ASYMPTOTIC BEHAVIOR OF BAYES' ESTIMATES[1]

By J. Fabius

Mathematisch Centrum, Amsterdam

0. Summary. This paper extends some of the results obtained by Freedman [2]. In Section 1 a class of prior distributions on the space of all substochastic distributions on the positive integers is given, such that along almost all sample sequences the corresponding posterior distributions of the expectations of all bounded functions on the positive integers are asymptotically normal. Section 2 shows that most of Freedman's results carry over to the case of distributions on the closed unit interval.

1. The discrete case. Let $(\Omega, \mathcal{A})$ be a measurable space and let $\{X_n, n \geqq 1\}$ be a sequence of measurable functions on it. As parameter space we use the space L of all substochastic distributions λ on the positive integers with the usual weak star topology and we put $\Lambda = \{\lambda \varepsilon L \mid \sum_{i=1}^{\infty} \lambda(i) = 1\}$, the dense subset of all proper probability distributions. We assume that there corresponds to each $\lambda \varepsilon \Lambda$ a probability P_λ on $\mathcal{A}$, under which $\{X_n, n \geqq 1\}$ is a sequence of independent identically distributed random variables with common distribution λ. The symbol π will always denote an element of Λ and will serve as the "true" value of the unknown parameter λ.

Let $\mathcal{L}$ denote the Borel σ-field in L, and let μ be a probability on $\mathcal{L}$. The topological carrier $C(\mu)$ of μ is defined to be the smallest compact set of μ-measure 1.

The measure μ will serve as prior distribution. The resulting posterior distribution given $\{X_1(\omega), \cdots, X_n(\omega)\}$ will be denoted by $\mu_{n,\omega}$, and is defined by

$$\mu_{n,\omega}(B) = \frac{\int_B \left\{\prod_{j=1}^{n} \lambda(X_j(\omega))\right\} \mu\,(d\lambda)}{\int_L \left\{\prod_{j=1}^{n} \lambda(X_j(\omega))\right\} \mu\,(d\lambda)} \tag{1.1}$$

for $B \varepsilon \mathcal{L}$ and nonvanishing denominator. Clearly, if defined, $\mu_{n,\omega}$ is a probability on $\mathcal{L}$ and $\mu_{n,\omega} \ll \mu$. If $\pi \varepsilon C(\mu)$, then the P_π-probability is one that $\mu_{n,\omega}$ is defined and $\pi \ll \pi'_{n,\omega}$, where $\pi'_{n,\omega}$ is the Bayes' estimate for π given $\{X_1(\omega), \cdots, X_n(\omega)\}$, defined by

$$\pi'_{n,\omega}(i) = \int_L \lambda(i)\mu_{n,\omega}\,(d\lambda), \qquad (i \geqq 1). \tag{1.2}$$

Received 5 June 1963; revised 22 November 1963.

[1] Revised version of the author's Ph. D. thesis at the University of California, Berkeley, prepared with the partial support of the Office of Naval Research, Contract Nonr-222-43. This paper in whole or in part may be reproduced for any purpose of the United States Government.

The following definition is a reformulation of Freedman's Definition 2.

DEFINITION 1.1. *A probability μ on $\mathcal{L}$ is tailfree if and only if there exists an integer $N \geqq 0$ and a sequence $\{\theta_k, k \geqq 1\}$ of independent random variables on some probability space independent of $(L, \mathcal{L}, \mu)$, such that $0 \leqq \theta_k \leqq 1$ for all $k \geqq 1$, and such that the conditional distribution under μ of $\{(1 - \sum_{i=1}^N \lambda(i))^{-1}\lambda(N + k), k \geqq 1\}$ given $\sum_{i=1}^N \lambda(i)$ on the set $\{\lambda \mid \sum_{i=1}^N \lambda(i) < 1\}$ coincides* a.s. $[\mu]$ *with the distribution of $\{\theta_k \prod_{i=1}^{k-1} (1 - \theta_i), k \geqq 1\}$.*

Freedman's Theorem 7 implies that under certain regularity conditions on the tailfree prior distribution μ the posterior distribution given $\{X_1(\omega), \cdots, X_n(\omega)\}$ of any finite linear combination of the form $n^{\frac{1}{2}} \sum_{i=1}^M a_i(\lambda(i) - \pi'_{n,\omega}(i))$ converges a.s. $[P_\pi]$ to the normal distribution with zero mean and variance $\sum_{i=1}^M a_i^2 \pi(i) - \{\sum_{i=1}^M a_i \pi(i)\}^2$ as $n \to \infty$. The following theorem specifies a class of tailfree prior distributions for which the same conclusion holds for infinite linear combinations with bounded coefficients.

THEOREM 1.1. *Let $\pi \varepsilon \Lambda$, and suppose that μ is a tailfree prior distribution with $N = 0$ and such that for every $k \geqq 1$, θ_k has a beta distribution with parameters r_k and s_k, where*

(i) $0 < r_k \leqq R < \infty$, $(k \geqq 1)$;

(ii) $0 < s_k \leqq r_{k+1} + s_{k+1}$, $(k \geqq 1)$;

(iii) $\sum_{k=1}^\infty r_k/(r_k + s_k) = \infty$.

Let $\{a_i, i \geqq 1\}$ be a sequence of real numbers with $\|a\| = \sup_i |a_i| < \infty$. Then the posterior distribution of $n^{\frac{1}{2}} \sum_{i=1}^\infty a_i(\lambda(i) - \pi'_{n,\omega}(i))$ given $\{X_1(\omega), \cdots, X_n(\omega)\}$ converges a.s. $[P_\pi]$ *to the normal distribution with zero mean and variance $\sum_{i=1}^\infty a_i^2 \pi(i) - \{\sum_{i=1}^\infty a_i \pi(i)\}^2$ as $n \to \infty$.*

Before proving the theorem we collect in the following lemma a few well-known facts concerning beta distributions, which we will use without further comment.

