ANNALES DE L'I. H. P., SECTION B

PERSI DIACONIS
DAVID FREEDMAN

A dozen de Finetti-style results in search of a theory

Annales de l'I. H. P., section B, tome 23, n° S2 (1987), p. 397-423.
<http://www.numdam.org/item?id=AIHPB_1987__23_S2_397_0>

Ann. Inst. Henri Poincaré,

Sup. au n° 2, 1987, Vol. 23, p. 397-423. *Probabilités et Statistiques*

A dozen de Finetti-style results in search of a theory

by

Persi DIACONIS (1)

Department of Statistics,
Stanford, California 94305

and

David FREEDMAN (2)

Department of Statistics,
University of California,
Berkeley, California 94720

ABSTRACT. — The first k coordinates of a point uniformly distributed over the n-sphere are independent standard normal variables, in the limit as $n \to \infty$ with k fixed. If $k \to \infty$ the theorem still holds, even in the sense of variation distance, provided $k = o(n)$. The main result of this paper is a fairly sharp bound on the variation distance. The bound gives another proof of the fact that orthogonally invariant probabilities on R^{∞} are scale mixtures of sequences of iid standard normals. Similar results are given for the exponential, geometric and Poisson distributions. We do not have the right general theorem.

Key words : de Finetti's theorem, exchangeable, symmetric, variation distance, binomial, multinomial, Poisson, geometric, normal, orthogonally invariant.

---

(1) Research partially supported by NSF Grant DMS86-00235.

(2) Research partially supported by NSF Grant DMS86-01634. I worked on this paper while visiting IIASA, Laxenburg, Austria, in 1983.

Classification A.M.S. : 60 *G* 10, 60 *J* 05.

RÉSUMÉ. — Les k premières coordonnées d'un point uniformément distribué sur la sphère de dimension n se comportent comme des variables gaussiennes réduites indépendantes quand $n \to \infty$ avec k fixé. Si $k \to \infty$ le théorème reste vrai même au sens de la distance de la variation, pourvu que $k = o(n)$. Le principal résultat de cet article est une borne assez précise sur la distance de la variation. Cette borne donne une autre dimension du fait que des probabilités invariantes par les transformations orthogonales sur $\mathbf{R}^\infty$ sont des mélanges de suites de variables gaussiennes réduites indépendantes. On donne des résultats analogues pour des distributions exponentielles géométriques et de Poisson. Nous ne savons pas de théorème représentant le bon cadre général.

1. INTRODUCTION

Let ξ be chosen at random on the surface of the sphere $\left\{\xi : \sum_{i=1}^{n} \xi_i^2 = n\right\}$. Then $\xi_1, \ldots, \xi_k$ are for k fixed, in the limit as $n \to \infty$, independent standard normal variables. This result is usually—but we think incorrectly—attributed to Poincaré (1912). The history, and the connection with Lévy's work, will be discussed in Section 6 below. We allow $k = o(n)$, a growth condition which is necessary as well as sufficient. We get a reasonably sharp bound on the variation distance between the law of $\xi_1, \ldots, \xi_k$ and the law of k independent standard normals: this bound is essentially $2k/n$.

More formally, let Q_{nrk} be the law of $\xi_1, \ldots, \xi_k$, when $(\xi_1, \ldots, \xi_k, \xi_{k+1}, \ldots, \xi_n)$ is uniformly distributed over the surface of the sphere $\left\{\xi : \sum_{i=1}^{n} \xi_i^2 = r^2\right\}$. Let P_σ^k be the law of $\sigma\zeta_1, \ldots, \sigma\zeta_k$ where the ζ's are independent standard normals. Section 2 proves

$$(1) \qquad \| Q_{nrk} - P_{r/\sqrt{n}}^k \| \leq 2(k+3)/(n-k-3) \qquad \text{for} \quad 1 \leq k \leq n-4.$$

The order k/n is right, although the 2 is not sharp. The inequality has content only when $k < (1/2)n - 3$.

The inequality is connected to a representation theorem of the de Finetti type. Let $X_1, X_2, \ldots$ be infinite sequence of random variables. Call this sequence *orthogonally invariant* if for every n, the law of $X_1, \ldots, X_n$ is invariant under all orthogonal transformations of R^n. A sequence is orthogonally invariant iff it is a scale mixture of iid standard normals, a result usually attributed to Schoenberg (1938); and see Freedman (1962). This theorem is false for finite sequences; indeed, Q_{nrn} is orthogonally invariant but not a mixture of normals.

Inequality (1), however, does lead to a finite version of the representation theorem, and then the infinite version follows by a passage to the limit. For the finite version, suppose $(X_1, \ldots, X_n)$ are n orthogonally invariant random variables. Let P_k be the law of $X_1, \ldots, X_k$; recall that $\zeta_1, \zeta_2, \ldots$ are independent standard normal variables and P_σ^k is the law of $\sigma\zeta_1, \ldots, \sigma\zeta_k$. Let $P_{\mu k} = \int P_\sigma^k \mu(d\sigma)$ where μ is a probability on $[0, \infty)$.

(2) THEOREM. — *If* $X_1, \ldots, X_n$ *are orthogonally invariant, there is a probability* μ *on* $[0, \infty)$ *such that for* $1 \leqq k \leqq n-4$,

$$\| P_k - P_{\mu k} \| \leqq 2(k+3)/(n-k-3).$$

In short, the first k of n orthogonally invariant variables are within about $2k/n$ of a scale mixture of iid standard normals. This is almost immediate from (1). Indeed, consider the class of orthogonally invariant probabilities $\mathbf{P}_n$ in R^n. This is convex, and the typical extreme point is the uniform distribution Q_{nrn} on the sphere of radius r. Clearly, if $P \in \mathbf{P}_n$,

$$P = \int Q_{nrn} \lambda(dr)$$

where λ is the P-law of $\sqrt{\sum_1^n X_i^2}$. So

$$P_k = \int Q_{nrk} \lambda(dr).$$

By convexity, it is enough to prove (2) for the extreme Q_{nrn}, and that is (1). The mixing measure μ can be taken as the law of $\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}$.

Again, the k/n rate is sharp, although the 2 is wrong. This is a bit more complicated to argue: if $P = Q_{n, \sqrt{n}, n}$, and k/n is bounded away from 1, then $\| P_k - P_{\mu k} \|$ is nearly minimized when μ is point mass at 1: compare (Diaconis and Freedman, 1980, sec. 4) on the binomial.

The infinite case, ie Schoenberg's representation theorem for orthogonal invariance, follows from the finite. Let $X_1, X_2, \ldots$ be an infinite sequence of orthogonally invariant random variables, with law P on R^∞. Let P_σ^∞ be the law of $\sigma\zeta_1, \sigma\zeta_2, \ldots$ where the ζ's are iid standard normals. Let

$$P_\mu = \int P_\sigma^\infty \mu(d\sigma).$$

(3) THEOREM. — *Let* $X_1, X_2, \ldots$ *be orthogonally invariant. Then there is a unique* μ *with*

$$P = \int P_\sigma^\infty \mu(d\sigma).$$

Proof. — For each n, let $X_1, \ldots, X_n$ be the first n variables in the sequence, with law P_n. Clearly, $P_{nk} = P_k$. Apply theorem (2), getting a probability μ_n on $[0, \infty)$ with

$$\| P_k - P_{\mu_n k} \| \leqq 2(k+3)/(n-k-3) \qquad \text{for} \quad 1 \leqq k \leqq n-4.$$

The sequence μ_n is tight, because

$$\int \text{Prob}\{ |\sigma\zeta_1| > \alpha \} \mu_n(d\sigma) = \text{Prob}\{ |X_1| > \alpha \} + O(1/n)$$

goes to zero if $n \to \infty$ and then $\alpha \to \infty$: but $\text{Prob}\{ |\sigma\zeta_1| > \alpha \} \to 1$ as $\sigma \to \infty$. For existence, let μ be a subsequential limit of μ_n: clearly, $P_{\mu_n} \to P_\mu$ weak star along the subsequence. For uniqueness,

$$E\{ e^{itX_1} \} = \int_0^\infty e^{-(1/2)\sigma^2 t^2} \mu(d\sigma)$$

determines μ, by the uniqueness theorem for Laplace transforms. For a more general and less analytic proof of uniqueness, *see* Dubins and Freedman (1979, Theorem 3.4). For an even more abstract version of uniqueness, *see* Diaconis and Freedman (1984, Theorem 4.15). ○

We have similar results in three other cases. Again, the rates are sharp, but not the constants. The argument from the bound to the finite version of de Finetti's theorem to be the infinite is the same in all cases, and will not be discussed in detail each time; nor will the sharpness of the rates.

