Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition)

by Lucien Le Cam*
University of California, Berkeley

and

Grace L. Yang
University of Maryland, College Park

Technical Report No. 68
July 1986

*Research partially supported by
National Science Foundation Grant DMS-8403239.

Department of Statistics
University of California
Berkeley, California

Introduction. This is a sequel to a paper [1] by R. B. Davies and a paper [4] be the authors. In [1] Davies addresses himself to a question that can be paraphrased as follows: Let $\mathbf{E}_n = \{P_{\theta,n}; \theta \in \Theta_n\}$ be experiments given by measures $P_{\theta,n}$ on σ-fields $\mathbf{A}_n$. Let $\mathbf{F}_n$ be the experiment formed by the restrictions $Q_{\theta,n}$ of the $P_{\theta,n}$ to a sub-field $\mathbf{B}_n \subset \mathbf{A}_n$. Suppose that Θ_n is a subset of a given Euclidean space and that the $\mathbf{E}_n$ satisfy the so called LAN assumptions.

What other conditions are needed to insure that the $\mathbf{F}_n$ also satisfy the LAN assumptions, or at least have log likelihoods approximable by a sum of two random terms, the first linear and the second quadratic in θ?

Davies gives sufficient conditions. They involve an asymptotic normality requirement for the conditional distributions given $\mathbf{B}_n$ of the log likelihoods of the $P_{\theta,n}$.

In [4] we treated cases where the observations $\{x_1, \cdots, x_n\}$ are independent or Markovian and where the passage to subfields is carried out on the individual x_j's. No conditional distributions were involved. However the technique could not yield the result obtained by Davies in the application to branching processes of his Theorem 4.1.

Since the study of approximations for conditional distributions is often delicate we took another look at the situation to see, among other things, whether one could use instead approximations for joint distributions. More precisely we looked for conditions under which one could take limits of joint distributions and then look at conditional distributions in these limits.

It turns out that this can be done if the statistics used are "distinguished" according to a definition of [3], Chapter 7. This property is often, but not always, readily verifiable. It held trivially for the likelihood ratio approximations of [4]. It must hold for the pair (μ_n, τ_n) of Theorem 4.1 of [1] and, as shown in Section 3 below, can be checked without much effort for the branching process case of [1].

Another aspect of the situation is that both [1] and [4] rely heavily on contiguity restrictions. These may be quite appropriate for the restricted experiments $\mathbf{F}_n$ on the subfields, but they cannot be defended easily for the parent experiments $\mathbf{E}_n$. A case in point would be the estimation of variances in the branching processes of [1], as will be shown in Section 3.

The paper is organized as follows. Section 2 contains our main result on passages to subfields in the form of Theorem 2. It says that, under appropriate restrictions on the statistics involved, one can take limits $\{D_{\theta,\infty}; \theta \in \Theta\}$ of families of distributions $\{D_{\theta,n}; \theta \in \Theta\}$ and deduce the asymptotic behavior of the likelihood ratios of the $D_{\theta,n}$ from that of the likelihod ratios for the $D_{\theta,\infty}$. The passage to subfields can also be carried out on the $D_{\theta,\infty}$. It is also shown (Lemma 2) that a requirement similar to that imposed in Davies' Theorem 4.1 on conditional distributions does imply the requirements of our Theorem 2.

The results do not make any reference to asymptotic normality or contiguity. They can be used more generally.

Section 3 returns to the branching processes. It uses arguments similar to those of Duby and Rouault in [2]. However these authors needed to let the size of the initial population increase to infinity. We show briefly that their results remain valid "locally" with initial population size $M_0 = 1$ and that the statistics of [1] are "distinguished" there. It is also shown that an apparently minor departure from the assumptions makes them lose this property. As already mentioned, the inappropriateness of contiguity restrictions on the parent experiments $\mathbf{E}_n$ is indicated.

Further applications of Theorem 2 will await another publication.

2. Stable and distinguished sequences of statistics.

The objects named in the title are described in some detail in [3], chapter 7, pages 99-108. After recalling their definitions we proceed to give a result relative to the possibility of interchanging two limit procedures as described in the introduction.

Let $\mathbf{E} = \{P_\theta; \theta \in \Theta\}$ be an experiment given by probability measures P_θ on a σ-field $\mathbf{A}$ of subsets of a set Ω. Let $\mathbf{X}$ be a compact space. By "statistics with values in $\mathbf{X}$" will be meant a Markov kernel $\omega \to f_\omega$ that assigns to each $\omega \in \Omega$ a probability measure f_ω on $\mathbf{X}$. These probability measures will be identified with positive linear functionals on the space $C(\mathbf{X})$ of continuous real functions on $\mathbf{X}$. This is sufficient for all purposes if the family $\mathbf{E}$ is dominated. Otherwise we shall extend the definition, admitting as statistics the more general "transitions" of [3]. However the main arguments involve only finite sets Θ. Hence no difficulty will occur.

The restriction to compact $\mathbf{X}$ is a matter of convenience. In applications, $\mathbf{X}$ will often be Euclidean and not compact. One can always compactify as follows:: Take a separating algebra Γ of bounded numerical functions on $\mathbf{X}$, imbed $\mathbf{X}$ in the cartesian product $\mathbf{R}^\Gamma$ and take its closure there. (The choice of Γ does matter).

