The Annals of Statistics
1977, Vol. 5, No. 1, 79-87

COMPOUND MULTINOMIAL LIKELIHOOD FUNCTIONS ARE UNIMODAL: PROOF OF A CONJECTURE OF I. J. GOOD

BY BRUCE LEVIN AND JAMES REEDS

Columbia University and Harvard University

I. J. Good's 1965 conjecture of the unimodality of the likelihood function of a symmetrical compound multinomial distribution is proved by the variation-diminishing property of the Laplace transform. The result is a special case of a several sample version with asymmetrical compounding Dirichlet distributions. The technique of proof is applied to yield similar results for the negative binomial distribution and a two point mixture of Poissons.

1. Introduction. Let $\mathbf{n} = (n_1, \cdots, n_t)$ be a sample of size $N = \sum_{j=1}^t n_j$ from a t-category multinomial distribution with parameters $\mathbf{p} = (p_1, \cdots, p_t)$. If $\mathbf{p}$ has a Dirichlet (generalized beta) distribution with parameters $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_t)$ then marginally the counts $(n_1, \cdots, n_t)$ have a compound multinomial distribution which we shall denote by $\mathbf{n} \sim CM(N, t, \boldsymbol{\alpha})$, or by $\mathbf{n} \sim CM(\boldsymbol{\alpha})$ when N and t are fixed. The probability mass function of the $CM(\boldsymbol{\alpha})$ distribution is

$$\Pr(\mathbf{n} \mid \boldsymbol{\alpha}) = \left(\frac{N!}{\prod_{j=1}^t n_j!}\right)\left(\frac{\Gamma(\sum_{j=1}^t \alpha_j)}{\prod_{j=1}^t \Gamma(\alpha_j)}\right)\left(\frac{\prod_{j=1}^t \Gamma(n_j + \alpha_j)}{\Gamma(N + \sum_{j=1}^t \alpha_j)}\right).$$

When $t = 2$ this is the beta-binomial, or Pólya, distribution which arises from the Pólya urn scheme (Feller (1968), page 120). The likelihood function for the $CM(\boldsymbol{\alpha})$ distribution is

$$L(\boldsymbol{\alpha} \mid \mathbf{n}) = (\Gamma(\textstyle\sum_{j=1}^t \alpha_j)/\prod_{j=1}^t \Gamma(\alpha_j))(\prod_{j=1}^t \Gamma(n_j + \alpha_j)/\Gamma(N + \sum_{j=1}^t \alpha_j)).$$

Assume $\boldsymbol{\alpha} = k\boldsymbol{\lambda}$ for known $\boldsymbol{\lambda}$ with $\lambda_1 + \cdots + \lambda_t = 1$. Given a sample of m independent observations $\mathbf{n}_1, \cdots, \mathbf{n}_m$ from the $CM(N, t, \boldsymbol{\alpha})$ distribution, we may view the likelihood function

$$L(k) = \textstyle\prod_{i=1}^m L(\boldsymbol{\alpha} \mid \mathbf{n}_i) \tag{1.1}$$

as a function of k, whose domain may be extended to $(0, \infty]$ by continuity:

$$L(\infty) = \lim_{k\to\infty} \textstyle\prod_{i=1}^m L(k\boldsymbol{\lambda} \mid \mathbf{n}_i) = \prod_{i=1}^m \prod_{j=1}^t \lambda_j^{n_{ij}}.$$

(Notation: $\mathbf{n}_i = (n_{i1}, \cdots, n_{it})$.) In the special case $m = 1$ and $\alpha_1 = \alpha_2 = \cdots = \alpha_t = k/t$ we have

$$\begin{aligned} L(k) &= (\Gamma(k)/\Gamma(k/t)^t)(\textstyle\prod_{j=1}^t \Gamma(n_j + k/t)/\Gamma(N + k)) \\ &= \textstyle\prod_{j=1}^t \prod_{h=0}^{n_j-1} (h + k/t)/\prod_{h=0}^{N-1} (h + k) \qquad \text{if} \quad k < \infty \\ &= t^{-N} \qquad \text{if} \quad k = \infty . \end{aligned} \tag{1.2}$$

Received January 1975; revised May 1976.
AMS 1970 *subject classifications.* Primary 62F10, 62F15; Secondary 44A10, 44A35.
Key words and phrases. Maximum likelihood estimate, unimodality, multinomial distribution, Dirichlet distribution, negative binomial distribution, variation-diminishing, Laplace transform.