The exponential

If $\xi=(\xi_1, \ldots, \xi_n)$ is uniform on the simplex $\left\{\xi_i \geqq 0 \text{ and } \sum_1^n \xi_i = s\right\}$, then $\xi_1, \ldots, \xi_k$ are nearly independent and exponential with parameter n/s: the variation error is at most

$$2(k+1)/(n-k-1).$$

This leads to a characterization of mixtures of independent exponential variables, as uniform on the simplex given the sum. Details on the bound are in Section 3.

The geometric

This follows the pattern for the exponential, restricting attention to nonnegative integer-valued variables. The exact bound is more complicated:

$$(2k+3)\,n/[(n-k-1)(n-k-2)].$$

Details on the bound are in Section 4.

The Poisson

The analog of Poincaré's theorem is a Poisson approximation to the multinomial: Drop s balls into n boxes. Count the number falling into box 1, box 2, ..., box k. These k counts are nearly iid Poisson variables with parameter s/n: the variation error at most $1.2\,k/n$, according to Kersten (1963); also see Vervaat (1970). Mixtures of iid Poisson variables can now be characterized as being conditionally multinomial given their sum. Details on the bound are in Section 5.

2. THE NORMAL CASE

We begin with some general remarks on variation distance. Let P and Q be probabilities on the measurable space $(\Omega, \mathbb{F})$. Then

$$(2.1) \qquad \| P - Q \|_{\mathbb{F}} = 2 \sup_{A \in \mathbb{F}} | P(A) - Q(A) |.$$

If $\mathbb{F}$ is understood, it may dropped. The 2 is a conventional nuisance factor. Clearly,

$$(2.2) \qquad \| P - Q \| = 2 \sup \left| \int \varphi \, dP - \int \varphi \, dQ \right|,$$

the sup being taken over all $\mathbb{F}$-measurable φ with $0 \leqq \varphi \leqq 1$.

If P and Q are absolutely continuous with respect to a σ-finite reference measure ρ, having densities p and q, then

$$(2.3) \qquad \| P - Q \| = \int | p - q | \, d\rho = \int \left| \frac{p}{q} - 1 \right| q d\rho$$
$$= 2 \int (p - q)^+ \, d\rho = 2 \int \left(\frac{p}{q} - 1 \right)^+ q d\rho$$

where $f^+ = f$ when $f > 0$ and 0 otherwise.

Let Σ be a sub σ-field of $\mathbb{F}$ which is *sufficient* in the the sense that for all $A \in \mathbb{F}$,

$$P(A \mid \Sigma) = Q(A \mid \Sigma) \text{ a. e. P and a. e. Q.}$$

(2.4) LEMMA. — *If* $\Sigma \subset \mathbb{F}$ *is sufficient, then* $\| P - Q \|_{\Sigma} = \| P - Q \|_{\mathbb{F}}$.

Proof. — Clearly, $\| P - Q \|_{\Sigma} \leqq \| P - Q \|_{\mathbb{F}}$. In the other direction, if $A \in \mathbb{F}$, then

$$| P(A) - Q(A) | = \left| \int \varphi \, dP - \int \varphi \, dQ \right| \leqq \frac{1}{2} \| P - Q \|_{\Sigma}.$$

where $\varphi = P(A \mid \Sigma) = Q(A \mid \Sigma)$ and (2.2) was used for the inequality. ○

We are now ready to prove inequality (1).

Proof of (1). — Since variation distance is invariant under 1-1 mappings, e. g. scaling, it suffices to take $r = \sqrt{n}$ so $\sigma = r/\sqrt{n} = 1$. Recall that $\xi = (\xi_1, \ldots, \xi_k, \xi_{k+1}, \ldots, \xi_n)$ is uniform on the sphere $\sum_1^n \xi_i^2 = n$ while $\zeta_1, \zeta_2, \ldots$ are iid N(0, 1). Let $\tilde{Q}$ be the Q_{nrk}-law of $\xi_1^2 + \ldots + \xi_k^2$ and $\tilde{P}$ the P_1^k- law of $\zeta_1^2 + \ldots + \zeta_k^2$. By lemma (2.4),

$$\|Q_{nrk}-P^k_{r\sqrt{n}}\|=\|\tilde{Q}-\tilde{P}\|. \tag{2.5}$$

We realize Q as the law of $\zeta_1/R, \ldots, \zeta_n/R$, where $R^2=\frac{1}{n}\sum_1^n \zeta_i^2$. So $\tilde{Q}$ is the law of

$$n\left(\sum_1^k \zeta_i^2\right)\Big/\left(\sum_1^n \zeta_i^2\right)$$

i.e., n times a beta $[k/2, (n-k)/2]$ variable (Cramer, 1946, sec. 18.4). Thus, $\tilde{Q}$ has density

$$f(x)=\frac{1}{n}\,\frac{\Gamma(n/2)}{\Gamma(k/2)\,\Gamma[(n-k)/2]}\left(\frac{x}{n}\right)^{(k/2)-1}\left(1-\frac{x}{n}\right)^{((n-k)/2)-1} \quad \text{for} \quad 0\leqq x\leqq n$$

$$=0 \qquad \text{for} \quad x>n$$

On the other hand, $\tilde{P}$ is χ_k^2 with density

$$g(x)=\frac{1}{2^{k/2}\,\Gamma(k/2)}e^{-x/2}\,x^{(k/2)-1} \qquad \text{for} \quad 0\leqq x<\infty$$

(Cramer, 1946, sec. 18.1). By (2.3)

$$\|Q-P\|=\|\tilde{Q}-\tilde{P}\|=2\int_0^\infty\left(\frac{f(x)}{g(x)}-1\right)^+ g(x)\,dx \tag{2.6}$$

Clearly,

$$f/g=A\,h$$

where

$$A=\left(\frac{2}{n}\right)^{k/2}\Gamma\left(\frac{n}{2}\right)\Big/\Gamma\left(\frac{n-k}{2}\right)$$

$$h(x)=e^{x/2}\left(1-\frac{x}{n}\right)^{((n-k)/2)-1} \qquad \text{for} \quad 0\leqq x\leqq n$$

$$=0 \qquad \text{for} \quad x\geqq n.$$

We must estimate h and A, and this is the core of the proof. We begin with h, and claim

$$\log h(x) \leqq \frac{1}{2}(k+2) + \frac{1}{2}(n-k-2)\log\left(1-\frac{k+2}{n}\right) \tag{2.7}$$
$$\text{for } 1 \leqq k \leqq n-3.$$

Indeed, an easy calculation shows that

$$\frac{\partial}{\partial x}\log h(x) = 0 \qquad \text{for} \quad x = k+2.$$

Next, we claim

$$\log\left[\left(1-\frac{k+2}{n}\right)\mathrm{A}\right] \leqq \frac{1}{2}\left\{-(n-k-2)\log\left(1-\frac{k+2}{n}\right)-k-2\right\} \tag{2.8}$$
$$\text{for } k \text{ even with } 1 \leqq k \leqq n-3.$$

Indeed, $\Gamma(z+1) = z\,\Gamma(z)$, so

$$\Gamma\left(\frac{n}{2}\right)/\Gamma\left(\frac{n-k}{2}\right) = \left(\frac{n}{2}-1\right)\ldots\left(\frac{n}{2}-\frac{k}{2}\right)$$
$$= (n-2)(n-4)\ldots(n-k)\,2^{-k/2}$$
$$= \left(\frac{n}{2}\right)^{k/2}\left(1-\frac{2}{n}\right)\left(1-\frac{4}{n}\right)\ldots\left(1-\frac{k}{n}\right)$$

Then

$$\log\left(1-\frac{k+2}{n}\right) + \log \mathrm{A} = \sum_{i=1}^{(k+2)/2} \log\left(1-\frac{2i}{n}\right)$$
$$\leqq \int_0^{(k+2)/2} \log\left(1-\frac{2x}{n}\right)dx$$
$$= \frac{1}{2}\left[-(n-k-2)\log\left(1-\frac{k+2}{n}\right)-k-2\right].$$

This proves (2.8). Combining (2.7-8) gives

(2.9) If k is even with $1 \leqq k \leqq n-3$, then $\left(1-\frac{k+2}{n}\right) f/g \leqq 1$.

