# Conditional Limit Theorems for Exponential Families and Finite Versions of de Finetti's Theorem

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 91
April, 1988

To appear in the Journal of Theoretical Probability

[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

**Abstract**

Consider an exponential family $P_\lambda$ which is maximal, smooth, and has uniformly bounded standardized fourth moments. Consider a sequence $X_1, X_2, \ldots$ of iid random variables with parameter $\lambda$. Let $Q_{nsk}$ be the law of $X_1, \ldots, X_k$ given that $S_n = X_1 + \ldots + X_n = s$. Choose $\lambda$ so $E_\lambda(X_1) = s/n$. If k and $n \to \infty$ but $k/n \to 0$,

$$||Q_{nsk} - P_\lambda^k|| = \gamma \frac{k}{n} + o(\frac{k}{n})$$

where $\gamma = {}^1/_2 E\{|1-Z^2|\}$ and Z is N(0,1). The error term is uniform in s, the value of $S_n$. Similar results are given for $k/n \to \theta$, and for mixtures of the $P_\lambda^k$. Versions of de Finetti's theorem follow.

**Keywords and phrases**

de Finetti's theorem, mixtures, exponential families, exchangeability.

## 1. Introduction

Let $X_1$, $X_2$, ..., $X_n$ be independent and identically distributed. Let $S_n = X_1 + X_2 + \ldots + X_n$. Under suitable regularity conditions, if k is small relative to n, the variables $X_1, \ldots, X_k$ are to a good approximation conditionally independent given $S_n$, with a common distribution depending on the value of $S_n$. Such theorems have been proved by Lanford (1973), Martin Lof (1970), Stam (1987), Tjur (1974), and Zabell (1980).

Our object is to prove such a theorem with an explicit error bound, uniform in the value of $S_n$; and we allow k to increase with n. Our interest in these refinements will be disclosed below, but first some examples. We work in the variation norm:

$$||P - Q|| = 2 \sup_A |P(A) - Q(A)|$$

Let Z be N(0,1) and define $\gamma$ as follows:

(1.1) $$\gamma = {}^1/_2 E\{|1-Z^2|\}$$

**i) The binomial.** Let $X_i$ be 0 or 1 and independent, with $P\{X_i=1\}=p$. The law of $X_1,\ldots,X_k$ will be denoted $P_p^{\,k}$. Let $Q_{nsk}$ be the law of $X_1,\ldots,X_k$ given $S_n=s$, namely, the law of k draws made at random without replacement from a box of n tickets, where s are marked "1" and the remaining n-s are marked "0". This law does not depend on the parameter p, so $S_n$ is said to be "sufficient."

If $k \to \infty$ but $k/n \to 0$, there is little difference between drawing with or without replacement. More precisely,

$$(1.2) \qquad \|Q_{nsk} - P_{s/n}^{\,k}\| = \gamma \frac{k}{n} + o\left(\frac{k}{n}\right)$$

This explicit rate is uniform in s. The theorem also covers the case where $k=O(1)$, but then the result is a little harder to state.

**ii) The normal.** Let $Z_i$ be independent $N(0,\sigma^2)$ variables. Write $P_\sigma^{\,k}$ for the law of $Z_1,\ldots,Z_k$. Let $X_i=Z_i^{\,2}$ and $S_n=X_1+\ldots+X_n$. Let $Q_{nsk}$ be the law of $Z_1,\ldots,Z_k$ given $S_n=s$. This time, $Q_{nsk}$ can be visualized as the law of the first k coordinates of a point drawn at random from the surface of a sphere of radius $\sqrt{s}$ in $R^n$. Again, $Q_{nsk}$ does not depend on the parameter $\sigma$, and $S_n$ is sufficient. The conditioned limit theorem takes the same form as before: if $k \to \infty$ but $k/n \to 0$, then uniformly in s,

$$(1.3) \qquad \|Q_{nsk} - P_{\sqrt{s}}^{\,k}\| = \gamma \frac{k}{n} + o\left(\frac{k}{n}\right)$$

The asymptotic rate $\gamma k/n$ in (1.3) is exactly the same as in (1.2). This rate also turns up for geometric and exponential variables (Diaconis and Freedman, 1987). The object in the present paper is to state and prove a general theorem, covering many of these special cases. The discrete case is a little easier, so the theorem is given in the absolutely continuous case.

In the rest of this section, we state the theorem carefully; proofs are deferred to sections 3 and 4. Section 2 makes the connection with de Finetti's theorem, and gives a relatively simple proof of a theorem of Kuchler and Lauritzen (1986), characterizing mixtures of exponential families; this proof is self-contained. Examples are given in section 5, including the gamma; other examples show that when the conditions of the theorem are violated (grossly enough), the rate of convergence depends on the value of $S_n$, so the conclusions of the theorem are false.

For the main result, consider an <u>exponential family</u> of probability measures $\{P_\lambda : \lambda \in \Lambda\}$ on the fixed interval I. Assume I=(a,b) is an open interval; a or b may be finite or infinite. Likewise for $\Lambda=(\alpha,\beta)$. Assume that the base measure for $\{P_\lambda\}$ is absolutely continuous, with a nonnegative, locally integrable density h on I. By definition,

$$P_\lambda(dx) = e^{\lambda x} h(x)\, dx / c(\lambda) \tag{1.4}$$

where

$$c(\lambda) = \int_I e^{\lambda x} h(x)\, dx \tag{1.5}$$

is finite for $\lambda \in \Lambda$. Let $X_\lambda$ have distribution $P_\lambda$. Let

$$m_\lambda = E\{X_\lambda\} \text{ and } \sigma_\lambda{}^2 = \text{var } X_\lambda$$

As is well known,

$$\left[\begin{aligned} m_\lambda &= \frac{d}{d\lambda} \log c(\lambda) \\ \sigma_\lambda^2 &= \frac{d^2}{d\lambda^2} \log c(\lambda) \\ \text{the } j^{th} &\text{ cumulant of } X_\lambda \text{ is } \frac{d^j}{d\lambda^j} \log c(\lambda) \end{aligned}\right. \tag{1.6}$$

As (1.6) implies:

(1.7) $m_\lambda$ is strictly increasing with $\lambda$

A standard reference on sufficiency and exponential families is Lehmann (1986, pp18 and 57).

Four regularity conditions will be needed.

(1.8) <u>$\Lambda$ is maximal:</u> as $\lambda\to\alpha$, the mass in $P_\lambda$ concentrates at a+; as $\lambda\to\beta$, the mass concentrates at b−. It follows that $m_\lambda\to$ a or b as $\lambda\to\alpha$ or $\beta$; see (1.16) below for details.

(1.9) <u>Fourth moments:</u> $E\{(X_\lambda - m_\lambda)^4\}/\sigma_\lambda^4$ is uniformly bounded for $\lambda\in\Lambda$.

(1.10) <u>Smoothness:</u> $\sup_{\lambda\in\Lambda}\sup_{|t|>\delta}|\psi_\lambda(t/\sigma_\lambda)| < 1$, where $\psi_\lambda(t)$ is the characteristic function of $P_\lambda$ and $\delta$ is any positive number. In effect, this says that h does not concentrate near a lattice, even after rescaling.

(1.11) <u>Integrability:</u> $\sup_{\lambda\in\Lambda}\int|\psi_\lambda(t/\sigma_\lambda)|^\nu\,dt < \infty$, for some $\nu\geq 1$. This too is a smoothness condition on h.

Let $X_1, X_2, \ldots$ be independent random variables with common distribution $P_\lambda$. Let $S_n = X_1 + \ldots + X_n$. We next define the regular conditional distribution $Q_{nsk}$ for $X_1, \ldots, X_k$ given $S_n = s$. (A reference on rcd's is Freedman, 1983, Appendix A10.)

(1.12) <u>Definition</u>. Let $t = x_1 + \ldots + x_k$. Then $Q_{nsk}$ is for $k<n$ the absolutely continuous distribution on $R^k$ with density

$$h(x_1)\ldots h(x_k)h^{(n-k)}(s-t)/h^{(n)}(s)$$

at $x_1, \ldots, x_k$, provided $s\in nI$ and $s-t\in(n-k)I$ and $0<h^{(n)}(s)<\infty$. If $k=n$, the distribution is singular; the $Q_{nsn}$-law of $X_1, \ldots, X_{n-1}$ is $Q_{nsn-1}$, and $X_n = s-(X_1+\ldots+X_{n-1})$.

In (1.12), $s \in nI$ means $s/n \in I$. Furthermore, $h^{(j)}$ is the j-fold convolution of h with itself. This must be finite: For example, take j=2. Fix any $\lambda \in \Lambda$. Then $e^{\lambda x} h(x)/c(\lambda)$ is an $L^1$ function, whose convolution with itself is another $L^1$ function, namely, $s \to e^{\lambda s}\, h^{(2)}(s)/c(\lambda)$. So $h^{(2)}$ is finite, at least ae. As will be seen in (3.6), for sufficiently large n, the function $h^{(n)}$ is positive everywhere; it will be continuous for $n \geq \nu$ by (1.11). For any n, however, $\{s: 0 < h^{(n)}(s) < \infty\}$ has measure 1 for all $P_\lambda$. It can be shown that $Q_{nsk}$ is a regular conditional distribution for $X_1, \ldots, X_k$ given $S_n = s$, relative to $P_\lambda$, simultaneously for all $\lambda \in \Lambda$.

Recall $\gamma$ from (1.1). Let Z be N(0,1). Define $\phi(\theta)$ as follows, for $0 < \theta < 1$:

(1.13) $$\phi(\theta) = E\{|1 - \sqrt{1-\theta}\, e^{1/2\,\theta Z^2}|\}$$

Let $P_\lambda^k$ denote the k-fold product of $P_\lambda$ with itself.