In the rest of this section the objects $\mathbf{X}$ and Θ will be kept fixed. One will consider nets or sequences $\{\mathbf{E}_\nu\}$ of experiments $\mathbf{E}_\nu = \{P_{\theta,\nu}; \theta \in \Theta\}$ and statistics S_ν with values in $\mathbf{X}$. This gives rise to families of distributions $\mathbf{D}_\nu = \{D_{\theta,\nu}; \theta \in \Theta\}$ where $D_{\theta,\nu}$ is the distribution of S_ν under $P_{\theta,\nu}$, that is the image $S_\nu P_{\theta,\nu}$ of $P_{\theta,\nu}$ by the kernel S_ν. The expectation of a $\gamma \in C(\mathbf{X})$ for this distribution will be written $\gamma D_{\theta,\nu}$.

Definition 1. Let $\{\mathbf{D}_\nu\}$ be indexed by the directed set $\{\nu\}$. The $\mathbf{D}_\nu$ converge vaguely to $\mathbf{D}_\infty = \{D_{\theta,\infty}; \theta \in \Theta\}$ if for every $\theta \in \Theta$ and $\gamma \in C(\mathbf{X})$ the expectations $\gamma D_{\theta,\nu}$ converge to $\gamma D_{\theta,\infty}$.

Note that, for the vague convergence, the set of all families of distributions indexed by Θ is a compact space. Thus every directed set of such families will have cluster points. That is why we have assumed compactness for $\mathbf{X}$.

A family $\mathbf{D}_\nu = \{D_{\theta,\nu}; \theta \in \Theta\}$ is also an experiment in its own right. In [3], we introduced a concept of "stability" for nets $\{\mathbf{D}_\nu\}$ in terms of a correspondence between vague convergence as families of distributions and weak convergence as experiments. It is a property that is not very strong in appearance, but is shown in [3] to be equivalent to the following approximation property, to be called finite approximability.

Let $\{s,t\}$ be a two point subset of Θ and let $\mu_\nu = D_{s,\nu} + D_{t,\nu}$. Let f_ν be the Radon-Nikodym density of $D_{t,\nu}$ with respect to μ_ν.

Definition 2. The net $\{\mathbf{D}_\nu\}$ is stable if for every two point set $\{s,t\} \subset \Theta$ and every $\epsilon > 0$, there is a finite subset $\{\gamma_j; j \in J\}$ of $C(\mathbf{X})$ and a $\nu(\epsilon)$ such that $\nu \geq \nu(\epsilon)$ implies

$$\inf_{j\in J} \int |f_\nu - \gamma_i| d\mu_\nu < \epsilon.$$

A formulation involving finite subsets of Θ instead of two point sets in given in [3], Theorem 1, page 101, together with several other equivalent conditions. For the two points sets $\{s,t\}$ another equivalent statement is as follows

Lemma 1. The net $\{D_{s,\nu}, D_{t,\nu}\}$ is stable if and only if for every subnet $\{\nu(\xi)\}$ such that $[D_{s,\nu(\xi)}, D_{t,\nu(\xi)}]$ converges vaguely to a limit $[D_{s,\infty}, D_{t,\infty}]$ one has

$$\liminf \int f_{\nu(\xi)}(1 - f_{\nu(\xi)}] d\mu_{\nu(\xi)} \geq \int f_\infty (1 - f_\infty) d\mu_\infty$$

with $\mu_\infty = D_{s,\infty} + D_{t,\infty}$ and $f_\infty = \dfrac{dD_{t,\infty}}{d\mu_\infty}$.

For a proof, see [3], Chapter 7 and below. (An equivalent statement can also be written in terms of Hellinger affinities $\int \sqrt{f_\nu(1 - f_\nu)} d\mu_\nu$).

To state our next result, it will be convenient to look at a particular pair $[D_{0,\nu}, D_{1,\nu}]$ and consider that the elements of $\mathbf{X}$ are random variables X_ν with possible distributions $D_{i,\nu}$, $i = 0,1$. Each $D_{i,\nu}$ yields a joint distribution for the pair (f_ν, X_ν). Convergence of such joint distributions will be the vague convergence on $C([0,1] \times \mathbf{X})$.

Theorem 1. Let $[D_{0,\nu}, D_{1,\nu}]$ converge vaguely to $[D_{0,\infty}, D_{1,\infty}]$. Let $\mu_\nu = D_{0,\nu} + D_{1,\nu}$ and $f_\nu = dD_{1,\nu}/d\mu_\nu$. The following conditions are all equivalent:

1) $\lim_\nu \inf \int f_\nu (1 - f_\nu) d\mu_\nu \geq \int f_\infty (1 - f_\infty) d\mu_\infty$.
2) For every $\epsilon > 0$ there is a continuous function ϕ such that $0 \leq \phi \leq 1$ and such that $\int |f_\nu - \phi| d\mu_\nu < \epsilon$ for all sufficiently large ν.
3) The joint distribution of (f_ν, X_ν) under $D_{i,\nu}$, $i = 0,1$, converges to that of (f_∞, X_∞) for $D_{i,\infty}$.