If k is even with $1 \leqq k \leqq n-3$, then (2.9) and (2.6) show

$$\| Q_{n,\sqrt{n},k} - P_1^k \| = \| \tilde{Q} - \tilde{P} \| \leqq \frac{2k+4}{n-k-2}.$$

If k is odd with $1 \leqq k \leqq n-4$, then $k+1$ is even, and

$$\| Q_{n,\sqrt{n},k} - P_1^k \| \leqq \| Q_{n,\sqrt{n},k+1} - P_1^{k+1} \| \leqq \frac{2k+6}{n-k-3}.$$

This completes the proof. ○

(2.10) *Remark.* – Lemma (2.4) is used only to simplify the calculations; indeed, with $t^2 = \sum_1^k x_i^2$, the density of Q_{nrk} at $(x_1 \ldots x_k)$ is

$$\left[\frac{1}{r\sqrt{\pi}}\right]^k \frac{\Gamma(n/2)}{\Gamma[(n-k)/2]} \left[1 - \frac{t^2}{r^2}\right]^{((n-k)/2)-1}$$

for $|t| \leqq r$; and 0 for $|t| > r$.

(2.11) *Remark on unique lifting.* – The uniqueness part of Theorem (3) is not surprising, because uniqueness always holds for infinite versions of de Finetti's theorem. For finite versions, the situation is more complicated: for example, if P_p^n makes $X_1, \ldots, X_n$ iid coin tosses with success probability p, then $\int P_p^n \mu(dp)$ only determines the first n moments of μ, so μ is not unique. For orthogonally invariant probabilities, however, uniqueness holds even for the finite version of de Finetti's theorem – and a little more. To state the result, let $\mathbf{C}_n$ be the convex set of all orthogonally invariant probabilities on R^n, and $\mathbf{C}_{nk} = \{ P_k : P \in \mathbf{C}_n \}$, i.e., $\pi \in \mathbf{C}_{nk}$ iff $\pi = P_k$ is the P-law of the first k coordinates, for some $P \in \mathbf{C}_n$. Then

(i) C_n is a simplex with extreme points the Q_{nrn}.
(ii) P is uniquely determined by P_k.
(iii) $\mathbf{C}_{nk}$ is a simplex.
(iv) The extreme points of $\mathbf{C}_{nk}$ are the Q_{nrk}.

Proof. – (i) $P = \int Q_{nrn} \lambda(dr)$, where λ is the P-law of $\sqrt{\sum_1^n X_i^2}$. (ii) It suffices to show that P_1 determines P. By orthogonal invariance, the characteristic function of P depends only on $|t| = \sqrt{\sum_1^n t_i^2}$, say as $\varphi(|t|)$. Now the characteristic function of P_1 is

$$E\{e^{it_1 X_1}\} = \varphi(|t_1|)$$

and this determines φ, so P, by Lévy's uniqueness theorem. That P_k determines P is *the unique lifting property*. Claims (iii) and (iv) follow by general arguments from the unique lifting property. ○

(2.12) *Remark on sharpness.* — The rates in (1) and (2) are sharp, but not the constants. Indeed, let k and n tend to ∞, with lim sup $k/n < 1$. We think we can prove that $\| Q_{n, \sqrt{n}, k} - P_{\mu, k} \|$ is minimized, at least asymptotically, when μ is point mass at 1. Now

(*a*) $\| Q_{n, \sqrt{n}, k} - P_1^k \| \approx \gamma\, k/n$ if $k = o(n)$

(*b*) $\| Q_{n, \sqrt{n}, k} - P_1^k \| \to \varphi(\theta)$ if $k/n \to \theta < 1$.

Here, $\gamma = \frac{1}{2} E\,|1 - Z^2|$ and

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

with Z a standard normal. Informally, $\sum_1^k \xi_i^2$ is n. beta $[k/2, (n-k)/2]$ which is asymptotically normal with mean k and variance nearly $2k\left(1 - \frac{k}{n}\right)$; while $\sum_1^k \zeta_i^2$ is χ_k^2 which is about $N(k, 2k)$. Multiplying a centered normal variable by $\sqrt{1-\theta}$ changes the distribution by $\varphi(\theta)$ in variation distance, and $\varphi(\theta) \approx \gamma\theta$ for small θ. Details are omitted, but see the next remark.

(2.13) *A more general result.* — We belive that we have proved the result in (2.12) for a fairly broad class of exponential families, including eg scale mixtures of gammas with fixed shape, or shape mixtures of gammas with fixed scale. We require uniformly bounded standardized fourth moment and some smoothness in the carrier measure, in order to get uniform bounds like (1) or (2). *See* Diaconis and Freedman (1986). More specifically, let I be an open interval (a, b), possibly infinite. Let $h \geqq 0$ on I be locally integrable. Let $\Lambda = (\alpha, \beta)$ be open, and for $\lambda \in \Lambda$ we have the exponential P_λ with density

$$\frac{1}{c(\lambda)} e^{\lambda x} h(x) \qquad \text{for} \quad x \in I$$

where

$$c(\lambda)=\int_{\mathrm{I}} e^{\lambda x} h(x)\, dx.$$

Let m_λ be the mean of P_λ, and σ_λ^2 the variance, and ψ_λ the characteristic function. As usual, m_λ is continuous and strictly increasing. We assume

(i) Λ is maximal, in the sense that $m_\lambda \to a$ as $\lambda \to \alpha$ while $m_\lambda \to b$ as $\lambda \to \beta$.

(ii) Fourth moments: $\sup_\lambda \frac{1}{\sigma_\lambda^4} \int (x-m_\lambda)^4\, \mathrm{P}_\lambda(dx) < \infty$.

(iii) Smoothness: $\sup_\lambda \sup_{|t|>\delta} |\psi_\lambda(t/\sigma_\lambda)| < 1$.

(iv) Integrability: $\sup_\lambda \int_{-\infty}^{\infty} |\psi_\lambda(t/\sigma_\lambda)|^\nu\, dt < \infty$ for some $\nu \geqq 1$.

Let $\mathrm{X}_1, \ldots, \mathrm{X}_n$ be iid P_λ. Then $\mathrm{S}=\mathrm{X}_1+\ldots+\mathrm{X}_n$ is sufficient for λ. Let Q_{nsn} be the usual version of the regular conditional distribution for $\mathrm{X}_1, \ldots, \mathrm{X}_n$ given $\mathrm{S}=s$. Let $s/n \in \mathrm{I}$. Let $\lambda^*=\lambda_{ns}^*$ be chosen so $m_{\lambda*}=s/n$. Let n and $k \to \infty$.

(*a*) $\|\mathrm{Q}_{nsk}-\mathrm{P}_{\lambda^*}^k\| = \gamma\, k/n + o(k/n)$ if $k=o(n)$

(*b*) $\|\mathrm{Q}_{nsk}-\mathrm{P}_{\lambda^*}^k\| = \varphi(\theta)+o(1)$ if $k/n \to \theta$ with $0<\theta<1$.

These estimates are uniform in s with $s/n \in \mathrm{I}$. Dropping the uniformity conditions, the estimates still hold provided λ^* is fixed in I and $s/n \to m_{\lambda^*}$; but the error depends on λ^*. Under such conditions, we can also prove the following.

(*c*) Suppose $k/n \to \theta$ with $0<\theta<1$. Fix $\lambda^* \in \mathrm{I}$; for each n, choose s as $s/n=m_{\lambda^*}$. Then $\|\mathrm{Q}_{nsk}-\mathrm{P}_{\mu k}\| \geqq \|\mathrm{Q}_{nsk}-\mathrm{P}_{\lambda^*}^k\| + o(1)$.

