DETERMINING BOUNDS ON INTEGRALS WITH APPLICATIONS TO CATALOGING PROBLEMS[1]

By Bernard Harris[2]

Stanford University

1. Introduction and summary. Assume that a random sample of size N has been drawn from a multinomial population with an unknown and perhaps countably infinite number of classes.

Hence, if X_j is the jth observation, and M_i the ith class, then

$$P\{X_j \varepsilon M_i\} = p_i \geqq 0 \qquad i = 1, 2, \cdots ; \text{for all } j,$$

and $\sum_1^\infty p_i = 1$.

If the number of classes is finite, then $p_i = 0$, for all $i > S$, where S is the number of classes.

We do not suppose the classes to have a natural ordering, since the classes may be species of insects, or chess openings.

Let n_r be the number of classes which occur exactly r times in the sample. Then

$$\sum_{r=0}^{\infty} rn_r = N \qquad \text{and} \qquad d = \sum_{r=1}^{N} n_r$$

where d is the number of distinct classes observed in the sample.

It is the purpose of this paper to present some techniques to aid the experimenter in answering the following kinds of questions.

1) Prediction of the number of distinct classes that will be observed in a second sample of size αN, $\alpha \geqq 1$.

2) Prediction of the number of additional classes that will be observed when the sample size is increased by $(\alpha - 1)N$ additional observations.

3) Estimation of the coverage of the sample, where coverage, denoted by C, is defined as follows:

$$C = \sum_j p_j . \tag{1}$$

The sum is to be taken over those classes for which at least one representative has been observed.

4) Prediction of the coverage of a second sample of size αN.

5) Prediction of the increased coverage to be obtained when the sample size is augmented by $(\alpha - 1)N$ additional observations.

Received May 20, 1958; revised January 9, 1959.

[1] This research was supported in part by the Office of Naval Research under Contract Nonr-225(21). Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at University of Nebraska.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

Let $d(\alpha)$ and $C(\alpha)$ denote the number of classes and coverage obtained from αN observations, either in the case of a second sample, or an augmented sample.

Subsequently, we will show that

$$Ed(\alpha) \backsim d + n_1 E\left\{\frac{1 - e^{-(\alpha-1)x}}{x}\right\} \tag{2}$$

and

$$EC(\alpha) \backsim 1 - \frac{n_1}{N} + \frac{n_1}{N} E\{1 - e^{-(\alpha-1)x}\} \tag{3}$$

where the expectations on the right hand side of (2) and (3) are taken with respect to $F(x)$, a constructed cumulative distribution function on $[0, \infty)$, which is unknown to the experimenter, but estimates of the moments of this distribution are available.

It will be shown that

$$\mu_r \backsim \frac{(r+1)!(n_{r+1})}{n_1} \tag{4}$$

where $\mu_r = \int_0^\infty x^r \, dF(x)$.

Hence, a reasonable procedure is to compute upper (lower) predictors for $d(\alpha)$ and $C(\alpha)$ by computing the supremum (infimum) of (2) and (3) respectively, where the supremum (infimum) is taken over all cumulative distribution functions whose first k moments are specified by (4).

It will be shown that

$$\varphi(x) = \frac{1 - e^{-(\alpha-1)x}}{x} \quad \text{and} \quad \psi(x) = 1 - e^{-(\alpha-1)x}$$

may be treated identically in the computation of extrema with the observation that the computation of the supremum of $E\{\varphi(x)\}$ will be identical with the computation of the infimum of $E\{\psi(x)\}$; and similarly for the computation of the infimum of $E\{\varphi(x)\}$. Thus, we will restrict the discussion to $\varphi(x)$.

The solution will be of the form

$$\sup(\inf) E\{\varphi(x)\} = \sum_{i=1}^{r} \lambda_i \varphi(x_i) \tag{5}$$

where $\lambda_i \geqq 0$, $\sum_{i=1}^r \lambda_i = 1$ and x_i belongs to the extended non-negative real numbers.

To determine the x_i and λ_i, the following system of equations must be solved.

$$\begin{aligned} \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_r x_r &= \mu_1 \\ \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_r x_r^2 &= \mu_2 \\ \cdots\cdots\cdots\cdots&\cdots\cdots \\ \lambda_1 x_1^k + \lambda_2 x_2^k + \cdots + \lambda_r x_r^k &= \mu_k \end{aligned} \tag{6}$$

where $x_1 \leqq x_2 \leqq \cdots \leqq x_r$.

If $k = 2c$, c an integer, the supremum will be obtained by solving (6), with $x_1 = 0$, $r \leqq (k+2)/2$, and $x_2, \cdots, x_r$ interior points of $[0, \infty)$. If $k = 2c + 1$, the solution of (6) will coincide with that for $k - 1$ with the addition of a mass point at infinity, with mass tending to zero at a rate which will satisfy the kth moment constraint. Since $\varphi(\infty) = 0$, no change in (5) is obtained by the addition of the kth moment constraint.

If $k = 2c + 1$, the infimum is obtained by solving (6) with $r \leqq (k+1)/2$ and the x_i are all interior points of $[0, \infty)$. If $k = 2c$, $c > 0$, the solution of (6) is obtained from that for $k - 1$ by the addition of a mass point at infinity with mass tending to zero at a rate which will satisfy the last moment constraint.

Explicit solutions are computed for $k \leqq 3$, and applied to several examples.

In addition, the low order moments of d, n_r, and C are computed, and the asymptotic sampling error of $\hat{C}(1) = 1 - (n_1/N)$ as an estimate of sample coverage is given.

It will be convenient before proceeding to the general problem to compute the low order moments of d, n_r, C.