Empty products in (1.2) and throughout this paper are to be interpreted as unity, and empty sums when they appear are to be interpreted as zero.

I. J. Good (1965, 1967, 1975) and Good and Crook (1974) have used these likelihood functions in estimating multinomial parameters. Good (1965, page 37) conjectured that the likelihood (1.2) is unimodal:

> "Given a Type I sample $(n_1, \cdots, n_t)$, the Type II log-likelihood of the flattening constant k $(k > 0)$, that is, of the symmetrical Dirichlet Type II distribution of parameter k when regarded as a function of k, has at most one local maximum. It takes its maximum at $k = \infty$ if $\chi^2 < t - 1$ and for a finite value of k if $\chi^2 > t - 1$, where
>
> $$\chi^2 = (t/N) \textstyle\sum_{i=1}^{t} (n_i - N/t)^2 \text{."}$$

As Good (1965) showed, the conjecture is true if $(d/dk) \log L(k) = 0$ has at most one root, but until now no complete proof guaranteeing this has appeared. Our proof, which uses the variation-diminishing property of the Laplace transform, shows the relevance of the χ^2 condition in a natural way.

Theorem 1 of this paper yields, as a special case, a proof of Good's conjecture. Theorem 2, an extension to a multivariate negative binomial distribution, answers Anscombe's question ((1950), page 367) about the uniqueness of the maximum likelihood estimate of the exponential parameter of the negative binomial distribution. Theorem 3 shows that the likelihood function for the mixing parameter of a mixture of two known Poissons is unimodal. An appendix gives an example showing that the "smoothing step" in the proof of Theorem 1 is indispensable.

2. Main result.

THEOREM 1. *The likelihood function* (1.1) *has at most one local maximum. It occurs for finite k if*

$$\chi^2 > \textstyle\sum_{j=1}^{t} \tau_j(n_{+j}/N) - m \tag{2.1}$$

and for $k = \infty$ *otherwise, where*

$$\chi^2 = \textstyle\sum_{i=1}^{m} \sum_{j=1}^{t} (n_{ij} - N\lambda_j)^2/N\lambda_j$$

and for $j = 1, \cdots, t$, $\tau_j = \lambda_j^{-1}$ *and* $n_{+j} = \sum_{i=1}^{m} n_{ij}$.

If (2.1) *is satisfied, every derivative of* $\log L(k)$ *has exactly one zero in* $(0, \infty)$; *if* (2.1) *is not satisfied, none of the derivatives of* $\log L(k)$ *have any zeros in* $(0, \infty)$.

PROOF OF THEOREM 1. Write

$$L(k) = \textstyle\prod_{i=1}^{m} \prod_{j=1}^{t} \prod_{h=0}^{n_{ij}-1} (h + k\lambda_j) / \prod_{h=0}^{N-1} (h + k)^m \,. \tag{2.2}$$

The logarithmic derivative of (2.2) is

$$\frac{d}{dk} \log L(k) = \textstyle\sum_{i=1}^{m} \sum_{j=1}^{t} \sum_{h=0}^{n_{ij}-1} 1/(k + h\tau_j) - \sum_{h=0}^{N-1} m/(k + h) \,. \tag{2.3}$$

This is the Stieltjes transform of a certain signed measure:

$$\frac{d}{dk}\log L(k) = \int_{0-}^{\infty} 1/(k+u)\,dG(u) - \int_{0-}^{\infty} 1/(k+u)\,dF(u)$$
$$= \int_{0-}^{\infty} 1/(k+u)\,d(G-F)(u)$$

where F is a measure placing mass m at each of the points $u = 0, 1, \cdots, N-1$ and G is a sum of measures G_{ij} placing unit mass at the points $u = h\tau_j$ for $0 \leqq h < n_{ij}$. (If $n_{ij} = 0$ then G_{ij} is the zero measure.) The lower limit of integration indicates any point to the left of zero; since $G - F$ assigns no mass for $u < 0$, this point is arbitrary.