LEMMA 1.1. *Suppose that, for every $n \geqq 1$, Y_n is a random variable with a beta distribution with parameters p_n and q_n. Then*

$$EY_n = p_n/(p_n + q_n), \qquad \operatorname{Var} Y_n = p_n q_n/(p_n + q_n)^2(p_n + q_n + 1).$$

If

$$\lim_{n\to\infty} p_n/(p_n + q_n) = \gamma \quad \textit{and} \quad \lim_{n\to\infty} n/(p_n + q_n) = l < \infty,$$

then $\mathcal{L}(n^{\frac{1}{2}}(Y_n - EY_n)) \to N(0, l\gamma(1 - \gamma))$ as $n \to \infty$.

PROOF OF THEOREM 1.1. Since $\sum_{i=1}^\infty E\theta_i = \infty$ by (iii), we have

$$E\prod_{i=1}^\infty (1 - \theta_i) = \prod_{i=1}^\infty (1 - E\theta_i) = 0,$$

and hence $\prod_{i=1}^\infty (1 - \theta_i) = 0$ a.s. Consequently, $\mu(\Lambda) = 1$. Since obviously $C(\mu) = L$, Formula (1.1) defines $\mu_{n,\omega}$ a.s. $[P_\pi]$ for all $n \geqq 1$, and, since $\mu_{n,\omega} \ll \mu$ a.s. $[P_\pi]$, it follows that

$$\mu_{n,\omega}(\Lambda) = 1 \quad \text{a.s.} \quad [P_\pi]. \tag{1.3}$$

For every $n \geqq 1$, $\omega \varepsilon \Omega$, let $\{\theta_{n,\omega}(i), i \geqq 1\}$ be a sequence of independent random variables on a suitable probability space, such that, for every $i \geqq 1$, $\theta_{n,\omega}(i)$ has a beta distribution with parameters $r_i + n_i(\omega)$ and $s_i + m_i(\omega)$, where

$$n_i = \sum_{j=1}^{n} I_{[X_j=i]}, \qquad m_i = \sum_{j=1}^{n} I_{[X_j>i]}. \tag{1.4}$$

Let

$$\rho_{n,\omega}(i) = \theta_{n,\omega}(i)\prod_{k=1}^{i-1}(1-\theta_{n,\omega}(k)), \qquad (n, i \geqq 1, \omega \varepsilon \Omega). \tag{1.5}$$

Then the posterior distribution of $\{\lambda(i), i \geqq 1\}$ given $\{X_1(\omega), \cdots, X_n(\omega)\}$ coincides a.s. $[P_\pi]$ with the distribution of $\{\rho_{n,\omega}(i), i \geqq 1\}$.

We now compute, for fixed $n \geqq 1$ and $\omega \varepsilon \Omega$, the means, variances and covariances of the random variables $\{\rho_{n,\omega}(i), i \geqq 1\}$, using (1.5), the independence of the random variables $\{\theta_{n,\omega}(i), i \geqq 1\}$, and Lemma 1.1. We have

$$\pi'_{n,\omega}(i) = \frac{r_i + n_i(\omega)}{r_1 + s_1 + n}\prod_{k=1}^{i-1}\frac{s_k + m_k(\omega)}{r_{k+1} + s_{k+1} + m_k(\omega)}, \qquad (i \geqq 1). \tag{1.6}$$

Putting

$$\pi''_{n,\omega}(i) = \frac{r_i + n_i(\omega)}{r_1 + s_1 + n + 1}\prod_{k=1}^{i-1}\frac{s_k + m_k(\omega) + 1}{r_{k+1} + s_{k+1} + m_k(\omega) + 1}, \qquad (i \geqq 1), \tag{1.7}$$

we obtain

$$E\rho_{n,\omega}(i)^2 = \frac{r_i + n_i(\omega) + 1}{r_i + n_i(\omega)}\pi'_{n,\omega}(i)\pi''_{n,\omega}(i), \qquad (i \geqq 1),$$

$$E\rho_{n,\omega}(i)\rho_{n,\omega}(j) = \pi''_{n,\omega}(i)\pi'_{n,\omega}(j), \qquad (j > i \geqq 1),$$

and hence by straightforward calculation,

$$\operatorname{Var}\rho_{n,\omega}(i) = t_{n,\omega}(i) + u_{n,\omega}(i), \qquad (i \geqq 1), \tag{1.8}$$

$$\operatorname{Covar}(\rho_{n,\omega}(i), \rho_{n,\omega}(j)) = v_{n,\omega}(i,j) + w_{n,\omega}(i,j), \qquad (j > i \geqq 1), \tag{1.9}$$

where

$$t_{n,\omega}(i) = \frac{r_1 - r_i + s_1 + n - n_i(\omega)}{(r_1 + s_1 + n)(r_i + n_i(\omega))}\pi'_{n,\omega}(i)\pi''_{n,\omega}(i), \qquad (i \geqq 1),$$

$$u_{n,\omega}(i) = \frac{r_1 + s_1 + n + 1}{r_1 + s_1 + n}\pi'_{n,\omega}(i)\pi''_{n,\omega}(i)(1 - p_{n,\omega}(i)), \qquad (i \geqq 1),$$

$$v_{n,\omega}(i,j) = \frac{-1}{r_1 + s_1 + n}\pi''_{n,\omega}(i)\pi'_{n,\omega}(j) \qquad (j > i \geqq 1),$$

$$w_{n,\omega}(i,j) = \frac{r_1 + s_1 + n + 1}{r_1 + s_1 + n}\pi''_{n,\omega}(i)\pi'_{n,\omega}(j)(1 - p_{n,\omega}(i)), \qquad (j > i \geqq 1),$$

with

$$p_{n,\omega}(i) = \prod_{k=1}^{i-1}\left(1 - \frac{r_{k+1} + s_{k+1} - s_k}{(s_k + m_k(\omega) + 1)(r_{k+1} + s_{k+1} + m_k(\omega))}\right), \quad (i \geqq 1).$$

