The Annals of Statistics
1974, Vol. 2, No. 5, 1029–1033

FISHER INFORMATION AND SPLINE INTERPOLATION[1]

BY PETER J. HUBER

Swiss Federal Institute of Technology, Zürich

It is shown that among all cumulative distribution functions passing through $k \geqq 2$ given points there is a unique one with minimal Fisher information; it is obtained by a curious type of spline interpolation. This answers some questions raised by D. G. Kendall and J. W. Tukey.

Problem. Estimate Fisher information $I(F) = \int (f'/f)^2 f\,dx$ from $k \geqq 2$ given points of the cumulative distribution function F.

This problem, which was posed by J. W. Tukey in [3], has a distinguished "minimal" solution—the smallest possible Fisher information $I(F_0)$ for distributions passing through the given points. The minimizing distribution function F_0 is obtained by a curious and rather non-trivial type of spline interpolation, as follows. At the same time, an existence problem, mentioned in [1], page 33 f., is solved.

Assume that the given values are $F(\xi_i) = t_i$ with $-\infty < \xi_1 < \xi_2 < \cdots < \xi_k < \infty$, $t_1 < t_2$, $t_{k-1} < t_k$. For convenience, we put $\xi_0 = -\infty$, $t_0 = 0$, $\xi_{k+1} = \infty$, $t_{k+1} = 1$. Then the solution F_0 can be described as follows.

(i) $F_0(\xi_i) = t_i$, $i = 0, \cdots, k+1$;

(ii) F_0 is two times continuously differentiable;

(iii) the density $f_0 = F_0'$ is strictly positive, except that it vanishes on those intervals $[\xi_i, \xi_{i+1}]$ for which $t_i = t_{i+1}$;

(iv) on each interval (ξ_i, ξ_{i+1}) the function $f_0^{\frac{1}{2}}{}''/f_0^{\frac{1}{2}}$ is constant $= \lambda_i$, i.e.

$$\begin{aligned}(f_0(x))^{\frac{1}{2}} &= a_i e^{\lambda_i x} + b_i e^{-\lambda_i x}\,, && \text{if} \quad \lambda_i > 0\\ &= a_i x + b_i\,, && \text{if} \quad \lambda_i = 0\\ &= a_i \cos|\lambda_i|x + b_i \sin|\lambda_i|x\,, && \text{if} \quad \lambda_i < 0\,.\end{aligned}$$

(v) There is one and only one F_0 satisfying (i) to (iv); it is the unique F_0 minimizing $I(F)$ subject to $F(\xi_i) = t_i$, $i = 1, \cdots, k$. The value of the minimum is $I(F_0) = -4 \sum (t_{i+1} - t_i)\lambda_i$.

The proof is somewhat involved, but as very similar proofs already occur in [2] it is hardly necessary to present all the details.

Let $\mathscr{F}$ be the set of all monotone functions satisfying

$$0 \leqq F(\xi_i - 0) = F(\xi_i) \leqq t_i \leqq F(\xi_i + 0) \leqq 1\,, \qquad i = 1, \cdots, k\,.$$

Received June 1973; revised November 1973.

[1] Work done at Cornell University, and supported by NSF Grant GP-9297.

AMS 1970 *subject classifications.* Primary 62G05; Secondary 41A15.

Key words and phrases. Fisher information, splines, interpolation, nonparametric estimation.

Define

$$I(F) = \sup \frac{(\int \psi' \, dF)^2}{\int \psi^2 \, dF} \tag{1}$$

where ψ ranges over the continuously differentiable functions with compact support. Note that $I(F)$ is the $L_2(F)$-norm of the functional $\psi \to -\int \psi' \, dF$, and that

$$I(F) = \int (f'/f)^2 f \, dx \tag{2}$$

whenever one of the two sides of (2) is finite (the density $f = F'$ then must be absolutely continuous; see [2]).

LEMMA 1. *There is an $F_0 \in \mathscr{F}$ which minimizes $I(F)$.*

PROOF. As $I(\cdot)$ is the supremum of a family of vaguely continuous functions, it is lower semicontinuous, and hence attains its minimum on the vaguely compact set $\mathscr{F}$.