(1.14) <u>Theorem</u>. Suppose conditions (1.8-11). Let k and $n \to \infty$. Let $s \in nI$. Choose $\lambda \in \Lambda$ so $m_\lambda = s/n$. Thus, $\lambda = \lambda_{ns}$ depends on n and s.

a) If $k/n \to 0$, then uniformly in s,

$$||Q_{nsk} - P_\lambda^k|| = \gamma \frac{k}{n} + o(\frac{k}{n})$$

b) If $k/n \to \theta$ with $0 \leq \theta < 1$, then uniformly in $\theta$ bounded away from 1, and uniformly in s,

$$||Q_{nsk} - P_\lambda^k|| \to \phi(\theta)$$

Part a) of the theorem shows that the conditional law $Q_{nsk}$ of $X_1, \ldots, X_k$ given $S_n = s$ merges in variation distance with $P_\lambda^k$, namely, the law of k independent variables having a common distribution drawn from the given exponential family. The parameter $\lambda$ is chosen to match the means: $m_\lambda = s/n$. That is the usual maximum likelihood estimate (Lehmann, 1986, p16). From another perspective, matching on the means is the Esscher tilting in disguise (Feller, 1971, sec XVI.7; Cover and Csisizar, 19xx).

The rates in the theorem-- $\gamma k/n$ and $\phi(\theta)$-- are the same for all the exponential families which satisfy conditions (1.8-11). The proof of a) works even if $k=O(1)$, and shows

$$(1.15) \qquad ||Q_{nsk} - P_\lambda^k|| = \gamma_k \frac{k}{n} + o\left(\frac{k}{n}\right)$$

where

$$\gamma_k = \frac{1}{2} E\{|1 - Z_k^2|\}$$

and $Z_k$ is the standardized version of $S_k$:

$$Z_k = (S_k - km_\lambda)/\sigma_\lambda \sqrt{k}$$

If $k \to \infty$, then $Z_k \to N(0,1)$ uniformly in $\lambda$ by (1.9), so $\gamma_k \to \gamma$. In any event, $\gamma_k \leq 1$.

The uniform 4th moment condition is stronger than it may appear at first glance: it rules out, for instance, the binomial. The assertion about $m_\lambda$ made in (1.8) is easily checked for finite endpoints; the next remark covers the infinities.

(1.16) <u>Remark</u>. Suppose $I=(-\infty,\infty)$ and $\Lambda=(-\infty,\infty)$ is maximal. Why does $m_\lambda \to \infty$ as $\lambda \to \infty$? By assumption, $P_\lambda$ drifts off to $\infty$, so $\int_0^\infty x\, P_\lambda(dx) \to \infty$. What remains to show is that $\int_{-\infty}^0 x\, P_\lambda(dx) = O(1)$. Fix $\delta>0$. Now $|x|\leq(e^{|x|}-1)$, so

$$\frac{1}{c(\lambda)} \int_{-\infty}^{0} |x|\, e^{\lambda x}\, h(x)\, dx \leq \frac{1}{c(\lambda)} \int_{-\infty}^{0} (e^{|x|}-1)\, e^{\lambda x}\, h(x)\, dx$$

$$= \frac{c(\lambda-1)}{c(\lambda)} P_{\lambda-1}(-\infty,0) - P_\lambda(-\infty,0)$$

This tends to 0 as $\lambda\to\infty$, provided we can bound $c(\lambda-1)/c(\lambda)$. Now $c(\lambda)=c_0(\lambda)+c_1(\lambda)$, where $c_0(\lambda)=\int_{-\infty}^0 e^{\lambda x}h(x)\, dx \to 0$ as $\lambda\to\infty$, while $c_1(\lambda)=\int_0^\infty e^{\lambda x}h(x)\, dx$ is monotone increasing with $\lambda$. Thus, $\limsup_{\lambda\to\infty} c(\lambda-1)/c(\lambda) \leq 1$. Also see Waterman (1971).

The page numbers skip from 8 to..... 11

## 2. de Finetti's theorem

Let $X_1, X_2, \ldots$ be an infinite exchangeable sequence of 0's and 1's (its law P is invariant under finite permutations). Then P is a mixture of coin-tossing processes:

$$(2.1) \qquad P = \int_{[0,1]} P_p^{\infty} \, \mu(dp)$$

Here, $P_p^{\infty}$ makes the X's independent, and $P_p\{X_i=1\}=p$.

For finite sequences, the theorem fails: for example, let $X_1, X_2, \ldots, X_n$ be the result of drawing n times at random without replacement from a box of n tickets, where some are marked "1", and the others, "0". (This distribution keeps turning up because it is a typical extreme point of the relevant convex set, as explained below.) Since $X_1+\ldots+X_n$ is constant, the law P of $X_1, \ldots, X_n$ cannot be a mixture of coin-tossing processes. However, if k is small relative to n, then $X_1, \ldots, X_k$ is nearly a mixture of coin-tossing processes.

To make this precise, let $P_k$ be the law of $X_1, \ldots, X_k$, and let $P_p^{k}$ be the law of k tosses of a p-coin. Then, for a suitable $\mu$,

$$(2.2) \qquad \left\| P_k - \int_{[0,1]} P_p^{k} \, \mu(dp) \right\| = \gamma \frac{k}{n} + o\left(\frac{k}{n}\right)$$

The argument: By symmetry, the law $Q_{nsk}$ of $X_1,...,X_k$ given that $S_n=s$ is the law of k draws made at random without replacement from a box with n tickets, where s are marked "1" and the remaining n-s are "0". (The computation is done relative to our exchangeable probability P; the result is the same as for the independent case, covered in example i of section 1.)

By the law of total probability,

$$P_k = \sum_{s=0}^{n} Q_{nsk} P\{S_n=s\} \tag{2.3}$$

And

$$||Q_{nsk} - P_{s/n}^{k}|| = \gamma \frac{k}{n} + o(\frac{k}{n}) \tag{2.4}$$

In principle, (2.2) follows from (2.3-4) using the convexity of the norm, provided (2.4) is uniform in s-- which it is. The mixing measure $\mu$ in (2.2) is discrete: it is the P-law of $S_n/n$. To get the infinite form of the theorem, let $n\to\infty$. For details, see Diaconis & Freedman (1980); for a general discussion of exchangeability, see Diaconis & Freedman (1984). The $\gamma$ in (2.4) is defined by (1.1), and is a universal constant-- of this paper anyway.

To set this argument up in greater generality, let $\{P_\lambda\}$ be an exponential family satisfying conditions (1.8-11). Let $X_1,...,X_n$ be the coordinate functions on $I^n$, and $S_n=X_1+...+X_n$.

Let $Q_{nsk}$ be the regular conditional $P_\lambda^n$-distribution for $X_1,\ldots,X_k$ given $S_n=s$, defined in (1.12).

(2.5) Definition. Let $C_n$ be the set of probabilities P on $I^n$ such that:

i) $P\{h^{(n)}(S_n) > 0)\} = 1$

ii) $Q_{nsn}$ is a regular conditional distribution for P given $S_n=s$.

Clearly, $P_\lambda^n \in C_n$. And so is $P_{\mu n}$, defined as $\int_\Lambda P_\lambda^n \mu(d\lambda)$. The set $C_n$ is convex, with extreme points $Q_{nsn}$. Any $P\in C_n$ is exchangeable, because the $Q_{nsn}$ are.

Write $P_k$ for the P-law of $X_1,\ldots,X_k$. If $k\le n$, as a matter almost of notation, $P_{nk}= P_k$. A finite version of de Finetti's theorem can now be stated, characterizing mixtures of the basic exponential family in terms of their sufficient statistics.

(2.6) Theorem. Suppose conditions (1.8-11). For $P\in C_n$, let $\mu=\mu_{nP}$ be the P-law of the $\lambda$ solving $m_\lambda=S_n/n$. Let k and $n \to \infty$ with $k/n \to 0$. Then

$$||P_k - P_{\mu k}||/\gamma\frac{k}{n} \to 1$$

Proof. As in the coin-tossing example, using Theorem (1.14a) to estimate the conditional probabilities $Q_{nsk}$. □

In this theorem, the class $C_n$ is defined as all probabilities which have the same conditionals given $S_n$ as the fixed exponential family $\{P_\lambda : \lambda \varepsilon \Lambda\}$. As far as the law $P_k$ of the first $k=o(n)$ coordinates is concerned, any $P \varepsilon C_n$ is nearly a mixture of the power probabilities $P_\lambda^k$.

The particular mixing measure $\mu$ constructed in (2.6) is nearly optimal, as shown by the next theorem, whose proof is deferred to section 4.

(2.7) Theorem. Fix $\lambda^* \varepsilon \Lambda$. Let k and n tend to $\infty$. Let $s=n \bullet m_{\lambda^*}$. Drop conditions (1.8-10), and assume (1.11) only at $\lambda^*$ rather than uniformly.

a) If $k/n \to 0$, then

$$\inf_\mu ||Q_{nsk} - P_{\mu k}|| = \gamma \frac{k}{n} + o(\frac{k}{n})$$

b) If $k/n \to \theta$ with $0<\theta<1$, then

$$\inf_\mu ||Q_{nsk} - P_{\mu k}|| \to \phi(\theta)$$

To see more explicitly why (2.6) is sharp, fix $\lambda^* \varepsilon \Lambda$ and let $s=n \bullet m_{\lambda^*}$. Now $Q_{nsn} \varepsilon C_n$, and this will be the test P in (2.6). If $k/n \to 0$, (2.7a) shows that no mixing measure can beat the one constructed in (2.6), by more than $o(k/n)$. On the other hand, if $k/n \to \theta > 0$, then (2.7b) shows that $Q_{nsk}$ is close to no mixture of $P_\lambda$'s, and our finite version of de Finetti's theorem cannot hold for such large k. The $\phi(\theta)$ is as in (1.13).