Keeping $i \geqq 1$ fixed and letting n tend to infinity, we have, by the strong law of large numbers.

$$\pi'_{n,\omega}(i) \to \pi(i) \quad \text{a.s.} \quad [P_\pi]$$

and, if $\pi(i) = 0$,

$$n^{\frac{1}{2}}\pi'_{n,\omega}(i) \to 0 \quad \text{a.s.} \quad [P_\pi],$$

since the same is true for $(r_i + n_i(\omega))/(r_1 + s_1 + n)$, while the product on the right side of (1.6) is bounded by 1 and converges to 1 a.s. $[P_\pi]$ if $\pi(i) > 0$. A similar argument shows that, for $i \geqq 1$, $\pi''_{n,\omega}(i) \to \pi(i)$ a.s. $[P_\pi]$ and, if $\pi(i) = 0$, $n^{\frac{1}{2}}\pi''_{n,\omega}(i) \to 0$ a.s. $[P_\pi]$ as $n \to \infty$. This implies that, for every $j > i \geqq 1$,

$$nt_{n,\omega}(i) \to \pi(i)(1 - \pi(i)) \quad \text{a.s.} \quad [P_\pi] \tag{1.10}$$

and

$$nv_{n,\omega}(i, j) \to -\pi(i)\pi(j) \quad \text{a.s.} \quad [P_\pi] \tag{1.11}$$

as $n \to \infty$. Moreover, for $j > i \geqq 1$,

$$nu_{n,\omega}(i) \to 0 \quad \text{a.s.} \quad [P_\pi] \tag{1.12}$$

and

$$nw_{n,\omega}(i, j) \to 0 \quad \text{a.s.} \quad [P_\pi] \tag{1.13}$$

as $n \to \infty$, since either $\sum_{k=i}^{\infty} \pi(k) = 0$ so that $n\pi'_{n,\omega}(i)\pi''_{n,\omega}(i)$ and $n\pi''_{n,\omega}(i)\pi'_{n,\omega}(j)$ tend to zero a.s. $[P_\pi]$ while $0 \leqq (1 - p_{n,\omega}(i)) \leqq 1$, or $\sum_{k=i}^{\infty} \pi(k) > 0$, in which case

$$\begin{aligned} 0 &\leqq n(1 - p_{n,\omega}(i)) \\ &\leqq \sum_{k=1}^{i-1} \frac{n(r_{k+1} + s_{k+1} - s_k)}{(s_k + m_k(\omega) + 1)(r_{k+1} + s_{k+1} + m_k(\omega))} \to 0 \quad \text{a.s.} \quad [P_\pi]. \end{aligned}$$

Hence, by (1.8) through (1.13), for all $i \geqq 1$,

$$\operatorname{Var}\left(n^{1/2} \sum_{k=1}^{i} a_k\, \rho_{n,\omega}(k)\right) \to \sum_{k=1}^{i} a_k^2\, \pi(k) - \left\{\sum_{k=1}^{i} a_k\, \pi(k)\right\}^2 \quad \text{a.s.} \quad [P_\pi] \tag{1.14}$$

as $n \to \infty$.

Since $\|a\| < \infty$,

$$\begin{aligned} \operatorname{Var}\left(\sum_{k=1}^{\infty} a_k\, \rho_{n,\omega}(k)\right) &= \sum_{k=1}^{\infty} a_k^2(t_{n,\omega}(k) + u_{n,\omega}(k)) \\ &\quad + 2\sum_{l=2}^{\infty}\sum_{k=1}^{l-1} a_k\, a_l(v_{n,\omega}(k, l) + w_{n,\omega}(k, l)) \end{aligned} \tag{1.15}$$

for all $n \geqq 1$, $\omega \varepsilon \Omega$, where both infinite sums on the right converge absolutely. Comparing (1.7) with (1.6), we see that $\pi''_{n,\omega}(i)$ is the Bayes' estimate for $\pi(i)$ given $\{X_1(\omega), \cdots, X_n(\omega)\}$ resulting from the prior distribution we obtain from μ on replacing s_k by $s_k + 1 (k \geqq 1)$. Hence

$$\sum_{k=1}^{\infty} \pi''_{n,\omega}(k) \leqq 1, \qquad (n \geqq 1, \omega \varepsilon \Omega).$$

Hence, by (i) and (ii), for $i \geqq 1$,

$$\begin{aligned} n\sum_{k=1}^{\infty} |t_{n,\omega}(k)| &\leqq (1 + R/n) \sum_{k=i+1}^{\infty} \pi''_{n,\omega}(k) \\ &\leqq (1 + R/n)(1 - \sum_{k=1}^{i} \pi''_{n,\omega}(k)) \to \sum_{k=i+1}^{\infty} \pi(k) \quad \text{a.s.} \quad [P_\pi] \end{aligned} \tag{1.16}$$

as $n \to \infty$, and

$$\begin{aligned} n \sum_{l=i+1}^{\infty} \sum_{k=1}^{l-1} |v_{n,\omega}(k, l)| &\leqq \sum_{l=i+1}^{\infty} \sum_{k=1}^{l-1} \pi''_{n,\omega}(k)\pi'_{n,\omega}(l) \\ &\leqq \sum_{l=i+1}^{\infty} \pi'_{n,\omega}(l) \to \sum_{k=i+1}^{\infty} \pi(k) \quad \text{a.s.} \quad [P_\pi] \end{aligned} \tag{1.17}$$

as $n \to \infty$. In view of (1.3), substitution of $a_k = 1 (k \geqq 1)$ in (1.15) gives

$$\sum_{k=1}^{\infty} (t_{n,\omega}(k) + u_{n,\omega}(k)) + 2\sum_{l=2}^{\infty} \sum_{k=1}^{l-1} (v_{n,\omega}(k, l) + w_{n,\omega}(k, l)) = 0 \quad \text{a.s.} \quad [P_\pi]$$

