On the Uniform Consistency of Bayes Estimates for Multinomial Probabilities

by

P Diaconis[1]
Mathematics Department
Harvard University
Cambridge, Mass 02138

and

D A Freedman[2]
Statistics Department
University of California
Berkeley, Calif 94720

Technical Report No 137
May 30, 1988

[1] Research partially supported by NSF Grant DMS 86-00235
[2] Research partially supported by NSF Grant DMS 86-01634

Department of Statistics
University of California
Berkeley, California

On the Uniform Consistency of Bayes Estimates for Multinomial Probabilities

by

P Diaconis[1]
Mathematics Department
Harvard University
Cambridge, Mass 02138

and

D Freedman[2]
Statistics Department
University of California
Berkeley, Calif 94720

Technical Report No 137
May 30, 1988

Abstract

A k-sided die is thrown n times, to estimate the probabilities $\theta_1,\ldots,\theta_k$ of landing on the various sides. The MLE for θ is the vector of empirical proportions $p=(p_1,\ldots,p_k)$. Consider a set of Bayesians who put uniformly positive prior mass on all reasonable subsets of the parameter space. Their posterior distributions will be uniformly concentrated near p. Sharp bounds are given, using entropy. These bounds apply to all sample sequences: there are no exceptional null sets.

---

[1] Research partially supported by NSF Grant DMS 86-00235
[2] Research partially supported by NSF Grant DMS 86-01634

1. Introduction

This paper is about the consistency of Bayes estimates. The usual statement is that for almost all sample sequences, as the sample size goes to infinity the posterior distribution piles up near the true value of the parameter. The objective is to reformulate this idea as a finite-sample result, without exceptional null sets or "true values" of parameters.

We begin with coin tossing, and develop an explicit inequality which shows that the posterior must concentrate near the observed fraction of heads. The inequality replaces the asymptotics and eliminates the null set; the observed fraction stands in for the true parameter.

To be a little more specific, suppose there are j heads in n tosses of a coin. Consider the posterior odds ratio for a parameter interval of fixed length centered at j/n. The posterior odds are bounded below by ab^n, where $a>0$ and $b>1$ are computable constants. So the odds go to infinity at an exponential rate.

If the prior assigns measure 0 to an interval, so will the posterior. Even if the prior assigns small positive mass to the interval, it may take a long time for the data to swamp the prior. The inequality must therefore take into account the degree to which the prior covers the parameter space.

The notion of "ϕ-positivity" is introduced, to measure coverage; ϕ is a positive function on $(0,1)$. A prior μ is said to be ϕ-positive if μ assigns mass $\phi(h)$ or more to every closed interval of length h in $[0,1]$. For example, if $\phi(h)= .1h$, then μ is ϕ-positive if and only if μ is bounded below by .1 x Lebesgue measure, setwise. Priors with densities which have zeros-- like betas-- can be handled using more complicated ϕ's; so can singular priors.

The inequality on the posterior odds ratio holds uniformly in ϕ-positive priors μ, and uniformly in the fraction j/n of heads. For any parameter interval $(j/n-h, j/n+h)$, the posterior odds ratio is bounded below by

$$\psi(h,\epsilon)\, e^{n(1-\epsilon)g(h)} \tag{1.1}$$

Here, $\epsilon>0$ is a nuisance of rigor; $\psi(h,\epsilon)>0$ is computed from ϕ and does not otherwise depend on the prior; $g(h)>0$ does not depend on the data or the prior.

The rest of this paper is organized as follows: section 2 gives a careful statement of the result for coin tossing; section 3 has a heuristic proof and section 4 the rigor. The extension to the multinomial is in sections 5-6, and the last section discusses the idea of ϕ-positivity.

History. In effect, we will estimate the posterior using the method of LaPlace (1774); he showed that the posterior piles up near the MLE, but only for the uniform prior. (An easy modern proof uses Chebychev's inequality, but that was not available to LaPlace.) Some modern references on the consistency of Bayes estimates include LeCam (1953), LeCam and Schwartz (1960), Schwartz (1965), Freedman (1963), Diaconis and Freedman (1986). Edwards-Lindman-Savage (1963) must be cited too; their idea was that the data eventually swamps a non-dogmatic prior-- the principle of stable estimation (pp201-8).

A closely related development is the asymptotic normality of the posterior, which is often called the Bernstein-von Mises theorem-- although LaPlace got there first; references include Johnson (1967, 1970), Ghosh-Sinha-Joshi (1982), LeCam (1986, secs 12.3, 12.4, 17.7).

2. The theorem for coin tossing

Let ϕ be a positive function on $(0,1)$. A prior probability μ on $[0,1]$ is "ϕ-positive" if $\mu[p,p+h]\geq\phi(h)$ for all p and h with $0\leq p<p+h\leq 1$.

Let H be the relative entropy function:

$$(2.1) \qquad H(p,\theta) = -p \log \theta - (1-p) \log (1-\theta)$$

Here, $p=j/n$ is the relative frequency of heads, and θ is the parameter-- the probability of heads. (The prior is a distribution over θ.) As is well known,

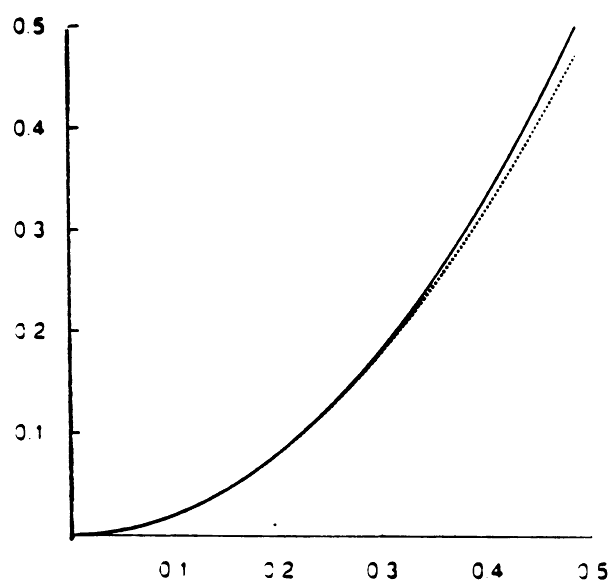
(2.2) $H(p,\bullet)$ is strictly convex, with a strict minimum at p.

For $0<h<{}^1/_2$, let

$$(2.3) \qquad g(h) = \inf_{p,\theta}\{H(p,\theta)-H(p,p): |\theta-p|\geq h\}$$

We will show later that $g(h)>0$, and the inf is attained. Clearly, g is monotone increasing. Its graph is shown in Figure 1; for details, see (5.12-19) below. Although g is defined on $(0,{}^1/_2)$, most of our results are proved only for $(0,{}^1/_4)$.

Figure 1. A graph of g(h), which is convex and increasing; $g(h) > 2h^2$, which is plotted for reference as a dashed line. The two curves are rather close.



To state the main result, suppose a coin is tossed n times, and $p=j/n$ is the fraction of heads. Let $0<h<1/2$. Let $R(n,p,h)$ be the posterior odds ratio for the inside of the parameter interval $(p-h,p+h)$ versus the outside, with respect to a ϕ-positive prior: the outside of the parameter interval is nonempty, because $h<1/2$. Let $0<\epsilon<1$. There is a $\psi(h,\epsilon)>0$, which depends on ϕ, h, and ϵ but not on n or p, such that the following inequality holds.