For the infinite version, in the setting of Theorem (2.6), let $X_1, X_2, \ldots$ be the coordinate functions on $I^\infty$, and $S_n = X_1 + \ldots + X_n$. Define $Q_{nsk}$ by (1.12). Let $P_n$ be the P-law of $X_1, \ldots, X_n$.

(2.8) Theorem. Suppose (1.8-11). Let P be a probability on $I^\infty$, such that $P_n \varepsilon C_n$ for all n. Then P is exchangeable, and

$$P = \int_\Lambda P_\lambda^{\infty} \mu(d\lambda)$$

The mixing measure $\mu$ is the weak-star limit of the law $\mu_n$ of $m^{-1}(S_n/n)$, as $n \to \infty$.

Proof. This follows by a limiting argument from (2.6), provided we can show $\mu_n$ is tight, and that is a consequence of (1.8). Suppose, for instance, that a and $\alpha$ are finite. Given $\epsilon > 0$ there is a $\delta > 0$ with $P_\lambda\{(a, a+\epsilon)\} > 1-\epsilon$ for $\alpha < \lambda < \alpha+\delta$. Let k and n approach infinity, with $k = o(n)$. By (2.6),

$$P\{a < X_1 < a+\epsilon\} \geq \int_{(\alpha,\alpha+\delta)} P_\lambda\{a < X_1 < a+\epsilon\} \ \mu_n(d\lambda) + O(\frac{k}{n})$$

$$\geq (1-\epsilon)\mu_n\{(\alpha, \alpha+\delta)\} + O(\frac{k}{n})$$

Therefore, $\lim_{\delta \to 0} \limsup_{n \to \infty} \mu_n\{(\alpha, \alpha+\delta)\} = 0$. □

This infinite version of de Finetti's theorem for exponential families is available under much weaker conditions: see Kuchler & Lauritzen (1986). The following simple argument for a special case of their theorem may be of interest. To set it up, and avoid irritating technicalities, drop (1.8-11) and assume (2.9-10) instead. (Half-finite or finite state spaces are easily accommodated; roughly the same argument works even for general, locally integrable h-- but the analysis is a little delicate.)

(2.9) Let h be a positive, continuous function on $(-\infty,\infty)$, with

(2.10) $c(\lambda) = \int_{-\infty}^{\infty} e^{\lambda x} h(x)\,dx < \infty$ for $\lambda$ inside the maximal interval $\Lambda$, which is nonempty.

In particular, $h^{(n)}$ is positive and continuous for all n.

The exponential family $\{P_\lambda : \lambda \in \Lambda\}$ is defined by (1.4-5), as before. Recall that $X_1, X_2, \ldots$ are the coordinate functions on $I^\infty$. Define $Q_{nsk}$ by (1.12), as usual. For any probability P on $I^\infty$, recall that $P_n$ is the law of $X_1, \ldots, X_n$. Define $C_n$ by (2.5). Define $M_Q$, a set of probabilities on $I^\infty$, as follows: $P \in M_Q$ iff $P_n \in C_n$ for all n. Informally, $P \in M_Q$ if it has the same conditionals given $S_n$ as the $P_\lambda^\infty$. In particular, P is exchangeable; the next theorem shows it is a mixture of $P_\lambda^{\infty}$.

(2.11) **Theorem.** Assume (2.9-10) rather than (1.8-11). Then $P \in M_Q$ iff

$$P = \int_\Lambda P_\lambda^\infty \mu(d\lambda)$$

Proof. The "if" part is easy, and $\mu$ is unique by standard arguments. For "only if", we use the general theory in Diaconis and Freedman (1984). If $P \in M_Q$, then $P = \int Q_\eta \nu(d\eta)$, where $Q_\eta$ is 0-1 on the $\sigma$-field $\hat{\Sigma} = \cap_n \sigma\{S_n, X_{n+1}, X_{n+2} \dots\}$ and $Q_\eta \in M_Q$. Especially, $Q_\eta$ is exchangeable; since it is 0-1 on $\hat{\Sigma}$, it makes the $X_i$ independent and identically distributed. It remains only to show that $Q_\eta = P_\lambda$ for some $\lambda \in \Lambda$, and that follows from (2.12), which writes $L(Y)$ for the law of Y, and $L(Y|X)$ for the law of Y given X.

(2.12) Suppose $X_1$ and $X_2$ are iid and $L\{X_1 | X_1 + X_2 = s\} = Q_{2s1}$. Then $L(X_1) = P_\lambda$ for some $\lambda \in \Lambda$.

Here is the argument for (2.12): $Q_{2s1}$ has the continuous, positive density $x \to h(x)h(s-x)/h^{(2)}(s)$. Now $L(X_1)$ is a mixture of $L(X_1 | X_1 + X_2 = s)$, so it too has a continuous positive density; call the latter f. Of course, $L(X_1 | X_1 + X_2 = s)$ can be computed directly from f, so

(2.13) $$f(x)\,f(s-x)/f^{(2)}(s) \;=\; h(x)\,h(s-x)/h^{(2)}(s)$$

Let

$$\lambda(x) = \log\frac{f(x)}{h(x)} - \log\frac{f(0)}{h(0)}$$

and

$$\phi(s) = \log\frac{f^{(2)}(s)}{h^{(2)}(s)} - 2\log\frac{f(0)}{h(0)}$$

Take logs in (2.13) and regroup:

$$\lambda(x)+\lambda(s-x) = \phi(s)$$

Put $s=x$: since $\lambda(0)=0$, we get $\phi(s)=\lambda(s)$, so

$$\lambda(x)+\lambda(y) = \lambda(x+y)$$

Now $\lambda(x)=\lambda x$ for some real number $\lambda$, proving (2.11) and the theorem. ○

(2.14) *Example.* An exponential family for which de Finetti's theorem fails: (2.9) does not obtain. Indeed, the base measure $\beta$ is discrete, assigning mass 1 each to $1, e, e^2, \ldots$ Consider the exponential family $\{P_\lambda\}$ through $\beta$. Now a finite sum $a_0+a_1e+\cdots+a_ke^k$ determines the integer coefficients $a_j$. Thus, $S_n=X_1+\cdots+X_n$ determines the order statistics of $X_1,\ldots,X_n$ and $Q_{ns}$ assigns equal weight to all permutations. If now $X_i$ are iid with values $1, e, e^2, \ldots$ then the law of $X_1,\ldots,X_n$ given $S_n=s$ is $Q_{ns}$--whether or not the law of $X_i$ is of the form $P_\lambda$. (It is in this sense that "de Finetti's theorem" fails; properly speaking, his theorem holds, but our variant of it fails.)

(2.15) *Example.* Another example for which de Finetti's theorem fails: the base measure $\beta$ is continuous, with a singular component. Let $\beta_0$ be uniform on the Cantor set, and $\beta_1=N(0,1)$. Let $\beta=\frac{1}{2}(\beta_0+\beta_1)$. Consider the exponential family $\{P_\lambda\}$ through $\beta$. The natural parameter space is the whole line. Now $\beta_0^{(n)}$ is supported on the Lebesgue-null set $C_n$, because $\limsup_{|t|\to\infty} |\hat{\beta}_0(t)| = 1$. Let $Q_{ns}$ be a regular conditional distribution for $X_1, \ldots, X_n$ given $S_n = s$, with respect to $\beta^\infty$; and $R_{ns}$ with respect to $\beta_0^\infty$. Now $Q_{ns} = R_{ns}$ for $\beta_0^{(n)}$-almost all s. In particular, $\beta_0 = P_\lambda$ for no $\lambda$. Yet, with respect to $\beta_0^\infty$, $Q_{ns}$ is a regular conditional distribution for $X_1, \ldots, X_n$ given $S_n = s$. This may seem like a cheat, since $Q_{ns}$ has a bigger domain than $R_{ns}$. If so, consider $P = \frac{1}{2}(\beta_0^\infty + \beta^\infty)$. This has $Q_{ns}$ for the law of $X_1, \ldots, X_n$ given $S_n = s$, but cannot be represented as $\int P_\lambda^\infty \mu(d\lambda)$.

For more discussion, see Diaconis & Freedman (1984).

## 3. The proof of Theorem (1.14)

This section will prove Theorem (1.14), starting from assumptions (1.8-11). We view s as variable and choose $\lambda = \lambda_{ns}$ to solve $m_\lambda = s/n$: the solution exists by (1.8) and is unique by (1.7). Let $\tilde{Q}$ be the $Q_{nsk}$-law of $S_k = X_1 + \cdots + X_k$ and $\tilde{P}$ the $P_\lambda^k$-law. (Dependence on n, s, k is not made explicit). By the sufficiency lemma (2.4) of Diaconis and Freedman (1987), we have $\| Q_{nsk} - P_\lambda^k \| = \| \tilde{Q} - \tilde{P} \|$. Let $f_k(t) = f_{k\lambda}(t)$ be the $P_\lambda^k$-density of $S_k$, namely,

(3.1) $$e^{\lambda t} h^{(k)}(t)/c(\lambda)^k$$

so $\hat{Q}$ has density

(3.2) $$\begin{aligned} q(t) &= f_k(t)\, f_{n-k}(s-t)/f_n(s) \\ &= h^{(k)}(t) h^{(n-k)}(s-t)/h_n(s) \end{aligned}$$

Now,

(3.3) $$\| Q_{nsk} - P_\lambda^k \| = \int \left| \frac{f_{n-k}(s-t)}{f_n(s)} - 1 \right| f_k(t)\, dt$$

We will estimate $f_{n-k}$ and $f_n$ using the Edgeworth expansion. Let $\tilde{t}$ be t standardized for $f_k$, that is,

(3.4) $$\tilde{t} = (t - k\frac{s}{n})/\sqrt{k}\,\sigma_\lambda$$

Let $\hat{t}$ be t standardized for $u \to f_{n-k}(s-u)$, that is,

(3.5) $$\hat{t} = -\sqrt{k/(n-k)}\ \tilde{t}$$

We claim the following.