for all $n \geqq 1$. Since $u_{n,\omega}(k) \geqq 0$ and $w_{n,\omega}(k, l) \geqq 0$ for all $n \geqq 1$, $l > k \geqq 1$, $\omega \varepsilon \Omega$, it follows from (1.10), (1.11), (1.16) and (1.17) that

$$n\sum_{k=1}^{\infty} u_{n,\omega}(k) \to 0 \quad \text{a.s.} \quad [P_\pi] \tag{1.18}$$

and

$$n\sum_{k=2}^{\infty} \sum_{k=1}^{l-1} w_{n,\omega}(k, l) \to 0 \quad \text{a.s.} \quad [P_\pi] \tag{1.19}$$

as $n \to \infty$. Hence (1.15) through (1.19) imply

$$\lim_{i\to\infty} \lim_{n\to\infty} \operatorname{Var}\left(n^{\frac{1}{2}} \sum_{k=i+1}^{\infty} a_k\, \rho_{n,\omega}(k)\right) = 0 \quad \text{a.s.} \quad [P_\pi].$$

Thus, by an obvious modification of Slutsky's theorem (cf. Cramér [1], Section 20.6), our assertion will follow if we can show that, for all $i \geqq 1$,

$$\mathfrak{L}\left(n^{\frac{1}{2}}\sum_{k=1}^{i} a_k(\rho_{n,\omega}(k) - \pi'_{n,\omega}(k))\right) \to N\left(0, \sum_{k=1}^{i} a_k^2\, \pi(k) - \left\{\sum_{k=1}^{i} a_k\pi(k)\right\}^2\right) \tag{1.20}$$

a.s. $[P_\pi]$ as $n \to \infty$. In fact it suffices to show this for all $i \geqq 1$ such that

$$\sum_{k=i}^{\infty} \pi(k) > 0,$$

since the preceding computations imply that

$$\lim_{n\to\infty} \operatorname{Var}\left(n^{\frac{1}{2}}\sum_{k=i}^{\infty} a_k(\rho_{n,\omega}(k) - \pi'_{n,\omega}(k))\right) = 0 \quad \text{a.s.} \quad [P_\pi]$$

if $\sum_{k=i}^{\infty} \pi(k) = 0$. For $i = 1$, (1.20) follows from Lemma 1.1. Now suppose that (1.20) holds for some i such that $\sum_{k=i+1}^{\infty} \pi(k) > 0$. Then, by (1.5),

$$n^{\frac{1}{2}}\sum_{k=1}^{i+1} a_k(\rho_{n,\omega}(k) - \pi'_{n,\omega}(k)) = a_{i+1}\, n^{\frac{1}{2}}(\theta_{n,\omega}(i+1) - E\theta_{n,\omega}(i+1))$$
$$\cdot\left(1 - \sum_{k=1}^{i} \rho_{n,\omega}(k)\right) + n^{\frac{1}{2}}\sum_{k=1}^{i}(a_k - a_{i+1}\, E\theta_{n,\omega}(i+1))(\rho_{n,\omega}(k) - \pi'_{n,\omega}(k)),$$

which, by the induction hypothesis, has the same limiting distribution, if any, as

$$a_{i+1}\, n^{\frac{1}{2}}(\theta_{n,\omega}(i+1) - E\theta_{n,\omega}(i+1)) \sum_{k=i+1}^{\infty} \pi(k)$$
$$+ n^{\frac{1}{2}}\sum_{k=1}^{i}\left(a_k - \frac{a_{i+1}\,\pi(i+1)}{\sum_{j=i+1}^{\infty} \pi(j)}\right)(\rho_{n,\omega}(k) - \pi'_{n,\omega}(k)).$$

The two terms in this last expression are independent and, by Lemma 1.1 and the induction hypothesis, a.s. $[P_\pi]$ asymptotically normal. Hence the sum of these terms is a.s. $[P_\pi]$ asymptotically normal, and it is a matter of straightforward calculation to show that the limiting variance is given by (1.20) with i replaced by $i + 1$. Thus the proof is complete.

It is of some interest to note that Condition (iii), which was used to insure that $\mu(\Lambda) = 1$, is in fact equivalent to this.

One might expect the conclusion of Theorem 1.1 to hold for a much wider class of prior distributions than the one described in the theorem. However, the method of the proof given here breaks down even in the comparatively simple case where the θ_i have a common distribution with a positive twice continuously differentiable density on [0, 1].

2. The continuous case. We now turn to the case where the observable random variables take their values in the closed unit interval I. Thus $\{X_n, n \geqq 1\}$ is a sequence of measurable functions on $(\Omega, \mathfrak{A})$ to I, Λ is the space of all probabilities on I, and we assume that there corresponds to each $\lambda \in \Lambda$ a probability P_λ on $\mathfrak{A}$, under which $\{X_n, n \geqq 1\}$ is a sequence of independent identically distributed random variables with common distribution λ. As usual, we use the weak star topology in Λ, which in this case coincides with the topology of complete convergence. Since Λ is compact in this topology, we will have no need to consider any strictly substochastic measures on I. The Borel σ-field in Λ will be denoted by $\mathfrak{L}$.

LEMMA 2.1. *Let D be any countable dense subset of I. Then the sets*

$$N(\lambda, d, \epsilon) = \{\lambda' \in \Lambda \mid \lambda[0, d) - \epsilon \leqq \lambda'[0, d) \leqq \lambda'[0, d] \leqq \lambda[0, d] + \epsilon\}$$

with $\lambda \varepsilon \Lambda$, $d \varepsilon D$ *and* $\epsilon > 0$, *form a subbase for the topology in* Λ, *and* $\mathcal{L}$ *coincides with the* σ*-field* $\mathcal{L}_D$ *induced in* Λ *by the functions* $\lambda \to \lambda[0, d]$ $(d \varepsilon D)$.