The general strategy of the proof is to apply the well-known

LEMMA 2.1. *If the function $A(u)$ has Laplace transform $\mathscr{L}_A(y) = \int_0^\infty e^{-uy}A(u)\,du$, then the number of roots of $\mathscr{L}_A(y)$ in $(0, \infty)$ is not greater than the number of sign changes of $A(u)$ in $(0, \infty)$.*

PROOF. For the general result, see Karlin (1968) or Pólya and Szegö (1964). We only make use of the special case where $A(u)$ has at most one sign change, in which case it is easy to verify the lemma making use of the monotone ratio property of e^{-uy}. □

As a matter of notation, if M is a finite, signed measure on $[0, \infty)$ we shall write

$$\mathscr{L}_{dM}(y) = \int_{0-}^{\infty} e^{-uy}\,dM(u)$$

and

$$\mathscr{L}_M(y) = \int_0^\infty e^{-uy}M(u)\,du\,, \qquad \text{where} \quad M(u) = \int_{0-}^{u} dM\,.$$

We freely use the relation $\mathscr{L}_{dM}(y) = y \cdot \mathscr{L}_M(y)$ for all $y > 0$.

Returning to the proof of Theorem 1, note that

$$\int_0^\infty e^{-(k+u)y}\,dy = 1/(k+u)$$

whence

$$\frac{d}{dk}\log L(k) = \int_0^\infty e^{-ky} \int_{0-}^{\infty} e^{-uy}\,d(G-F)(u)\,dy$$
$$= \int_0^\infty e^{-ky}\phi(y)\,dy = \mathscr{L}_\phi(k)$$

where

$$\phi(y) = \int_{0-}^{\infty} e^{-uy}\,d(G-F)(u) = \mathscr{L}_{d(G-F)}(y)\,,$$

which is the Laplace transform of the signed measure $G - F$. Once we show that $\phi(y)$ has at most one sign change in $(0, \infty)$, we apply Lemma 2.1 and we are done. The "distribution" function $G(u) - F(u) = \int_{0-}^{u} d(G-F)$ has many sign changes, however, and in order to prove that $\phi(y)$ behaves as desired, we must smooth $G(u) - F(u)$.

First convolute $G(u) - F(u)$ with the distribution function of a uniform random variable on $[0, 1]$ and call the result $(G^* - F^*)(u)$. Thus

$$(G^* - F^*)(u) = U(u) * (G-F)(u) \qquad \text{where} \quad U(u) = \begin{cases} 0 & u < 0 \\ u & 0 \leqq u \leqq 1 \\ 1 & u > 1\,. \end{cases}$$

Since F^* is mN times the distribution function of a random variable whose integer part is uniformly distributed over $\{0, 1, \cdots, N-1\}$ and whose fractional part is uniform on $[0, 1]$, we see that F^* is just mN times uniform measure on $[0, N]$ with

$$\begin{aligned} dF^*(u) &= 0 && u \leqq 0 \\ &= m\,du && 0 \leqq u \leqq N \\ &= 0 && u > N\,. \end{aligned}$$

Similarly G^* is the sum of measures which are each Lebesgue measure on intervals of the form $[j\tau, j\tau + 1]$, $j = 0, 1, \cdots, \beta$ (for nonnegative integer β), and zero elsewhere. We claim G^* is described by

LEMMA 2.2. *For all* $u \geqq 0$,

$$\int_{0-}^{u} x\,dG^*(x) \leqq (u/2) \cdot \int_{0-}^{u} dG^*(x)\,. \tag{2.4}$$