(3.6) *Lemma.* Let $0 < \theta_1 < 1$. Then

$$f_{n-k}(s-t)/f_n(s) = \sqrt{n/(n-k)}\, e^{-\frac{1}{2}\hat{t}^2} + O(\sqrt{k}/n)|\tilde{t}| + O(1/n)$$

uniformly in n, k, s, t with $k < \theta_1 n$.

*Proof.* Recall that $m_\lambda = s/n$, so $\lambda$ depends on s and n. Abbreviate $\sigma = \sigma_\lambda$. By the Edgeworth expansion, (3.7) $$f_n(s) = \frac{1}{\sigma\sqrt{2\pi n}}\,[1+O(\tfrac{1}{n})]$$

(3.8) $$f_{n-k}(s-t) = \frac{1}{\sigma\sqrt{2\pi(n-k)}}\,e^{-\frac{1}{2}\hat{t}^2}[1+\frac{q(\lambda)}{\sqrt{n-k}}\,H_3(\hat{t})]+O(1/\sigma n^{3/2})$$

The O-terms are uniform by (1.9-11): see (3.12) below. As a matter of notation, $H_3(x) = x^3-3x$, and $q(\lambda) = \frac{1}{6}\,E_\lambda[(X-m_\lambda)^3]/\sigma_\lambda^3$. The latter is uniformly bounded by (1.9). By (3.7), if $n \geq n_0$ then $f_n(s) = f_{n\lambda_{ns}}(s)$ is positive for all s. Therefore $h^{(n)}(s)>0$ for all s: see (3.1). As a result, $f_{n\lambda'}(s)>s$ for all $\lambda'$ and s, even for $\lambda'\neq\lambda_{ns}$. Now

$$f_{n-k}(s-t)/f_n(s) = \sqrt{n/(n-k)}\,e^{-\frac{1}{2}\hat{t}^2}[1+O(\frac{1}{\sqrt{n-k}})\,|\,H_3(\hat{t})\,|\,]+O(1/n)$$

But

$$\begin{aligned} e^{-\frac{1}{2}\hat{t}^2}H_3(\hat{t}) &\leq |\hat{t}|\,(\hat{t}^2+3)\,e^{-\frac{1}{2}\hat{t}^2} \\ &= O(|\hat{t}|) \\ &= O(\sqrt{k/n})\cdot|\tilde{t}| \end{aligned}$$

○

(3.9) *Lemma.* If $k = o(n)$, and $|\hat{t}|<\theta_2<\infty$, then uniformly in s and t,

$$f_{n-k}(s-t)/f_n(s) = 1+\frac{1}{2}\frac{k}{n}(1-\tilde{t}^2)+O(\sqrt{k}/n)|\tilde{t}|$$

$$+O(k^2/n^2)(\tilde{t}^2+\tilde{t}^4)+O(1/n)+O(k^2/n^2)$$

*Proof.* From (3.6),

(3.10) $$f_{n-k}(s-t)/f_n(s) = (1+\frac{1}{2}\frac{k}{n})\,e^{-\frac{1}{2}\tilde{t}^2}+O(\sqrt{k}/n)|\tilde{t}|+O(1/n)+O(k^2/n^2).$$

Now

$$e^{-\frac{1}{2}\hat{t}^2} = 1 - \frac{1}{2}\hat{t}^2 + O(\hat{t}^4)$$

$$= 1 - \frac{1}{2}\hat{t}^2 + O(\frac{k^2}{n^2})\tilde{t}^4$$

$$= 1 - \frac{1}{2}\frac{k}{n}\tilde{t}^2 + O(\frac{k^2}{n^2})(\tilde{t}^2 + \tilde{t}^4)$$

○

(3.11) *Lemma.* $\int |\tilde{t}|^\nu f_k(t)\,dt = O(1)$ uniformly in $\lambda \in \Lambda$ for $\nu = 1,2,3,4$, under condition (1.9).

*Proof.* Only the case $\nu = 4$ need by proved. By an elementary calculation,

$$\mathrm{var}_\lambda(S_k) = k\sigma_\lambda^2$$

$$E_\lambda\{(S_k - km_\lambda)^4\} = kE_\lambda\{(X_1 - m_\lambda)^4\} + 3k(k-1)\,\sigma_\lambda^4$$

So

$$\int \tilde{t}^4 f_k(t)\,dt = E_\lambda\{(S_k - km_\lambda)^4\}/[\mathrm{var}_\lambda(S_k)]^2$$

$$= \frac{1}{k}E_\lambda\{(X_1 - m_\lambda)^4\}/\sigma_\lambda^4 + 3\frac{k-1}{k}$$

○

*Proof of Theorem (1.14b).* We compute as follows:

$$\| Q_{nsk} - P_\lambda^k \| = \int \left| \frac{f_{n-k}(s-t)}{f_n(s)} - 1 \right| f_k(t)\ dt$$

$$= \int \left| \sqrt{\frac{n}{n-k}}\, e^{-\frac{1}{2}\frac{k}{n-k}\tilde{t}^2} - 1 \right| f_k(t)\ dt + o(1)$$

$$= E\left\{ \left| \sqrt{\frac{1}{1-\theta}}\, e^{-\frac{1}{2}\frac{\theta}{1-\theta}Z^2} - 1 \right| \right\} + o(1)$$

The first line is (3.3). The second is (3.6), with (3.11) to control the error term in $\tilde{t}$, and (3.5) to evaluate $\hat{t}$. The third is the central limit theorem. Changing variables gives $\phi(\theta)$. ○

*Proof of Theorem (1.14a).* As before,

$$\| Q_{nsk} - P_\lambda^k \| = \frac{1}{2}\frac{k}{n}\int_{|\hat{t}|\le\theta_2} 5\ |1-\tilde{t}^2|\ f_k(t)\ dt$$

$$+\int_{|\hat{t}|>\theta_2} \left|\sqrt{\frac{n}{n-k}}\, e^{-\frac{1}{2}\hat{t}^2} - 1\right| f_k(t)\ dt + o(k/n)$$

Now $|\hat{t}| > \theta_2$ implies $|\tilde{t}| > \frac{1}{2}\sqrt{n/k}\,\theta_2$ by (3.3), an event of probability $O(k^2/n^2)$ by (3.9). This eliminates the 2nd term, and the first is asymptotic to $\frac{1}{2}\frac{k}{n}E\{|1-Z^2|\}$. ○

(3.12) *Remark.* The Edgeworth expansion can be done by following the argument in (Feller, 1971, sec XVI.2). Let $X_\lambda$ have law $P_\lambda$. We work on the standardized variable $(X_\lambda - m_\lambda)/\sigma_\lambda$, and make the estimates uniform in $\lambda$, to approximate the density for $(S_n - nm_\lambda)/\sigma_\lambda\sqrt{n}$. The density for $S_n$ itself comes out by a change of scale. In Feller's equation (2.4) on p. 533, $\sigma = 1$ by the standardization. Next, Feller's $q_\delta$ comes from (1.10), and the $L^\nu$ - bound on the characteristic function from (1.11). The contribution near 0 can be estimated uniformly in $\lambda$ by condition (1.9).

## 4. Proof of Theorem (2.7)

This section will prove Theorem (2.7). We drop conditions (1.8-10), fix $\lambda^* \in \Lambda$ and assume (1.12) only for $\lambda = \lambda^*$, that is, we assume, $\psi_{\lambda^*}(t/\sigma_{\lambda^*}) \in L^\nu$. Condition (1.11) holds for $\lambda = \lambda^*$ by the Riemann-Lebesgue lemma. Condition (1.9) holds for $\lambda = \lambda^*$ by an elementary argument: $P_{\lambda^*}$ has a fourth moment. In particular, the Edgeworth expansion is available.

There is a shift in viewpoint. In the previous section, s varied and $\lambda$ followed. Here, the main $\lambda$ of interest is $\lambda^*$, and $s^* = n \cdot m_{\lambda^*}$. The first result is the analog of (2.6). To state it, let

(4.1) $$\tilde{t} = (t - k\frac{s^*}{n})/\sqrt{k}\sigma_{\lambda^*}$$

(4.2) $$\hat{t} = -\sqrt{k/(n-k)}\ \tilde{t}$$

These are the two standardizations of t.

(4.3) *Lemma*. Let $0 < \theta_1 < 1$ and $\theta_2 < \infty$. Then

$$f_{n-k,\lambda^*}(s^* - t)/f_{n,\lambda^*}(s^*) = \sqrt{n/(n-k)}\, e^{-\frac{1}{2}\hat{t}^2} + O(\sqrt{k}/n) + O(1/n)$$

uniformly in $n, k, t$ with $k < \theta_1 n$ and $|\tilde{t}| \le \theta_2$.

*Proof* As in (3.6). ○

The next result is the analog of (3.9).

(4.4) *Lemma*. If $k = o(n)$ then uniformly in t with $|\tilde{t}| \le \theta_2$, $f_{n-k}(s^* - t)/f_n(s^*) = 1 + \frac{1}{2}\frac{k}{n}(1 - \tilde{t}^2) + (\frac{k}{n})$

*Proof* As in (3.9): if $\tilde{t}$ is bounded then $\hat{t}$ is very small, by (4.2). ○

Some additional estimates will be presented.

(4.5) *Lemma*. $\frac{1}{2}(e^u + e^{-u}) \le e^{\frac{1}{2}u^2}$, with equality only at $u = 0$.