2. Moments of d, n_r and C. To compute Ed and Ed^2, define a sequence of random variables $\{Y_j\}$ as follows: Let

$$Y_j = \begin{cases} 1, & \text{if } j\text{th class occurs in sample,} \\ 0, & \text{otherwise.} \end{cases}$$

Then $Ed = E\sum_{j=1}^{\infty} Y_j = \sum_{j=1}^{\infty} EY_j$. Since $EY_j = 1 - (1 - p_j)^N$, we have

$$E(d) = \sum_{j=1}^{\infty} [1 - (1 - p_j)^N]. \tag{7}$$

Similarly,

$$\begin{aligned} Ed^2 = E\left(\sum_{j=1}^{\infty} Y_j\right)^2 = E\left(\sum_{j=1}^{\infty} Y_j^2\right) + E\left(\sum_{i \neq j} Y_i Y_j\right) = \sum_{j=1}^{\infty} [1 - (1 - p_j)^N] \\ + \sum_{i \neq j} [1 - (1 - p_j)^N - (1 - p_i)^N + (1 - p_i - p_j)^N]. \end{aligned} \tag{8}$$

Thus,

$$\begin{aligned} \sigma_d^2 &= Ed^2 - (Ed)^2 \\ &= \sum_{j=1}^{\infty} [1 - (1 - p_j)^N] + \sum_{i \neq j} [1 - (1 - p_j)^N - (1 - p_i)^N \\ &\qquad + (1 - p_i - p_j)^N] - \left(\sum_{j=1}^{\infty} [1 - (1 - p_j)^N]\right)^2 \end{aligned}$$

$$\begin{aligned} &= \sum_{j=1}^{\infty} [(1 - p_j)^N - (1 - p_j)^{2N}] \\ &\qquad + \sum_{i \neq j} [(1 - p_i - p_j)^N - (1 - p_j)^N (1 - p_i)^N] \end{aligned} \tag{9}$$

$$= \sum_{j=1}^{\infty} [(1 - p_j)^N - (1 - 2p_j)^N]$$

$$+ \sum_{i,j} [(1 - p_i - p_j)^N - (1 - p_i)^N(1 - p_j)^N].$$

It is easily seen that the second term in σ_d^2 is always negative.

For, when $N = 1$, $(1 - p_i - p_j) - (1 - p_i)(1 - p_j) = -p_i p_j < 0$ implies $(1 - p_i)^N(1 - p_j)^N \geqq (1 - p_i - p_j)^N$.

To compute En_r, define random variables $Z_j^{(r)}$ as follows:

$$Z_j^{(r)} = \begin{cases} 1, & \text{if } j\text{th class occurs } r \text{ times in sample,} \\ 0, & \text{otherwise.} \end{cases}$$

Clearly,

$$E(n_r) = \sum_{j=1}^{\infty} EZ_j^{(r)}, \qquad E(Z_j^{(r)}) = \binom{N}{r} p_j^r (1 - p_j)^{N-r}$$

Hence,

$$E(n_r) = \sum_{j=1}^{\infty} \binom{N}{r} p_j^N (1 - p_j)^{N-r}. \tag{10}$$

To find $\sigma_{n_r}^2$ and cov (n_r, n_s), we first compute $E(n_r, n_s)$.

$$E(n_r n_s) = E\left(\sum_{i=1}^{\infty} Z_i^{(r)} \sum_{j=1}^{\infty} Z_j^{(s)}\right)$$

$$\begin{aligned} &= \sum_{i \neq j} \frac{N!}{r!\, s!(N - r - s)!} p_i^r p_j^s (1 - p_i - p_j)^{N-r-s} \qquad \text{if } r \neq s \\ &= \sum_{j=1}^{\infty} \binom{N}{r} p_j^r (1 - p_j)^{N-r} \\ &\quad + \sum_{i \neq j} \frac{N!}{(r!)^2(N - 2r)!} p_j^r p_i^r (1 - p_i - p_j)^{N-2r} \qquad \text{if } r = s. \end{aligned} \tag{11}$$

Thus,

$$\begin{aligned} \text{cov}(n_r, n_s) &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{N!}{r!\, s!} p_i^r p_j^s \\ &\quad \cdot \left(\frac{(1 - p_i - p_j)^{N-r-s}}{(N - r - s)!} - \frac{N!(1 - p_i)^{N-r}(1 - p_j)^{N-s}}{(N - r)!(N - s)!}\right) \\ &\quad - \sum_{i=1}^{\infty} \frac{N!}{r!\, s!(N - r - s)!} p_i^{r+s}(1 - 2p_i)^{N-r-s} \end{aligned} \tag{12}$$

$$\sigma_{n_r}^2 = \sum_{j=1}^{\infty} \left[\binom{N}{r} p_j^r (1 - p_j)^{N-r} - \binom{N}{r}^2 p_j^{2r} (1 - p_j)^{2N-2r}\right]$$

$$+ \sum_{i \neq j} \left[\frac{N!}{(r!)^2(N-2r)!} p_j^r p_i^r (1 - p_i - p_j)^{N-2r} - \binom{N}{r}^2 p_i^r p_j^r (1-p_i)^{N-r}(1-p_j)^{N-r} \right]$$

$$(13) \quad = \sum_{i,j} \frac{N!}{(r!)^2} p_i^r p_j^r \left(\frac{(1-p_i-p_j)^{N-2r}}{(N-2r)!} - \frac{N!(1-p_i)^{N-r}(1-p_j)^{N-r}}{(N-r)!^2} \right) + \sum_{j=1}^{\infty} \left[\binom{N}{r} p_j^r (1-p_j)^{N-r} - \frac{N!}{(r!)^2(N-2r)!} p_j^{2r}(1-2p_j)^{N-2r} \right].$$

To compute EC, the random variables Y_j are used in the computation of Ed are again employed.