(2.4) **Theorem.** $R(n,p,h) \geq \psi(h,\epsilon)\, e^{n(1-\epsilon)g(h)}$ for $0<h<1/4$

The first factor on the right does not depend on the data. It depends on the prior only through ϕ; it depends on h and ϵ. The second factor depends on h and ϵ too; but it depends on the data only through the sample size n. In particular, p is not involved on the right. The bound grows exponentially fast as $n\to\infty$. As it turns out, $\psi(h,\epsilon)$ is the minimal prior mass in an interval of length about ϵh^2: more rigorously, $\psi(h,\epsilon)=\phi(h^*)$, where $h^*=\min\{1/2\,\epsilon g(h),h\}$.

The unattainable ideal version of the theorem has $\psi(h,\epsilon)$ replaced by $\phi(h)$, and $\epsilon=0$ in the exponent. On the log scale, these blemishes vanish, as the corollary shows.

(2.5) Corollary. $\liminf_{n\to\infty} \inf_{p,\mu} \frac{1}{n} \log R(n,p,h) \geq g(h)$

In (2.5), the prior μ is restricted to be ϕ-positive; $0<h<1/4$;

and $g(h)$ is best possible.

As will be seen, $g(h)>2h^2$; so (2.5) implies that for suitable

$\psi(h)>0$, depending only on ϕ,

(2.6) Corollary. $R(n,p,h)\geq\psi(h)e^{2nh^2}$ for all n, all $p\in[0,1]$,

all $h\in(0,1/4)$, and all ϕ-positive priors μ.

3. Heuristics

Entropy comes into the argument when you compute the posterior odds ratio:

$$R(n,p,h) = \frac{\int_{(p-h,p+h)} \theta^{pn}(1-\theta)^{(1-p)n}\ \mu(d\theta)}{\int_{[0,p-h]\cup[p+h,1]} \theta^{pn}(1-\theta)^{(1-p)n}\ \mu(d\theta)}$$

The integrand is $e^{-nH(p,\theta)}$.

The numerator must be bounded from below, and the denominator from above. The signs may cause a little confusion: for example, the numerator is large when $H(p,\bullet)$ is small, that is, close to its minimum $H(p,p)$.

To bound the numerator, let h^* be small and positive.
For θ within h^* of p, an argument by uniform continuity will show that $H(p,\theta)$ is within $\epsilon g(h)$ of the minimal value $H(p,p)$. And the μ-measure of these θ's is at least $\psi(h,\epsilon)$. So the numerator is at least $e^{-n[H(p,p)+\epsilon g(h)]}\ \psi(h,\epsilon)$.

For the denominator: $H(p,\theta)$ is at its minimum in θ when $\theta=p$. so the worst θ in the denominator is $\theta=p+h$ or $\theta=p-h$. That suggests trying to minimize $H(p,p+h)-H(p,p)$ and $H(p,p-h)-H(p,p)$, leading to the study of $g(h)$. In the end, convexity arguments show that the denominator is at most $e^{-n[H(p,p)+g(h)]}$.

Now

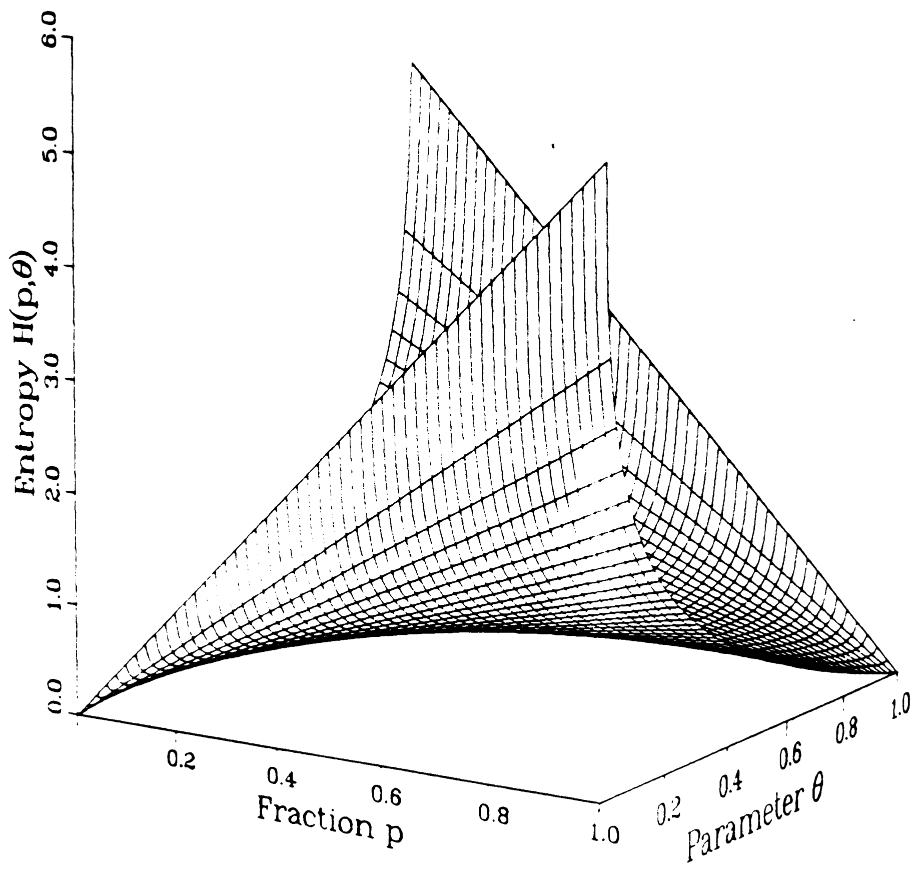
$$R(n,p,h) \geq \frac{e^{-n[H(p,p)+\epsilon g(h)]} \quad \psi(h,\epsilon)}{e^{-n[H(p,p)+g(h)]}}$$

$$= \quad \psi(h,\epsilon) \; e^{n(1-\epsilon)g(h)}$$

That is Theorem (2.4).

The entropy function $H(p,\theta)$ is naturally defined on the closed unit square, except for the corners $p=\theta=0$ and $p=\theta=1$. We set it to 0 there, but lim $H(p,\theta)$ is undefined as (p,θ) converges to either of those corners; although $H(p,p)$ tends to 0 as p tends to 0 or 1. (The discontinuity is a tedious point of difficulty in the background.) By comparison, it is safe to set $H(p,\theta)=\infty$ if $p<1$ and $\theta=1$, or $p>0$ and $\theta=0$, since ∞ is the limiting value on those edges of the unit square. A fishnet plot of $H(p,\theta)$ is shown in Figure 2.

Figure 2. Fishnet plot of $H(p,\theta)$. The singularities at (0,0) and (1,1) are not visible. $H(\bullet,\theta)$ is linear; $H(p,\bullet)$ is convex; $p \to H(p,p)$ is concave; $p \to H(p,p+h)$ is neither convex nor concave.



4. Proofs for the coin.

Let $0<h<{}^1/_4$. Recall $g(h)$ from (2.3). Confirm that

$$(4.1) \qquad g(h) = \min_p \{H(p,p+h)-H(p,p):\ 0\le p<1-h\} = \min_p \{H(p,p-h)-H(p,p):\ h<p\le 1\} > 0$$

For example, $p\to H(p,p+h)-H(p,p)$ is continuous on $(0,1-h)$; positive by (2.2); tends to $\log {}^1/_{1-h}>0$ as $p\to 0^+$; tends to ∞ as $p\to(1-h)^-$. And $H(1-p,1-\theta)=H(p,\theta)$. For details, see the next section.