*Proof.* The left hand side is $\sum_{j=0}^{\infty} u^{2j}/(2j)!$, the right is $\sum_{j=0}^{\infty} u^{2j}/2^j j!$, and $(2j)! \ge 2^j j!$ with equality only at $j = 0$ or 1. ○

(4.6) *Lemma*. Let Z be a symmetric random variable. Fix $\varepsilon$ with $0 < \varepsilon < 1$. Let $0 < T < 1$ with $T^2 \le 1 - \varepsilon$. Then

$$P\{|Z|<T\} \geq e^{-\frac{1}{2}(1-\varepsilon)u^2} \int_{|Z|<T} e^{uZ}\, dP$$

with equality only at u = 0.

*Proof.* By symmetry and (4.1),

$$\int_{|Z|<T} e^{uZ}\, dP = \int_{|Z|<T} e^{-uZ}\, dP$$

$$= \int_{|Z|<T} \tfrac{1}{2}(e^{uZ}+e^{-uZ})\, dP$$

$$< e^{\frac{1}{2}u^2T^2}\, P\{|Z|<T\}$$

○

Abbreviate $m = m_{\lambda^*}$ and $\sigma = \sigma_{\lambda^*}$. We now consider $\lambda$ near $\lambda^*$ and t near km.

(4.7) *Lemma.* Fix $\varepsilon > 0$. For all $k \geq k_\varepsilon$ and $|\lambda - \lambda^*| \leq 17\sigma/\sqrt{k}$,

$$c(\lambda)^k \geq c(\lambda^*)^k\, e^{km(\lambda-\lambda^*) + \frac{1}{2}(1-\varepsilon)k\sigma^2(\lambda-\lambda^*)^2}$$

Equality holds only at $\lambda = \lambda^*$.

*Proof.* Use Taylor's theorem on $\log c(\lambda)$, with (1.6) to identify the first two derivatives at $\lambda^*$. ○

The $P_\lambda^k$ density of $S_k = X_1 + \cdots + X_k$ at t is $f_{k\lambda}(t) = e^{\lambda t} h^{(k)}(t)/c(\lambda)^k$.

(4.8) *Lemma.* For all $k \geq k_\varepsilon$ and $|\lambda - \lambda^*| \leq 17\sigma\sqrt{k}$,

$$f_{k\lambda}(t)/f_{k\lambda^*}(t) \leq e^{u\tilde{t} - \frac{1}{2}(1-\varepsilon)u^2}$$

where $u = (\lambda - \lambda^*) \cdot \sigma\sqrt{k}$ and $\tilde{t} = (t - km)/\sigma\sqrt{k}$. Equality holds only at u = 0, that is, $\lambda = \lambda^*$.

*Proof.* This is immediate from (4.7). ○

(4.9) *Lemma.* For $t/k \in I$, let $\phi_{kt}(\lambda) = e^{\lambda t}/c(\lambda)^k$. Then $\lambda \to \log \phi_{kt}(\lambda)$ is strictly concave, with its maximum at $\lambda = \lambda_{kt}$, the solution to $m_\lambda = t/k$.

*Proof.* From (1.6),

$$\frac{d}{d\lambda} \log \phi_{kt}(\lambda) = t - km_\lambda$$

$$\frac{d^2}{d\lambda^2} \log \phi_{kt}(\lambda) = -k\sigma_\lambda^2. \qquad \circ$$

Recall that $0 < \varepsilon < 1$.

(4.10) *Lemma.* Let $\varepsilon < \delta < 1$. For $k \geq k_\delta$, for all t with $|\tilde{t}| \leq 1 - \delta$,

a) $|\lambda_{kt} - \lambda^*| \leq 1/\sigma\sqrt{k}$

b) $\lambda \geq \lambda^+ = \lambda^* + 2/\sigma\sqrt{k}$ entails $\phi_{kt}(\lambda) \leq \phi_{kt}(\lambda^+)$

c) $\lambda \leq \lambda^- = \lambda^* - 2/\sigma\sqrt{k}$ entails $\phi_{kt}(\lambda) \leq \phi_{kt}(\lambda^-)$

d) $|\lambda - \lambda^*| \geq 2/\sigma\sqrt{k}$ entails $\phi_{kt}(\lambda) < \phi_{kt}(\lambda^*)$

*Proof. Claim a).* This is so because $\frac{d}{d\lambda} m_\lambda = \sigma_\lambda^2 \to \sigma^2$ as $\lambda \to \lambda^*$.

*Claims b) & c).* These follow from a) and (4.9).

*Claim d).* This follows from b) & c), once it is established that $\phi_{kt}(\lambda^\pm) < \phi_{kt}(\lambda^*)$. But, for example,

$$\phi_{kt}(\lambda^+)/\phi_{kt}(\lambda^*) < e^{u\tilde{t} - ½(1-\varepsilon)u^2}$$

by (4.8) on $\lambda = \lambda^+$, with $u = (\lambda^+ - \lambda^*) \cdot \sigma\sqrt{k} = 2$. Now

$$|u\tilde{t}| \leq 2(1-\delta) < ½(1-\varepsilon)u^2 \qquad \circ$$

*Proof of Theorem (2.7b).* Let q be the $Q_{nsk}$-density of $X_1 + \cdots + X_k$: see (3.2). Recall that $f_{k\lambda}$ is the $P_\lambda^k$ density of $S_k$: see (3.1). Let $f_\mu = \int f_{k,\lambda}\, \mu(d\lambda)$ be the $P_\mu$-density. Abbreviate f for $f_{k\lambda^*}$. Then

(4.11)
$$\begin{aligned}\|Q-P_\mu\| &= 2\int (q-f_\mu)^+ \\ &\geq 2\int_J (q-f_\mu)^+ \\ &\geq 2\int_J (q-f_\mu) \\ &= 2\int_J \left(\frac{q}{f}-1\right)f + 2\int_J \left(1-\frac{f_\mu}{f}\right)f\end{aligned}$$

Of course,

(4.12)
$$q(t)/f(t) = f_{n-k}(s^* - t)/f_n(s^*)$$

For J we choose the approximate interval where $q>f$, namely, $\{t: |\tilde{t}| \leq \theta_2\}$; where

(4.13)
$$\theta_2^2 = \frac{1-\theta}{\theta}\log\frac{1}{1-\theta} < 1$$

(To see where $\theta_2$ comes from, check that

$$\sqrt{\frac{1}{1-\theta}}\, e^{-\frac{1}{2}\frac{\theta}{1-\theta}z^2} \geq 1$$

exactly for $|z| \leq \theta_2$.)

The first term at the end of (4.11) is $\phi(\theta)+o(1)$ by (4.12) and (4.3). It is only left to show that

(4.14)
$$\int_J \left(1-\frac{f_\mu}{f}\right) f \geq o(1)$$

This will be so for any interval J of the form $\{|\tilde{t}|<T<1\}$, where T is now fixed. Indeed, the left side of (4.14) is linear in $\mu$, so we need only take $\mu = \delta_\lambda$. As a matter almost of notation, when $\mu = \delta_\lambda$,

$$f_\mu(t)/f(t) = \phi_{kt}(\lambda)/\phi_{kt}(\lambda^*)$$

There are two cases in the proof of (4.14) for $\mu = \delta_\lambda$.

*Case 1:* $|\lambda-\lambda^*| \leq 17/\sigma\sqrt{k}$. Now by (4.8) the integral is for $k>k_\varepsilon$ at least

(4.15)
$$\int_{|\tilde{t}|<T} [1-e^{u\tilde{t}-\frac{1}{2}(1-\varepsilon)u^2}]\, f(t)\, dt$$

In this case, $u = (\lambda-\lambda^*)\cdot\sigma\sqrt{k}$ is at most 17 in absolute value. As $k\to\infty$, the expression in

(4.15) converges uniformly in u to

$$\int_{|Z|<T} [1 - e^{uZ - \frac{1}{2}(1-\varepsilon)u^2}]$$

which is positive by (4.6). This completes the proof of (4.14) in Case 1.

*Case 2:* $|\lambda - \lambda^*| \geq 2/\sigma\sqrt{k}$. In this case, Lemma (4.10d) completes the proof of (4.14). ○

*Proof of Theorem (2.7a).* This is quite similar, but a little more delicate. Let $n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ be the normal density. Recall that $f = f_{k\lambda^*}$ is the $P_\lambda^k$ density of $X_1 + \cdots + X_k$. For $k > k_\delta$, by the Edgeworth expansion,

$$f(t) \geq (1-\delta)\, n(\tilde{t})/\sigma\sqrt{k} \quad \text{for all } t \text{ with } |\tilde{t}| \leq 1$$

Then

(4.16)
$$\begin{aligned}
\| Q - P_\mu \| &\geq 2\int_J (q - f_\mu)^+ \\
&= 2\int_J \left(\frac{q}{f} - \frac{f_\mu}{f}\right)^+ f(t)\, dt \\
&\geq 2(1-\delta)\int_J \left(\frac{q}{f} - 1 + 1 - \frac{f_\mu}{f}\right)^+ \frac{n(\tilde{t})}{\sigma\sqrt{k}}\, dt \\
&\geq 2(1-\delta)\int_J \left(\frac{q}{f} - 1 + 1 - \frac{f_\mu}{f}\right) \frac{n(\tilde{t})}{\sigma\sqrt{k}}\, dt \\
&\geq 2(1-\delta)\int_J \left(\frac{q}{f} - 1\right) \frac{n(\tilde{t})}{\sigma\sqrt{k}}\, dt + 2(1-\delta)\int_J \left(1 - \frac{f_\mu}{f}\right) \frac{n(\tilde{t})}{\sigma\sqrt{k}}\, dt
\end{aligned}$$

For J choose the interval $\{|\tilde{t}| \leq \sqrt{1-\varepsilon}\}$. By (4.12) and (4.4), the first term at the end of (4.16) is at least

$$\frac{1}{2}\frac{k}{n}(1-\delta)\, 2\int_{Z^2 \leq 1-\varepsilon} (1 - Z^2) + o\left(\frac{k}{n}\right)$$

The second term is positive, as before: in (4.15), the density f(t) should be replaced by the normal $n(\tilde{t})/\sigma\sqrt{k}$, so (4.15) is exactly $\int_{|Z|<\sqrt{1-\varepsilon}} 1 - e^{uZ - \frac{1}{2}(1-\varepsilon)u^2}$. This is positive by (4.6) or direct calculation. Approximate normality or symmetry is not good enough, since we must estimate to $o(k/n)$ not $o(1)$.