Then, we have $C = \sum_{j=1}^{\infty} p_j Y_j$ and

$$(14) \quad EC = \sum_{j=1}^{\infty} p_j EY_j = \sum_{j=1}^{\infty} p_j[1 - (1-p_j)^N] = 1 - \sum_{j=1}^{\infty} p_j(1-p_j)^N.$$

Similarly,

$$EC^2 = E\left(\sum_{j=1}^{\infty} p_j Y_j\right)^2 = E(\sum_{i,j} p_i p_j Y_i Y_j)$$

$$= \sum_{j=1}^{\infty} p_j^2(1-(1-p_j)^N) + \sum_{i \neq j} p_i p_j [1 - (1-p_j)^N - (1-p_i)^N + (1-p_i-p_j)^N] \quad (15)$$

$$= \sum_{j=1}^{\infty} p_j^2[(1-p_j)^N - (1-2p_j)^N] + 1 - 2\sum_{j=1}^{\infty} p_j(1-p_j)^N + \sum_{i,j} p_i p_j (1-p_i-p_j)^N.$$

Thus

$$\sigma_C^2 = EC^2 - (EC)^2 = EC^2 - \left[1 - \sum_{j=1}^{\infty} p_j(1-p_j)^N\right]^2$$

$$(16) \quad = \sum_{j=1}^{\infty} p_j^2[(1-p_j)^N - (1-2p_j)^N] + \sum_{i,j} p_i p_j [(1-p_i-p_j)^N - (1-p_j)^N(1-p_i)^N].$$

As in the case of σ_d^2, the second term in σ_C^2 is always negative.

In Appendix A, it is shown that for large N, satisfactory exponential approximations may be used for Ed, En_r, EC. Employing these, we obtain for Ed, En_r, and EC:

$$(17) \quad E(d) \sim \sum_{j=1}^{\infty} [1 - e^{-Np_j}]$$

$$E(n_r) \sim \frac{1}{r!} \sum_{j=1}^{\infty} (Np_j)^r e^{-Np_j} \tag{18}$$

$$E(C) \sim 1 - \sum_{j=1}^{\infty} p_j e^{-Np_j}. \tag{19}$$

From (16), we get

$$E(C) \sim 1 - \frac{E(n_1)}{N}. \tag{20}$$

Henceforth, these approximations will be employed throughout.

It should be noted that the first two moments of n_r and (20) are obtained in papers by Good [3] and Good and Toulmin [4].

3. Prediction of number of classes observed and coverage obtained in enlarged and second samples. From (7) and (17), we have

$$E(d(\alpha)) = \sum_{j=1}^{\infty} [1 - (1 - p_j)^{\alpha N}] \tag{21}$$

and

$$\begin{aligned} E(d(\alpha)) &\sim \sum_{j=1}^{\infty} [1 - e^{-\alpha N p_j}] \\ &= E(d) + \sum_j [e^{-Np_j} - e^{-\alpha N p_j}]. \end{aligned} \tag{22}$$

Then

$$E(d(\alpha)) \sim E(d) + \sum_j \left[\frac{1 - e^{-(\alpha-1)Np_j}}{Np_j} Np_j e^{-Np_j}\right]. \tag{23}$$

In exactly the same manner,

$$E(C(\alpha)) = 1 - \sum_{j=1}^{\infty} p_j (1 - p_j)^{\alpha N} \tag{24}$$

and

$$\begin{aligned} E(C(\alpha)) &\sim 1 - \sum_{j=1}^{\infty} (p_j e^{-\alpha N p_j}) \\ &= 1 - \sum_j [p_j e^{-\alpha N p_j} + p_j e^{-Np_j} - p_j e^{-Np_j}] \\ &= E(C) + \sum_j [1 - e^{-(\alpha-1)Np_j}] p_j e^{-Np_j}. \end{aligned} \tag{25}$$

Then

$$E(C(\alpha)) \sim E(C) + \sum_j \left[\frac{1 - e^{-(\alpha-1)Np_j}}{N}\right] Np_j e^{-Np_j}. \tag{26}$$

Formulas (23) and (26) have an interesting interpretation. The second term on the right hand side in each case is the expected increment when the sample size is increased by $(\alpha - 1)N$; or equivalently the expected increment over the number of classes, or coverage of the first sample, that will be obtained in a second sample of size αN.

Define

$$F(c) = \frac{\sum_{Np_j \leqq c} Np_j e^{-Np_j}}{\sum_j Np_j e^{-Np_j}} . \tag{27}$$

One readily observes that $F(c)$ is a cumulative distribution function, and since it depends on the unknown parameters $(p_1, p_2, \cdots)$, it is unknown to the experimenter.

Consider μ_r the rth moment of $F(c)$.

$$\mu_r = \int_0^\infty x^r \, dF(x) = \frac{\sum_j (Np_j)^{r+1} e^{-Np_j}}{\sum_j Np_j e^{-Np_j}} .$$

Then, from (18), we have

$$\mu_r \sim \frac{(r+1)! \, E(n_{r+1})}{E(n_1)} . \tag{28}$$

Also, from (23) and (27), we have

$$E(d(\alpha)) \sim E(d) + E(n_1) \int_0 \frac{1 - e^{-(\alpha-1)x}}{x} \, dF(x) \tag{29}$$

and, from (26) and (27),

$$\begin{aligned} E(C(\alpha)) &\sim E(C) + E\left(\frac{n_1}{N}\right) \int_0^\infty (1 - e^{-(\alpha-1)x}) \, dF(x) \\ &\sim 1 - \frac{E(n_1)}{N} + \frac{E(n_1)}{N} \int_0^\infty (1 - e^{-(\alpha-1)x}) \, dF(x). \end{aligned} \tag{30}$$

Replacing the expected values by the observed quantities in (28), (29) and (30), we obtain

$$m_r = \frac{(r+1)! \, n_{r+1}}{n_1} \tag{31}$$

$$\hat{d}(\alpha) = d + n_1 E(\varphi(x)) \tag{32}$$

$$\hat{C}(\alpha) = 1 - \frac{n_1}{N} + \frac{n_1}{N} E(\psi(x)) \tag{33}$$

where the expected values are computed with respect to $F(x)$ as defined by (27).