The ψ-function and the proof of Theorem (2.4)

Assume without real loss of generality that $0\le p\le {}^1/_2$. Keep $h\in(0,{}^1/_4)$. Clearly,

$$(4.2) \qquad \frac{\partial}{\partial\theta} H(p,\theta) = -\frac{p}{\theta} + \frac{1-p}{1-\theta}.$$

If $p<\theta<p+h$, the first term on the right in (4.2) is trapped in $(-1,0)$. The second term is positive, and at most $(1-p)/(1-p-h)$; this bound increases with p, to a maximum of $({}^1/_2)/({}^1/_2-h)$: the latter is at most 2. This is where we use the condition that $h<{}^1/_4$. Thus

$$(4.3) \qquad \left|\frac{\partial}{\partial\theta} H(p,\theta)\right| < 2, \quad \text{provided } p<\theta<p+h,\ 0\le p\le {}^1/_2,\ 0<h<{}^1/_4$$

Fix $\epsilon>0$. Let $h^*=\min\{{}^1/_2\epsilon g(h),h\}$. Let $\psi(h,\epsilon)=\phi(h^*)$, a positive lower bound on the prior μ-mass in $(p,p+h^*)$. By (4.3), $p<\theta<p+h^*$ entails $H(p,\theta)<H(p,p)+\epsilon g(h)$. The odds ratio for the inside of $(p-h,p+h)$ vs the outside can now be estimated, as follows.

$$R(n,p,h) = \frac{\int_{(p-h,p+h)} e^{-nH(p,\theta)}\mu(d\theta)}{\int_{[0,p-h]\cup[p+h,1]} e^{-nH(p,\theta)}\mu(d\theta)}$$

Since $h^* \leq h$, the numerator is bounded below by

$$\int_{(p,p+h^*)} e^{-nH(p,\theta)} \mu(d\theta) \geq e^{-n[H(p,p)+\epsilon g(h)]} \mu(p,p+h^*)$$

$$\geq e^{-n[H(p,p)+\epsilon g(h)]} \psi(h,\epsilon)$$

The denominator is bounded above by $e^{-n[H(p,p)+g(h)]}$. For example, suppose $p+h \leq \theta \leq 1$. Then $H(p,\theta) \geq H(p,p+h)$ by (2.2); and $H(p,p+h) \geq H(p,p)+g(h)$ because $g(h)$ is the worst-case entropy differential. (Since $p<1$, $H(p,1)=\infty$; and the inequalities are trivial if $p+h<1$ but $\theta=1$; or if $p+h=\theta=1$.) In consequence, and this completes the proof of the theorem,

$$R(n,p,h) \geq \psi(h,\epsilon)\, e^{-n(1-\epsilon)g(h)} \qquad \square$$

Proof of (2.5). The inequality is immediate from (2.4). To see that $g(h)$ is best possible, fix h. For now, fix j and n too. Abbreviate $p=j/n$. We must bound $R(n,p,h)$ from above. As (2.2) shows, the numerator is bounded above by $e^{-nH(p,p)}$.

Let $\delta>0$. The denominator is bounded below by the integral over $[p+h,p+h+\delta]$. For θ in that interval, $H(p,\theta)$ is at most $H(p+h+\delta)$, by (2.2). So the denominator is at least

$$\mu(p+h,p+h+\delta)\bullet e^{-n[H(p,p+h+\delta)]} .$$

If $p+h+\delta<1$, then,

$$\frac{1}{n}\log R(n,p,h) \leq O(\frac{1}{n}) + H(p,p+h+\delta) - H(p,p)$$

To complete the argument, let $n\to\infty$; let $p=j/n$ tend to a point where $H(p,p+h)-H(p,p)$ takes its minimum value $g(h)$; and let $\delta\to 0$. (Eventually $p+h<1$, because $H(p,1-)=\infty$; for details on the minimization, see the next section.) □

5. The function g(h).

The outline of this section is as follows. The entropy differential $D_{+h}(p)=H(p,p+h)-H(p,p)$ is introduced, and its derivative computed in terms of the auxiliary "left hand" and "right hand" functions L_h and R_h. The point p_h where D_{+h} achieves its minimum is computed by solving $L_h(p)=R_h(p)$. The g(h) of (2.3) is the worst-case entropy differential, $D_{+h}(p_h)$; some properties of g(h) are developed.

Entropy differentials

Let $h\in(0,{}^1/_2)$, so $h<1-h$. For $0\leq p<1-h$, let

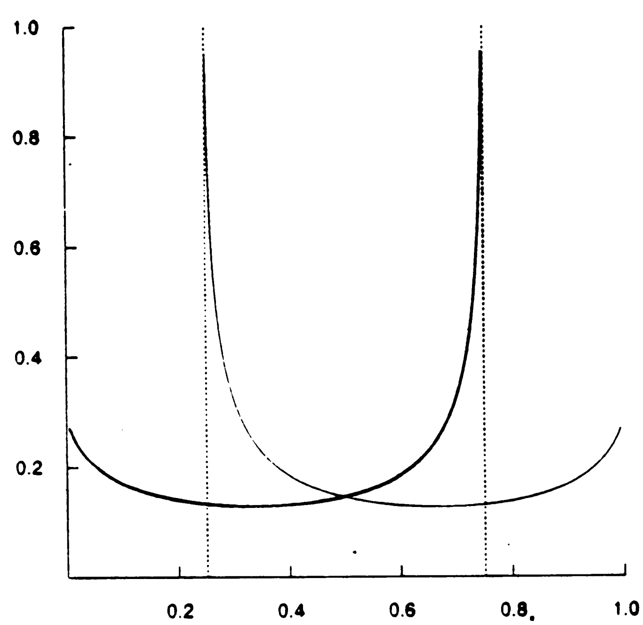
$$D_{+h}(p)=H(p,p+h)-H(p,p)$$

For $h<p\leq 1$, let

$$D_{-h}(p)=H(p,p-h)-H(p,p)$$

These are the "entropy differentials," which are to be bounded below. They are plotted in Figure 3.

Figure 3. The entropy differentials are convex functions of p, and $D_{+h}<D_{-h}$ on $(h,{}^1/_2)$. The figure shows $h={}^1/_4$. The heavier curve is D_{+h}; the ligher one, D_{-h}. Vertical lines are shown at h and 1-h.



The main object of this section is to give some description of g and the entropy differentials, using convexity arguments; and the next lemma is basic to the development.

(5.1) **Lemma.** Let $p>0$ and let δ be real, with $p+\delta>0$ and $\delta\neq 0$. Let $\Phi(p)=p\log[p/(p+\delta)]$. Then Φ is strictly convex.

Proof. First, suppose $\delta<0$. Then $p>|\delta|$ and

$$\Phi(p) = -p\log\left(1-\frac{|\delta|}{p}\right) = \sum_{j=1}^{\infty} \frac{1}{j}|\delta|^{j}\frac{1}{p^{j-1}}$$

Each term is convex.

Next, suppose $\delta>0$. Let $u=p+\delta>\delta$. Now

$$\Phi(p) = (u-\delta)\log\left(1-\frac{\delta}{u}\right)$$

$$= -(u-\delta)\sum_{j=1}^{\infty} \frac{1}{j}\delta^{j}\frac{1}{u^{j}}$$

$$= -\delta + \sum_{j=1}^{\infty}\left\{\frac{1}{j}-\frac{1}{j+1}\right\}\delta^{j+1}\frac{1}{u^{j}}$$

Again, each term is convex. □

(5.2) **Lemma.**

a) $D_{-h}(p)=D_{+h}(1-p)$.

b) $D_{-h}(^1/_2)=D_{+h}(^1/_2)$.

c) $D_{+h}(0+)<\infty$, but $D_{+h}(1-h-)=\infty$.

d) $D_{-h}(1-)<\infty$, but $D_{-h}(h+)=\infty$.

e) The entropy differentials are strictly convex functions of p.

f) $D_{-h}(p)-D_{+h}(p)$ is strictly decreasing as p increases from h to 1-h.

g) $D_{-h}(p)>D_{+h}(p)$ for $h<p<^1/_2$.

Proof. Claims a)-d) are clear.