PROOF. It suffices by linearity to prove the lemma for G^* such that

$$\begin{aligned} dG^*(u) &= du && \text{if } j\tau \leqq u \leqq j\tau + 1\,, \quad j = 0, \cdots, \beta \\ &= 0 && \text{otherwise.} \end{aligned}$$

Define $b = b(u) = \min([u/\tau], \beta)$ where $[x]$ denotes the greatest integer $\leqq x$. Then

$$\begin{aligned} G^*(u) = \int_{0-}^{u} dG^*(x) = b + u - \tau b && \text{if } u \leqq \tau b + 1 \\ + 1 && \text{if } u > \tau b + 1\,. \end{aligned} \tag{2.5}$$

Similarly

$$\begin{aligned} \int_{0-}^{u} x\,dG^*(x) = \tau b(b-1)/2 + b/2 + \tfrac{1}{2}(u^2 - (\tau b)^2) && \text{if } u \leqq \tau b + 1 \\ + \tau b + \tfrac{1}{2} && \text{if } u > \tau b + 1\,. \end{aligned} \tag{2.6}$$

In the first case $u \leqq \tau b + 1$, (2.4) holds if and only if

$$\tau b(b-1) + b + u^2 - (\tau b)^2 \leqq ub + u^2 - u\tau b$$

if and only if

$$(1 + \tau b - u) \cdot b(\tau - 1) \geqq 0\,, \quad \text{which is true.}$$

In the second case $u > \tau b + 1$, (2.4) holds if and only if

$$\tau b(b-1) + b + 2\tau b + 1 \leqq u(b+1)$$

which is also true. □

Now by the convolution theorem for Laplace transforms,

$$\mathscr{L}_{d(G^*-F^*)}(y) = \mathscr{L}_{dU}(y) \cdot \mathscr{L}_{d(G-F)}(y) = \left(\frac{1-e^{-y}}{y}\right) \mathscr{L}_{d(G-F)}(y)$$

so that it suffices to show that $\mathscr{L}_{d(G^*-F^*)}(y)$ has at most one zero on $(0, \infty)$. Consider the convolution of $G^*(u) - F^*(u)$ with the function

$$\begin{aligned} I(u) &= u && u \geqq 0 \\ &= 0 && u < 0\,, \end{aligned}$$

which is

$$H(u) = I(u) * (G^*(u) - F^*(u)) = \int_{0-}^{u} (u - x)\, d(G^* - F^*)(x)$$
$$= \int_{0-}^{u} (G^*(x) - F^*(x))\, dx\,.$$

Then

$$\mathscr{L}_H(y) = \mathscr{L}_I(y) \cdot \mathscr{L}_{d(G^*-F^*)}(y) = \left(\frac{1 - e^{-y}}{y^3}\right)\phi(y)\,,$$

and another application of Lemma 2.1 shows that $((1 - e^{-y})/y^3)\phi(y)$, and hence $\phi(y)$, has no more sign changes than $H(u)$. By Lemma 2.3 below, if $H(u_0) < 0$ for any u_0, then $H(u) < 0$ for all $u \geqq u_0$. Thus H has at most one sign change.

LEMMA 2.3. *$u \geqq 0$ and $H(u) < 0$ implies $H'(u) \leqq 0$.*

PROOF. Suppose $H(u) < 0$ for some $u > 0$. If $u \geqq N$ we have immediately $F^*(u) = mN \geqq G^*(u)$ since the total mass in G^* is mN. If $u < N$ then $F^*(u) = mu$ and $\int_0^u F^*(x)\, dx = mu^2/2 = (u/2)F^*(u)$. Thus $H(u) < 0$ implies

$$\int_{0-}^{u} (u - x)\, dF^*(x) > \int_{0-}^{u} (u - x)\, dG^*(x)$$

which implies

$$(u/2)F^*(u) > uG^*(u) - \int_{0-}^{u} x\, dG^*(x) \geqq (u/2)G^*(u)$$

by Lemma 2.2, and thus $H'(u) = G^*(u) - F^*(u) < 0$. □

Summarizing, H has at most one sign change on $(0, \infty)$, hence the same is true for ϕ, hence $(d/dk) \log L(k)$ has at most one zero on $(0, \infty)$.