*Remark.* For part a) of the theorem,

$$\left(\frac{q}{f}-1\right) \doteq \frac{1}{\sqrt{1-\theta}}\, e^{-\frac{1}{2}\frac{\theta}{1-\theta}\tilde{t}^2} - 1$$

which is positive for

$$\tilde{t}^2 < \frac{1-\theta}{\theta} \log \frac{1}{1-\theta}$$

and negative for $\tilde{t}^2$ larger. Furthermore,

$$E\left\{\sqrt{\frac{1}{1-\theta}}\, e^{-\frac{1}{2}\frac{\theta}{1-\theta}Z^2}\right\} = 1$$

For part b) of the theorem,

$$\frac{q}{f} - 1 \doteq \frac{1}{2}\frac{k}{n}(1-\tilde{t}^2)$$

which is positive for $\tilde{t}^2 < 1$ and negative for $\tilde{t}^2$ larger. Of course, $\int \tilde{t}^2 = 1$.

*An interesting identity.*

$$f_{k,\lambda}(t)/f_{k,\lambda^*}(t) = e^{u\tilde{t}}/\phi_k(u)$$

where $\tilde{t} = (t - km_{\lambda^*})/\sigma_{\lambda^*}\sqrt{k}$, $u = (\lambda-\lambda^*)\,\sigma_{\lambda^*}\sqrt{k}$, and $\phi_k(u) = E_{\lambda^*}(e^{u\tilde{X}})$, namely, the $P^*_\lambda$-Laplace transform of the standardized X: see (4.1). Indeed, the left side is by algebra $e^{u\tilde{t}}$ times

$$e^{(\lambda-\lambda^*)km_{\lambda^*}}\, c(\lambda^*)^k/c(\lambda)^k$$

Now integrate over t against $f_{k,\lambda^*}(t)$; or expand $\log c(\lambda)$ around $\lambda^*$.

## 5. Examples

The first two examples (gamma with scale or shape parameter) are well known exponential families, which satisfy the conditions of theorem (1.14). We believe the resulting estimates are new, as are the implied forms of de Finetti's theorem characterizing mixtures of these families. A little more generally, our conditions (1.8-11) hold if h on $(0,\infty)$ satisfies $h(x)/x^{\alpha-1} \to A$ as $x \to 0$ and $h(x)/x^{\beta-1} \to B$ as $x \to \infty$, for some positive, finite $\alpha$, $\beta$, A and B, not necessarily equal.

(5.1) *Gamma with scale parameter.* To put this in canonical form, fix the shape parameter $\rho > 0$. Let $I = (0,\infty)$ and $\Lambda = (-\infty, 0)$. The carrier density is $h(x) = x^{\rho-1}$. The $P_\lambda$ density is

$$e^{\lambda x}\, h(x)/c(\lambda)$$

with

$$c(\lambda) = |\lambda|^{\rho}/\Gamma(\rho)$$

This is the law of $-X/|\lambda|$, where X is $\Gamma_\rho$. The conditions (1.8-11) are obvious.

(5.2) *Gamma with shape parameter.* If X is $\Gamma_\lambda$, the law of $\log X$ is in canonical form with $I = (-\infty,\infty)$, $\Lambda = (0,\infty)$, $h(x) = e^{-e^x}$ and $c(\lambda) = \Gamma(\lambda)$. Here, condition (1.8) is easy to check, but (1.9-11) are not so obvious. The following relationship will be helpful (Ahlfors, 1966, p 198):

$$\frac{d^2}{d\lambda^2}\log\Gamma(\lambda) = \sum_{n=0}^{\infty}\frac{1}{(\lambda+n)^2} \tag{5.3}$$

This gives $\sigma_\lambda^2$ by (1.6). Differentiation of (5.3) gives the value for $\kappa_4(\lambda)$, the 4$^{th}$ cumulant, by (1.6):

$$\kappa_4(\lambda) = 6\sum_{n=0}^{\infty}\frac{1}{(\lambda+n)^4} \tag{5.4}$$

Of course, $\kappa_4(\lambda) = E\{(X_\lambda - m_\lambda)^4\} - 3\sigma_\lambda^4$, when $X_\lambda$ is distributed as $P_\lambda$.

There are two cases to consider.

*Case 1:* $\lambda \to 0$. Then $\sigma_\lambda^2 = \lambda^{-2} + O(1)$ by (5.3). And $\kappa_4(\lambda) = 6\lambda^{-4} + O(1)$, so $\kappa_4(\lambda)/\sigma_\lambda^4 \to 6$. This settles (1.9) near 0. If X follows the $\Gamma_\lambda$ distribution, an elementary argument shows that the distribution of $-\sigma_\lambda^{-1} \log X$ tends to the exponential. Indeed, the density converges in $L^1$, proving (1.10) for $\lambda$ near 0. It also converges in $L^2$, proving (1.11) for $\lambda$ near 0 by Plancherel's identity.

*Case 2:* $\lambda \to \infty$. Now $\sigma_\lambda^2$ is between $\int_\lambda^\infty u^{-2}\,du = 1/\lambda$ and $\int_{\lambda+1}^\infty u^{-2}\,du = 1/(\lambda+1) = (1/\lambda) + O(1/\lambda^2)$. Likewise, $\kappa_4(\lambda) = (2/\lambda^3) + O(1/\lambda^4)$, so $\kappa_4(\lambda)/\sigma_\lambda^4 = O(1/\lambda) \to 0$. This proves (1.9) for large $\lambda$. For (1.10), if X follows the $\Gamma_\lambda$ distribution, then X is about $N(\lambda, \lambda)$, and $\sqrt{\lambda}(\log X - \log \lambda) \to N(0,1)$. In more detail, let Y be $\log \Gamma_\lambda$. Then $(Y - \lambda) \cdot \sqrt{\lambda}$ has density

$$f_\lambda(z) = \frac{\gamma(\lambda)}{\sqrt{2\pi}} e^{-\lambda g(z/\sqrt{\lambda})}$$

where $\gamma(\lambda) \to 1$ as $\lambda \to \infty$ and $g(u) = e^u - 1 - u$. See (Diaconis and Freedman, 1986). Clearly, $f_\lambda(z) \to n(z)$ uniformly in $|z| \le L$ as $\lambda \to \infty$, because $g(u) = \tfrac{1}{2}u^2 + O(u^3)$ as $u \to 0$. Here, n(z) is the normal density.

We claim $f_\lambda \to n$ in $L^1$, proving (1.10) in Case 2. This reduces to showing that $\int_{|z|>L} f_\lambda(z)\,dz$ is small for L large, uniformly in $\lambda > \lambda_0$. Only the upper tail will be done. Now

$$f_\lambda'(z) = -f_\lambda(z) g'(z/\sqrt{\lambda})\sqrt{\lambda}$$

and $g'(u) = e^u - 1$ is monotone, so

$$\begin{aligned} \int_L^\infty f_\lambda(z)\,dz &\le \int_L^\infty -f_\lambda'(z)\,dz/g'(L/\sqrt{\lambda})\sqrt{\lambda} \\ &= f_\lambda(L)/g'(L/\sqrt{\lambda})\sqrt{\lambda} \\ &\to \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}L^2}/L. \end{aligned}$$

This completes the proof of (1.10).

For (1.11), let $\hat{f}_\lambda(t) = \int_{-\infty}^\infty e^{itz} f_\lambda(z)\,dz$ be the characteristic function. Integrating by parts,

$$(it)^2 \hat{f}_\lambda(t) = \int_{-\infty}^\infty e^{itz} f_\lambda''(z)\,dz$$

So all we need is

(5.5) $$f_\lambda''(z) \text{ vanishes at } \pm\infty$$

(5.6) $$f_\lambda''(z) \to n''(z) \text{ uniformly on compacts}$$

(5.7) $$\int |f_\lambda''(z)|\, dz = O(1) \text{ as } \lambda \to \infty$$

Relations (5.5-6) are easy to check. For (5.7),

$$f_\lambda''(z) = f_\lambda(z) \cdot G_\lambda(z)$$

where

$$G_\lambda(z) = \lambda g''(z/\sqrt{\lambda}) - [g'(z/\sqrt{\lambda})]^2.$$

For $\lambda > \lambda_0$, by differentiation, $G_\lambda(z)$ is strictly convex in z with its minimum at $z \approx \frac{1}{2}\sqrt{\lambda}$. Especially, $G_\lambda(z)$ will be positive for $|z| \geq 2$ if we can make $G_\lambda(\pm 2) > 0$. But $G_\lambda(z) \to z^2 - 1$ as $\lambda \to \infty$.

Now

$$\begin{aligned} \int_2^\infty |f_\lambda''(z)|\, dz &= \int_2^\infty f_\lambda''(z)\, dz \\ &= -f_\lambda'(2) \\ &= f_\lambda(z)\, g'(2/\sqrt{\lambda})\sqrt{\lambda} \\ &\to 2n(2) \end{aligned}$$

This completes the argument for (1.11). ○

(5.8) *Remark.* $m_\lambda$ is not needed for these arguments. However,

$$m_{\lambda_2} - m_{\lambda_1} = \int_{\lambda_1}^{\lambda_2} \frac{d^2}{d\lambda^2} \log\Gamma(\lambda)\, d\lambda$$

so $m_\lambda = -\lambda^{-1} + O(1)$ as $\lambda \to 0$ and $m_\lambda = \log\lambda + O(1)$ as $\lambda \to \infty$. A more careful argument shows $m_\lambda = \log\lambda + O(1/\lambda)$ as $\lambda \to \infty$.