Claim e) follows from (5.1).

Claim f). Clearly,

$$\begin{aligned} D_{-h}(p)-D_{+h}(p) &= H(p,p-h)-H(p,p+h) \\ &= p\log\frac{p+h}{p-h} - (1-p)\log\frac{1-p+h}{1-p-h} \\ &= p\log\left[\frac{1+\frac{h}{p}}{1-\frac{h}{p}}\right] - (1-p)\log\left[\frac{1+\frac{h}{1-p}}{1-\frac{h}{1-p}}\right] \\ &= 2\sum_{j=0}^{\infty} h^{2j+1}\left\{\frac{1}{p^{2j}} - \frac{1}{(1-p)^{2j}}\right\} \end{aligned}$$

This is legitimate because $0<h/p<1$ and $0<h/(1-p)<1$. Finally,

$$p \to \left\{\frac{1}{p^{2j}} - \frac{1}{(1-p)^{2j}}\right\}$$

is monotone decreasing as p increases, at least for $j>0$.

Claim g) is immediate from b) and f). □

Next, we have to differentiate the entropy differentials, and it is enough to consider $D_{+h}(\bullet)$. For $0<p<1-h$ let

(5.3) $$L_h(p) = \log(p+h) - \log p + \log(1-p) - \log(1-p-h)$$

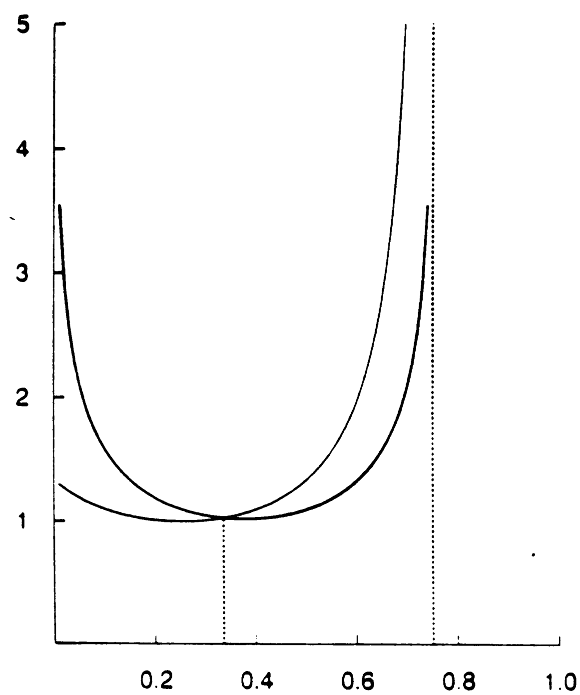
(5.4) $$R_h(p) = \frac{1-p}{1-p-h} - \frac{p}{p+h} = \frac{h}{(p+h)(1-p-h)}$$

As is easily seen,

(5.5) $$\frac{\partial}{\partial p} D_{+h}(p) = -L_h(p) + R_h(p)$$

To minimize D_{+h}, we have to solve $L_h(p) = R_h(p)$; hence the "L" and "R", for "left" and "right" hand sides of the equation. Some facts about D, L and R are developed as lemmas: see Figure 4.

Figure 4. The functions $L_h(\bullet)$ and $R_h(\bullet)$ are convex. Their graphs cross at p_h. The figure shows $h=1/4$ and $p_h=.3345$. The heavier curve is L_h; the lighter one, R_h. A vertical line is shown at $1-h$.



(5.6) **Lemma.** Fix α with $0 \leq \alpha \leq 1$. Abbreviate $\beta = 1 - \alpha$.

a) $$L_h(1/2 - \alpha h) = 4h + 2\sum_{j=1}^{\infty} \frac{\alpha^{2j+1} + \beta^{2j+1}}{2j+1} (2h)^{2j+1}$$

b) $$R_h(1/2 - \alpha h) = 4h + 2\sum_{j=1}^{\infty} \beta^{2j} (2h)^{2j+1}$$

Proof. As is easily verified,

$$L_h(1/2 - \alpha h) = \log(1+2\beta h) - \log(1-2\beta h) + \log(1+2\alpha h) - \log(1-2\alpha h)$$

$$R_h(1/2 - \alpha h) = \frac{4h}{1 - 4\beta^2 h^2}$$ □

(5.7) **Lemma.**

a) $R_h(0+) < \infty = L_h(0+)$.

b) $R_h(1/2) > L_h(1/2)$.

c) $$\left. \frac{\partial}{\partial p} D_{+h}(p) \right|_{p=1/2} > 0$$

d) $$\left. \frac{\partial}{\partial p} D_{-h}(p) \right|_{p=1/2} < 0$$

e) $D_{+h}(\bullet)$ is decreasing near 0.

f) $D_{-h}(\bullet)$ is increasing near 1.

Proof. Claim a) is easy.

Claim b) follows from (5.6), with $\alpha = 0$.

Claim c) is immediate from b), and d) is symmetric.

Claim e) follows from a), and f) is symmetric. □

Let p_h be the unique point where $D_{+h}(\bullet)$ attains its minimum.

(5.8) **Lemma.**

a) $L_h(\bullet)$ and $R_h(\bullet)$ are strictly convex.

b) $R_h(1-h-\delta)/L_h(1-h-\delta)\to\infty$ as $\delta\to 0^+$.

c) $L_h(p)-R_h(p)$ is strictly decreasing as p increases from 0 to 1-h.

d) p_h is the unique solution to $L_h(p)=R_h(p)$.

e) L_h is minimum at $^1/_2-^1/_2h$; and R_h, at $^1/_2-h$.

f) $^1/_2-h < p_h < ^1/_2-^1/_2h$.

Proof. Claim a). As is easily seen,

$$R_h(p) = h\left\{\frac{1}{p+h} + \frac{1}{1-p-h}\right\}$$

Each term is convex. Another easy calculation shows

$$\frac{\partial^2}{\partial p^2}L_h(p) = \frac{1}{p^2} - \frac{1}{(p+h)^2} + \frac{1}{(1-p-h)^2} - \frac{1}{(1-p)^2} > 0$$

Claim b) is easy.

Claim c). $R_h(p)-L_h(p)=\frac{\partial}{\partial p}D_{+h}(p)$, which increases because D_{+h} is convex.

Claim d) follows.

Claim e) is easy.

Claim f). This is equivalent to the pair of inequalities

$$L_h(^1/_2-h) > R_h(^1/_2-h)$$

$$L_h(^1/_2-^1/_2h) < R_h(^1/_2-^1/_2h)$$

which follow from (5.6). □

Recall that D_{+h} takes its minimum at p_h. We now make some remarks about the function $h \to p_h$. Abbreviate

$$p' = \frac{\partial}{\partial h} p_h .$$

(5.9) **Remark.** $p'<0$ for $0<h<{}^1/_2$

Proof. The implicit function theorem and (5.8d) imply that p_h is a smooth function of h. Take $p=p_h$ in (5.3-4) and differentiate with respect to h; equate the results to see

$$\frac{p'+1}{p+h} - \frac{p'}{p} - \frac{p'}{1-p} + \frac{p'+1}{1-p-h} = \frac{(p+h)(1-p-h) - h(p'+1)(1-2p-2h)}{[(p+h)(1-p-h)]^2}$$

This simplifies to

$$p'\left\{ \frac{1}{p+h} - \frac{1}{p} - \frac{1}{1-p} + \frac{1}{1-p-h} \right\} = \frac{h(p'+1)(2p-1+2h)}{[(p+h)(1-p-h)]^2}$$

and then

$$(5.10) \qquad p'\frac{2p-1+h}{p(1-p)} = (p'+1)\frac{2p-1+2h}{(p+h)(1-p-h)}$$

Now $2p-1+h<0$ because $p_h<{}^1/_2-{}^1/_2 h$ while $2p-1+2h>0$ because $p_h>{}^1/_2-h$; see (5.8f). If $p'+1>0$ then $p'<0$ by (5.10); on the other hand, if $p'+1\le 0$ then $p'<0$. $\square$

A graph of p_h is shown in Figure 5. Numerical calculations suggest that p_h is a convex function of h, but practically linear on $(0,{}^1/_2)$:

$$ {}^1/_2 - {}^2/_3 h < p_h < {}^1/_2 - .64h $$

Some of this can be proved; also see (5.8f).