To show that the higher derivatives of $\log L(k)$ have at most one root, we need only notice that

$$\frac{d^p}{dk^p} \log L(k) = \mathscr{L}_\Psi(k) \qquad \text{where} \quad \Psi(y) = (-y)^{p-1}\phi(y)\,.$$

The preceding argument carries over exactly.

To show the necessity of the χ^2 condition (2.1) we will prove that if $\chi^2 \leqq \sum \tau_j(n_{+j}/N) - m$ then $H(u) \geqq 0$ for all u. First, for sufficiently large values of u, we may evaluate $H(u)$ using (2.5) and (2.6) as

$$u(G^*(u) - F^*(u)) + \int_{0-}^{u} x\, dF^*(x) - \int_{0-}^{u} x\, dG^*(x)$$
$$= 0 + \frac{mN^2}{2} - \sum_{i=1}^{m} \sum_{j=1}^{t} (\tfrac{1}{2}\tau_j n_{ij}(n_{ij} - 1) + n_{ij}/2)$$
$$= N/2\left(\sum_{j=1}^{t} \tau_j(n_{+j}/N) - m - \chi^2\right)$$

which is $\geqq 0$ by assumption. Thus $H(u) \geqq 0$ for all $u > 0$ since Lemma 2.3 showed that otherwise $\lim_{u\to\infty} H(u) < 0$. Now $H(u)$ is not identically zero since F and G are not identical, and therefore $\mathscr{L}_H(y)$ is strictly positive. Thus $\phi(y)$, $\Psi(y)$ and $(d^p/dk^p) \log L(k)$ all have no roots in $(0, \infty)$.

For sufficiency, it is clear that unless we have the trivial case in which all but one of the n_{ij} ($j = 1, \cdots, t$) equal zero, for each $i = 1, \cdots, m$, we must have

$$\lim_{k\to 0} (-1)^p \frac{d^p}{dk^p} \log L(k) = \infty .$$

An elementary asymptotic expansion of (2.3) shows

$$\lim_{k\to\infty} (-1)^p \frac{d^p}{dk^p} \log L(k) < 0$$

if (2.1) is satisfied. Thus continuity ensures a root of $(d^p/dk^p) \log L(k)$ in $(0, \infty)$. □

REMARKS. 1. When $m = 1$ and $\lambda_j = 1/t$ for $j = 1, \cdots, t$, the condition (2.1) reduces to $\chi^2 > t - 1$ given in Good's conjecture. This condition for the existence of a local maximum can be rephrased in terms of the "sample repeat rate" or "index of coincidence" of the sample $\mathbf{n}$. The minimum variance unbiased estimate of the repeat rate $\rho = \sum_{j=1}^{t} p_j^2$ is

$$\hat{\rho} = \sum_{j=1}^{t} n_j(n_j - 1)/N(N - 1) ,$$

and the condition $\chi^2 > t - 1$ is exactly $\hat{\rho} > 1/t$. The population repeat rate ρ must satisfy $\rho \geqq 1/t$ by the Cauchy-Schwarz inequality, with strict inequality if the p_j are not all equal. Thus to observe $\chi^2 \leqq t - 1$ is to observe $\rho \leqq 1/t$ which should signal the anomalous situation.

2. A somewhat simpler proof can be given for the special case of Good's conjecture ($m = 1$, $\lambda_j = 1/t$). The initial convolution of $G - F$ with U is not necessary in this case, and Lemma 2.2 can be simplified. In the example given in the appendix, only the final convolution with $I(u)$ is used.

3. Extensions to other distributions. Analogous results hold for the negative binomial distribution and the Poisson mixture.