If any one of those three conditions hold and if, in addition, the set of discontinuities of f_∞ has μ_∞ measure zero, then $\int |f_\nu - f_\infty| d\mu_\nu$ tends to zero.

Proof. Consider a decision problem in which it is desired to select a point z in the interval $[0,1]$ with a loss function $W_i(z) = (i - z)^2$. A decision procedure is a Markov kernel $K(dz,x)$. For $[D_{0,\nu}, D_{1,\nu}]$ it has a risk sum equal to

$$\int\int [z^2(1 - f_\nu) + (1 - z)^2 f_\nu] K(dz,x) \mu_\nu(dx) = \int f_\nu (1 - f_\nu) d\mu_\nu + \int\int (z - f_\nu)^2 K(dz,x) \mu_\nu(dx).$$

It follows that the minimum risk sum is $r_\nu = \int f_\nu(1 - f_\nu)d\mu_\nu$, attained by the nonrandomized estimate f_ν.

Let $\phi \in C(\mathbf{X})$ be such that $0 \leq \phi \leq 1$. It can be used to estimate z yielding a risk sum

$$\begin{aligned} R_\nu(\phi) &= \int \phi^2 dD_{0,\nu} + \int (1 - \phi)^2 dD_{1,\nu} \\ &= r_\nu + \int |f_\nu - \phi|^2 d\mu_\nu. \end{aligned}$$

Do this also for the limit pair $[D_{0,\infty}, D_{1,\infty}]$ obtaining a risk sum

$$R_\infty(\phi) = r_\infty + \int |f_\infty - \phi|^2 d\mu_\infty.$$

Note that, according to the first expression for $R_\nu(\phi)$ one will have $\lim_\nu R_\nu(\phi) = R_\infty(\phi)$ for $\phi \in C(\mathbf{X})$. Now select a $\phi \in C(\mathbf{X})$, $0 \leq \phi \leq 1$, so that $\int |f_\infty - \phi|^2 d\mu_\infty < \epsilon^2$. One can write

$$\int |f_\nu - \phi|^2 d\mu_\nu = [R_\nu(\phi) - R_\infty(\phi)] - (r_\nu - r_\infty) + \int |f_\infty - \phi|^2 d\mu_\infty.$$

Since $R_\nu(\phi) - R_\infty(\phi) \to 0$ and since $\liminf(r_\nu - r_\infty) \geq 0$, this implies that for ν large one will have $\int |f_\nu - \phi)^2 d\mu_\nu < \epsilon^2$. Thus (1) $\Longrightarrow$ (2). To show that (2) $\Longrightarrow$ (3) it is enough to show that for all $\gamma_1 \in C[0,1]$ and $\gamma_2 \in C(\mathbf{X})$, the expectations $E_{i,\nu}\gamma_1(f_\nu)\gamma_2(X_\nu)$ for $D_{i,\nu}$ converge to the corresponding expectations for $D_{i,\infty}$. However, take a ϕ as done above and consider $E_{i,\nu}[\gamma_1(f_\nu) - \gamma_1(\phi)]\gamma_2(x_\nu)$. This does not exceed

$$||\gamma_2|| \int |\gamma_1(\phi) - \gamma_1(f_\nu)| d\mu_\nu \leq \mu_\nu[|f_\nu - \phi| > \epsilon^{1/2}] + \mathrm{Osc}(\gamma_1, \epsilon)$$

where $\mathrm{Osc}(\gamma_1,\epsilon)$ is the maximum oscillation of γ_1 on sets of diameter $\epsilon^{1/2}$.

By Chebyshev's inequality, the first term does not exceed ϵ for all ν sufficiently large. By taking ϵ sufficiently small, one can make $\mathrm{Osc}(\gamma_1,\epsilon)$ as small as one please. Thus $\int |\gamma_1(f_\nu) - \gamma_1(\phi)| d\mu_\nu$ can be made arbitrarily small. The same applies to $\int |\gamma_1(f_\infty) - \gamma_1(\phi)| d\mu_\infty$ and the conclusion follows by the triangle inequality.

That (3) implies (1) is immediate. It even implies $\lim r_\nu = r_\infty$. Hence the equivalence of (1) (2) and (3).

To obtain the final statement, note that if the set of discontinuities of f_∞ has μ_∞ measure zero, then $\int |f_\infty - \phi| d\mu_\nu$ tends to $\int |f_\infty - \phi| d\mu_\infty < \epsilon$. Thus $\int |f_\nu - f_\infty| d\mu_\nu \leq \int |f_\nu - \phi| d\mu_\nu + \int |\phi - f_\infty| d\mu_\nu$ will eventually be inferior to 2ϵ.

This concludes the proof of the theorem.

Remark. Life would be more pleasant if, under the equivalent conditions (1) (2) (3) of the theorem, the integrals $\int |f_\nu - f_\infty| d\mu_\nu$ would always converge to zero. This is not the case as shown by the following example.