(5.11) **Remarks.**

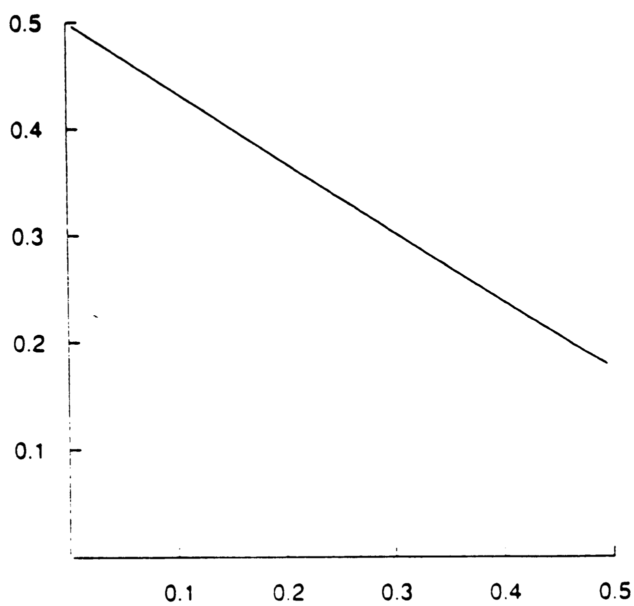
a) $p_h > {}^1/_2 - {}^2/_3 h$ for $0<h<{}^1/_2$, by (5.6).

b) $p_h < {}^1/_2 - ({}^2/_3-\epsilon)h$ for $0<h<\delta$, again by (5.6).

c) The derivative of p_h at 0 is ${}^2/_3$.

d) $p' \to {}^2/_3$ as $h \to 0^+$, by (5.10) and c).

Figure 5. A graph of $h \to p_h$; this function is virtually a straight line on $(0,{}^1/_2)$.



The construction of g(h) and its properties

Let g(h) be the worst-case entropy differential:

$$g(h) = D_{+h}(p_h)$$
$$= H(p_h, p_h+h) - H(p_h, p_h)$$
$$= H(1-p_h, 1-p_h-h) - H(1-p_h, 1-p_h)$$

Numerical calculations suggest that g(h) is convex, that $g(h)/(2h^2)$ is convex, and the latter is practically constant, increasing from 1 at h=0 to 1.07 at $h={}^1/_2$. Some of this can be proved, and will be discussed now.

(5.12) **Remarks.**

a) $g(h) \approx 2h^2$ as $h \to 0^+$. Indeed,

$$H(p,p+h) - H(p,p) = \frac{1}{2}\,\frac{h^2}{p(1-p)} + O(h^3)$$

as $h \to 0^+$, uniformly in p bounded away from 0 and 1;

recall from (5.11) that $p_h \to {}^1/_2$ as $h \to 0^+$.

b) g(h) is monotone. Indeed,

$$\text{(5.13)} \qquad \frac{\partial}{\partial h} D_{+h}(p) = \frac{h}{(p+h)(1-p-h)} > 0$$

Since the whole curve $D_{+h}(\bullet)$ shifts up when h increases, so does its minimum value, which is g(h).

(5.14) Proposition. Let $0<p<1$ and $0<h<1-p$. Then $D_{+h}(p)>2h^2$. Indeed, there is equality at $h=0$; the difference is strictly increasing (but not convex).

Proof. Equation (5.13) shows

$$(5.15) \qquad \frac{\partial}{\partial h} D_{+h}(p) = \frac{h}{(p+h)(1-p-h)} \geq 4h$$

because $(p+h)(1-p-h)\leq{}^1/_4$, with equality only at $p+h={}^1/_2$. This proves strict monotonicity. $\square$

(5.16) Corollary. Let $0<h<{}^1/_2$. Then $g(h)>2h^2$.

Proof. $g(h) = \inf\,\{D_{+h}(p)\colon 0<p<{}^1/_2\}$. If $0<h<{}^1/_2$ and $0<p<{}^1/_2$ then $0<h<1-p$, and (5.14) applies. $\square$

(5.17) Corollary. Let $0<h<{}^1/_2$. Then $g(h)-2h^2$ is strictly increasing in h.

Proof. Use (5.15). $\square$

(5.18) Remark. $h\to D_{+h}(p)/2h^2$ is not increasing in h. For example, with $p=.3$, the min occurs near $h=.27$; the function decreases from 0 to .27, then increases.

(5.19) Proposition. Let $0<h<{}^1/_2$. Then $g(h)<3h^2$.

Proof. Choose $p={}^1/_2-h$. Then $g(h)<D_{+h}(p)$, so

$$\begin{aligned} g(h) &< ({}^1/_2-h)\log(1-2h) + ({}^1/_2+h)\log(1+2h) \\ &= \sum_{j=1}^{\infty}\left[\frac{1}{2j-1} - \frac{1}{2j}\right](2h)^{2j} \\ &= 2h^2 + 4h^2\left\{\sum_{j=2}^{\infty}\left[\frac{1}{2j-1} - \frac{1}{2j}\right](2h)^{2(j-1)}\right\} \end{aligned}$$

But $2h<1$, so $g(h)/2h^2$ is bounded above by

$$1 + 2\left\{\sum_{j=2}^{\infty}\left[\frac{1}{2j-1} - \frac{1}{2j}\right]\right\} < 1 + 2\left\{\sum_{j=2}^{4}\left[\frac{1}{2j-1} - \frac{1}{2j}\right] + \frac{1}{9}\right\} < 1.5 \quad \square$$

If $0<h<{}^1/_4$, a similar argument starting from $p={}^1/_2$ gives a much tighter upper bound, $g(h)<2.1h^2$.

6. The theorem for the multinomial.

Let S_k be the simplex of all k-vectors θ with non-negative coordinates θ_i adding to 1. Consider a die with k sides, labelled 1,...,k. In n tosses, the relative frequencies with which these sides land form a vector $p=(p_1,...,p_k)$ in S_k.

(6.1) For $0<h<1/k$, let $N_k(h,p)$ be the polyhedral neighborhood of p consisting of the $\theta \in S_k$ with $|\theta_i - p_i|<h$ for all i.

Plainly, $N_k(h,p)$ is the sphere around p of radius h-- in the sup norm.

(6.2) Let $H_k(p,\theta)$ be the relative entropy:

$$H_k(p,\theta) = -\sum_{i=1}^{k} p_i \log \theta_i$$

This can be defined everywhere by the convention $0\times\infty=0$, but the limit of $H_k(p,\theta)$ is not well defined if eg p_1 and θ_1 both tend to zero.

The first main result in this section, to be proved later, shows that the minimum entropy differentials do not depend on the dimension k. Recall g from (2.3).

(6.3) **Proposition.**

$$\inf_{p\in S_k,\ \theta\notin N_k(h,p)} [H_k(p,\theta) - H_k(p,p)] = g(h)$$

A prior μ on the simplex S_k is "ϕ-positive" if it is uniformly positive on a certain class of sub-simplexes. More specifically, let v be a vector of length k, whose entries are either +1 or -1.