3.1. Let $(n_1, \cdots, n_t)$ be a vector of observations whose components n_j are independently distributed as negative binomial random variables with parameters p and α_j, with probability mass function

$$\Pr(n_j \mid p, \alpha_j) = \frac{1}{n_j!} \Gamma(n_j + \alpha_j)/\Gamma(\alpha_j) p^{n_j}(1 - p)^{\alpha_j}$$
$$= \binom{n_j + \alpha_j - 1}{n_j} p^{n_j}(1 - p)^{\alpha_j}$$

for $n_j = 0, 1, 2, \cdots$ and $\alpha_j > 0$, $0 < p < 1$. Define $k = \sum_{j=1}^{t} \alpha_j$, $\boldsymbol{\lambda} = k^{-1}\boldsymbol{\alpha}$ and assume $\boldsymbol{\lambda}$ known as above. With a sample of m such observation vectors $\mathbf{n}_1, \cdots, \mathbf{n}_m$ we obtain the likelihood function

$$L(p, k) = L(p, k \mid \boldsymbol{\lambda}, \mathbf{n}_1, \cdots, \mathbf{n}_m) = \prod_{i=1}^{m} \prod_{j=1}^{t} \binom{n_{ij} + k\lambda_j - 1}{n_{ij}} p^N (1 - p)^{mk} \tag{3.1}$$

where $N = \sum_{i=1}^{m} \sum_{j=1}^{t} n_{ij}$.

THEOREM 2. *The likelihood function* (3.1) *has at most one local maximum at* $(\hat{p}, \hat{k})$ *where* $\hat{p} = N(N + m\hat{k})$. *The maximum occurs for* $\hat{k} < \infty$ *if and only if*

$$\sum_{j=1}^{t} \tau_j \sum_{i=1}^{m} n_{ij}(n_{ij} - 1) > N^2/m .$$

(The solution $\hat{k} = \infty$ corresponds to independent Poisson $(N\lambda_j/m)$ random variables.)

PROOF. The likelihood equations yield $\hat{p} = N/(N + m\hat{k})$ trivially and $\hat{k}$ satisfies the equation

$$\sum_{i=1}^{m} \sum_{j=1}^{t} \sum_{h=0}^{n_{ij}-1} 1/(k + h\tau_j) - m \log (1 + N/mk) = 0$$

with $\tau_j = \lambda_j^{-1} > 1$. Here the measure F is simply m times Lebesgue measure on $(0, N/m)$ since for $k > 0$,

$$\int_0^{N/m} 1/(k + u)\, du = \log (1 + N/mk) .$$

The measure G is identical with that described in Section 2. We now verify directly that the function $H(u) = \int_{0-}^{u} (u - x)\, d(G - F)(x)$ obeys the conclusion of Lemma 2.3 using the formulae

$$\begin{aligned} F(x) &= mx \qquad \text{for} \quad 0 \leqq x \leqq N/m \\ &= N \qquad \text{for} \quad x \geqq N/m \end{aligned}$$

and

$$\begin{aligned} \int_0^u F(x)\, dx &= mu^2/2 & \text{if} \quad u \leqq N/m \\ &= \frac{m}{2}(N/m)^2 + N(u - N/m) & \text{if} \quad u > N/m . \end{aligned}$$

For let $H(u) > 0$ for some $u \geqq 0$. If $u \leqq N/m$ we have

$$\begin{aligned} 0 > H(u) &= \int_{0-}^{u} G(x)\, dx - \int_{0-}^{u} F(x)\, dx \geqq (u/2)G(u) - \int_{0-}^{u} F(x)\, dx \\ &= (u/2)(G(u) - F(u)) \qquad \text{implies} \quad H'(u) < 0 . \end{aligned}$$