However, since $F(x)$ is unknown, a reasonable procedure is to compute the supremum and infimum of the expected values using the m_r as estimates of the moments.

Since this technique, the method of moment inequalities, is of application in other problems, we will obtain some general results relative to computation of extrema of expected values of functions with respect to unknown cumulative distribution functions.

4. Method of moment inequalities. We proceed now to investigate the computation of extrema of expected values. The methods used in this section are similar to those used in Chernoff and Reiter [1] and Karlin and Shapley [6].

Let $\mathfrak{F}$ be the class of cumulative distribution functions on $[0, \infty)$, and $\mathfrak{F}_b$ that subset of $\mathfrak{F}$ all of whose elements have $F(b) = 1$.

Let $\mathfrak{F}^{(\mu_1, \mu_2, \cdots, \mu_k)}(\mathfrak{F}_b^{(\mu_1, \mu_2, \cdots, \mu_k)})$ be the class of cumulative distribution functions on $[0, \infty)$ ($[0, b]$) whose first k moments are $\mu_1, \mu_2, \cdots, \mu_k$, respectively. We will suppose throughout that $\mu_1, \mu_2, \cdots, \mu_k$ is a legitimate moment sequence, i.e., that there exists a cumulative distribution function $F(x) \,\varepsilon\, \mathfrak{F}(\mathfrak{F}_b)$, whose first k moments are $\mu_1, \mu_2, \cdots, \mu_k$.

Let $g(x)$ be a function continuous and bounded on $[0, b]$.

Designate the subset of Euclidean $k + 1$ space, whose coordinates are $(\int_0^\infty g(x)dF(x), \mu_1, \cdots, \mu_k)$, $F \,\varepsilon\, \mathfrak{F}_b$, by X_{k+1}.

THEOREM 1. *X_{k+1} is closed, convex, and bounded.*

PROOF: X_{k+1} is clearly bounded.

To demonstrate convexity, note that any convex linear combination of elements of $\mathfrak{F}_b$ is an element of $\mathfrak{F}_b$ and the mapping $F \rightarrow (Eg(x), \mu_1, \cdots, \mu_k) = T \cdot F$ is linear and thus preserves convexity.

To see that X_{k+1} is closed, note that $\mathfrak{F}_b$ is compact (in the topology of convergence in distribution). The conclusion follows upon application of the Helly-Bray Theorem.

The point

$$\left(\min_{F \varepsilon \mathfrak{F}_b} (\max) \int_0^\infty g(x)\, dF(x), \mu_1, \mu_2, \cdots, \mu_k\right)$$

is then easily characterized as the boundary point $\tilde{x}^* \,\varepsilon\, X_{k+1}$ whose first coordinate is a minimum (maximum), with fixed second, third, $\cdots$, $k + 1$th coordinates.

We also remark that X_{k+1} is a $k + 1$ dimensional convex body as long as $g(x)$ is not linearly dependent on the monomials $1, x, x^2, \cdots, x^k$. Henceforth, we will assume that $g(x)$ satisfies this condition.

THEOREM 2. *The extreme points of X_{k+1}, $k > 1$, are exactly those points which correspond to the moment sequences of degenerate cumulative distribution functions, i.e.,*

$$\begin{aligned} F(x) &= 0, & x &< a, \\ &= 1, & x &\geqq a, a \,\varepsilon\, [0, b]. \end{aligned}$$

PROOF: One can easily show that any extreme point of X_{k+1} is the image under T of an extreme point of $\mathfrak{F}_b$. It is easily seen that the extreme points of $\mathfrak{F}_b$ are the degenerate cumulative distribution functions.

The points in X_{k+1} corresponding to degenerate cumulative distribution functions have the form $\tilde{x} = (g(t), t, t^2, \cdots, t^k)$, $t \varepsilon [0, b]$.

Consider the hyperplane $H(g(t), \mu_1, \cdots, \mu_k) = t_0^2 - 2\mu_1 t_0 + \mu_2 = 0$ where t_0 is fixed and $t_0 \varepsilon [0, b]$.

Then,

$$(t - t_0)^2 = t_0^2 - 2tt_0 + t^2 > 0 \qquad t \neq t_0$$
$$= 0 \qquad t = t_0.$$

Thus, $t_0^2 - 2\mu_1 t_0 + \mu_2 = 0$ is a supporting hyperplane at $\tilde{x}_0 = (g(t_0), t_0, t_0^2, \cdots, t_0^k)$ and $\tilde{x}_0$ is not attainable as a non-trivial convex linear combination of points of X_{k+1} corresponding to degenerate distributions.

Hence for $k > 1$, points in X_{k+1} corresponding to degenerate distributions are the extreme points of X_{k+1}.

COROLLARY. *If $g(x)$ is strictly convex (concave), the set of extreme points of X_{k+1}, $k \geqq 1$, are exactly those points which correspond to the moment sequences of degenerate cumulative distribution functions.*

Since $\tilde{x}^*$ is a boundary point of X_{k+1}, it can be represented as a convex linear combination of at most $k + 1$ extreme points of X_{k+1}, i.e.,

$$\tilde{x}^* = \sum_{j=0}^{k} \lambda_j \tilde{x}_j = \sum_{j=0}^{k} \lambda_j (g(t_j), t_j, t_j^2, \cdots, t_j^k). \tag{34}$$

Hence, the minimizing (maximizing) distribution is discrete with positive probability concentrated on at most $k + 1$ points in $[0, b]$.

We will now show that the maximum number of points of positive probability can be reduced still further in the specific cases of interest to us.

Since $\tilde{x}^*$ is a boundary point of X_{k+1}, there is a supporting hyperplane at $\tilde{x}^*$, which also contains the extreme points of X_{k+1} of which $\tilde{x}^*$ is a convex linear combination.