(6.4) Definition. Let $T_k(v)$ be the set of k-vectors $x=(x_1,...,x_k)$ satisfying:

$$\sum_{i=1}^{k} x_i = 0 \quad \text{and} \quad 0 \le v_i x_i \le 1 \text{ for } i=1,...,k$$

For $p \in S_k$ and $0<h<{}^1/_k$ let

$$T_k(p,h,v) = p+hT_k(v) = \{p+hx: x \in T_k(v)\}$$

Say μ is ϕ-positive iff $\mu\{T_k(p,h,v)\}>\phi(h)$ whenever $T_k(p,h,v) \subset S_k$.

To state the next result, suppose the k-sided die is tossed n times, and p is the vector of empirical frequencies. Let $0<h<{}^1/_k$. Let $R(n,p,h)$ be the posterior odds ratio for the inside of $N_k(h,p)$ versus the outside, with respect to a ϕ-positive prior: the outside is nonempty, because $h<{}^1/_k$. Let $\epsilon>0$. There is a $\psi(h,\epsilon)>0$, which depends on ϕ, h, and ϵ but not on n or p, such that the following inequality holds.

(6.5) **Theorem.** $R(n,p,h) \ge \psi(h,\epsilon)\, e^{n(1-\epsilon)g(h)}$ for $0<h<{}^1/_{2k}$.

(6.6) Corollary. $\liminf_{n\to\infty} \inf_{p,\mu} \frac{1}{n} \log R(n,p,h) \ge g(h)$

In (6.6), the prior μ is restricted to be ϕ-positive, $0<h<{}^1/_{2k}$, and $g(h)$ is best possible.

7. Proofs for the multinomial.

The proof of (6.3)

Suppose $k \geq 3$. Since the entropy function (6.2) is convex in θ with its minimum at p, the infimum outside the convex polyhedron $N_k(h,p)$ is attained on the boundary. Consider eg the intersection of the boundary with

$$F = \{\theta:\ \theta \in S_k \text{ and } \theta_k = p_k + h\}$$

Assume for the sake of argument that this face is nonempty, so $p_k + h \leq 1$. Consider

$$\inf_{\theta \in F} H_k(p,\theta) - H_k(p,p) \tag{7.1}$$

Now

$$\begin{aligned} H_k(p,\theta) &= -\sum_{i=1}^{k} p_i \log \theta_i \\ &= -\sum_{i=1}^{k-1} p_i \log \theta_i - p_k \log(p_k+h) \\ &= -\sum_{i=1}^{k-1} p_i \log(\theta_i/1-p_k-h) \\ &\qquad - p_k \log(p_k+h) - (1-p_k)\log(1-p_k-h) \end{aligned}$$

The last sum in the display can be written as $(1-p_k)H_{k-1}(\tilde{p},\tilde{\theta})$ where $\tilde{p}_i = p_i/1-p_k$ and $\tilde{\theta}_i = \theta_i/1-\theta_k$ for $i=1,\ldots,k-1$. So $\tilde{p},\ \tilde{\theta} \in S_{k-1}$. Now $(1-p_k)H_{k-1}(\tilde{p},\tilde{\theta})$ is minimized in $\tilde{\theta}$ at $\tilde{\theta}=\tilde{p}$, and the value of the minimum is

$$-\sum_{i=1}^{k-1} p_i \log(p_i/1-p_k) = -\sum_{i=1}^{k-1} p_i \log p_i + (1-p_k)\log(1-p_k)$$

We pause to confirm that our minimum is on the boundary of $N_k(h,p)$: by construction, it is only on the hyperplane $\theta_k = p_k + h$. Switching back to the original coordinate system, the minimum is at θ with

$$\theta_i = \frac{(1-p_k-h)}{1-p_k} p_i \quad \text{for } i=1,\ldots,k-1 \quad \text{and} \quad \theta_k = p_k + h$$

Clearly, the θ_i are non-negative and sum to 1. For $i<k$, θ_i falls below p_i by the amount

$$\frac{p_i}{1-p_k} h < h.$$

Coming back to the main line of argument, the infimum in (7.1) is obtained by subtracting $H_k(p,p)$, and equals

$$-p_k \log(p_k+h) - (1-p_k)\log(1-p_k-h) + p_k \log p_k + (1-p_k)\log(1-p_k)$$

which can be recognized as $D_{+h}(p_k)$. The latter is minimized when $p_k = p_h$, and the minimum value is $g(h)$. This completes the proof of (6.3). □

(7.2) Remark. If h is near 0, then p_h is near $^1/_2$. So the $p \in S_k$ with the worst entropy differentials have one or two coordinates near $^1/_2$. By renumbering, suppose $p_1 \leq \ldots \leq p_k$. There are two possibilities for the worst-case p's:

i) $p_k = 1-p_h > {}^1/_2$ and $p_1, \ldots, p_{k-1}$ are free;

ii) $p_k = p_h < {}^1/_2$ and $p_1, \ldots, p_{k-1}$ are free.

Case i) includes eg the possibility that $p_k = 1-p_h$ and $p_{k-1} = p_h$ and $p_1 = \ldots = p_{k-2} = 0$.

Recall the definition (6.4) of the simplex $T_k(v)$. The proof of the next lemma is omitted as standard.

(7.3) Lemma. The extreme points of $T_k(v)$ consist of all k-vectors x which sum to zero, with $x_i = v_i$ or $x_i = 0$ for all i.

The proof of (6.5)

Suppose by renumbering the sides that $p_1 \leq \ldots \leq p_k$. Let $v_1 = \ldots = v_{k-1} = 1$ and $v_k = -1$. We work on the simplex $T_k(p,h,v)$, which is wholly in the interior of S_k, because $p_k \geq {}^1/_k > h$. Indeed, the simplex has k extreme points by (7.3):

$$\begin{array}{llllll}
p_1 & p_2 & p_3 & \ldots\ldots & p_{k-1} & p_k \\
p_1+h & p_2 & p_3 & \ldots\ldots & p_{k-1} & p_k-h \\
p_1 & p_2+h & p_3 & \ldots\ldots & p_{k-1} & p_k-h \\
p_1 & p_2 & p_3+h & \ldots & p_{k-1} & p_k-h \\
\cdot & & & & & \\
\cdot & & & & & \\
\cdot & & & & & \\
p_1 & p_2 & p_3 & \ldots\ldots & p_{k-1}+h & p_k-h
\end{array}$$

And each extreme point is in S_k.

The rest of the argument is as for the coin. Indeed, as a function of $\theta \varepsilon T_k(p,h,v)$, $H_k(p,\theta)$ still has Lipschitz constant 2. For the proof, set $p_k = 1 - p_1 - \ldots - p_{k-1}$ and $\theta_k = 1 - \theta_1 - \ldots - \theta_{k-1}$; then differentiate with respect to θ_i for $i<k$:

$$\frac{\partial}{\partial \theta_i} H(p,\theta) = -\frac{p_i}{\theta_i} + \frac{p_k}{\theta_k}$$

The first term is trapped in $[-1,0]$, because $\theta_i \geq p_i$ for $\theta \varepsilon T_k(p,h,v)$. The second term is at most 2; indeed, $\theta_k \geq p_k - h$, so the second term is bounded above by $p_k/(p_k-h)$: but $p_k \geq {}^1/_k$ and $h < {}^1/_{2k}$. This completes the Lipschitz estimate.

To estimate the odds ratio, bound the numerator below by integrating over $T_k(p,h,v)$, using the Lipschitz estimate. Bound the denominator above using (6.3).

$\square$

8. Some facts about ϕ-positivity

This section has some remarks and examples on the idea of ϕ-positivity; we hope to explore the theory more systematically in the future. Recall that ϕ is a positive function on $(0,1)$; and the prior μ is ϕ-positive iff it assigns mass $\phi(h)$ or more to every closed interval of length h in $[0,1]$.