(The inequality $\int_{0-}^{u} G(x)\, dx \geqq (u/2)G(u)$ follows from a simple modification of Lemma 2.2.). If $u > N/m$ then already $F(u) = N \geqq G(u)$. Thus unimodality is established. The necessary and sufficient condition is obtained from

$$\begin{aligned} 0 > \lim_{u\to\infty} H(u) &= \sum_{i=1}^{m} \sum_{j=1}^{t} \sum_{h=0}^{n_{ij}-1} (u - h\tau_j) - (Nu - \tfrac{1}{2}N^2/m) \qquad \text{for large} \quad u \\ &= \tfrac{1}{2}(N^2/m - \sum_{i=1}^{m} \sum_{j=1}^{t} \tau_j n_{ij}(n_{ij} - 1)) . \end{aligned}$$

□

3.2. Consider two Poisson distributions $\mathscr{P}(\lambda_i)$ with known means λ_i, $i = 1, 2$ and $0 < \lambda_1 < \lambda_2$. Suppose we have observations $n_1, \cdots, n_m$ where the n_i are i.i.d. from $p\mathscr{P}(\lambda_1) + (1 - p)\mathscr{P}(\lambda_2)$ with unknown mixing proportion p, $0 \leqq p \leqq 1$. The likelihood function is

$$L(p \mid \mathbf{n}) = \prod_{i=1}^{m} (pe^{-\lambda_1}\lambda_1^{n_i}/n_i! + (1 - p)e^{-\lambda_2}\lambda_2^{n_i}/n_i!) , \tag{3.2}$$

and the likelihood equation may be reduced to

$$\sum_{i=1}^{m} 1/(r_i + x) - m/(1 + x) = 0$$

where $x = (1 - p)/p$ and for $i = 1, \cdots, m$, $r_i = e^{\lambda_2 - \lambda_1}(\lambda_1/\lambda_2)^{n_i}$.

THEOREM 3. *The likelihood function* (3.2) *has at most one local maximum as a function of p in* (0, 1). *The necessary and sufficient condition that the maximum occur in* (0, 1) *is that both* $\sum_{i=1}^m r_i > m$ *and* $\sum_{i=1}^m r_i^{-1} > m$. *Otherwise the maximum of* (3.2) *occurs at* $p = 0$ *or* $p = 1$ *as* $\bar{n} = (1/m) \sum_{i=1}^m n_i$ *is* $>$ *or* $<$ $\lambda_2 - \lambda_1/\log(\lambda_2/\lambda_1)$, *respectively. If equality obtains,* (3.2) *is constant for all p.*

PROOF. Consider

$$f(x) = \sum_{i=1}^m 1/(r_i + x) - m/(1 + x) = \int_{0-}^{\infty} 1/(x + u)\, d(G - F)(u)$$

where F puts mass m at $u = 1$ and G is a sum of m measures G_i placing unit mass at $u = r_i$. Let $H(u) = \int_0^u (G(x) - F(x))\, dx$ so that $f(x) = \mathscr{L}_\phi(x)$ where

$$\phi(y) = \mathscr{L}_{d(G-F)}(y) = y^2 \mathscr{L}_H(y)\,.$$

The conclusion of Lemma 2.3 applies to H: if $H(u) < 0$ then $\int_0^u F(x)\, dx > 0$ and $u \geqq 1$ so that $F(u) = m \geqq G(u)$ and $H'(u) \leqq 0$. Hence H, and consequently ϕ and f all have at most one sign change.

If f has exactly one root in $(0, \infty)$ then by the foregoing $H(u)$ must have a sign change, so that $H(\infty) = m - \sum_{i=1}^m r_i < 0$. But $m - \sum_{i=1}^m r_i = \lim_{x\to\infty} xf(x)$ showing $f(x) < 0$ for large x so that $f(0) = \sum_{i=1}^m r_i^{-1} - m > 0$ is also necessary.