Thus, $\tilde{\alpha} \cdot \tilde{x}^* = c$ and $\tilde{\alpha} \cdot \tilde{x} \geqq c$, $\tilde{x} \varepsilon X_{k+1}$. In particular;

$$\sum_{i=1}^{k} \alpha_i t^i + \alpha_{k+1} g(t) \geqq c \qquad \text{for all } t \varepsilon [0, b] \tag{35}$$

with equality holding for those extreme points of which $\tilde{x}^*$ can be written as a convex linear combination. Rewriting (35) we have

$$P(t) = \sum_{i=0}^{k} \alpha_i t^i + \alpha_{k+1} g(t) \geqq 0 \qquad t \varepsilon [0, b]. \tag{36}$$

Then, to find the relevant extreme points, we have to find the roots of $P(t) = 0$, $t \varepsilon [0, b]$.

We note at once that all interior roots of $P(t)$ are multiple roots, since by (36), $P(t) \geqq 0$ for all $t \varepsilon [0, b]$.

Let r be the number of distinct real roots of $P(t)$ in $[0, b]$.

Define r' as follows:

$$r' = \begin{cases} r & \text{if } 0, b \text{ are not roots of } P(t) \\ r - \frac{1}{2} & \text{if one of } 0, b \text{ is a root of } P(t) \\ r - 1 & \text{if both } 0, b \text{ are roots of } P(t) \end{cases}$$

Let $\mathcal{G}_b^{(k)}$ be the collection of continuous, bounded, and monotonic functions on $[0, b]$, whose first k derivatives exist and are monotonic in $(0, b)$. In addition, we require that $\mathcal{G}_b^{(k)}$ contain only functions not linearly dependent on the monomials $1, t, t^2, \cdots, t^k$.

THEOREM 3. *If* $P(t) = Q(t) + \alpha g(t)$, *where* α *is a real number,* $Q(t)$ *is a polynomial of degree* k, *and* $g(t) \varepsilon \mathcal{G}_b^{(k)}$, *then, there are at most* $k + 1$ *real roots of* $P(t)$ *in* $(0, b)$.

PROOF: We proceed by induction.

When $k = 0$, there is clearly at most one real root of $P(t)$ in $(0, b)$.

Then, suppose the conclusion holds when $k = n, n \geqq 0$ for any function $P(t)$ satisfying the hypotheses of the theorem.

Then, if $P_1(t)$ satisfies the hypotheses of the theorem for $k = n + 1$, $P_1'(t)$ satisfies the hypotheses for $k = n$. In addition, between every root of $P_1(t)$, there is a root of $P_1'(t)$, but $P_1'(t)$ has at most $n + 1$ roots in $(0, b)$, hence $P_1(t)$ has at most $n + 2$ roots in $(0, b)$.

THEOREM 4. *If* $P(t)$ *satisfies the hypotheses of Theorem* 3 *and in addition* $P(t) \geqq 0$ *for all* $t \varepsilon [0, b]$, *then*

$$r' \leqq \frac{k+1}{2}. \tag{37}$$

PROOF: Let γ be the number of distinct roots of $P(t)$ at 0 and b, i.e., the number of distinct boundary roots.

Then $r' = r - (\gamma/2)$.

By Theorem 3, $P'(t)$ has at most k distinct roots in $(0, b)$. Since $P(t)$ has only multiple roots in $(0, b)$, whenever $P(t) = 0, t \varepsilon (0, b)$, we have $P'(t) = 0$.

In addition, if $t_0, t_1 (t_0 < t_1)$ are two distinct roots of $P(t)$, then there exists a t^* such that $P'(t^*) = 0, t_0 < t^* < t_1$.

Hence, $(r - 1) + (r - \gamma) \leqq k$, $(r' + (\gamma/2) - 1) + (r' - (\gamma/2)) \leqq k$, $r' \leqq (k + 1)/2$.

Theorem 4 provides an extension of results contained in Chernoff and Reiter [1] and Rustagi [7]. The use of r' is similar to Wald's [8] notion of the degree of a cumulative distribution function.

For any discrete cumulative distribution function $F(x) \varepsilon \mathfrak{F}_b$, let $r'(F)$ be defined as follows:

Let $r_1'(F)$ = number of saltuses of $F(x)$ in $(0, b)$

$2r_2'(F)$ = number of saltuses of $F(x)$ at 0 or b.

Then,

$$r'(F) = r_1'(F) + r_2'(F).$$

$r'(F)$ is called the degree of the discrete cumulative distribution function $F(x)$.

Thus, we have seen that the degree of the minimizing (maximizing) distribution $\varepsilon\, \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_k)}$ is at most $(k+1)/2$ for functions in $\mathcal{G}_b^{(k)}$.

We now employ some results due to Wald [8] to establish the following theorem:

THEOREM 5. *There are exactly two cumulative distribution functions* $F_1^*(x)$ *and* $F_2^*(x)\ \varepsilon\, \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_k)}$ *with* $r'(F) \leqq (k+1)/2$ *where* $F_1^*(x)$ *is continuous at* b, *and* $F_2^*(x)$ *has a saltus at* b.

PROOF: From Theorem 4, there exist at least two cumulative distribution functions of degree $\leqq (k+1)/2$. They are the minimizing and maximizing distributions respectively for functions in $\mathcal{G}_b^{(k)}$.

Let $w(F) = r'(F)$ if $F(x)$ has no saltus at b
$\qquad = r'(F) + \frac{1}{2}$ if $F(x)$ has a saltus at b.

Wald has shown the following:

Let $F(x)$ and $G(x)$ be two cumulative distribution functions belonging to $\mathfrak{F}$. Then,

A) If $w(F)$ and $w(G)$ are both $\leqq q$, then $F(x) - G(x)$ changes sign at most $2q - 2$ times.

B) If $w(F)$ and $w(G)$ are both $\leqq q$, $q > 1$; and both $F(x)$ and $G(x)$ have a saltus at $\alpha > 0$, then $F(x) - G(x)$ changes sign at most $2q - 3$ times.