(8.1) Remark. If $\phi(h) > ah$ for all h, and μ is ϕ-positive, then μ is bounded setwise below by a times Lebesgue measure.

It is natural to conjecture that a ϕ-positive class of measures is bounded below (setwise) by a positive measure, but that turns out to be wrong; ϕ-positivity is a more general idea.

(8.2) Example. There is a ϕ-positive class of probability measures $M = \{\mu\}$ on $[0,1]$ such that if α is a measure and $\alpha \leq \mu$ setwise for all $\mu \in M$, then $\alpha = 0$.

Construction. The class M will be countable. Let λ be Lebesgue measure on $[0,1]$. Let λ_n assign mass $1/n+1$ to each of $0/n$, $1/n$, $2/n$, ..., n/n. Let

$$\mu_n = \frac{n+1}{n+2}\lambda_n + \frac{1}{n+2}\lambda$$

Let $R=\{r\}$ be the rationals in $[0,1]$, and Q the irrationals. If $\alpha\le\mu_n$, then $\alpha\{r\}\le 1/n+2$ and $\alpha(Q)\le 1/n+2$, so in the end $\alpha\{r\}=0$, $\alpha(R)=0$, and $\alpha(Q)=0$.

We claim that $\{\mu_n\}$ is ϕ-positive, with $\phi(h)=h^2/4$. To verify this, consider the interval $[x,x+h]$. Suppose $\frac{a-1}{n} < x \le \frac{a}{n}$ and $\frac{b}{n} \le x+h < \frac{b+1}{n}$. Clearly, $\frac{b-a}{n} \ge h - \frac{2}{n}$; so $b-a \ge nh-2$. So, there are at least $b-a+1$ rationals of order n in $[x,x+h]$, and

$$\lambda_n[x,x+h] \ge \frac{nh-1}{n+1}$$

Now

$$\begin{aligned}\mu_n[x,x+h] &\ge \frac{nh-1}{n+2}\\ &\ge \frac{1}{2}h - \frac{1}{n+2}\\ &\ge \frac{1}{4}h \quad \text{if } n+2 \ge \frac{4}{h}\\ &\ge \frac{1}{4}h^2\end{aligned}$$

If $n+2 < \frac{4}{h}$, a lower bound on $\mu_n[x,x+h]$ is still $\frac{1}{4}h^2$, from the λ-term only. In fact, $\phi(h)$ is of order h^2, as one sees by taking n of order $1/h$. $\square$

There is a connection with monotone rearrangements (Hardy-Littlewood-Polya, 1934).

(8.3) Remark. Let ϕ be convex, with derivative f, and $\phi(1)=1$. So f is monotone nondecreasing, and its integral is 1. All rearrangements of f are ϕ-positive. Some rearrangements have bigger (and nonconvex) ϕ's; for such a ϕ, all rearrangements of its density will no longer be ϕ-positive. If $\phi(h)=ah^2$, the rearrangements can be bounded below only be a trivial measure.

We endow the space of probabilities on [0,1] with the weak-star topology, which is compact and metrizable.

(8.4) Remark Let M be a ϕ-positive class. Then the closed convex hull of M is ϕ-positive too.

If M consists of one prior, or finitely many priors, then M is ϕ-positive; the next result is a small generalization.

(8.5) Remark Let M be a closed, convex class of probabilities on [0,1]. Suppose that each element of M assigns positive mass to every open interval. Then M is ϕ-positive.

Proof. Fix h with $0<h<1$. Let $0\leq x\leq 1-h$. Let the continuous function f_x on [0,1] vanish to the left of x and to the right of x+h; let $f_x=1$ at $x+\frac{1}{2}h$; complete f_x by linear interpolation. Now $\mu(f_x)$ is a continuous positive function of $\mu \in M$ and x; so it has a positive minimum: $\phi(h)$ can be defined as this minimum, over μ and x. □

Let M_ϕ be the class of ϕ-positive μ. When is M_ϕ nonempty? When is ϕ the exact inf, that is, $\phi(h) = \inf\{\mu[x,x+h]: \mu \in M_\phi$ and $0\leq x<x+h\leq 1\}$? What are the extreme points of M_ϕ? At this point, we only have some scattered remarks as partial answers.

(8.6) Example. Let $\phi(h)=h/10$, for $0<h<1$. One compact convex class M of ϕ-positive μ is the set of μ of the form

$$.1*\text{Lebesgue} + .9*\nu,$$

where ν is any probability. The extreme points have $\nu=\delta_x$. This class is maximal, by a standard extension argument off intervals. There seem to be two other compact convex ϕ-positive classes M, which are minimal: take $\nu=\delta_0$ or δ_1. To get intermediate classes, mix over any compact set of δ_x's containing x = 0 or 1.

(8.7) Example. Let $\phi(h)={}^{1}/_{2}h$ for $h<{}^{2}/_{3}$ and $\phi(h)=2h$ for ${}^{2}/_{3}<h<1$. The extreme points of the class of ϕ-positive μ seem to be as follows:

$$ {}^{1}/_{2}\text{ Lebesgue} + {}^{1}/_{2}\ \delta_a \quad \text{with } {}^{1}/_{3}\le a\le {}^{2}/_{3} $$

$$ {}^{1}/_{2}\text{ Lebesgue} + {}^{1}/_{2}\{3a\ \delta_a + \text{density 3 on } ({}^{2}/_{3}+a,1)\} \quad \text{for } a<{}^{1}/_{3} $$

(8.8) Remark. Let $M=\{\mu\}$ be ϕ-positive. Then $\phi(1/n)\le 1/n$, otherwise μ has mass greater than 1. Likewise, if ϕ is the exact inf of M, then $\phi(h/n)\le\phi(h)/n$.

(8.9) Example. $a_n = 2^n\phi(1/2^n)$ can decrease arbitrarily rapidly.

Construction. Let $a_1 < 1/2$, and $a_{n+1} < a_n$. Let μ_n have density equal to a_n on $[0,1/2^n]$ and equal to b_n on $(1/2^n,1]$. So b_n can be computed from a_n, and $b_n > 1$. Let $M = \{\mu_n\}$. We claim that M is ϕ-positive for suitable ϕ; and if ϕ is the exact inf, $\phi(1/2^n) = a_n/2^n$. Indeed, if $m \le n$, then

$$\mu_m[0,1/2^n] = a_m/2^n$$

On the other hand, if $m>n$,

$$\mu_m[0,1/2^n] > a_n/2^n$$

Indeed,

$$\begin{aligned}\mu_m[0,1/2^n] &> b_m(1/2^n - 1/2^m)\\ &> b_m/2^{n+1}\\ &> 1/2^{n+1}\\ &> a_n/2^n\end{aligned}$$

□

References

P Diaconis & D Freedman (1986). On the consistency of Bayes estimates. Ann Statist 14 1-67 (with discussion)

W Edwards, H Lindman & LJ Savage (1963). Bayesian statistical inference for psychological research. Psychol. Rev. 70 193-242

D Freedman (1963). On the asymptotic behavior of Bayes estimates in the discrete case. Ann Math Statist 34 1386-1403

JK Ghosh, BK Sinha, SN Joshi (1982). Expansions for posterior probability and integrated Bayes risk. In S Gupta & JO Berger, Statistical Decision Theory and Related Topics, III, Vol I, pp 403-456. Academic Press, New York.

GH Hardy, JE Littlewood, G Polya (1934). Inequalities. Cambridge University Press.