LEMMA 2. *If F_0 minimizes $I(F)$ and if $f_0 > 0$ except on the intervals $[\xi_i, \xi_{i+1}]$ where $t_i = t_{i+1}$, then F_0 is the unique member of $\mathscr{F}$ minimizing $I(F)$.*

PROOF. Assume that $f_0 > 0$ for all x (the general case is treated analogously). Let $F_1 \in \mathscr{F}$ be any distribution having finite Fisher information; put

$$F_\varepsilon = (1 - \varepsilon)F_0 + \varepsilon F_1 , \qquad 0 \leqq \varepsilon \leqq 1 .$$

We may assume without loss of generality that $f_1 > 0$ (otherwise replace F_1 by, say, $(F_0 + F_1)/2$, which has this property). As $I(\cdot)$ is convex (see [2]), monotone convergence and Fatou's lemma yield, respectively,

$$\frac{d}{d\varepsilon} I(F_\varepsilon) = \int \left[2 \frac{f_\varepsilon'}{f_\varepsilon} (f_1' - f_0') - \left(\frac{f_\varepsilon'}{f_\varepsilon} \right)^2 (f_1 - f_0) \right] dx \tag{3}$$

and

$$\frac{d^2}{d\varepsilon^2} I(F_\varepsilon) \geqq 2 \int \left(\frac{f_1'}{f_1} - \frac{f_0'}{f_0} \right)^2 \frac{f_0^2 f_1^2}{f_\varepsilon^3} \, dx \tag{4}$$

(primes always denote differentiation with respect to x).

If also F_1 minimizes Fisher information, then, by convexity, $I(F_\varepsilon) = I(F_0)$ for all $\varepsilon \in [0, 1]$ and (4) implies

$$\frac{f_1'}{f_1} = \frac{f_0'}{f_0} \qquad \text{a.e.} \tag{5}$$

We integrate this and obtain

$$f_1 = c f_0 \tag{6}$$

for some constant c (it was overlooked in [2], that this is false unless $f_0 > 0$). Since

$$I(F_0) = I(F_1) = \int (f_1'/f_1)^2 f_1 \, dx = \int (f_0'/f_0)^2 c f_0 \, dx = c I(F_0) ,$$

we must have $c = 1$.

Now consider the following auxiliary problems.

Problem A. Let $-\infty < \xi < \eta < \infty$ and assume that $F(\xi)$, $F(\eta)$, $f(\xi)$, $f(\eta)$ are given and fixed. The problem is to minimize the convex functional

$$I_{\xi\eta}(F) = \int_\xi^\eta (f'/f)^2 f\, dx \,. \tag{7}$$

If this problem has a solution F_0, then (3) implies that

$$\left[\frac{d}{d\varepsilon} I_{\xi\eta}(F_\varepsilon)\right]_{\varepsilon=0} = \int_\xi^\eta \left\{2\frac{f_0'}{f_0}(f_1' - f_0') - \left(\frac{f_0'}{f_0}\right)^2 (f_1 - f_0)\right\} dx \geqq 0 \tag{8}$$

for all F_1 with finite $I_{\xi\eta}(F_1)$ and satisfying the side conditions. Conversely, if (8) holds for all these F_1, then F_0 minimizes (7).

If $f_0 > 0$ on $[\xi, \eta]$, then (8) can be integrated by parts to give

$$\begin{aligned}\left[\frac{d}{d\varepsilon} I_{\xi\eta}(F_\varepsilon)\right]_{\varepsilon=0} &= -\int_\xi^\eta \left\{2\left(\frac{f_0'}{f_0}\right)' + \left(\frac{f_0'}{f_0}\right)^2\right\}(f_1 - f_0)\, dx \\ &= -4\int_\xi^\eta \frac{f_0^{\frac{1}{2}\prime\prime}}{f_0^{\frac{1}{2}}}(f_1 - f_0)\, dx \,.\end{aligned} \tag{9}$$

Provided $f_0 > 0$ on $[\xi, \eta]$, it follows that F_0 minimizes (7) iff f_0 satisfies the differential equation

$$\frac{f_0^{\frac{1}{2}\prime\prime}}{f_0^{\frac{1}{2}}} = \lambda = \text{const. on} \quad [\xi, \eta] \,. \tag{10}$$

It is fairly easy to see that the side conditions can always be met with a solution of (10) which is strictly positive in (ξ, η). If $f(\xi) = f(\eta) = 0$, this is trivial to show. Assume therefore that $f(\xi)$ and $f(\eta)$ are not both 0. For $\lambda > 0$, the solution of (10) is

$$(f_0(x))^{\frac{1}{2}} = ae^{\lambda x} + be^{-\lambda x} \tag{11}$$

with a, b not both being negative. If $ab \geqq 0$, then $f_0(x) > 0$ for all x; if $ab < 0$, then $f_0(x)$ is monotone and hence > 0 in (ξ, η). For each value of λ in the range $(-\pi/(\eta - \xi), \infty)$ there is a unique positive solution to (10) which takes the given values of f_0 at ξ and η, and as λ decreases from ∞ to $-\pi/(\eta - \xi)$, this solution increases from 0 to ∞, hence it is also possible to obtain the given value of $F_0(\eta) - F_0(\xi)$.