PROOF. The first assertion follows from the definition of complete convergence. To prove the second assertion we note that $N(\lambda, d, \epsilon) \varepsilon \mathcal{L}_D$ for all $\lambda \varepsilon \Lambda$, $d \varepsilon D$, $\epsilon > 0$, so that $\mathcal{L}_D \supset \mathcal{L}$ (cf. Halmos [3], Theorem 51.C). On the other hand, the functions $\lambda \to \lambda[0, d]$ $(d \varepsilon D)$ are easily seen to be upper semicontinuous and hence $\mathcal{L}$-measurable. Thus $\mathcal{L} \supset \mathcal{L}_D$.

Let μ be a probability on $\mathcal{L}$ and let $C(\mu)$ be its topological carrier, defined as before. Again, μ will serve as prior distribution, but here the definition of the resulting posterior distribution given $\{X_1(\omega), \cdots X_n(\omega)\}$ is much more delicate than in the discrete case. We denote the product measurable space $(\Lambda \times \Omega, \mathcal{L} \times \mathcal{A})$ by $(\tilde{\Omega}, \tilde{\mathcal{A}})$, and we define a probability $\tilde{P}_\mu$ on it by

$$\tilde{P}_\mu(B \times A) = \int_B P_\lambda(A)\mu(d\lambda), \qquad (A \varepsilon \mathcal{A}, \quad B \varepsilon \mathcal{L}). \tag{2.1}$$

We shall assume that $\mathcal{A}$ coincides with the σ-field induced in Ω by $\{X_n, n \geqq 1\}$. Then the right side of (2.1) is defined since the integrand is $\mathcal{L}$-measurable.

For any $\tilde{\omega} = (\lambda, \omega) \varepsilon \tilde{\Omega}$ and any function ξ on Ω, we write

$$\tilde{\lambda}_{\tilde{\omega}} = \lambda, \qquad \tilde{\xi}(\tilde{\omega}) = \xi(\omega),$$

and, for any class $\mathcal{C}$ of subsets of Λ or Ω,

$$\tilde{\mathcal{C}} = \{C \times \Omega \mid C \varepsilon \mathcal{C}\} \quad \text{or} \quad \{\Lambda \times C \mid C \varepsilon \mathcal{C}\}$$

respectively.

DEFINITION 2.1. *A function* $\mu_{n,\omega}(B)$ *on* $(\Omega \times \mathcal{L})$ *to* I *is a posterior distribution given* $\{X_1(\omega), \cdots, X_n(\omega)\}$, *if and only if*

(i) *for every* $\omega \varepsilon \Omega$, *the function* $\mu_{n,\omega}(\cdot): B \to \mu_{n,\omega}(B)$ *on* $\mathcal{L}$ *is a probability,*

(ii) *for every* $B \varepsilon \mathcal{L}$, *the function* $\mu_{n,\cdot}(B): \omega \to \mu_{n,\omega}(B)$ *on* Ω *is* $\{X_1, \cdots, X_n\}$*-measurable,*

(iii) *for every* $B \varepsilon \mathcal{L}$,

$$\mu_{n,\cdot}(B) = \tilde{P}_\mu(B \times \Omega \mid \tilde{X}_1, \cdots, \tilde{X}_n) \quad \text{a.s.} \quad [\tilde{P}_\mu].$$

LEMMA 2.2. *There always exists a posterior distribution given* $\{X_1(\omega), \cdots, X_n(\omega)\}$.

PROOF. By Definition 2.1 the existence of such a posterior distribution is equivalent to the existence of a mixed conditional distribution relative to $\tilde{P}_\mu$ of $\tilde{\lambda}$ given $\{\tilde{X}_1, \cdots, \tilde{X}_n\}$, which is guaranteed, since by Lemma 2.1 the σ-field $\tilde{\mathcal{L}}$ is induced by a countable family of random variables on the probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}_\mu)$ (cf. Loève [4], 27.2.A).

Although the preceding lemma asserts the existence, it by no means asserts uniqueness of posterior distributions. Usually there will be many different posterior distributions, and, for each $n \geqq 1$ the statistician will have to select a particular one.

We shall use the following notation. $\mathcal{A}_n$ denotes the σ-field induced in Ω by

$\{X_1, \cdots, X_n\}$. $\sum_n$ is the set of all permutations of the integers $\{1, \cdots, n\}$, and $\mathcal{S}_n$ is the σ-field of all symmetric sets in $\mathcal{A}_n$, i.e., a set $A \subset \Omega$ is in $\mathcal{A}_n$ if and only if $A = [(X_1, \cdots, X_n) \varepsilon S]$ for some Borel set S in the closed n-dimensional unit cube, and $A \varepsilon \mathcal{S}_n$ if and only if in addition $\sigma A = A$ for all $\sigma \varepsilon \sum_n$, where $\sigma[(X_1, \cdots, X_n) \varepsilon S] = [(X_{\sigma(1)}, \cdots, X_{\sigma(n)}) \varepsilon S]$.

