal processes of the likelihood ratio process (*not* the original process). Note that the likelihood ratio process will, if truncation might occur, be restricted to a subset of E_n.

It will frequently happen that one is interested in inference with an invariant loss function. In that case there will exist an optimal invariant procedure which consists of assuming a uniform a priori "distribution" on E_n for ξ. This will give an a posteriori distribution which has a density proportional to $\exp L(\xi, \xi_0)$, for any ξ_0 in the range permitted by the 0 and 1 values of the ϕ.

Note that this procedure is not the maximum likelihood procedure, even for large samples. What is the effect of a sample of size N? It is to multiply the parameters and measure of the processes involved by N, and thus to change the scale by N. That is, the estimation procedure is hyperefficient, which we have already seen in the discontinuity problem which leads to the inference problem. In the one-dimensional pure Gaussian case, the ratio of the variances of maximum likelihood to the minimum variance estimate is 3.25, for an efficiency of 1.803 for the best estimate with quadratic loss function.

An interesting observation is that from the computational viewpoint the maximum likelihood estimate may be preferable because of the greatly increased labor to compute the best estimate. This is so because the constant B above may be quite small, causing the numerical computations to be extended over a considerable range, and also because the integration of the likelihood function can be computationally expensive. If the loss of the maximum likelihood estimate is small, the increased efficiency of better estimates may not compensate for the cost of computation.

REFERENCES

[1] H. Chernoff and H. Rubin, "The estimation of the location of a discontinuity in density," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1956, Vol. 1, pp. 19–37.

[2] H. Rubin, "Estimation of the location of the change of rate of a differential process," unpublished technical report, University of Oregon, 1957.

[3] H. Rubin and H. G. Tucker, "Estimating the parameters of a differential process," *Ann. Math. Statist.*, Vol. 30 (1959), pp. 641–658.