The exponential family in the next example is non-standard. We were looking for something which violated (1.8-11), but the gaps were not large enough.

(5.9) *Example.* Let $h(x)=1$ for $2j< x\leq 2j+1$ and 0 for $2j+1<x\leq 2j+2$, for each nonnegative integer j. So $I = (0,\infty)$ and $\Lambda = (-\infty,0)$. Conditions (1.8-11) are satisfied. Indeed,

$$c(\lambda) = \int_0^\infty e^{\lambda x} h(x)\,dx$$

$$= \sum_{j=0}^\infty \int_{2j}^{2j+1} e^{\lambda x}\,dx$$

$$= \frac{1}{\lambda}\,\frac{e^\lambda - 1}{1-e^{2\lambda}}$$

$$= -\frac{1}{\lambda}\,\frac{1}{1+e^\lambda}$$

Of course, $c(\lambda)>0$ because $\lambda<0$. By painful calculation,

$$m_\lambda = -\frac{1}{\lambda} + \frac{1}{1+e^\lambda}$$

$$\sigma_\lambda^2 = \frac{1}{\lambda^2} - \frac{e^\lambda}{(1+e^\lambda)^2}$$

More easily,

$$\psi_\lambda(t) = \frac{\lambda}{\lambda+it}\,\frac{1+e^\lambda}{1+e^{\lambda+it}}$$

Abbreviate

$$f_\lambda(x) = e^{\lambda x} h(x)/c(\lambda)$$

Case 1: $\lambda\to\infty$. Now $m_\lambda\to 0$ and $\sigma_\lambda \approx 1/|\lambda|$. Consider the density of $(-\lambda)X_\lambda$, where X is distributed as $P_\lambda$:

$$g_\lambda(z) = (1+e^\lambda)\,e^{-z} h(z/|\lambda|) \quad \text{for } z>0$$

Let $g_0(z) = e^{-z}$. Now $g_\lambda \to g_0$ in $L^1$ and in $L^2$: conditions (1.10) and (1.11) follow. Furthermore, $\int z^\nu g_\lambda(z)\,dz \to \int z^\nu g_0(z)\,dz$ for any positive $\nu$, proving (1.9). Condition (1.8) is easy.

Case 2: $\lambda \to 0$. Now $m_\lambda \to \infty$; again, $\sigma_\lambda \approx 1/|\lambda|$. As before, the normalized density of $(-\lambda)X_\lambda$ is $g_\lambda(z)$. As $\lambda \to 0$, however, $(1+e^\lambda)\,h(z/|\lambda|) \to 1$ weakly but not pointwise or in norm. So $g_\lambda \to g_0$ only weakly. This complicates the argument appreciably. For (1.11), by a stroke of luck,

$$\limsup_{\lambda\to 0} \int_0^\infty [g_\lambda(z) - g(z)]^2\,dz < \infty$$

Plancherel's identity works again. For (1.9), $\int z^\nu g_\lambda(z)\,dz \to \int z^\nu g_0(z)\,dz$.

The argument for (1.10) is harder. Write $\lambda = -\lambda'$ where $\lambda' > 0$, so $\lambda' \to 0^+$.

$$1/\psi_\lambda(\lambda' t) = (1-it)\frac{1+e^{-\lambda'(1-it)}}{1+e^{-\lambda'}}$$

$$1/\psi_\lambda(\lambda' t) = (1-it)\frac{1+e^{-\lambda'(1-it)}}{1+e^{-\lambda'}}$$

We must bound the norm from below, for $|t| \geq \delta$. The denominator $1+e^{-\lambda'} \to 2$ and is immaterial. Now

$$\begin{aligned}|(1-it)[1+e^{-\lambda'(1-it)}]|^2 &= (1+t^2)[(1+e^{-\lambda'}\cos\lambda' t)^2 + e^{-2\lambda'}(\sin\lambda' t)^2] \\ &= (1+t^2)[1+e^{-2\lambda'} + 2e^{-\lambda'}\cos\lambda' t].\end{aligned}$$

Heuristically, the critical t is $\pi/\lambda'$, giving a value of $\pi^2$. More carefully, we estimate for t in the interval $2k\pi/\lambda' \leq t \leq (2k+2)\pi/k'$. We divide Gaul into 3 parts:

a) $2k\pi \leq \lambda' t \leq (2k+\frac{3}{4})\pi$

b) $(2k+\frac{3}{4})\pi \leq \lambda' t \leq (2k+\frac{5}{4})\pi$

c) $(2k+\frac{5}{4})\pi \leq \lambda' t \leq (2k+2)\pi$

On the 1st and 3rd interval, the cosine factor in the norm is at least

$$1-\sqrt{2}\,e^{-\lambda'} + e^{-2\lambda'} = 2-\sqrt{2} + O(\lambda')$$

so the squared norm is for small $\lambda'$ at least

$$(1+t^2)(2-\sqrt{2}-\varepsilon)$$

In the middle interval, we get for our lower bound

$$[1+(2k+\frac{3}{4})^2\frac{\pi^2}{\lambda'^2}]\frac{8}{9}\lambda'^2 \geq \frac{1}{2}\pi^2$$

since for small $\lambda'$, crudely,

$$(1-e^{-\lambda'})^2 \geq \frac{8}{9}\lambda'^2$$

On a) + c), we get

$$|\psi_\lambda(\lambda' t)|^2 \leq \frac{4}{2-\sqrt{2}-\varepsilon}\cdot\frac{1}{1+t^2}$$

which is tolerable for $|t|>3$ say. If $|t|\leq 3$, there is no problem because $\psi_\lambda(\lambda' t)\to 1-it$ uniformly. On b), we get

(5.10) $$|\psi_\lambda(\lambda' t)|^2 \leq \frac{4}{\pi^2/2} < 1$$ ○

(5.11) *Example.* In this example, condition (1.8) will fail, but (1.9-11) hold, and the conclusions of theorem (1.14) fail, for a trivial reason. Let $h(x)=e^{-x}$ on $I=(0,\infty)$ and take $\Lambda=(-\infty,0)$, which is not maximal. By a direct calculation,

(5.12) $$\frac{d}{dx}Q_{ns1} = \frac{n-1}{s}\left(1-\frac{x}{s}\right)^{n-2} \text{ for } 0\leq x\leq s$$

Let $n\to\infty$. If $s/n\to 1/\theta>1$, the density in (5.12) goes to $\theta e^{-\theta x}$, which is not of the form $(\lambda+1)e^{-(\lambda+1)x}$ for $\lambda\in\Lambda$. ○

Conditions (1.8-11) are far from necessary, but something like them is needed to generate a uniform bound like that in (1.14). That is the point of the next example, which involves large gaps.

(5.13) *Example.* In this example, condition (1.8) will hold, but (1.9-11) all fail. Furthermore, for suitable $s_n$, (5.14) $$\lim_{n\to\infty} \inf_\mu \| Q_{ns_n1} - P_\mu \| = 2$$

For the construction, let $f(j)=j^j$, $j=1,2\ldots$ Let $N(\alpha,\sigma^2)$ be the normal distribution with mean $\alpha$ and variance $\sigma^2$; its density at x will be written $\phi(\alpha,\sigma^2,x)$. The carrier density $h(x)$ is defined for $x \in I=(-\infty,\infty)$ by

$$h(x)=\sum_{j=1}^{\infty} \phi(f(j), 1, x)$$

As is easily verified, h is continuous and strictly positive; however, $\int_{-\infty}^{\infty} h(x)\,dx=\infty$. Changing notation let

(5.15) $$c(\lambda) = \int_{-\infty}^{\infty} e^{-\lambda x} h(x)\,dx \quad \text{for } \lambda \in \Lambda = (0,\infty)$$

Of course,

(5.16) $$c(\lambda) = e^{\frac{1}{2}\lambda^2} \sum_{j=1}^{\infty} e^{-\lambda f(j)}$$

This converges for $\lambda \in \Lambda$, but $c(0+)=\infty$. Indeed, associate the integer $\nu$ to small $\lambda$ by

(5.17) $$1/f(\nu) \geq \lambda > 1/f(\nu+1)$$

Then

(5.18) $$c(\lambda)/\nu \to 1$$

To prove (5.18), we begin with a lower bound:

$$c(\lambda) \geq \sum_{j=1}^{\nu-1} e^{-\lambda f(j)}$$

and in this range $\lambda f(j) \leq f(j)/f(\nu) \leq f(\nu-1)/f(\nu) \leq 1/\nu$. So, $c(\lambda) \geq (\nu-1)e^{-1/\nu}$. For a good upper bound, we need only estimate

(5.19) $$\sum_{j=\nu}^{\infty} e^{-\lambda f(j)}$$

In this range, $\lambda f(j) \geq f(j)/f(\nu+1) \geq j-\nu-1$. So the sum in (5.19) is $O(1)$, and this completes the proof of (5.18).

We now define $P_\lambda$ by its density

(5.20) $$h_\lambda(x) = \frac{1}{c(\lambda)} e^{-\lambda x} h(x) \quad \text{for } x \in I$$

As will be clear, $\Lambda=(0,\infty)$ is maximal, so (1.8) holds.