The idea of the proof for e.g. (*a*) is this: let $f_{k,\lambda}$ be the P_λ-density of $\mathrm{X}_1+\ldots+\mathrm{X}_k$. Then the norm in (*a*) is

$$\int \left| \frac{f_{n-k,\lambda^*}(s-t)}{f_{n,\lambda^*}(s)} - 1 \right| f_{k,\lambda^*}(t)\, dt$$

and $f_{n-k,\lambda^*}(s-t)/f_{n,\lambda^*}(s)$ can be estimated by Edgeworth. For $k=o(n)$ this turns out, to a good approximation, to be

$$\frac{1}{2}\,\frac{k}{n}(1-\tilde{t}^2), \qquad \text{where} \quad \tilde{t}=(t-km_{\lambda^*})/\sigma_{\lambda^*}\sqrt{k}$$

For (*c*), let $p_\mu=\int f_{k,\lambda}\,\mu(d\lambda)$ be the P_μ-density of $\mathrm{X}_1+\ldots+\mathrm{X}_k$. Let $\tilde{\mathrm{Q}}$ be the Q_{nsk}-law of $\mathrm{X}_1+\ldots+\mathrm{X}_k$, with density q. Let $f=f_{k,\lambda^*}$. Then

$$
\begin{aligned}
\|Q-P_\mu\| &= 2\int (q-p_\mu)^+ \\
&\geqq 2\int_K (q-p_\mu)^+ \\
&\geqq 2\int_K (q-p_\mu) \\
&= 2\int_K \left(\frac{q}{f}-1\right) f + 2\int_K \left(1-\frac{p_\mu}{f}\right) f.
\end{aligned}
$$

Choose for K the interval

$$\mathfrak{z}^2 < \frac{1-\theta}{\theta}\log\frac{1}{1-\theta},$$

essentially where $Q > P_\mu$. The first integral is $\varphi(\theta)+o(1)$. The second is positive, up to $o(1)$: only point masses need be considered for μ. We have a similar bound for $k=o(n)$.

These estimates apply to the normal and exponential distributions. They do not apply fully to the geometric or Poisson cases, because the standardized fourth moments blow up near zero. However, the estimates do apply locally, and demonstrate the sharpness of the k/n rate.

3. THE EXPONENTIAL CASE

The main result of this section shows that the uniform distribution on a simplex has approximately independent exponential coordinates. Some of the reasoning in the previous section will be useful in this connection and can be abstracted as a lemma. This is a more direct version of (Diaconis and Freedman, 1980, p. 757).

(3.1) LEMMA:

(*a*) *For* $0 \leqq x < N$, *let* $\varphi(x) = \log\left(1-\dfrac{x}{N}\right)$. *Then* $\varphi(0)=0$, φ *is strictly decreasing, strictly concave, and* $\varphi(N-) = -\infty$.

(*b*) *For* $0 \leqq x < 1$, *let* $f(x) = -(1-x)\log(1-x)-x$. *Then* $f(0)=0$, f *is strictly decreasing, strictly concave, and* $f(1-) = -1$.

(*c*) $\sum_{i=1}^{m} \varphi(i) \leqq \int_0^m \varphi(x)\, dx$

(*d*) $\sum_{i=1}^{m-1} \varphi(i) \geqq \int_1^m \varphi(x)\, dx \geqq \int_0^m \varphi(x)\, dx$

(*e*) $\int_0^y \varphi(x)\, dx = -(\mathrm{N}-y)\,\varphi(y) - y = \mathrm{N} f(y/\mathrm{N})$

(*f*) $\left(1-\frac{1}{\mathrm{M}}\right)\ldots\left(1-\frac{a}{\mathrm{M}}\right)\left(1-\frac{1}{\mathrm{N}}\right)\ldots\left(1-\frac{b}{\mathrm{N}}\right)$

$$\leqq \left(1-\frac{1}{\mathrm{M}+\mathrm{N}}\right)\ldots\left(1-\frac{a+b-1}{\mathrm{M}+\mathrm{N}}\right).$$

Proof. — Claims (*a*) and (*b*) are elementary; (*c*) and (*d*) follow. Claim (*e*) is elementary. For (*f*), the log of the left side is bounded above by

$$\mathrm{M} f\left(\frac{a}{\mathrm{M}}\right) + \mathrm{N} f\left(\frac{b}{\mathrm{N}}\right). \tag{3.2}$$

The log of the right side is bounded below by

$$(\mathrm{M}+\mathrm{N}) f\left(\frac{a+b}{\mathrm{M}+\mathrm{N}}\right). \tag{3.3}$$

The expression in (3.2) is smaller than that in (3.3), by Jensen's inequality. ○

We are now ready to state and prove the analog of inequality (1) for the uniform. Let P_λ^n be the law of $\zeta_1, \ldots, \zeta_n$ which are iid exponentials with parameter λ, so $\mathrm{P}_\lambda^n\{\zeta_i > y\} = e^{-\lambda y}$. Given $\zeta_1 + \ldots + \zeta_n = s$, the ζ's are conditionally uniform on the simplex. Let Q_{nsk} be the law of $\xi_1, \ldots, \xi_k$ where $\xi = (\xi_1, \ldots, \xi_k, \xi_{k+1}, \ldots, \xi_n)$ is uniform on the simplex $\left\{\xi_i \geqq 0 \text{ for all } i \text{ and } \sum_1^n \xi_i = s\right\}$. The next result shows that Q_{nsk} is nearly $\mathrm{P}_{n/s}^k$, provided $k = o(n)$. Informally, this is because $\zeta_1 + \ldots + \zeta_n$ is practically constant, so the conditioning is immaterial.

(3.4) THEOREM. — $\| \mathrm{Q}_{nsk} - \mathrm{P}_{n/s}^k \| \leqq 2(k+1)/(n-k+1)$ *for* $1 \leqq k \leqq n-2$.

Proof. — It suffices to do the case $s = n$. The sum is sufficient so lemma (2.4) applies, and

$$\|Q_{nsk}-P^k_{n/s}\|=\|\tilde{Q}-\tilde{P}\| \tag{3.5}$$

where $\tilde{Q}$ is the Q_{nsk} law of $\xi_1+\ldots+\xi_k$ and $\tilde{P}$ is the P^k_1 law of $\zeta_1+\ldots+\zeta_k$. We realize Q as the law of $n\zeta_1/S, \ldots, n\zeta_k/S$ where $S=\zeta_1+\ldots+\zeta_n$ and the ζ's are iid standard exponentials. So, $\tilde{Q}$ is the law of

$$n\sum_1^k \zeta_i/\sum_1^n \zeta_i$$

i. e., n times a beta $(k, n-k)$ variable, with density

$$f(x)=\frac{1}{n}\,\frac{\Gamma(n)}{\Gamma(k)\,\Gamma(n-k)}\left(\frac{x}{n}\right)^{k-1}\left(1-\frac{x}{n}\right)^{n-k-1} \qquad \text{for} \quad 0\leqq x\leqq n$$

$$=0 \qquad \text{for} \quad x>n.$$

On the other hand, $\tilde{P}$ is gamma, with density

$$g(x)=\frac{1}{\Gamma(k)}e^{-x}x^{k-1} \qquad \text{for} \quad 0\leqq x<\infty.$$

As before

$$\|Q-P\|=\|\tilde{Q}-\tilde{P}\|=2\int\left(\frac{f}{g}-1\right)^+ g\,dx \tag{3.6}$$

and we must estimate $f/g=A\,h$, where

$$A=\frac{\Gamma(n)}{n^k\,\Gamma(n-k)}=\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k}{n}\right)$$

$$h(x)=e^x\left(1-\frac{x}{n}\right)^{n-k-1} \qquad \text{for} \quad 0\leqq x\leqq n$$

$$=0 \qquad \text{for} \quad x>n.$$

We claim

$$\left(1-\frac{k+1}{n}\right)\frac{f}{g}\leqq 1 \qquad \text{for} \quad 1\leqq k\leqq n-2. \tag{3.7}$$

Indeed, the log of the left side is

$$x+(n-k-1)\log\left(1-\frac{x}{n}\right)+\sum_{i=1}^{k+1}\log\left(1-\frac{i}{n}\right)$$

$$\leqq x+(n-k-1)\log\left(1-\frac{x}{n}\right)-(n-k-1)\log\left(1-\frac{k+1}{n}\right)-k-1$$

by lemma (3.1). The upper bound is maximized at $x=k+1$ where it vanishes, proving (3.7). Then (3.6) proves the theorem. ○

(3.8) *Remark.* – The rate but not the constant is sharp. Lemma (2.4) is only to avoid tedious calculation, since the Q_{nsk} density is

$$\left(\frac{n}{s}\right)^k\left(1-\frac{1}{n}\right)\dots\left(1-\frac{k}{n}\right)\left(1-\frac{t}{s}\right)^{n-k-1}$$

where $t=x_1+\dots+x_k\leqq s$ and all $x_i\geqq 0$.