Let ψ be a function defined on [0,1], equal to zero at zero and to $\psi(x) = \sin\left(\frac{c}{x}\right)$ for $x \in (0,1]$. Select c so that $\int_0^1 \sin\left(\frac{c}{x}\right) dx = 0$. This is possible. There is a value close to 2π with that property.

Let δ_0 be the probability measure concentrated at zero and let λ be the Lebesgue measure on [0,1]. Let $\mu_\infty = \delta_0 + \lambda$ and let $f_\infty = \frac{1}{2} + \frac{1}{4}\psi$. For integer values ν, let $\epsilon_\nu = \frac{C}{2\pi\nu}$. Take $\phi_\nu(x) = \frac{1}{2}$ for $0 \leq x \leq \epsilon_\nu$ and let $\phi_\nu(x) = f_\infty(x)$ for $\epsilon_\nu < x \leq 1$. Let $\mu_\nu = \lambda_\nu + \lambda$ where λ_ν is the Lebesgue measure on $[0,\epsilon_\nu]$ multiplied by ϵ_ν^{-1}. Then $\int \phi_\nu d\mu_\nu$ may differ from unity but one can make an f_ν such that $\int f_\nu d\mu_\nu = 1$ by letting $f_\nu = \phi_\nu$ on $[0,\epsilon_\nu]$ and $f_\nu = k_\nu\phi_\nu$ in $(\epsilon_\nu,1]$ for constants k_ν that tend to unity. Then the conditions (1) (2) (3) of the theorem are satisfied. However

$$\int |f_\nu - f_\infty| d\mu_\nu \geq \frac{1}{4}\int_1^\infty \left|\sin\left(\frac{2\pi\nu y}{c}\right)\right| \frac{dy}{y^2}$$

does not tend to zero.

To go further we need an additional definition as follows.

Consider experiments $\mathbf{E}_\nu$ as before with statistics S_ν whose distributions from families $\mathbf{D}_\nu$.

Definition 3. The statistics S_ν are pairwise asymptotically sufficient if for every two point set $\{s,t\} \subset \Theta$ the distance between the binary experiments $\{P_{s,\nu},P_{t,\nu}\}$ and $\{D_{s,\nu},D_{t,\nu}\}$ tends to zero as ν increases indefinitely. The S_ν are called distinguished, for $\mathbf{E}_\nu$, if they are pairwise asymptotically sufficient and if their distributions are stable.

With this we can pass to our main goal which is to obtain relations between pairs of statistics (S_ν,T_ν).

Consider then experiments $\mathbf{E}_\nu = \{P_{\theta,\nu};\ \theta \in \Theta\}$, a pair $(\mathbf{X},\mathbf{Y})$ of compact spaces and pairs (S_ν,T_ν) where S_ν is a statistic with values in $\mathbf{X}$ and T_ν is statistic with values in $\mathbf{Y}$. Both are defined on $\mathbf{E}_\nu$. Under $P_{\theta,\nu}$ the pair (S_ν,T_ν) has a joint distribution $D_{\theta,\nu} = (S_\nu,T_\nu)P_{\theta,\nu}$ considered as a positive linear functional on the space $C(\mathbf{X}\times\mathbf{Y})$. Vague convergence of distributions will be pointwise convergence on $C(\mathbf{X}\times\mathbf{Y})$.

For simplicity we shall consider only the case where Θ is the two point set $\{0,1\}$. Then one can let $m_\nu = P_{0,\nu} + P_{1,\nu}$ and $\mu_\nu = D_{0,\nu} + D_{1,\nu}$. Each joint measure on $\mathbf{X}\times\mathbf{Y}$ has a marginal on $\mathbf{X}$. This will be noted by a "prime" so that for instance the marginal of μ_ν on $\mathbf{X}$ is μ_ν'. The corresponding marginals on $\mathbf{Y}$ will have double primes, so that the $\mathbf{Y}$-marginal of μ_ν is μ_ν''.

With these notations one can introduce likelihood ratios such as

$$f_\nu = \frac{dD_{1,\nu}}{d\mu_\nu},\ f_\nu' = \frac{dD_{1,\nu}'}{d\mu_\nu'},\ f_\nu'' = \frac{dD_{1,\nu}''}{d\mu_\nu''}$$

and $f_\nu^* = \dfrac{dP_{1,\nu}}{dm_\nu}$.

Note that f_ν is a function of the pair (x,y) and that, for instance, f_ν' is a function of x only, conditional expectation of f_ν given x for the measure μ_ν.

If the $D_{i,\nu}$ have vague limits $D_{i,\infty}$, the pair $(D_{0,\infty},D_{1,\infty})$ defines corresponding densities f_∞, f_∞', f_∞''. It will be convenient to say that $D_{i,\infty}$ is the distribution of a pair (X,Y) on $\mathbf{X}\times\mathbf{Y}$.

With this notation one can state the following.

Theorem 2. Assume that $D_{i,\nu}$ converges vaguely to $D_{i,\infty}$. Assume in addition the following:

(A) For the experiments $\mathbf{E}_\nu$ the S_ν are distinguished.

(B) The net of pairs $(D_{0,\nu}'', D_{1,\nu}'')$ is stable.