By an elementary calculation,

(5.21) $$P_\lambda = \sum_{j=1}^{\infty} w_j(\lambda) N(f(j)-\lambda, 1)$$

the probability weights being given by

$$w_j(\lambda) = e^{\frac{1}{2}\lambda^2 - \lambda f(j)}/c(\lambda)$$

This presents $P_\lambda$ as a location mixture of normal distributions.

The argument for (5.18) shows that the w's become uniform on $\{1,2,\ldots,\nu\}$ as $\lambda \to 0$. As a result,

(5.22) $$\| P_\lambda - P_\nu^* \| \to 0$$

where

$$P_\nu^* = \frac{1}{\nu} \sum_{j=1}^{\nu} N(f(j),1)$$

In particular, as $\lambda \to 0^+$, the mass in $P_\lambda$ drifts off to $+\infty$ and spreads out:

(5.23) $$\lim_{\lambda \to 0} \sup_x P_\lambda\{[x, x+K]\} = 0$$

for each fixed K.

We choose $s_n = nf(n)$, and claim

(5.24) $$\| Q_{ns_n 1} - N(f(n), (n-1)/n) \| \to 0$$

This is the hardest part of the argument. The idea: if $X_1 + \cdots + X_n = nf(n)$, then each $X_i$ must come from the $N(f(n), 1)$ part of the mixture. If any $X_i$ comes from a bigger part, the sum will be too big. Conversely, if all come from $N(f(n), 1)$ or less and if any one comes from a smaller part, the sum will be too small.

For the rigor,

(5.25) $$\frac{d}{dx} Q_{ns1} = h(x)\, h^{(n-1)}(s-x)/h^{(n)}(x)$$

We claim

(5.26) $$h^{(n)}(s_n) \approx 1/\sqrt{2\pi n} \quad \text{as } n \to \infty$$

To start the proof of (5.26),

(5.27) $$h^{(n)}(s_n) = \sum_{j_1,\dots,j_n} \phi[f(j_1) + \cdots + f(j_n), n, s_n]$$

We consider 3 types of n-tuples:

A) $\max j_i \le n$ and $\min j_i < n$

B) $j_1 = \cdots = j_n = n$

C) $\max j_i > n$

Plainly, the contribution to the sum in (5.27) from B) is exactly $1/\sqrt{2\pi n}$. To estimate the contribution from C), let $k = \max j_i - n$. There are at most $(n+k)^n$ n-tuples whose max is $n+k$. And $f(n+k) \ge s_n$, so each term in C is at most

$$\begin{aligned}\phi[f(n+k), n, s_n] &= \phi[0, n, s_n - f(n+k)] \\ &= \phi[0, n, f(n+k) - nf(n)]\end{aligned}$$

Overall, the contribution from C) is at most

(5.28) $$\sum_{k=1}^{\infty} (n+k)^n \phi[0, n, f(n+k) - nf(n)]$$

Crudely,

$$f(n+k) \ge (n+k)n^n$$

So

(5.29) $$\phi[0, n, f(n+k) - nf(n)] \le \frac{1}{\sqrt{2\pi n}} e^{-n(n+k)}$$

provided $n \ge 3$. Now

$$\sum_{k=1}^{\infty}(n+k)^n e^{-n(n+k)} \le \sum_{k=1}^{\infty}(n+k)e^{-(n+k)}$$

$$\le \int_n^{\infty} xe^{-x}\,dx$$

$$= (n+1)e^{-n}$$

In sum, the contribution from C) is at most

(5.30) $$\frac{1}{\sqrt{2\pi n}} \cdot (n+1)e^{-n}$$

Likewise, the contribution from A) is at most

(5.31) $$n^n \phi(0, n, f(n) - f(n-1)) \le \frac{1}{\sqrt{2\pi n}} \cdot n^n e^{-n^2}$$

for $n \ge 4$. This completes the proof of (5.26).

In principle, we must now estimate the numerator on the right of (5.25). However,

(5.32) $$\int_{-\infty}^{\infty} h(x)\, h^{(n-1)}(s_n - x)\,dx = h^{(n)}(s_n)$$

The integral on the left in (5.32) is

(5.33) $$\sum_{j_1, j_2 \cdots j_n} \int_{-\infty}^{\infty} \phi[f(j_1), 1, x]\, \phi[f(j_2) + \cdots + f(j_n), n-1, s_n - x]\,dx$$

The term with $j_1 = j_2 = \cdots = j_n = f(n)$ already contributes $\phi[nf(n), n, s_n] = 1/\sqrt{2\pi n}$, so all other terms amount to $o(1/\sqrt{n})$ and contribute $o(1)$ in variation distance to $Q_{ns_n 1}$. Up to $o(1)$, then, $Q_{ns_n 1}$ is $L(X_1 | X_1 + \cdots + X_n = ns_n)$, where the $X_i$ are iid $N[f(n), 1]$, and this proves (5.24), with a bound on the error of

(5.34) $$(n+1)e^{-n} + n^n e^{-n^2}$$

For our purposes, what counts is that $Q_{ns_n 1}$, when centered at $f(n)$, is tight. We can now prove (5.14). Start with $\mu = \delta_\lambda$. In view of (5.23), $Q_{ns_n 1}$ becomes orthogonal to all $P_\lambda$: if $\lambda$ stays away from 0, this is because $f(n) \to \infty$; if $\lambda \to 0$, because $P_\lambda$ spreads out and $Q_{ns_n 1}$ does not. The argument for $P_\mu$ is similar.

We will now argue that conditions (1.9-11) fail. For simplicity, we let $\nu \to \infty$ with $\lambda = 1/f(\nu)$. Then $c(\lambda)/\nu \to 1$ by (5.18). The same technique shows

(5.35) $$c^{(k)}(\lambda) = e^{-1}f(\nu)^k + e^{-f(\nu+1)/f(\nu)} f(\nu+1)^k + O(1)$$

The first term on the right is dominant, so

(5.36) $$m_\lambda \approx e^{-1}f(\nu)/\nu$$

(5.37) $$\sigma_\lambda^2 \approx e^{-1}f(\nu)^2/\nu$$

and the $4^{\text{th}}$ standardized moment is of order $\nu$. Thus (1.9) fails.

For (1.10) and (1.11), recall that $\psi_\lambda$ is the characteristic function of $P_\lambda$. We claim

(5.38) $$\psi_\lambda(t) \to 1 \text{ uniformly in } t \text{ with } |t| < K/\sigma_\lambda$$

Indeed, from (5.22),

(5.39) $$\psi_\lambda(t) - \frac{1}{\nu}\sum_{j=1}^{\nu} e^{-\frac{1}{2}t^2 + itf(j)} \to 0$$

uniformly in t. From (5.37): if $|t| < K/\sigma_\lambda \approx K\sqrt{e\nu}/f(\nu)$, and $j = 1, .., \nu - 1$ then $|tf(j)| \leq [1 + o(1)] K\sqrt{e\nu} f(\nu - 1)/f(\nu) \to 0$. So, all but the $\nu^{\text{th}}$ term in (5.39) is practically 1. This proves (5.38), and shows (1.11-12) to fail. ○

(5.40) *Remark.* If $f(j) = j^\alpha$, then conditions (1.8-11) hold. If $f(j) = \alpha^j$ for $\alpha > 1$, then (1.8-11) all fail, but so does the estimate for part C) of (5.37), so we do not know what happens.

(5.41) *Remark.* In these examples, (1.9-11) all hold or all fail. It would be interesting to have examples which separate the conditions.

**References**

Ahlfors, L.V. (1966). *Complex Analysis.* 2nd ed. New York: McGraw Hill.

Cover, T. and ?. Csisizar (19--).

Diaconis, P. and Freedman, D. (1980). Finite exchangeable sequences. *Ann. Prob.* **8** 745-64.

Diaconis, P. and Freedman, D. (1984). Partial exchangeability and sufficiency. In *Statistics: Applications and New Directions.* J.K. Ghosh and J. Roy (Eds) 205-236. Indian Statistical Institute, Calcutta.

Diaconis, P. and Freedman, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. Henri Poincaré* **23** 397-423

Feller, W. (1971). *Probability Theory and Its Applications Vol II.* 2nd ed. New York: Wiley.

Freedman, D. A. (1983). *Markov Chains.* Springer, New York.

Küchler, U. and Lauritzen, S. (1986). Exponential families, extreme point models and minimal space-time invariant functions for stochastic processes with stationary and independent increments. Technical Report, Aalborg Universitet.

Lanford, O. (1973). Entropy and equilibrium states in classical statistical mechanics, p1-107 (???) in *Springer Lecture Notes in Physics No. 20,* A. Lenard (Ed).

Lehmann, E. (1986). *Testing Statistical Hypotheses.* 2nd ed. Wiley, New York.

Martin-Löf, P. (1970). Statistika Modeller. Institutet för försäkringsmatematik och matematisk statistik vid Stockholms Universitet.

Stam, A. J. (1987). On the convergence of the distribution of independent random variables given their sum. Technical Report No. TW 279, Department of Mathematics and Computer Science. Groningen University, The Netherlands.

Tjur, T. (1974). Conditional probability distributions. Institute of Mathematical Statistics, University of Copenhagen.

Waterman, M.S. (1971). A note on the reparametrization of an exponential family. *Ann. Math. Statis.* **42**, 752-754.

Zabell, S. L. (1980). Rates of convergence for conditional expectations. *Ann. Prob.* **8**, 928-41.

# 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). 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, 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.

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.

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. RIZZARDI, F. (Aug 1987). Two-Sample t-tests where one population SD is known.

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.

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.

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.

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. (August 1987). 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.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987). 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. (September 1987). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (September 1987). Unachievable 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. and MACGIBBON, B. (November 1987). 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). Optimal stopping of two-parameter processes on hyperfinite 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. (November 1987). 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. (January 1988). On the consistency of Bayes estimates.

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. and DOKSUM, K.A. (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. (February 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.

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.