R Johnson (1967). An asymptotic expansion for posterior distributions. Ann Math Statist 38 1899-1906

R Johnson (1970). Asymptotic expansions associated with posterior distributions. Ann Math Statist 41 851-864

PS LaPlace (1774). Memoire sur la probabilite des causes par les evenements. Memoires de mathematique et de physique presentes a l'academie royale des sciences, par divers savants, & lus dans ses assembles 6. Reprinted in LaPlace's Oeuvres Completes 8 27-65. English translation by S Stigler (1986) Statistical Science 1 359-78

L LeCam (1953). On some asymptotic properties of the maximum likelihood estimates and related Bayes estimates. Univ Calif Publ Statist 1 277-330

L LeCam (1986). Asymptotic Methods in Statistical Decision Theory. Springer, New York

L LeCam & L Schwartz (1960). A necessary and sufficient condition for the existence of consistent estimates. Ann Math Statist 31 140-50.

L Schwartz (1965). On Bayes' procedures. Z Wahr verw Geb 4 10-26

TECHNICAL REPORTS

Statistics Department

University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.

2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosciences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.

3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.

4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11. No. 1, 185-214.

5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.

6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.

7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.

8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.

9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.

10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.

11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.

12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.

14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.

15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.

16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.

17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review.

18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.

19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).

20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting. 1985, Vol. 4, 251-262.

21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.

22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.

23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984.

25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.

26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.

27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.

29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.

30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES. 1985 Vol 3 pp. 1-13.

31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.

32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.

33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.

34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.

35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science. Feb 1986, Vol. 1, No. 1, 3-39.

36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.

37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly. Feb 1986, Vol. 93, No. 2, 123-125.

38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.

39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.

40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.

41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.

42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.

43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?

44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.

45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.

46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.

47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data.

48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.

49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.

50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada. January, 1986.

51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.

53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.

54. BLACKWELL, D. (November 1985). Approximate normality of large products.

55. FREEDMAN, D. A. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics. 12, 101-128.

56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.

57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.

58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.

59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.

60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.

61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.

62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.

64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.

65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.

66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data.

68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).

69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.

72. LEHMANN, E.L. (July 1986). Statistics - an overview.

73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.

74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.

75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.

76. O'SULLIVAN, F. (September 1986). Relative risk estimation.

77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.

78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.

79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks. To appear in Statistical Science.

80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.

81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.

82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.

83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.

85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory. Ann. Inst. Henri Poincaré, 1987, 23, 397-423.

86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.

87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.

88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.

89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.

91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. To appear in the Journal of Applied Probability.

92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.

92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.

93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.

94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.

95. CANCELLED

96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in *Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.*

97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for carcinogenesis. To appear in Environmental Health Perspectives.

98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.

99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.

100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.

101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators. Annals of Statistics, June, 1988.

102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.

103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.

104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer. IEEE Computer Graphics and applications, June, 1988.

105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.

106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.

107. CHENG, C-S. (Aug 1987, revised Oct 1988). Some orthogonal main-effect plans for asymmetrical factorials.

108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.

109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \le k \le n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.

110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals. Annals of Statistics, June, 1988.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987, revised June 1988). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.

113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.

114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.

115. BICKEL, P.J. and RITOV, Y. (Sept. 1987, revised Aug 1988). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (Sept.1987, revised Aug. 1988). Achieving information bounds in non and semiparametric models.

117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.

118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.

119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.

120. DONOHO, D. and LIU, R. (October 1987, revised March 1988). Geometrizing rates of convergence, II.

121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. To appear in *Statistics a Guide to the Unknown.*

122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.

123. DONOHO, D.L., MACGIBBON, B. and LIU, R.C. (Nov.1987, revised July 1988). Minimax risk for hyperrectangles.

124. ALDOUS, D. (November 1987). Stopping times and tightness II.

125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.

126. DALANG, R.C. (December 1987, revised June 1988). Optimal stopping of two-parameter processes on nonstandard probability spaces.

127. Same as No. 133.

128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.

129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Červonenkis classes of index 1.

130. STONE, C.J. (Nov.1987, revised Sept. 1988). Uniform error bounds involving logspline models.

131. Same as No. 140

132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes

133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.

134. DUBINS, L.E. and SCHWARZ, G. (December 1987). A sharp inequality for martingales and stopping-times.

135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.

136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.

137. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities.

137a. DONOHO, D.L. and LIU, R.C. (1987). Geometrizing rates of convergence, I.

138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.

139. BERAN, R. (January 1988). Refining simultaneous confidence sets.

140. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.

141. BRILLINGER, D.R. (January 1988). Two reports on trend analysis: a) An Elementary Trend Analysis of Rio Negro Levels at Manaus, 1903-1985 b) Consistent Detection of a Monotonic Trend Superposed on a Stationary Time Series

142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (February 1988). Randomization in the two-armed bandit problem.

144. DABROWSKA, D.M., DOKSUM, K.A. and SONG, J.K. (February 1988). Graphical comparisons of cumulative hazards for two populations.

145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.

146. BICKEL, P.J. and RITOV, Y. (Feb.1988, revised August 1988). Estimating integrated squared density derivatives.

147. STARK, P.B. (March 1988). Strict bounds and applications.

148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.

149. NOLAN, D. (March 1988). Asymptotics for a multivariate location estimator.

150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.

151. NOLAN, D. (March 1988). Limit theorems for a random convex set.

152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On a theorem of Kuchler and Lauritzen.

153. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the problem of types.

154. DOKSUM, K.A. (May 1988). On the correspondence between models in binary regression analysis and survival analysis.

155. LEHMANN, E.L. (May 1988). Jerzy Neyman, 1894-1981.

156. ALDOUS, D.J. (May 1988). Stein's method in a two-dimensional coverage problem.

157. FAN, J. (June 1988). On the optimal rates of convergence for nonparametric deconvolution problem.

158. DABROWSKA, D. (June 1988). Signed-rank tests for censored matched pairs.

159. BERAN, R.J. and MILLAR, P.W. (June 1988). Multivariate symmetry models.

160. BERAN, R.J. and MILLAR, P.W. (June 1988). Tests of fit for logistic models.

161. BREIMAN, L. and PETERS, S. (June 1988). Comparing automatic bivariate smoothers (A public service enterprise).

162. FAN, J. (June 1988). Optimal global rates of convergence for nonparametric deconvolution problem.

163. DIACONIS, P. and FREEDMAN, D.A. (June 1988). A singular measure which is locally uniform.

164. BICKEL, P.J. and KRIEGER, A.M. (July 1988). Confidence bands for a distribution function using the bootstrap.

165. HESSE, C.H. (July 1988). New methods in the analysis of economic time series I.

166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.

167. BREIMAN, L., STONE, C.J. and KOOPERBERG, C. (August 1988). Confidence bounds for extreme quantiles.

168. LE CAM, L. (August 1988). Maximum likelihood an introduction.

169. BREIMAN, L. (August 1988). Submodel selection and evaluation in regression-The conditional case and little bootstrap.

170. LE CAM, L. (September 1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge.

171. STONE, C.J. (September 1988). Large-sample inference for logspline models.

172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.

173. MILLAR, P.W. (October 1988). Optimal estimation in the non-parametric multiplicative intensity model.

174. YOR, M. (October 1988). Interwinings of Bessel processes.

175. ROJO, J. (October 1988). On the concept of tail-heaviness.

176. ABRAHAMS, D.M. and RIZZARDI, F. (September 1988). BLSS - The Berkeley interactive statistical system: An overview.

177. MILLAR, P.W. (October 1988). Gamma-funnels in the domain of a probability, with statistical implications.

178. DONOHO, D.L. and LIU, R.C. (October 1988). Hardest one-dimensional subfamilies.

179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signals from data missing low frequencies.

180. FREEDMAN, D.A. and PITMAN, J.A. (Nov. 1988). A measure which is singular and uniformly locally uniform.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.