C) If $F(x)$ and $G(x)$ both $\varepsilon\, \mathfrak{F}^{(\mu_1,\mu_2,\cdots,\mu_k)}$, then $F(x) - G(x)$ changes sign at least k times, unless $F(x) = G(x)$.

Now, suppose there exist two cumulative distribution functions, $F(x)$ and $G(x)$, both $\varepsilon\, \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_k)}$, both of degree $\leqq (k+1)/2$, and both continuous at b.

Then, by A,

$$F(x) - G(x) \text{ changes sign at most } 2r' - 2 \text{ times.}$$

By C,

$$F(x) - G(x) \text{ changes sign at least } k \text{ times.}$$

But, by hypothesis $2r' - 2 \leqq k - 1$. Hence, $F(x) = G(x)$.

Suppose there exist two cumulative distribution functions $F(x)$ and $G(x)\ \varepsilon\, \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_k)}$ both of which have a saltus at b, and both of which have degree $\leqq (k+1)/2$.

Then, by B,

$$F(x) - G(x) \text{ changes sign at most } 2q - 3 \text{ times.}$$

By C,

$$F(x) - G(x) \text{ has at least } k \text{ changes in sign.}$$

But, $2q - 3 = 2r' - 2 \leqq k - 1$. Thus, $F(x) = G(x)$.

We note that B requires $q > 1$, implies $r > \frac{1}{2}$. However, when $r = \frac{1}{2}$, the theorem holds trivially by the monotonicity of $g(x)$.

This establishes that there are exactly two cumulative distribution functions $\varepsilon\, \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_k)}$ with degree $\leqq (k+1)/2$, one of which has a saltus at b, and the other is continuous at b. These are the minimizing and maximizing distributions.

We can now obtain the following theorem.

THEOREM 6. *The degree of the minimizing (maximizing) cumulative distribution function,*

$$r'(F_i^*) = \frac{k+1}{2}, \qquad i = 1, 2, \tag{38}$$

$$F_i^*(x) \ \varepsilon \ \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_k)},$$

unless

$$\mu_k = \min_{F \varepsilon \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_{k-1})}} (\max) \int_0^\infty x^k \, dF(x).$$

PROOF: If in Theorem 5, we replace k by $k - 1$, then since $x^k \ \varepsilon \ \mathcal{G}_b^{(k-1)}$, there are exactly two cumulative distribution functions $\varepsilon \ \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_{k-1})}$ with degree $\leqq k/2$, and these distributions have the property that they determine the extrema of μ_k for $F(x) \ \varepsilon \ \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_{k-1})}$.

Hence, we may conclude that if $r'(F) < (k + 1)/2$ and $F(x) \ \varepsilon \ \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_k)}$, then μ_k is an extremum of $E(X^k)$, where the extremization is over all $F(x) \ \varepsilon \ \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_{k-1})}$. Thus, the inequality of Theorem 4 is an equality, whenever μ_k is not an extremum determined by $\mu_1, \mu_2, \cdots, \mu_{k-1}$.

COROLLARY. *The* max(min) μ_{k+v}, $v > 0$, *for* $F \ \varepsilon \ \mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_k)}$ *is attained by choosing* $F(x)$ *to be one of the two cumulative distribution functions with* $r'(F) \leqq (k + 1)/2$, *one of which has a saltus at* b, *and the other of which is continuous at* b.

Thus, we have shown that this technique will enable us to obtain the sharpest possible bounds on higher moments given the first k moments.

If an extremizing cumulative distribution function is of degree $< (k + 1)/2$, we will call the moment sequence $(\mu_1, \mu_2, \cdots, \mu_k)$ a degenerate moment sequence.

We can now characterize the two extremizing distributions as follows:

THEOREM 7. *If* $(\mu_1, \mu_2, \cdots, \mu_k)$ *is non-degenerate and the two extremizing cumulative distribution functions are* $F_1^*(x)$ *and* $F_2^*(x)$ *respectively,*

a) *if* $k = 2q + 1$, q *a non-negative integer,*
then,

$F_1^*(x)$ *has saltuses at* $\frac{k+1}{2}$ *points in* $(0, b)$

$F_2^*(x)$ *has saltuses at* $\frac{k-1}{2}$ *points in* $(0, b)$ *and at both* 0 *and* b.

b) *if* $k = 2q$, *then,*

$F_1^*(x)$ *has saltuses at* $\frac{k}{2}$ *points in* $(0, b)$ *and at* 0

$F_2^*(x)$ *has saltuses at* $\frac{k}{2}$ *points in* $(0, b)$ *and at* b.

PROOF: The proof is immediate from Theorems 5 and 6 and by noting that these are the only ways in which cumulative distribution functions of degree $(k + 1)/2$ can be obtained.

We note that whether $F_1^*(x)$ will be a maximizing or minimizing distribution depends on the particular choice of $g(x) \varepsilon \mathcal{G}_b^{(k)}$.

We now extend the preceding result to cumulative distribution functions in $\mathfrak{F}^{(\mu_1, \mu_2, \cdots, \mu_k)}$.

THEOREM 8. *If* $(\mu_1, \mu_2, \cdots, \mu_k)$ *is non-degenerate, and* $g(x) \varepsilon \mathcal{G}_\infty^{(k)}$, *then*

$$\sup_{F \varepsilon \mathfrak{F}^{(\mu_1, \cdots, \mu_k)}} (\inf) \int_0^\infty g(x)\, dF(x)$$

is obtained by one of the following:

$$\int_0^\infty g(x)\, dF_1^*(x), \qquad F_1^*(x) \varepsilon \mathfrak{F}^{(\mu_1, \mu_2, \cdots, \mu_k)} \tag{39}$$

where $F_1^*(x)$ *has saltuses at* $(k + 1)/2$ *points in* $(0, \infty)$ *for* $k = 2q + 1$, q *an integer* $\geqq 0$, *and* $F_1^*(x)$ *has saltuses at* $k/2$ *points in* $(0, \infty)$ *and at* 0, *for* $k = 2q$; *or,*

$$\lim_{b \to \infty} \int_0^b g(x)\, dF_{2b}^*(x), \qquad F_{2b}^*(x) \varepsilon \mathfrak{F}_b^{(\mu_1, \mu_2, \cdots, \mu_k)} \tag{40}$$

where $F_{2b}^*(x)$ *has saltuses at* $k - 1/2$ *points in* $(0, b)$ *and at both* 0 *and* b *if* $k = 2q + 1$, *and* $F_{2b}^*(x)$ *has saltuses at* $k/2$ *points in* $(0, b)$ *and at* b *for* $k = 2q$.