(3.9) *Remark.* – Heuristic versions of (3.4) are known: The uniform distribution on the simplex can be represented as the joint law of the spacings of $n-1$ points dropped at random into the unit interval, and the spacings are approximately independent exponentials for many purposes. *See* Feller (1971, p. 74) or Diaconis and Efron (1986) for further discussion. Rigorous versions and applications are given by LeCam (1958), Pyke (1965), and Holst (1979).

(3.10) *de Finetti's Theorem.* – Let $R_+=[0, \infty)$; Let $X_1, \dots, X_n$ be the coordinate variables on R_+^n and $S=X_1+\dots+X_n$. Let $\mathbf{C}_n$ be the class of probabilities on R_+^n which share with the iid exponentials the property that given $S=s$, the conditional joint distribution of $X_1, \dots, X_n$ is uniform on the simplex. If P is a probability on R_+^∞, let P_n be the P-law of the first n coordinates. The infinite form of de Finetti's theorem asserts that if a probability P on R_+^∞ has $P_n\in\mathbf{C}_n$ for every n, then P is a unique scale mixture of iid exponential variables. The infinite theorem follows from the finite version, which is given in the next remark.

(3.11) *Finite de Finetti.* – Recall $\mathbf{C}_n$ from (3.10). Clearly, $P_\lambda^n\in\mathbf{C}_n$; so is $P_{\mu n}=\int P_\lambda^n\,\mu(d\lambda)$ for any probability μ on $[0, \infty)$. If $P\in\mathbf{C}_n$, i.e., P is conditionally uniform give the sum, then there is a μ such that for all $k\leqq n-2$,

$$\|P_k-P_{\mu k}\|\leqq 2(k+1)/(n-k+1).$$

In other words, if n nonnegative random variables are conditionally uniform on the simplex given their sum, the first $k=o(n)$ are to within about $2k/n$ a scale mixture of iid exponentials. As in the normal case, the k/n rate is sharp, but not the constant 2.

(3.12) *Another characterization of* $\mathbf{C}_n$. – With previous notation, $P\in\mathbf{C}_n$ iff $P(A)=P(A+x)$ for all Borel sets A and all n-tuples $x_1, \dots, x_n$ of real numbers with $\sum_1^n x_i=0$, provided $A\subset R^n_+$ and $A+x\subset R^n_+$. In one direction, suppose $P\in C_n$. Then $P=\int Q_{nsn}\lambda(ds)$, so it is enough to prove that $Q(A)=Q(A+x)$, where Q stands for Q_{nsn}, the uniform distribution on the simplex. But $A\cap\{S=s\}$ and $(A+x)\cap\{S=s\}$ are congruent, and hence of equal Lebesgue measure. In the other direction, suppose $P(A)=P(A+x)$. Let V_s be a regular conditional distribution for P given $S=s$, so $V_s(S=s)=1$ for PS^{-1}-almost all s. For any particular B and x,

$$V_s(B)=V_s(B+x) \qquad \text{for} \quad PS^{-1}\text{-almost all } s,$$

as one sees by integrating over $S>s$: take $A=B\cap\{S>s\}$. The invariance must hold for e. g. all spheres B with rational center and radius, and all rational x, forcing V_s to be Lebesgue measure on $\{S=s\}$. ○

Some lemmas on the beta will be developed in order to prove the unique lifting property.

(3.13) LEMMA. — *If* X *is beta* (p, q) *and independently* Y *is beta* $(p+q, r)$ *then* XY *is beta* $(p, q+r)$.

Proof. — Let U, V, W be independent gamma variables with parameters p, q, r respectively. Then $U/(U+V)$ is beta (p, q) independently of $U+V$ which is gamma $(p+q)$, so the trivial identity

$$\frac{U}{U+V}\cdot\frac{U+V}{U+V+W}=\frac{U}{U+V+W}$$

proves the lemma. ○

(3.14) COROLLARY. — *If* X *has a beta distribution, then log* X *is infinitely divisible.*

Proof. — If X is beta (p, q), then e. g. X can be represented as beta $(p, q-\varepsilon)$. beta $(p+q-\varepsilon, \varepsilon)$. ○

(3.15) COROLLARY. — *If* X *is beta with given parameters, and* $Y\geq 0$ *is independent of* X, *then the law of* XY *determines the law of* Y.

Proof. — By (3.14), the characteristic function of log X never vanishes. ○

(3.16) *Remark on unique lifting.* — The analog of (2.11) holds in this context too. As before, let C_n be the convex set of all probabilities on the positive orthant of R^n which are conditionally uniform given the sums of the coordinate variables $X_1 \ldots X_n$. Let $\mathbf{C}_{nk} = \{ P_k : P \in \mathbf{C}_n \}$, i. e., $\pi \in \mathbf{C}_{nk}$ iff $\pi = P_k$ is the P-law of $X_1 \ldots X_k$ for some $P \in \mathbf{C}_n$. Now

$$P = \int Q_{nsn} \lambda(ds)$$

so

$$P_k = \int Q_{nsk} \lambda(ds)$$

and it is enough to compute λ from P_k. To avoid trivialites, suppose $1 \leqq k \leqq n-2$. The critical case is $k=1$, and it is enough to compute λ from P_1. Now Q_{ns1} is s. beta $(1, n-1)$. So P_1 is the law of SX, where S and X are independent, S has the law λ, and X is beta $(1, n-1)$. Finally, P_1 determines λ by (3.15).

Another argument for unique lifting starts from the characterization (3.12) of $\mathbf{C}_n$. Take $A = \{ X_1 \geqq y_1, \ldots, X_n \geqq y_n \}$ with $y_i \geqq 0$; take $x_1 = y_2 + \ldots + y_n$, $x_2 = -y_2, \ldots, x_n = -y_n$: so

$$A + x = \{ X_1 \geqq y_1 + \ldots + y_n, X_2 \geqq 0, \ldots, X_n \geqq 0 \} = \{ X_1 \geqq y_1 + \ldots + y_n \}.$$

The upshot is that for $P \in \mathbf{C}_n$,

$$P \{ X_1 \geqq y_1, \ldots, X_n \geqq y_n \} = P_1 \{ X_1 \geqq y_1 + \ldots + y_n \},$$

and P_1 determines P.

4. THE GEOMETRIC CASE

Let P_p^n be law of $\zeta_1 \ldots \zeta_n$, which are iid geometric variables with parameter p, so $P_p^n \{ \zeta_i = j \} = (1-p) p^j$ for $j = 0, 1, \ldots$ Given $\zeta_1 + \ldots + \zeta_n = s$, the ζ's are uniform on the simplex. Let Q_{nsk} be the law of $\xi_1 \ldots \xi_k$ where $\xi = (\xi_1, \ldots, \xi_k, \xi_{k+1}, \ldots \xi_n)$ is uniform on the simplex

$$\left\{ \xi_i \text{ is a nonnegative integer for all } i \text{ and } \sum_1^n \xi_i = s \right\}.$$

The analog of (1) is the following theorem.

(4.1) THEOREM. — *Let* $p = s/(n+s)$. *For* $1 \leqq k \leqq n-3$,

$$\| Q_{nsk} - P_p^k \| \leq 2 \left\{ \frac{n^2}{(n-k-1)(n-k-2)} - 1 \right\}$$

Proof. — Here scaling is not feasible, so all values of s must be considered. Let $t = j_1 + \ldots + j_k$. Clearly,

$$(4.2) \qquad P_p^k \{ j_1, \ldots, j_k \} = (1-p)^k p^t.$$

By the "stars and bars" lemma (Feller, 1968, sec. II.5), there are $\binom{n+s-1}{s}$ n-tuples $j_1 \ldots j_n$ of nonnegative integers with sum s. Then for $t = j_1 + \ldots + j_k \leq s$,

$$(4.3) \qquad Q_{nsk}(j_1, \ldots, j_k) = \binom{n-k+s-t-1}{s-t} \Big/ \binom{n+s-1}{s}.$$

Dividing (4.3) by (4.2), the ratio $Q_{nsk}(j_1, \ldots, j_k)/P_p^k(j_1, \ldots, j_k)$ is seen to equal

$$\frac{[(n+s-k-t-1)!/((s-t)!(n-k-1)!)]\,[s!(n-1)!/(n+s-1)!]}{(n^k/(n+s)^k)\,(s^t/(n+s)^t)}$$

This is N/D, where

$$(4.4) \qquad \begin{aligned} N &= \left(1-\frac{1}{s}\right)\left(1-\frac{2}{s}\right)\cdots\left(1-\frac{t-1}{s}\right)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k}{n}\right) \\ D &= \left(1-\frac{1}{n+s}\right)\left(1-\frac{2}{n+s}\right)\cdots\left(1-\frac{k+t}{n+s}\right) \end{aligned}$$

Now $N\left(1-\frac{k+1}{n}\right)\left(1-\frac{k+2}{n}\right) \leq D$ by Lemma (3.1f). The balance of the argument is omitted. ○

(4.5) *Remark.* — The usual remark on sharpness of rates is omitted.