Alternatively, one can also show that a solution of (10) which has an isolated zero in (ξ, η) cannot correspond to a minimum of (7); see below (end of Problem B).

Problem B. Let $-\infty < \xi < \eta < \zeta < \infty$ and assume that $F(\xi)$, $F(\eta)$, $F(\zeta)$, $f(\xi)$ and $f(\zeta)$ are given and fixed. The problem is to minimize $I_{\xi\zeta}(\cdot)$. It is evident from (2) that f_0 must be continuous at η, and it must satisfy (10) in each of the intervals (ξ, η), (η, ζ). I assert now that f_0' must be continous at η.

Proof. Assume first that $f_0(\eta) > 0$. Then a discontinuity in f_0' corresponds to a Dirac δ in $f_0^{\frac{1}{2}\prime\prime}/f_0^{\frac{1}{2}}$ at η, and it is easy to verify that by choosing f_1 such that $f_1 - f_0$ is symmetric around η and nonzero at η it is possible to achieve a strictly negative value for (9). If $f_0(\eta) = 0$, then a glance at the solutions of (10) shows that $f_0'(\eta) = 0$ and f_0' is trivially continuous.

Furthermore, $f_0(\eta) = 0$ can happen only when either $F(\xi) = F(\eta)$ or $F(\eta) = F(\zeta)$. The proof is based on the following idea: if $f_0(\eta) = 0$ and f_0 satisfies (10), then $f_0'(x)/f_0(x) \sim 2/(x - \eta)$ near η, and the integral (8) diverges to $-\infty$ if f_1 is smooth and nonzero at η. This allows to show that F_0 then does not correspond to a minimum.

Problem C. Minimization of $I_{-\infty,\xi_1}(\cdot)$. The unbounded intervals $(-\infty, \xi_1)$ and (ξ_k, ∞) need a special treatment; it suffices to consider one of them. To exclude trivialities, assume $0 < t_1 < t_2$ (here $k \geqq 2$ is essential!). If $f_0(\xi_1) = 0$, then F_0 does not correspond to a minimum (see the preceding paragraph); if $f_0(\xi_1) > 0$, the density must be of the form

$$f_0(x) = a^2 e^{-2\lambda|x|} \tag{12}$$

on $(-\infty, \xi_1)$, with $\lambda > 0$. Hence (9) reads

$$\left[\frac{d}{d\varepsilon} I_{-\infty,\xi_1}(F_\varepsilon)\right]_{\varepsilon=0} = -4\lambda \int_{-\infty}^{\xi_1} (f_1 - f_0)\,dx\,. \tag{13}$$

In order that (13) is always $\geqq 0$, we must make the total mass $F_0(\xi_1) - F_0(-\infty)$ as large as possible, that is $F_0(-\infty) = 0$, and similarly $F_0(+\infty) = 1$; hence F_0 is a genuine probability distribution.

We are now ready to collect and put together these many pieces of evidence. Take a distribution F_0 minimizing $I(\cdot)$; its existence is asserted in Lemma 1. The auxiliary problems A, B, C show that F_0 satisfies (10) in each of the intervals (ξ_i, ξ_{i+1}) and that it satisfies the assumptions of Lemma 2; hence F_0 is the unique distribution minimizing $I(\cdot)$. Moreover, it satisfies (i) to (iv). On the other hand, any F_0 satisfying (i) to (iv) has the property that

$$\left[\frac{d}{d\varepsilon} I(F_\varepsilon)\right]_{\varepsilon=0} = -4 \int \frac{f_0^{\frac{1}{2}''}}{f_0^{\frac{1}{2}}} (f_1 - f_0)\,dx = 0\,;$$

hence $I(\cdot)$ is stationary at F_0; since $I(\cdot)$ is convex, F_0 corresponds to a minimum. This terminates the proof of the main assertion of this paper.

When the grid $(\xi_1, \cdots, \xi_k)$ is refined, then $I(F_0)$ converges to the true value $I(F)$. This follows at once from the remark that $I(\cdot)$ is lower semicontinuous and that F_0 converges weakly to F, thus $\liminf I(F_0) \geqq I(F)$, but $I(F_0) \leqq I(F)$. I have not yet been able to find the rate of convergence.

REFERENCES

[1] Boneva, L. I., Kendall, D. and Stefanov, I. (1971). Spline transformations: Three new diagnostic aids for the statistical data-analyst (with discussion). *J. Roy. Statist. Soc. Ser. B* 33 1–70.

[2] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

[3] TUKEY, J. W. (1972). On calculating intrinsic accuracies from % points. (Mimeographed manuscript).

DEPARTMENT OF MATHEMATICS
E.T.H.
CLAUSIUSSTR 55
8006 ZURICH, SWITZERLAND