LEMMA 2.3. *There always exists a posterior distribution $\mu_{n,\omega}$ given $\{X_1(\omega), \cdots, X_n(\omega)\}$ which is invariant under all permutations of $\{X_1, \cdots, X_n\}$.*

PROOF. Let $B \varepsilon \mathcal{L}$, and let g be a bounded Borel function on the n-dimensional closed unit cube. Then we have, indicating expectations relative to $\tilde{P}_\mu$ by $\tilde{E}_\mu$, and setting $\tilde{B} = B \times \Omega$,

$$\begin{aligned} \tilde{E}_\mu(I_{\tilde{B}} g(\tilde{X}_1, \cdots, \tilde{X}_n)) &= \tilde{E}_\mu\{I_{\tilde{B}} \tilde{E}_\mu(g(\tilde{X}_1, \cdots, \tilde{X}_n) \mid \tilde{\mathcal{L}})\} \\ &= (1/n!)\tilde{E}_\mu\Big\{I_{\tilde{B}} \sum_{\sigma \varepsilon \Sigma_n} \tilde{E}_\mu(g(\tilde{X}_{\sigma(1)}, \cdots, \tilde{X}_{\sigma(n)}) \mid \tilde{\mathcal{L}})\Big\} \\ &= (1/n!)\tilde{E}_\mu\Big(I_{\tilde{B}} \sum_{\sigma \varepsilon \Sigma_n} g(\tilde{X}_{\sigma(1)}, \cdots, \tilde{X}_{\sigma(n)})\Big) \\ &= (1/n!)\tilde{E}_\mu\Big\{\tilde{P}_\mu(\tilde{B} \mid \tilde{\mathcal{S}}_n) \sum_{\sigma \varepsilon \Sigma_n} g(\tilde{X}_{\sigma(1)}, \cdots, \tilde{X}_{\sigma(n)})\Big\} \\ &= \tilde{E}_\mu\{\tilde{P}_\mu(\tilde{B} \mid \tilde{\mathcal{S}}_n) g(\tilde{X}_1, \cdots, \tilde{X}_n)\}. \end{aligned}$$

Hence

$$\tilde{P}_\mu(\tilde{\mathcal{B}} \mid \tilde{\mathcal{A}}_n) = \tilde{P}_\mu(\tilde{B} \mid \tilde{\mathcal{S}}_n) \quad \text{a.s.} \quad [P_\mu],$$

for all $B \varepsilon \mathcal{L}$. Consequently, the argument used in the proof of Lemma 2.2 shows the existence of a posterior distribution $\mu_{n,\omega}$ given $\{X_1(\omega), \cdots, X_n(\omega)\}$ such that, for every $B \varepsilon \mathcal{L}$, the function $\mu_{n,\cdot}(B) : \omega \to \mu_{n,\omega}(B)$ is $\mathcal{S}_n$-measurable.

Next we single out a class of prior distributions which have a special structure, similar to that of the tailfree prior distributions in the discrete case. First a few auxiliary definitions and conventions.

DEFINITION 2.2. *A tree of partitions is a sequence $\{T_s, s \geqq 0\}$ of finite partitions of I in nonempty disjoint intervals, such that*

(i) $T_0 = \{I\}$,

(ii) T_{s+1} *is a refinement of* $T_s (s \geqq 0)$,

(iii) $\max_{J \varepsilon T_s} |J| \to 0$ *as* $s \to \infty$, *where* $|J|$ *denotes the length of the interval* J.

If $\{T_s, s \geqq 0\}$ is a tree of partitions, we define $T_{s,J'} = \{J \varepsilon T_s \mid J \subset J'\}$ $(s \geqq 1, J' \varepsilon T_{s-1})$, and we denote the σ-field induced in Λ by the functions $\lambda \to \lambda(J)(J \varepsilon T_s)$, by $\mathfrak{I}_s (s \geqq 1)$.

DEFINITION 2.3. *A probability μ on $\mathcal{L}$ is tailfree if and only if there exists a tree of partitions $\{T_s, s \geqq 0\}$ and a family of nonnegative random variables $\{\theta_{s,J}, s \geqq 1, J \varepsilon T_s\}$ on some probability space independent of $(\Lambda, \mathcal{L}, \mu)$, such that*

(i) $\sum_{J \varepsilon T_{s,J'}} \theta_{s,J} = 1, (s \geqq 1, J' \varepsilon T_{s-1})$,

(ii) *the families* $\{\theta_{s,J}, J \varepsilon T_s\}$, $(s \geqq 1)$ *are independent,*

(iii) *for every $s \geqq 1$, the distribution of $\{\lambda(J), J \varepsilon T_s\}$ under μ, coincides with the distribution of $\{\rho_{s,J}, J \varepsilon T_s\}$, where $\rho_{s,J} = \prod_{r=1}^{s} \theta_{r,J_r}$, with $J_r \varepsilon T_r$, $J_r \supset J$ for $1 \leqq r \leqq s$.*

Just as in the discrete case (cf. Freedman [2], Section 6), it is possible to give an alternative definition of a description nature, using the notation

$$n_J(\omega) = \sum_{j=1}^{n} I_{[X_j \varepsilon J]}(\omega) \quad \text{for } J \subset I,\ \omega \varepsilon \Omega,\ n \geqq 1.$$

DEFINITION 2.4. *A probability μ on $\mathfrak{L}$ is tailfree if and only if there exists a tree of partitions $\{T_s,\ s \geqq 0\}$ and, for every $n \geqq 1$, a posterior distribution $\mu_{n,\omega}(B)$ given $\{X_1(\omega), \cdots, X_n(\omega)\}$, such that, for every $s \geqq 1$, $B \varepsilon \mathfrak{J}_s$, $\mu_{n,\omega}(B)$ depends on $\omega \varepsilon \Omega$ only through $\{n_j,\ J \varepsilon T_s\}$.*