(4.6). *Remark.* — In physics, the conditional distribution of iid geometric variables given their sum (viz, the uniform on the simplex) is referred to as "Bose-Einstein"; the conditional distribution of iid Poissons given their sum (the multinomial) is "Maxwell-Boltzman". *See* Feller (1968, pp. 40 ff).

(4.7) *de Finetti's theorem.* — Let Z_+ denote the nonnegative integers. Let $X_1, \ldots, X_n$ be the coordinate variables on Z_+^n and $S = X_1 + \ldots + X_n$. Let C_n be the class of probabilities P on Z_+^n which share with the iid geometrics the property that given $S = s$, the conditional joint distribution of $X_1, \ldots, X_n$ is Q_{nsn}, the uniform on the simplex. If P is a probability on Z_+^∞, let P_n be the law of the first n coordinates. The infinite form of de Finetti's theorem asserts that if a probability P on Z_+^∞ has $P_n \in C_n$ for every n, then P is a unique mixture of iid geometric variables. The infinite theorem follows from the finite version, given in the next remark.

(4.8) *Finite de Finetti.* – Clearly, $P_p^n \in C_n$; and so is $P_{\mu n} = \int P_p^n \mu(dp)$ for any probability μ on [0, 1]. If $P \in C_n$, there is a μ such that for $k \leqq n-3$,

$$\| P_k - P_{\mu k} \| \leqq 2 \left\{ \frac{n^2}{(n-k-1)(n-k-2)} - 1 \right\}.$$

In other words, if n nonnegative integer-valued random variables are conditionally uniform on the simplex given their sum, the first $k = o(n)$ are to within about $2k/n$ a mixture of iid geometrics. The rate is sharp, but not the constant.

(4.9) *Another characterization of* C_n. – With previous notation, $P \in C_n$ iff $P(A) = P(A+x)$ for all sets A and all n-tuples $x = x_1 + \ldots + x_n$ of integers with $\sum_1^n x_i = 0$, provided $A \subset Z_+^n$ and $A + x \subset Z_+^n$. By $A+x$, of course, we mean $\{a + x: a \in A\}$. The proof is omitted as elementary.

(4.10) *Unique lifting.* – Continue with the notation of (4.7). Let $P \in C_n$. Then P_k, the P-law of $X_1, \ldots, X_k$, determines P. To avoid trivialities, suppose $n \geqq 2$ and $k = 1$. Let λ be the P_n-law of S. The problem is to compute λ from P_1. But

$$P_1(j) = \sum_{s=j}^{\infty} \frac{\binom{n-1+s-j-1}{s-j}}{\binom{n+s-1}{s}} \lambda(s)$$

for $j = 0, 1, \ldots$ By taking successive differences $n-1$ times, one recovers $\lambda(s) \Big/ \binom{n+s-1}{s}$ and hence $\lambda(s)$. A less-algebraic proof starts from (4.9).

5. THE POISSON CASE

Let P_λ^n be the law of $\zeta_1 \ldots \zeta_n$ which are iid Poisson variables with parameter λ. Given $S = \zeta_1 + \ldots + \zeta_n = s$, the ζ's are multinomial. Let Q_{nsk} be the law of $\zeta_1, \ldots, \zeta_k$ given $S = s$.

(5.1) THEOREM. — $\| Q_{nsk} - P_\lambda^k \| \leqq 1.2\, k/n$ *with* $\lambda = s/n$, *for* $k < \frac{1}{2} n$.

Proof. — By sufficiency, $\| Q - P \| = \| \tilde{Q} - \tilde{P} \|$, where $\tilde{Q}$ is binomial with s trials and success probability k/n, while $\tilde{P}$ is Poisson with parameter ks/n. Now appeal to Kersten (1963). ○

(5.2) *Finite de Finetti.* — Let Z_+ denote the nonnegative integers; let $X_1 \dots X_n$ be the coordinate function on Z_+^n, and $S = X_1 + \dots + X_n$. Then $P \in C_n$ iff Q_{nsn} is the P-law of $X_1, \dots, X_n$ given $S = s$. Clearly, $P_\lambda^n \in C_n$; and so is $P_{\mu n} = \int P_\lambda^n \mu(d\lambda)$ for any probability μ on $[0, \infty)$. If $P \in C_n$, there is a μ such that for $k < \frac{1}{2} n$,

$$\| P_k - Q_{\mu k} \| \leqq 1.2\, k/n.$$

In other words, if n nonnegative integer-valued random variables are conditionally multinomial $(n, s/n, \dots, s/n)$ given their sum, the first $k = o(n)$ are to within about k/n a mixture of iid Poissons. The rate is sharp but not the constant.

(5.3) *Unique lifting.* — Let $P \in C_n$. Then P_k, the P-law of $X_1, \dots, X_k$, determines P. To avoid trivialities, suppose $n \geqq 2$ and $k = 1$. Let λ be the P-law of S. The problem is to compute λ from P_1. But

$$P_1(j) = \sum_{s=j}^{\infty} \binom{s}{j} \left(\frac{1}{n}\right)^j \left(\frac{n-1}{n}\right)^{s-j} \lambda(s).$$

Let

$$\theta(s) = \left(\frac{n-1}{n}\right)^s \lambda(s)$$

and

$$\varphi(x) = \sum_{s=0}^{\infty} \theta(s) x^s$$

with radius of convergence $n/n-1$. Clearly,

$$\varphi^{(j)}(1) = j!\, (n-1)^j P_1(j)$$

and since φ is analytic on the disk $\{x : |x| < n/n-1\}$ it is determined by its derivatives at 1.

(5.4) *More on unique lifting.* — We consider (4.10) and (5.3) a bit more generally. Let $M(i, j)$ be an infinite stochastic matrix with entries which are positive for $j \leqq i$ and vanish for $j > i$, i.e., M is a lower triangular matrix on the nonnegative integers Z_+. Let λ be a probability distribution on Z_+. In essence, the unique lifting property is that λM determines λ, for certain M. More specifically, let $X_1, \ldots, X_n$ be the coordinate process on Z_+^n. Let $M(i, .)$ be the law of X_1 when $X_1 + \ldots + X_n$ is uniformly distributed on the simplex $\{X_1 + \ldots + X_n = i\}$, that is, $M(i, .)$ is the Q_{ni1} corresponding to the geometric. Then λM determines λ by (4.10). Likewise for the M corresponding to the multinomial, by (5.3). For general M, a formal inverse always exists. Indeed, let M_n be the upper $n \times n$ submatrix of M: then $M_{n+1}^{-1} = M_n^{-1}$ when it can. However, M is not necessarily 1-1. The counterexample M, with domain the positive integers, is

$$\begin{bmatrix}
1 & & & & & & & \\
\frac{1}{2} & \frac{1}{2} & & & & & & \\
\frac{1}{2} & \frac{1}{2}-\varepsilon_3 & \varepsilon_3 & & & & & \\
\frac{1}{4} & \frac{1}{4}-2\varepsilon_3 & \frac{1}{4}+2\varepsilon_3 & \frac{1}{4} & & & & \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}-\varepsilon_5 & \varepsilon_5 & & & \\
\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6}-2\varepsilon_5 & \frac{1}{6}+2\varepsilon_5 & \frac{1}{6} & & \\
\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6}-\varepsilon_7 & \varepsilon_7 & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{bmatrix}$$

Algebraically, for $i = 3, 5, 7, \ldots$

$$\begin{aligned}
M(i, j) &= \frac{1}{i-1} && \text{for } j = 1, \ldots, i-2 \\
&= \frac{1}{i-1} - \varepsilon_i && \text{for } j = i-1 \\
&= \varepsilon_i && \text{for } j = i.
\end{aligned}$$

For $i = 4, 6, 8, \ldots$

$$\begin{aligned}
M(i, j) &= \frac{1}{i} && \text{for } j = 1, \ldots\ i-3 \text{ or } j = i \\
&= \frac{1}{i} - 2\varepsilon_{i-1} && \text{for } j = i-2
\end{aligned}$$

$$= \frac{1}{i} + 2\,\varepsilon_{i-1} \qquad \text{for} \quad j = i-1.$$

Let

$$\begin{aligned} \lambda(i) &= 0 \quad \text{for odd } i \\ &= 1/2^{i/2} \quad \text{for even } i. \end{aligned}$$

Let

$$\begin{aligned} &\lambda'(1) = 0 \\ &\lambda'(i) = 0 \quad \text{for even } i \\ &\lambda'(i) = 1/2^{(i-1)/2} \quad \text{for odd } i \geqq 3. \end{aligned}$$

Then $\lambda\,M = \lambda'\,M$.