Conversely, if both $\sum_1^m r_i > m$ and $\sum_1^m r_i^{-1} > m$, then $f(0) > 0$ and for all sufficiently large x, $f(x) < 0$, so that f has a root in $(0, \infty)$. The reader may examine the values of (3.2) at the endpoints $p = 0$ and $p = 1$ to derive the remaining assertions in the theorem. □

APPENDIX

The following graphs illustrate the need to smooth. Figure 1(a) gives the graph of the mass function $d(G - F)(u)$ for the multinomial data $t = 4$, $N = 16$, $\mathbf{n} = (1, 3, 5, 7)$, $m = 1$. Here $\chi^2 = 5$. Figure 1(b) gives $G(u) - F(u)$ and Figure 1(c) gives $H(u) = I(u) * (G - F)(u)$. Each graph is the indefinite integral of the previous one. Only $H(u)$ exhibits just one sign change.

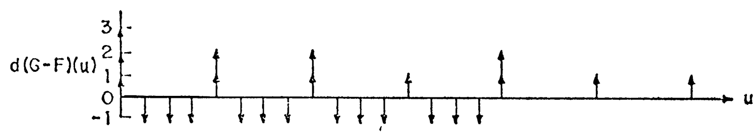

FIG. 1(a). Each arrow ↑ represents one delta-function. Graph displays 8 sign changes.

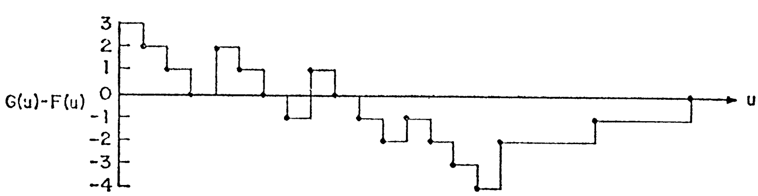

FIG. 1(b). Vertical lines added for clarity. Graph displays 3 sign changes.

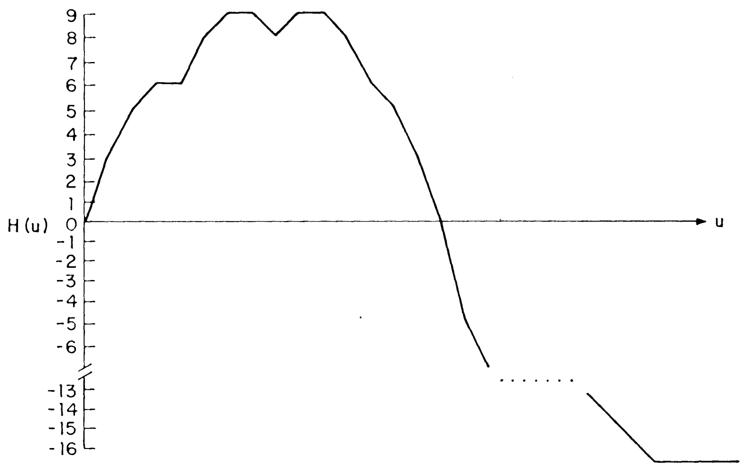

FIG. 1 (c). Graph compressed for space. Graph displays 1 sign change.

REFERENCES

[1] ANSCOMBE, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika* **37** 358–382.

[2] FELLER, W. (1968). *An Introduction to Probability Theory and its Applications* **1** (3rd ed.). Wiley, New York.

[3] GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* MIT Press.

[4] GOOD, I. J. (1967). A Bayesian significance test for multinomial distributions (with discussion). *J. Roy. Statist. Soc. Ser. B* **29** 399–431.

[5] GOOD, I. J. (1975). The Bayes factor against equiprobability of a multinomial population assuming a symmetric Dirichlet prior. *Ann. Statist.* **3** 246–250.

[6] GOOD, I. J. and CROOK, J. F. (1974). The Bayes/non-Bayes compromise and the multinomial distribution. *J. Amer. Statist. Assoc.* **69** 711–720.

[7] KARLIN, S. (1968). *Total Positivity* **1**. Stanford Univ. Press.

[8] PÓLYA, G. and SZEGÖ, G. (1964). *Aufgaben und Lehrsatze aus der Analysis.* Springer-Verlag, Berlin.

DEPARTMENT OF MATHEMATICAL STATISTICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720