THEOREM 2.1. *Definitions* 2.3 *and* 2.4 *are equivalent.*

PROOF. Let μ be tailfree according to Definition 2.3. For $s \geqq 1$, we write

$$C_s = \{x = (x_J,\ J \varepsilon T_s) \mid x_J \geqq 0,\ J \varepsilon T_s;\ \sum_{J \varepsilon T_s} x_j = 1\},$$

i.e., C_s is the simplex in which the random vector $\{\rho_{s,J},\ J \varepsilon T_s\}$ takes its values. Taking C to be a Borel set in C_s, we have then, for every nonnegative integer r and all $s \geqq 1$,

$$\tilde{P}_\mu((\tilde{\lambda}(J),\ J \varepsilon T_s) \varepsilon C \mid \tilde{n}_{J'},\ J' \varepsilon T_{s+r}) = \tilde{P}((\tilde{\lambda}(J),\ J \varepsilon T_s) \varepsilon C \mid \tilde{n}_J,\ J \varepsilon T_s) \quad \text{a.s.} \quad [\tilde{P}_\mu] \tag{2.2}$$

and hence, since $\tilde{\mathfrak{S}}_n$ is the σ-field induced in $\tilde{\Omega}$ by $\{\tilde{n}_J,\ J \varepsilon \bigcup_{s=1}^{\infty} T_s\}$, by letting $r \to \infty$, we obtain

$$\tilde{P}_\mu((\tilde{\lambda}(J),\ J \varepsilon T_s) \varepsilon C \mid \tilde{\mathfrak{S}}_n) = \tilde{P}_\mu((\tilde{\lambda}(J),\ J \varepsilon T_s) \varepsilon C \mid \tilde{n}_J,\ J \varepsilon T_s). \tag{2.3}$$

But this implies that μ is tailfree according to Definition 2.4, by virtue of Lemma 2.3 and the argument used in the proof of Lemma 2.2.

Now let μ be tailfree according to Definition 2.4. Then (2.3) and hence also (2.2) holds for all $s \geqq 1$, $r \geqq 0$ and all Borel sets C in C_s. Thus, taking $r = 1$ and putting $C' = [(\lambda(J),\ J \varepsilon T_s) \varepsilon C]$, we have for all $y \varepsilon C_{s+1}$, $z \varepsilon C_s$ such that $z_J = \sum_{J' \varepsilon T_{s+1,J}} y_{J'}$, $(J \varepsilon T_s)$,

$$\tilde{P}_\mu[\tilde{\lambda} \varepsilon C;\ \tilde{n}_{J'} = ny_{J'},\ J' \varepsilon T_{s+1}]\tilde{P}_\mu[\tilde{n}_J = nz_J,\ J \varepsilon T_s] = \tilde{P}_\mu[\tilde{\lambda} \varepsilon C;\ \tilde{n}_J = nz_J,\ J \varepsilon T_s]\tilde{P}_\mu[\tilde{n}_{J'} = ny_{J'},\ J' \varepsilon T_{s+1}].$$

Hence

$$\int_{C'} \{\prod_{J' \varepsilon T_{s+1}} \lambda(J')^{ny_{J'}}\}\ \mu(d\lambda)$$
$$= \int_\Lambda \{\prod_{J' \varepsilon T_{s+1}} \lambda(J')^{ny_{J'}}\}\ \mu(d\lambda) \int_{C'} \{\prod_{J \varepsilon T_s} \lambda(J)^{nz_J}\}\ \mu(d\lambda) \Big/ \int_\Lambda \{\prod_{J \varepsilon T_s} \lambda(J)^{nz_J}\} \mu(d\lambda)$$

provided the denominator on the right side does not vanish. If it does vanish, then

$$\prod_{J' \varepsilon T_{s+1}} \lambda(J')^{ny_{J'}} = \prod_{J \varepsilon T_s} \lambda(J)^{nz_J} = 0 \quad \text{a.s.} \quad [\mu].$$

Therefore the conditional expectation relative to μ of $\prod_{J' \varepsilon T_{s+1}} \lambda(J')^{ny_{J'}}$ given

$\mathfrak{I}_s$ is a.s. $[\mu]$ proportional to $\prod_{J \varepsilon T_s} \lambda(J)^{nz_J}$. But this is exactly Definition 2.3, restated in terms of moments.

If μ is a tailfree prior distribution, then (2.3) and Lemma 2.3 assert that there exists a posterior distribution $\mu_{n,\omega}$ satisfying

$$(2.4) \quad \mu_{n,\omega}(B) = \int_B \{\prod_{J \varepsilon T_s} \lambda(J)^{n_J(\omega)}\} \mu(d\lambda) \Big/ \int_\Lambda \{\prod_{J \varepsilon T_s} \lambda(J)^{n_J(\omega)}\} \mu(d\lambda)$$

for $s \geqq 1$, $B \varepsilon \mathfrak{I}_s$, provided the denominator does not vanish. In the discrete case, for any $\pi \varepsilon \Lambda$, (1.1) actually defined $\mu_{n,\omega}(B)$ for all $B \varepsilon \mathfrak{L}$ a.s. $[P_\pi]$, provided $\pi \varepsilon C(\mu)$. Here the requirement that $\pi \varepsilon C(\mu)$ is not sufficient to insure that (2.4) determines $\mu_{n,\omega}$ on $\mathfrak{L}$ a.s. $[P_\pi]$. If, for any $\lambda \varepsilon \Lambda$, $s \geqq 1$, λ_s denotes the probability on T_s defined by $\lambda_s(J) = \lambda(J)$ $(J \varepsilon T_s)$, and if m_s is the distribution of λ_s under μ, so that $m_s(C) = \mu(C')$ for any Borel set C in C_s, then, for any $s \geqq 1$, the right side of (2.4) is well defined a.s. $[P_\pi]$ if $\pi_s \varepsilon C(m_s)$, the topological carrier of m_s in C_s. Hence, if $\pi_s \varepsilon C(m_s)$ for all $s \geqq 1$, then (2.4) defines $\mu_{n,\omega}(B)$ a.s. $[P_\pi]$ for $B \varepsilon \bigcup_{s=1}^\infty \mathfrak{I}_s$, and hence it determines $\mu_{n,\omega}(B)$ a.s. $[P_\pi]$ for $B \varepsilon \mathfrak{L}$. From now on, we shall always assume that this is the case, and we shall refer to $\mu_{n,\omega}$ determined by (2.4) as "the" posterior distribution given $\{X_1(\omega), \cdots, X_n(\omega)\}$.