6. HISTORY

1. Poincaré, Borel, and Maxwell

The asymptotic normality of the first k coordinates of a random point on the n-sphere is a theorem usually attributed to Poincaré (1912): *see* e. g. McKean (1973, p. 197), Letac (1981, p. 412), or Billingsley (1979, p. 342). However, after a diligent search, we were unable to find the result in Poincaré. The closest we came was a brief mention of statistical mechanics at p. 43, and Liouville's theorem at pp. 145 ff.

The earliest reference to the theorem we could find in the probability literature was Borel (1914, Chapter V). In equation (12) on p. 66, Borel gives a sharp statement of the theorem for $k=1$. On pp. 90-93, he makes the connection with the kinetic theory of gas. (We continue with our notation, rather than switching to his.) Consider m particles, each with 3 velocity coordinates, all denoted $x_1, \ldots, x_n$, with $n = 3\,m$. Each particle has the same mass, to be denoted by c. The system is constrained to lie on the surface of constant kinetic energy, $\frac{1}{2} c \sum_{i=1}^{n} x_i^2 = h^2$. A uniform distribution on this energy surface is assumed (Liouville's theorem is the usual —but partial—justification). Now x_1 tends in distribution to $N(0,\, 2\,h^2/nc)$

as $n \to \infty$, uniformly in h and c. Borel asserts asymptoic independence and normality for $x_1, \ldots, x_{n-k}$, provided $n-k \to \infty$. This may be true in some sense, but for variation distance $k = o(n)$ is required.

Borel was aiming for "the usual form of Maxwell's theorem", that the empirical distribution of $x_1, \ldots, x_n$ tends to normal. From a modern perspective, this is quite easy. If $Z_1, \ldots, Z_n$ are iid $N(0,1)$ variables, then their empirical distribution tends to $N(0,1)$. So must the empirical distribution of $Z_1/R_n, \ldots, Z_n/R_n$, where $R_n^2 = \frac{1}{n}\sum_{i=1}^{n} Z_i^2 \to 1$ a.e. One reference of historical interest is Maxwell (1875, p. 309); a second, Maxwell [1878, eqn (49-55)], is more technical and focused on the law of the individual x's.

The application to kinetic theory is an example of what mathematical physicists now call *the equivalence of ensembles.* If $H(x)$ is the Hamiltonian, one can work with the "microcanonical distribution", the uniform distribution on $\{x : H(x) = h^2\}$. Alternatively, one can work with the Gibb's distribution $G(dx)$ on R^n. This is a probability whose density is proportional to $e^{-cH(x)}$. The constant c is chosen so $\int H(x)\,G(dx) = h^2$. The equivalence of ensembles obtains when

$$\int g(x)\,G(dx) \approx \int g(x)\,U(dx)$$

for some wide class of functions g, and for some well-specified meanning of the approximate equality.

To get Borel's example, take $H(x) = \frac{1}{2}c(x_1^2 + \ldots + x_n^2)$. Theorem (1) gives the equivalence of ensembles as n tends to infinity, for the g's depending on $o(n)$ coordinates. Physicists tend to assume that the equivalence always holds, although currently there is a move towards rigor. The leading researchers are Dobrushin, Lanford, and Ruelle. A convenient reference is Ruelle (1978, Chapter 1).

2. Paul Lévy

Lévy made extensive use of versions of theorem (1) for k finite in discussing means on function spaces. The material is first presented in

Chapter III of Lévy (1922). This is reprinted without essential change in Lévy (1951). Also see McKean (1973) who describes Brownian motion as "the uniform distribution on the sphere of radius $\sqrt{\infty}$".

We will attempt here to sketch the connection to theorem (1). Lévy's idea was to define the mean value M of U as the expected value of U (f), where U is a functional, and f is chosen at random on the unit sphere of $L_2[0,1]$. This is clearly insane, because there is no rotationally invariant countably additive probability on that sphere. Nothing daunted, Lévy defines M(U) for certain U by a limiting process. Following Gâteaux (1919), he discretizes [0,1] as $\left[0, \frac{1}{n}\right], \left(\frac{1}{n}, \frac{2}{n}\right], \ldots, \left(\frac{n-1}{n}, n\right]$ and considers the approximation f_n to f obtained by averaging f over these intervals. So f_n is constant on each interval. Let $a_1, \ldots, a_n$ be the values of f_n on the n intervals. Now $M_n(U)$ is defined as $E\{U(f_n)\}$, when f_n is chosen at random on the n-sphere $\frac{1}{n}\sum_{i=1}^{n} a_i^2 = 1$, i. e., $\int_0^1 f_n^2(t)\,dt = 1$. If $M_n(U)$ converges as $n \to \infty$, the limit is declared to be M(U). For example, suppose at least formally that $U(f) = \rho\left[f\left(\frac{1}{n}\right)\right]$ where ρ is a real-valued function. Then $M_n(U) = \rho\left[f_n\left(\frac{1}{2}\right)\right]$, whose expectation by (1) tends to

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(1/2)z^2}\,\rho(z)\,dz.$$

Similar things can be be done if U depends on f at $o(n)$ coordinates.

3. de Finetti

The original theorem (de Finetti, 1931) states that an exchangeable process of 0's and 1's is a unique mixture of coin-tossing processes. The move from $\{0,1\}$ to a compact Hausdorff space is due to Hewitt and Savage (1955), which also gives some history. The result is false for abstract spaces (Dubins and Freedman, 1979).

4. Recent literature on Poincaré's theorem

Stam (1982) proved a version of theorem (1) with an error bound in variation distance, assuming $k=o(\sqrt{n})$. He gives some interesting applications to geometrical probability theory. Gallardo (1983) and Yor (1985) gave an argument (with k fixed) using Brownian motion in n dimensions.

Freedman and Lane (1980, Lemmas 1 and 2) showed how to derive the convergence of an empirical distribution of dependent random variables ("the usual form of Maxwell's theorem") from information on the limiting behavior of pairwise joint distributions. Indeed, in the general setup of (2.13), suppose $s/n \to m_\lambda$. Then the empirical distribution of $x_1, \ldots, x_n$ converges weakly in Q_{sns}-probability to P_λ. This is because Q_{nsn} is exchangeable, and $Q_{ns2} \to P_\lambda^2$ by (2.13).

5. Recent literature on de Finetti's theorem

The infinite representation theorems were discussed in Freedman (1962, 1963), who gave characterizations for mixtures of the various exponential families. Diaconis and Freedman (1980) gave a finite form of de Finetti's theorem for $0-1$ variables, with an error bound. The present note is a sequel, carrying out the analysis in four additional families of distributions. Zaman (1986) has results for Markov chains. *See* Eaton (1981) on the normal case, and Diaconis-Eaton-Lauritzen (1986) for vector-valued random variables. Partial exchangeability is discussed from various perspectives in Diaconis and Freedman (1984), Aldous (1985), Lauritzen (1984), Ressell (1985). Finite versions of these results do not seem to be available in any degree of generality. Local central limit theorems are relevant, as in Martin-Lof (1970); or conditioned limit theorems, as in Csiszár (1984) and Zabell (1980).