Then for $(D_{0,\infty},D_{1,\infty})$ the X coordinate is sufficient and the joint distributions of $(f_\nu,f_\nu',f_\nu'',S_\nu,T_\nu)$ for $D_{i,\nu}$ converge to that of

$$(f_\infty,f_\infty',f_\infty'',X,Y) \text{ for } D_{i,\infty}.$$

Proof. The proof follows closely the argument used for Theorem 1. Thus we shall just give a short sketch. A first remark is that one may as well assume that S_ν is sufficient for $\mathbf{E}_\nu$. Indeed let $\bar{f}_\nu = f_\nu' S_\nu$ be the image of f_ν' by S_ν. Let $Q_{1,\nu}$ be the measure whose density with respect to m_ν is $\bar{f}_\nu$. Since S_ν is asymptotically sufficient the inequalities of [3], page 72, show that $\int |f_\nu^* - \bar{f}_\nu| d\mu_\nu \to 0$. Equivalently $\|P_{1,\nu} - \|Q_{1,\nu}\|^{-1} Q_{1,\nu}\| \to 0$. Then replace $P_{0,\nu}$ by $Q_{0,\nu} = m_\nu - \|Q_{1,\nu}\|^{-1} Q_{1,\nu}$. For the pair $(Q_{0,\nu},Q_{1,\nu})$ the statistics S_ν are sufficient and the properties A and B are preserved.

Now, as in the proof of Theorem 1, find functions $\phi' \in C(\mathbf{X})$ and $\phi'' \in C(\mathbf{Y})$ such that $\int |f_\infty' - \phi'| d\mu_\infty' < \epsilon^2$ and $\int |f_\infty'' - \phi''| d\mu_\infty'' < \epsilon^2$. Then $\int |f_\nu' - \phi'| d\mu_\nu' < \epsilon^2$ and $\int |f_\nu'' - \phi''| d\mu_\nu'' < \epsilon^2$ for all sufficiently large ν and one can repeat the argument that yielded property (3) in Theorem 1. This gives the result for the systems $(f_\nu',f_\nu'',S_\nu,T_\nu)$. The extra part f_ν need no special attention since $\int |f_\nu - \bar{f}_\nu| d\mu_\nu \to 0$ for a function $\bar{f}_\nu$ defined on $\mathbf{X}\times\mathbf{Y}$ by writing $\bar{f}_\nu(x,y) = f_\nu'(x)$.

The result as stated follows.

Remark 1. The theorem can be extended without difficulty, except for more complex notation, to finite sets Θ instead of two point sets.

Remark 2. Note well that the result is not only that the $(f_\nu, f_\nu', f_\nu'', S_\nu, T_\nu)$ have limiting distributions but that the distribution are those given by the likelihood ratios in the limits $(D_{0,\infty}, D_{1,\infty})$.

Remark 3. The result does not remain valid under condition (A) alone, even if T_ν is also exactly sufficient. The stability of the pairs $(D_{0,\nu}, D_{1,\nu})$ is not inherited by their marginals $(D_{0,\nu}'', D_{1,\nu}'')$.

An example can be constructed as follows. Let $\mathbf{X}$ be the two point set $\{0,1\}$ and let $\mathbf{Y}$ be the interval $[0,1]$. Let Ω be $[0,1]$ with its Lebesgue measure λ. Let λ_i be the Lebesgue measure on the segment $i \times \mathbf{Y}$ in $\mathbf{X} \times \mathbf{Y}$. Let $h_\nu(\omega)$ be the ν^{th} binary digit of ω in Ω. Consider an experiment $\mathbf{E}_\nu$ where for $\theta = 0$ the distribution of $\omega \in \Omega$ is λ and where for $\theta = 1$ the distribution of ω has a density $2h_\nu$ with respect to λ. Map the experiment on $\mathbf{X} \times \mathbf{Y}$ by the statistic $\omega \rightarrow [h_\nu(\omega), \omega]$. Here our $D_{i,\nu}$ have limits on $\mathbf{X} \times \mathbf{Y}$ in the sense that $\int \gamma dD_{i,\nu} \rightarrow \int \gamma dD_{i,\infty}$ for every bounded measurable function γ on $\mathbf{X} \times \mathbf{Y}$. Each coordinate map is sufficient. Yet both $D_{0,\infty}''$ and $D_{1,\infty}''$ are identical to the Lebesgue measure on $[0,1]$. Condition A is satisfied. Condition B is not. The systems (f_ν, f_ν', f_ν'') have a limiting distribution, but it is not that of $(f_\infty, f_\infty', f_\infty'')$.

A natural example of a similar situation will be encountered in Section 3.

If one can establish that our condition (B) holds, our Theorem 2 can be used to bypass the requirement of convergence of conditional distributions used by Davies' in Theorem 4.1 of [1]. There the role of our S_ν is played by Davies' $\tilde{\Delta}_n$ and the role of T_ν is played by the pair (μ_n, τ_n).

In the set up of Theorem 2 conditional distributions can be used to establish (B) as follows.