In order to give a definition of consistency similar to Freedman's definition for the discrete case, we introduce the weak star topology in the space of all probability measures on $\mathfrak{L}$, so that a sequence $\{\mu_n, n \geqq 1\}$ of such measures converges to a probability μ_0 on $\mathfrak{L}$ and only if

$$\int_\Lambda f(\lambda)\mu_n(d\lambda) \to \int_\Lambda f(\lambda)\mu_0(d\lambda)$$

as $n \to \infty$, simultaneously for all continuous functions f on Λ.

DEFINITION 2.5. *If μ is a probability on $\mathfrak{L}$ and $\pi \varepsilon \Lambda$, the pair (π, μ) is said to be consistent if and only if, as $n \to \infty$, a.s. $[P_\pi]$, $\mu_{n,\omega} \to \delta_\pi$, the probability on $\mathfrak{L}$ which has all its mass concentrated at π.*

THEOREM 2.2. *Let $\pi \varepsilon \Lambda$ and let μ be a tailfree prior distribution, such that $\pi_s \varepsilon C(m_s)$ for all $s \geqq 1$. Then (π, μ) is consistent.*

PROOF. We have to show that as $n \to \infty$ a.s. $[P_\pi]$,

$$(2.5) \qquad \int_\Lambda f(\lambda)\mu_{n,\omega}(d\lambda) \to f(\pi),$$

simultaneously for all continuous functions f on Λ. For any $i \geqq 0$, the function $\phi_i(\lambda) = \int_I x^i \lambda(dx)$ is continuous on Λ by the very definition of the topology in Λ. Since the family of functions $\{\phi_i, i \geqq 0\}$ separates the points of Λ and contains the constant function $\phi_0 \equiv 1$, the algebra generated by these functions is dense in the sense of the uniform norm in the space of all continuous functions on Λ. Thus it suffices to prove that a.s. $[P_\pi]$

$$(2.6) \qquad \int \left\{\prod_{k=1}^K \phi_{i_k}(\lambda)\right\} \mu_{n,\omega}(d\lambda) \to \prod_{k=1}^K \phi_{i_k}(\pi)$$

as $n \to \infty$, for all finite K and all nonnegative integers $i_1, \cdots, i_K$.

For any fixed $K < \infty, i_1, \cdots, i_K \geqq 0$, and any $\epsilon > 0$, there exists an $s_0 \geqq 0$, such that

$$(2.7) \qquad \left|\prod_{k=1}^{K} \phi_{i_k}(\lambda) - \prod_{k=1}^{K} \{\sum_{J \varepsilon T_s} x_J^{i_k} \lambda(J)\}\right| < \epsilon,$$

uniformly for $s \geqq s_0$, $\lambda \varepsilon \Lambda$, $x_J \varepsilon J$ $(J \varepsilon T_s)$. For any $s \geqq 1$, $\mu_{n,\omega}$ restricted to $\mathfrak{J}_s$ may be regarded as the posterior distribution of λ_s given $\{n_J(\omega), J \varepsilon T_s\}$ resulting from the prior distribution m_s on the Borel sets of C_s. Thus, applying Freedman's results for the finite discrete case (cf. [2], Theorem 1), we obtain, since $\pi_s \varepsilon C(m_s)$, that a.s. $[P_\pi]$

$$(2.8) \qquad \int_\Lambda \left[\prod_{k=1}^{K} \{\sum_{J \varepsilon K_s} x_J^{i_k} \lambda(J)\}\right] \mu_{n,\omega}(d\lambda) \to \prod_{k=1}^{K} \{\sum_{J \varepsilon T_s} x_J^{i_k} \pi(J)\}$$

as $n \to \infty$. Together with (2.7) this gives: for every $K < \infty$ and all $i_1, \cdots, i_K \geqq 0$, and for any $\epsilon > 0$

$$(2.9) \qquad \limsup_{n\to\infty} \left|\int_\Lambda \left\{\prod_{k=1}^{K} \phi_{i_k}(\lambda)\right\} \mu(d\lambda) - \prod_{k=1}^{K} \phi_{i_k}(\pi)\right| \leqq 2\epsilon \quad \text{a.s.} \quad [P_\pi],$$

and hence (2.6) follows if we let $\epsilon \downarrow 0$ along a countable sequence, since the set of K-tuples of nonnegative integers with K finite is countable.

The crucial point of the preceding proof is of course the fact that we can apply Freedman's results for the finite discrete case to $\mu_{n,\omega}$ restricted to $\mathfrak{J}_s$. This actually makes it possible to carry over all of Freedman's results. Thus, under the assumptions of Theorem 2.2 and suitable regularity conditions on m_s, the posterior distribution of $\{n^{\frac{1}{2}}(\lambda(J) - n^{-1}n_J(\omega)), J \varepsilon T_s\}$ is a.s. $[P_\pi]$ asymptotically normal, and

$$n^{\frac{1}{2}}(\pi'_{n,\omega}(J) - n_J(\omega)/n) \to 0 \quad \text{a.s.} \quad [P_\pi]$$

for all $J \varepsilon T_s$ as $n \to \infty$, where

$$\pi'_{n,\omega}(J) = \int_\Lambda \lambda(J)\mu_{n,\omega}(d\lambda), \qquad (J \varepsilon T_s)$$

is the usual Bayes' estimate for $\pi(J)$.

In view of the results of Section 1 one might conjecture that also the posterior distribution of $\int_I g(x)\lambda(dx)$ is a.s. $[P_\pi]$ asymptotically normal if g is any continuous function on I and if μ belongs to some special class of tailfree prior distributions. This however remains an open question.

The author wishes to express his gratitude to Professor D. Blackwell and Professor D. A. Freedman for their interest and many helpful and challenging discussions.

REFERENCES

[1] Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.
[2] Freedman, D. A. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. *Ann. Math. Statist.* **34** 1386–1403.
[3] Halmos, P. R. (1950). *Measure Theory*. Van Nostrand, New York.
[4] Loève, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, New York.