6. Corrections

We would like to correct some errors in Diaconis and Freedman (1980). On p. 757, in equation (36) replace $O(k/n)$ by $o(k/n)$. In that equation and (41), assume $k \to \infty$, although the argument does give a result for $k=O(1)$. On p. 764, in the last remark, the urn is to contain $r_n=1$ red balls and $b_n=n-1$ black balls. Let H be the hypergeometric distribution for the number of red balls in k draws made at random without replacement from this urn. Let B_p be the binomial distribution with k trials and success probability p. Then $\|H - B_{1/n}\| \approx 2k^2/n^2$. The minimal value for $\|H - B_p\|$ is essentially k^2/n^2, for

$$p = \frac{1}{n} + \frac{k-1}{n^2} \quad \text{or} \quad p = \frac{1}{n} + \frac{k-1}{2n^2}.$$

REFERENCES

D. ALDOUS, Exchangeability and Related Topics, *Springer Lecture Notes in Math.*, No. 1196, 1985.

E. BOREL, *Introduction géométrique à quelques théories physiques*, Gauthier-Villars, Paris, 1914.

P. C. BILLINGSLEY, *Probability and Measure*, Wiley, New York, 1979.

I. CSISZÁR, Sanov Property, Generalized I Projections, and a Conditioned Limit Theorem, *Ann. Prob.*, Vol. **12**, 1984, p. 768-793.

B. DE FINETTI, Funzioni caratteristica di un fenomeno aleatorio, *Atti della R. Academia Nazionale dei Lincei Ser* 6, *Memorie Classe di Scienze Fisiche, Mathematiche e Naturali*, Vol. **4**, 1931, pp. 251-299.

B. DE FINETTI, La prévision : ses lois logiques, ses sources subjectives, *Ann. Inst. Henri Poincaré*, Vol. **7**, 1937, pp. 1-66.

P. DIACONIS, Finite Forms of de Finetti's Theorem on Exchangeability, *Synthèse*, Vol. **36**, 1978, pp. 271-281.

P. DIACONIS, M. EATON and S. LAURITZEN, (1986). *Finite de Finetti Theorems*, 1986, Unpublished manuscript.

P. DIACONIS and B. EFRON, Probabilistic-Geometric Theorems Arising from the Analysis of Contingency Tables, *Contributions to the Theory and Application of Statistics*, A. GELFAND Ed. Academic Press, Boston, pp. 103-125.

P. DIACONIS and D. FREEDMAN, Finite Exchangeable Sequences, *Ann. Prob.*, Vol. **8**, 1980, pp. 745-764.

P. DIACONIS and D. FREEDMAN, Partial Exchangeability and Sufficiency, in *Statistics, Applications, and New Directions*, J. K. GHOSH and J. ROY Eds., 1984, pp. 205-236, Indian Statistical Institute, Calcutta.

P. DIACONIS and D. FREEDMAN, *A Finite Version of de Finetti's Theorem for Exponential Families with Uniform Asymptotic Estimates*, Technical report, Department of Statistics, University of California, Berkeley, 1986.

L. DUBINS and D. FREEDMAN, Exchangeable Processes Need Not be Mixtures of Independent Identically Distributed Random Variables, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, Vol. **48**, 1979, pp. 115-132.

M. EATON, On the Projections of Isotropic Distributions, *Ann. Stat.*, **9**, 1981, pp. 391-400.

W. FELLER, *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd Edition, Wiley, New York, 1968.

W. FELLER, *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd Edition, Wiley, New York, 1971.

D. FREEDMAN, Invariants Under Mixing which Generalize de Finetti's Theorem, *Ann. Math. Statist.*, Vol. **33**, 1962, pp. 916-923.

D. FREEDMAN, Invariants Under Mixing which Generalize de Finetti's Theorem: Coninuous Time Parameter, *Ann. Math. Statist.*, Vol. **34**, 1963, pp. 1194-1216.

D. FREEDMAN and D. LANE, The Empirical Distribution of Fourier Coefficients, *Ann. Stat.*, Vol. **8**, 1980, pp. 1244-51.

L. GALLARDO, Au sujet du contenu probabiliste d'un lemma d'Henri Poincaré, *Annales de l'Université de Clermont*, Vol. **69**, 1983, pp. 192-197.

R. GÂTEAUX, Sur la notion d'intégrale dans le domaine fonctionnel et sur la théorie du potentiel, *Bull. Soc. Math. France*, Vol. **47**, 1919, pp. 47-67.

E. HEWITT and L. J. SAVAGE, Symmetric Measures on Cartesian Products, *Trans. Amer. Math. Soc.*, Vol. **80**, 1955, pp. 470-501.

L. HOLST, Two Conditional Limit Theorems with Applications, *Ann. Prob.*, Vol. **7**, 1979, pp. 1066-1072.

N. JOHNSON and N. L. KOTZ, *Urn Models and their Applications*, Wiley, New York, 1977.

J. KERSTEN, Verallgemeinerung eines Satzes von Prochrow und LeCam, *Zeitschrift für Wahrscheinlichkeistheorie und verwandte Gebiete*, Vo. **2**, 1963, pp. 173-179.

S. L. LAURITZEN, Extreme Point Models in Statistics, *Scand. Jour. Statist.*, Vol. **11**, 1984, pp. 65-91.

L. LECAM, Une théorème sur la division d'un intervalle par des points pris au hasard, *Publ. Inst. Statist. Universitaire Paris*, Vol. **7**, 1958, pp. 7-16.

G. LETAC, Isotropy and Sphericity: Some Characterizations, *Ann. Stat.*, Vol. **9**, 1981, pp. 408-417.

P. LÉVY, *Leçons d'Analyse Fonctionnelle*, Gauthier-Villars, Paris, 1922.

P. LEVY, *Analyse Fonctionnelle*, Mémorial des Sciences Mathématiques, Vol 5, Gauthier-Villars, Paris, 1925.

P. LÉVY, *Problèmes concrets d'analyse fonctionnelle*, Gauthier-Villars, Paris, 1951.

P. MARTIN-LOF, *Statistika Modeller Anteckningar fran seminarier läsåret*, 1969-70 urarbetade av R. Sunberg, Institutionen för matematisk statistik, Stockholm's universitet.

J. C. MAXWELL, *Theory of Heat*, 4th ed. Longmans, London, 1875.

J. C. MAXWELL, On Boltzmann's theorem on the average distribution of energy in a system of material points, *Cambridge Phil. Soc. Trans.*, Vol. **12**, 1878, p. 547.

H. MCKEAN, Geometry of differential space, *Ann. Prob.*, Vol. **1**, 1973, pp. 197-206.

H. POINCARÉ, *Calcul des probabilités*, Gauthier-Villars, Paris, 1912.

R. PYKE, Spacings, *J. Royal Statist. Soc.*, B, Vol. **7**, 1965, pp. 395-449.

P. RESSELL, de Finetti Type Theorems: an Analytic Approach, *Ann. Prob.*, Vol. **13**, 1985, pp. 898-922.

D. RUELLE, *Thermodynamic Formalism*, Addison Wesley, Reading, Mass., 1978.

I. J. SCHOENBERG, Metric Spaces and Positive Definite Functions, *Trans. Amer. Math. Soc.*, Vol. **44**, 1938, pp. 522-536.

A. J. STAM, Limit Theorems for Uniform Distributions on High Dimensional Euclidean Spaces, *J. Appl. Prob.*, Vol. **19**, 1982, pp. 221-228.

V. VERVAAT, Upper Bounds for the Distance in Total Variation Between the Binomial or Negative Binomial and the Poisson Distribution, *Statistica Neerlandica*, Vol. **23**, 1970, pp. 79-86.

M. YOR, Inégalités de martingales continus arrêtés à un temps quelconques I, *Springer Lecture Notes in Math.*, No. 1118, 1985.

S. ZABELL, Continuous Versions of Regular Conditional Distributions, *Ann. Prob.*, Vol. **7**, 1979, pp. 159-165.

S. ZABELL, Rates of Convergence for Conditional Expectations, *Ann. Prob.*, Vol. **8**, 1980, pp. 928-941.

A. ZAMAN, A Finite Form of de Finetti's Theorem for Stationary Markov Chains, *Ann. Prob.*, Vol. **14**, 1986, pp. 1418-27.

(*Manuscrit reçu le* 23 *décembre* 1986.)