PROOF: For any cumulative distribution function $F(x) \varepsilon \mathfrak{F}^{(\mu_1, \mu_2, \cdots, \mu_k)}$ and any $b_1 > 0$ we have $\int_{b_1}^\infty x\, dF(x) \leqq \mu_1$ and $b_1 \int_{b_1}^\infty dF(x) \leqq \int_{b_1}^\infty x\, dF(x)$. Hence, $\int_{b_1}^\infty dF(x) \leqq \mu_1/b_1$. Therefore by choosing b_1 sufficiently large, we have for any $\epsilon > 0$, $\int_{b_1}^\infty dF(x) < \epsilon$.

Further, Wald [8] has shown that if $\mu_1, \mu_2, \cdots, \mu_k$ is a legitimate moment sequence, we can find a discrete cumulative distribution function with no more than $k + 1$ saltuses in $[0, \infty)$ which has these moments. We suppose the last saltus is at b_2. Choose $b > \max(b_1, b_2)$. Then since $g(x)$ is bounded, i.e. $|g(x)| < M$, we have $\int_b^\infty g(x) dF(x) < M\epsilon$ for all cumulative distribution functions in $\mathfrak{F}^{(\mu_1, \mu_2, \cdots, \mu_k)}$. Hence we can employ Theorem 7 and the conclusion follows.

THEOREM 9. $\lim_{b \to \infty} F_{2b}^*(x) = \hat{F}_2^*(x)$, *where* $\hat{F}_2^*(x)$ *is the extremizing cumulative distribution function* $F_1^*(x)$ *computed for* $k - 1$ *moment constraints.*

PROOF: From Theorem 7, for every b, $F_{2b}^*(x)$ is uniquely determined.

Further, as $b \to \infty$, $F_{2b}^*(b) - F_{2b}^*(b - 0) = 0(1/b^k)$, since otherwise $\mu_k \to \infty$. Thus $\lim_{b \to \infty} \int_0^{b-} x^i\, dF_{2b}^*(x) = \mu_i$, $i = 1, 2, \cdots, k - 1$. Let

$$\hat{F}_{2b}^*(x) = \begin{cases} F_{2b}^*(x), & x < b, \\ F_{2b}^*(b^-), & x \geqq b. \end{cases}$$

Then, as $b \to \infty$, $\hat{F}_{2b}^*(x)$ converges to $F_1^*(x)$, computed for $k - 1$ moment constraints. This is readily seen, as follows:

$\hat{F}_{2b}^*(x)$ converges at all continuity points to a limit function $\hat{F}_2^*(x)$, which is a cumulative distribution function, $r'(\hat{F}_2^*(x)) = k/2$, and satisfies $k - 1$ moment constraints, hence must be $F_1^*(x)$ computed for $k - 1$ moment constraints.

Corollary. *$\int_0^\infty g(x)\,dF_1^*(x) = \lim_{b\to\infty}\int_0^\infty g(x)\,dF_{2b}^*(x)$ where F_{2b}^* is computed for one more moment constraint than $F_1^*(x)$.*

We now obtain the following theorem due to Wald [8], as a consequence, of the preceding results.

Theorem 10. *If $(\mu_1, \mu_2, \cdots, \mu_k)$ is a non-degenerate moment sequence, then for $(\mu_1, \mu_2, \cdots, \mu_{k+1})$ to be a legitimate moment sequence, it is necessary and sufficient that*

$$\mu_{k+1} \geqq \int_0^\infty x^{k+1}\,dF_1^*(x) = \mu_{k+1}^*.$$

If equality holds, then $(\mu_1, \mu_2, \cdots, \mu_{k+1})$ is a degenerate moment sequence.

Proof: Let $\mu_{k+1}^* = \int_0^\infty x^{k+1}\,dF_1^*(x)$, $F_1^*(x)\,\varepsilon\,\mathfrak{F}^{(\mu_1,\mu_2,\cdots,\mu_k)}$. Since $F_1^*(x)$ is a unique extremal solution, it determines $2r$ quantities, x_i, λ_i; $i = 1, 2, \cdots, r$ such that $\lambda_1 x_1^i + \lambda_2 x_2^i + \cdots + \lambda_r x_r^i = \mu_i (i = 1, 2, \cdots, k)$ and $0 \leqq x_1 < x_2 < \cdots < x_r < \infty$, $\lambda_i \geqq 0$, $\sum_{i=1}^r \lambda_i = 1$. From Theorem 9,

$$F_{2b}^*(x)\,\varepsilon\,\mathfrak{F}_b^{(\mu_1,\mu_2,\cdots,\mu_{k+1})}$$

satisfies

$$\lambda_1' x_1'^i + \lambda_2' x_2'^i + \cdots + \lambda_r' x_r'^i + \lambda_{r+1}' b^i = \mu_i (i = 1, 2, \cdots, k+1)$$

where $\lambda_j' \to \lambda_j$, $x_j' \to x_j$, and $\lambda_{r+1}' = 0(1/b^k)$ as $b \to \infty$.

Since $F_1^*(x)$ provides an extremum of μ_{k+1}, and we have shown the construction of $F_{2b}^*(x)$, and since $\lambda_{r+1}' b^{k+1} \geqq 0$ we can conclude $\mu_{k+1} \geqq \mu_{k+1}^*$, with equality if $(\mu_1, \mu_2, \cdots, \mu_{k+1})$ is degenerate.