Consider the joint distributions $D_{i,\nu}$ as distributions of the pair of coordinates (X_ν, Y_ν) in $\mathbf{X} \times \mathbf{Y}$. Let us say that the conditional distributions of X_ν given Y_ν are <u>finitely approximable</u> if for every $\gamma \in C(\mathbf{X})$ and every $\epsilon > 0$ there is a finite set $\{\psi_j; j \in J\}$ of elements of C(Y) such that, taking conditional expectations for μ_ν one has

$$\inf_{j \in J} E \mid E[\gamma(X_\nu) \mid Y_\nu] - \psi_j(Y_\nu) \mid < \epsilon$$

for all sufficiently large ν.

Lemma 2. Consider the situation described for Theorem 2. Let (A) be satisfied and assume that the conditional distributions of X_ν given Y_ν for μ_ν are finitely approximable. Then condition (B) holds.

Proof. As in Theorem 2, one can assume that X_ν is exactly sufficient. Let then $\bar{f}_\nu(x,y) = f_\nu'(x)$. For μ_ν the conditional expectations satisfy

$$E(\bar{f}_\nu | Y_\nu) = E(f_\nu' | Y_\nu) = f_\nu''(Y_\nu).$$

By assumption (A) and Theorem 1, for each $\epsilon > 0$ there is a $\phi \in C(\mathbf{X})$ such that $\int |f_\nu' - \phi| d\mu_\nu' < \epsilon$ for all large ν this will give

$$E \mid E f_\nu'' (Y_\nu) - E(\phi | Y_\nu)| < \epsilon.$$

By the assumption of the lemma, the $E(\phi|Y_\nu)$ can be approximated by elements of C(Y). Therefore the finite approximability also holds for f_ν'' and the f_ν'' satisfy the condition given in Definition 2. This gives the desired result.

According to this lemma, the conditions (A) (B) of Theorem 2 are implied by those used by Davies in his Theorem 4.1. Note, however, that under the other restrictions imposed there, the type of convergence of conditional distributions required by Davies is a necessary requirement for his conclusion. This is so because of the properties of Laplace transforms defined on open sets. It would not be necessary if one would work with a finite Θ, as done here.

In conclusion, Theorem 2 offers an alternate approach to that of Davies in a fairly general framework. The stability property (B) is often easy to verify (as in the cases treated in [4], where it is entirely obvious). In the branching process example used by Davies in [1], the stability is obvious once one has established the required asymptotic sufficiency. Some additional remarks on this case follow.

3. Some remarks on branching processes

In this section we consider on branching process constructed as follows. There is a probability measure p on the integers $\{1,2,3,...\}$. If the n^{th} generation contains m_n individuals, each one of them, independently of the rest, produces a progeny whose size is taken at random from the distribution p, leading to a total population M_{n+1} for the $(n+1)^{st}$ generation.

If ξ is a variable with distribution p, it will be assumed that $E\xi = \mu > 1$ and that $\sigma^2 = E(\xi - \mu)^2$ is finite and non zero. There is a huge literature on inference for such processes. See for instance the references in [1] and [2]. Thus we shall content ourselves with a few simple remarks.

In [1] Davies assume that p depends on a single parameter $\theta = \mu$ and studies the asymptotic behavior of the experiment f_n in which one observes $M_0 = 1$, $M_1, \cdots, M_n$. This is done through his Theorem 4.1 by introducing a larger experiment $\mathbf{E}_n$ in which one observes not only the total size of the population but also the entire genealogical tree for each individual in it. Note that here the initial population has size $M_0 = 1$. One can also start with a population of size z as done in [2] and let both n and z go to infinity. Naturally, the asymptotics will be different for such a case.

Davies did include a short remark about estimating parameters other than the mean μ. As will be explained below a technique different from that of his Theorem 4.1 becomes necessary.

Technically it should be noted that the results of Davies depend very strongly on the fact that he works in contiguous neighborhoods of a fixed distribution p. By contrast the results of [2] are uniform over a certain class $\mathbf{P}$ of measures p, one of the important features being that it is proved that, asymptotically, the experiment can be parametrized through the pair (μ,σ^2) independently of all other features of p.

As said above the results of [2] depend materially on the fact that the initial population size z goes to infinity. Here are some simple remarks on what happens if M_0 is kept equal to unity.

To begin consider numbers μ, σ^2, b_0 and ϵ_0 with $0 < \sigma^2 \leq b_0 < \infty$ and $\epsilon_0 > 0$ with $\mu > 1$. Let $\mathbf{P}_0$ be the class of all probability measures on the integers $\{1,2,...\}$ that satisfy the following requirements

1) $E\xi = \mu$
2) $E(\xi - \mu)^2 = \sigma^2$

3) $E|\xi - \mu|^3 \leq b_0$

4) $\sup_k p(k)p(k+1) \geq \epsilon_0$.

Lemma 3. Let μ, σ^2, b_0 and ϵ_0 be fixed. Consider an experiment f_ν in which one observes the entire process $M_0 = 1, M_1, M_2, \cdots, M_n \cdots$. Assume that the possibilities for the distributions of progeny are either p_ν or q_ν, both in $\mathbf{P}_0$. Let P_ν and Q_ν be the corresponding distributions for the process.

Then if $||p_\nu - q_\nu|| \to 0$ so does $||P_\nu - Q_\nu||$.

Proof. Consider the behavior of M_{n+1} conditionally given the size M_n of the n^{th} generation. According to local limit theorems, for which see for instance Petrov [5], Chapter 7, pages 187 to 216, there are constants $b < \infty$ and $c \in (0,1)$ such that for any $p \in \mathbf{P}_0$ one has

$$\sum_k \left| P[M_{n+1} = k \mid M_n] - \frac{1}{\sigma\sqrt{2\pi M_n}} \exp\left\{-\frac{1}{2}\frac{(k-\mu M_n)}{\sigma^2 M_n}\right\}\right|$$

$$\leq b\left[\frac{1}{M_n^{1/3}} + M_n^{2/3}c^{M_n}\right].$$

For any $p \in \mathbf{P}_0$ the series $\sum_n M_n^{-1/3}$ and $\sum_n M_n^{2/3}c^{M_n}$ are almost surely convergent, with uniform bounds as p varies in $\mathbf{P}$. Thus, for pairs (p,q) in $\mathbf{P}_0$ the tail of the process $\{M_n\}$ will be uninformative. For our f_ν, if $||p_\nu - q_\nu||$ tends to zero, the distance between the two possible distributions of $\{M_1, \cdots, M_n\}$ also tends to zero for each fixed n. The result follows.

Remark 1. With $M_0 = 1$ if $||p_\nu - q_\nu||$ does not tend to zero the experiment f_ν can retain a great deal of information. For instance it could happen that p_ν and q_ν are disjoint so that the first generation allows one to find out which one of the two is operating.

However, Lemma 3 or a slight modification can still be used to show that the asymptotic parametrizability by μ and σ^2 only, described in Duby and Rouault [2] or Swensen [6], remains, valid "locally", see below.

If one maintains condition (1) (2) and (3) but remove condition (4) of the definition of $\mathbf{P}_0$ the situation can be entirely different. There are pairs (p,q) satisfying (1) (2) and (3) such that $p(k) > 0$ for all $k \geq 1$ but $q(2k+1) \equiv 0$ for all integers k. (For the measure q every birth is a twin birth). What happens then in the process $M_0 = 1, M_1, \cdots, M_n$ may be complex, but what happens in one generation is easy to see.

Consider experiments $\mathbf{G}_n$ where one observes M_{n+1} conditionally given the size $M_n = m_n$ of the n^{th} generation. Assume that the distribution of the progeny is one of two possibilities $f_n = (1-\alpha_n)q + \alpha_n p$ or q, where α_n is some number $\alpha_n \in [0,1]$.

If $m_n\alpha_n \to \infty$, then the (conditional) distribution of M_{n+1} under f_n differs little from what it would be under p, even if $\alpha_n \to 0$. The conditional experiment $\mathbf{G}_n$ does not degenerate, nor does it become perfect. The same is true if $m_n \to \infty$ and $m_n\alpha_n$ tends to a finite non zero limit. However if $m_n \to \infty$ and $m_n\alpha_n \to 0$, the experiment $\mathbf{G}_n$ degenerates.

This suggests, but does not quite prove, that experiments f_n in which one observes $M_0 = 1, M_1, \cdots, M_n$ and where the distribution of the progeny is either f_n or q will become perfect if $\mu^{r_n}\alpha_n \to \infty$ for some sequence r_n such that $n - r_n \to \infty$. This would be so even if $\alpha_n \to 0$ so that $||f_n - q|| \to 0$.

A proof that the f_n do tend to the perfect experiment is easy. Assume $M_n = m_n$ and let ξ_j; $j = 1,2...,m_n$ be the sizes of the individual progenies from M_u. Let $\eta_j = \xi_j$ modulo 2. Then $\Sigma\eta_j$ is a binomial variable with m_n trials and probability of success $\alpha_n\pi$ with $\pi = p\{\xi$ is odd). If $m_n\alpha_n\pi$ is large the probability that $\sum_j \eta_j$ is odd will be close to 1/2. Thus, if $\mu^{r_n}\alpha_n \to \infty$, with $n - r_n \to \infty$, the probability that some M_k, $k \le n$ will be odd will tend to unity under f_n but will be exactly zero under q.

Under such conditions, note that the conditional distribution of $\dfrac{M_{n+1} - \mu M_n}{\sigma\sqrt{M_n}}$ is asymptotically normal $\mathbb{N}(0,1)$ for all f_n. Thus the statistics $(M_{n+1} - \mu m_n)/\sigma\sqrt{m_n}$ cannot be stable for the experiments $\mathbf{G}_n$. This is very much the same thing as the behavior of the Y_ν in the example of Remark 3, following Theorem 2.

In the class $\mathbf{P}_0$ the expectation μ and variance σ^2 of the progeny distribution were kept constant. One can use the local limit theorems to obtain results in which μ and σ^2 vary. To do this, consider fixed numbers $0 < \epsilon_0 < b_0 < \infty$ and a sequence $\{\epsilon_n\}$, $\epsilon_n > 0$ $\epsilon_n \to 0$. Let $\mathbf{P}_1$ be the class of distributions $p = \mathbf{L}(\xi)$ such that

1) $1 + \epsilon_0 \le E\xi \le b_0$
2) $\epsilon_0 \le E(\xi - \mu)^2 \le b_0$
3) $E|\xi - \mu|^3 \le b_0$
4) $\sup_k p(k)p(k+1) \ge \epsilon_0$.

Let $\mathbf{P}_n$ be any subset of $\mathbf{P}_1$ that has a total variation norm diameter at most ϵ_n.

The approximation to likelihood ratios described by Duby and Rouault in [2] is still valid on $\mathbf{P}_n$ if one observes $M_0 = 1, M_2, \cdots, M_n$. That approximation involves only the parameters μ and σ^2 and the expressions

$$\sum_{k=0}^{n-1} \frac{(M_{k+1} - \mu M_k)^2}{\sigma^2 M_k}.$$

We shall not enter into the details, for which see [2], but end on a remark concerning what happens in a class such as $\mathbf{P}_n$ defined here if one keeps μ fixed but vary σ.

Let p_n and q_n be two elements of P_n. With the same expectation μ but respective variances σ_n^2 and τ_n^2. Let h_{k+1}^2 be the conditional square Hellinger distance

$$h_{k+1}^2 = \frac{1}{2}\sum_j \mid [P(M_{k+1} = j \mid M_k,p_n)]^{1/2} - [P(M_{k+1} = j \mid M_k,q_n)]^{1/2}\Big|^2.$$

Up to terms of order $M_k^{-1/3}$ this is approximately $1 - \left[1 - \dfrac{(\sigma_n - \tau_n)^2}{\sigma_n^2 + \tau_n^2}\right]^{1/2}$.

Let P_n be the distribution of $M_0 = 1, M_1, \cdots, M_n$ under p_n. Let Q_n be the corresponding distribution under q_n. It follows as in Lemma 3 that if $n(\sigma_n - \tau_n)^2$ remains bounded then the sequences $\{P_n\}$ $\{Q_n\}$ will be contiguous.

The experiment (P_n,Q_n) behaves asymptotically as if one observed a variable $T_n = \theta\chi_n^2$ where χ_n^2 is a chi-square with n degrees of freedom and θ is allowed to take value σ_n^2 or τ_n^2.

To apply Davies' Theorem 4.1 one would look at the bigger experiment in which the genealogies of the individuals in the n^{th} generation are all observed. This would give other measures, say P_n^* and Q_n^*. To insure contiguity for $\{P_n^*\}$, $\{Q_n^*\}$ one needs to have that $\mu^n(\sigma_n - \tau_n)^2$ remain bounded. Davies' Theorem 4.1 could then be applied, but the reduced experiment (P_n,Q_n) would tend to the trivial uninformative experiment. Thus the note about estimating other parameters at the end of Section 5 of [1] should be given more weight than its short length seems to suggest.

We would also like to attract the reader's attention to the fact that the conditions (4) imposed on our classes $\mathbf{P}_0$ and $\mathbf{P}_1$ are not visible in the formulations of [1] or [6], but an analogous condition is implied.

The apparent lack of need for our condition (4) in the paper cited is due to the fact that the authors work with alternatives q_n such that $||q_n - p|| \to 0$ for a fixed p, independent of n. A similar remark applies to the third moment requirement. As we have seen above a condition such as (4) is needed to obtain uniform results on classes such as $\mathbf{P}_n$. It is also needed for the uniform results of [2].

References

[1] Davies, R. B. "Asymptotic inference when the amount of information is random" *Proc. Berkeley Conference in Honor of J. Neyman and J. Kiefer* Vol II, Wadsworth Inc. (1985) pp. 841-864.

[2] Duby, C. and Rouault, A. "Estimation non paramétrique de l'espérance et de la variance de la loi de reproduction d'un processus de ramification" *Ann. Inst. Henri Poincaré,* Vol. XVIII #2 (1982) pp. 149-163.

[3] Le Cam, L. *Asymptotic methods in statistical decision theory.* Springer Verlag, N. Y. (1986) xxvi + 735 pages.

[4] Le Cam, L. and Yang, G. L. "On the preservation of local asymptotic normality under information loss" Technical report No. 53, Nov 1985 submitted to *Ann. Statist.*

[5] Petrov, V. V. *Sums of independent random variables.* Springer Verlag, N. Y. (1975).

[6] Swensen, A. R. "Asymptotic inference for a class of stochastic processes" PhD. Thesis U.C. Berkeley (1980).

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. To appear in 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. To appear in J. of Forecasting.

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). Proportional hazards, transformation models, partial likelihood, the order bootstrap, and adaptive inference, I.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimates. 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. To appear in Statistical Science.

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. To appear in Amer. Math Monthly.

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 May 1986). 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). On the preservation of local asymptotic normality under information loss.

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

55. FREEDMAN, D. A. (December 1985, revised Feb. 1986). As others see us: A case study in path analysis. Prepared for the Journal of Educational Statistics.

56. LE CAM, L. and YANG, G. L. (January 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies.

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). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986). 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). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 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.

70. LEHMANN, E.L. (July 1986). Statistics - An overview.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). On estimating the number of unseen speices and system reliability.

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.