On the other hand, by choosing b sufficiently large, we can produce a distribution $F_{2b}^*(x)$ for any choice of $\mu_{k+1} \geqq \mu_{k+1}^*$.

Many of the above results may be extended to other unbounded functions by similar limiting arguments.

5. Computation of extremizing cumulative distribution functions on $[0, \infty)$ for $k = 0, 1, 2, 3$.

To find the extremizing cumulative distribution functions, we have to solve the following system of equations:

$$\begin{aligned} &\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_r x_r = \mu_1 \\ &\lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_r x_r^2 = \mu_2 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\lambda_1 x_1^k + \lambda_2 x_2^k + \cdots + \lambda_r x_r^k = \mu_k \end{aligned} \tag{41}$$

$$\lambda_i \geqq 0, \qquad \sum_{i=1}^r \lambda_i = 1, \qquad 0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_r < \infty.$$

Assuming throughout that $(\mu_1, \mu_2, \cdots, \mu_k)$ is a non-degenerate moment sequence, we can then apply Theorem 8.

When $k = 0$, we have trivially

$$F_1^*(x) = \begin{cases} 0, & x < 0, \\ 1, & 0 \leqq x. \end{cases}$$

(42)
$$F^*_{2b}(x) = \begin{cases} 0, & x < b, \\ 1, & b \leqq x. \end{cases}$$

For $k = 1$,

(43)
$$F^*_1(x) = \begin{cases} 0, & x < \mu, \\ 1, & \mu \leqq x. \end{cases}$$

To determine $F^*_{2b}(x)$, applying Theorem 8, (41) becomes $\lambda_2 b = \mu$. Hence

(44)
$$F^*_{2b}(x) = \begin{cases} 0, & x < 0, \\ \dfrac{b - \mu}{b}, & 0 \leqq x < b, \\ 1, & b \leqq x. \end{cases}$$

When $k = 2$, to determine $F^*_1(x)$, we consider $\lambda_2 x_2 = \mu_1$, and $\lambda_2 x_2^2 = \mu_2$. Thus

$$\lambda_1 = \frac{\mu_2 - \mu_1^2}{\mu_2}, \qquad \lambda_2 = \frac{\mu_1^2}{\mu_2}, \qquad x_2 = \frac{\mu_2}{\mu_1},$$

(45)
$$F^*_1(x) = \begin{cases} 0, & x < 0, \\ \dfrac{\mu_2 - \mu_1^2}{\mu_2}, & 0 \leqq x < \dfrac{\mu_2}{\mu_1}, \\ 1, & \dfrac{\mu_2}{\mu_1} \leqq x. \end{cases}$$

For $F^*_{2b}(x)$, we solve (41), which becomes:

$$\lambda_1 x_1 + \lambda_2 b = \mu_1, \qquad \lambda_1 x_1^2 + \lambda_2 b^2 = \mu_2$$

obtaining

$$\lambda_1 = \frac{(\mu_1 - b)^2}{(\mu_1 - b)^2 + (\mu_2 - \mu_1^2)}, \qquad \lambda_2 = \frac{\mu_2 - \mu_1^2}{(\mu_1 - b)^2 + (\mu_2 - \mu_1^2)},$$

$$x_1 = \frac{\mu_1 b - \mu_2}{b - \mu_1}.$$

Hence,

(46)
$$F^*_{2b}(x) = \begin{cases} 0, & x < \dfrac{\mu_1 b - \mu_2}{b - \mu_1}, \\ \dfrac{(\mu_1 - b)^2}{(\mu_1 - b)^2 + (\mu_2 - \mu_1^2)}, & \dfrac{\mu_1 b - \mu_2}{b - \mu_1} \leqq x < b, \\ 1, & b \leqq x. \end{cases}$$

Note that as $b \to \infty$, $\lambda_1 \to 1$, $\lambda_2 \to 0$, $x_1 \to \mu$, $\lambda_2 b \to 0$, and $\lambda_2 b^2 \to \mu_2 - \mu_1^2$, and $F_{2b}^*(x)$ is $F_1^*(x)$ for $k = 1$ with the addition of the infinitesimal mass at ∞.

When $k = 3$, for $F_1^*(x)$ consider

$$\lambda_1 x_1 + \lambda_2 x_2 = \mu_1 \tag{1}$$

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 = \mu_2 \tag{2}$$

$$\lambda_1 x_1^3 + \lambda_2 x_2^3 = \mu_3 . \tag{3}$$

From (1) and (2) we have

$$\lambda_1 = \frac{(\mu_1 - x_2)^2}{(\mu_1 - x_2)^2 + (\mu_2 - \mu_1^2)}, \qquad \lambda_2 = \frac{\mu_2 - \mu_1^2}{(\mu_1 - x_2)^2 + (\mu_2 - \mu_1^2)},$$

$$x_1 = \frac{\mu_1 - (1 - \lambda_1)x_2}{\lambda_1}.$$

Substituting these in (3), we get

$$\frac{\{\mu_1(\mu_1 - x_2) + (\mu_2 - \mu_1^2)\}^3}{(\mu_1 - x_2)\{(\mu_1 - x_2)^2 + (\mu_2 - \mu_1^2)\}} + \frac{(\mu_2 - \mu_1^2)x_2^3}{(\mu_1 - x_2)^2 + (\mu_2 - \mu_1^2)} = \mu_3 .$$

Setting $\mu_1 - x_2 = y$ and $\mu_2 - \mu_1^2 = \sigma^2$, we have

$$\frac{(\mu y + \sigma^2)^3}{y(y^2 + \sigma^2)} + \frac{\sigma^2(\mu - y)^3}{y^2 + \sigma^2} = \mu_3 .$$

Thus

$$-\sigma^2 y^4 + (\mu^3 + 3\mu\sigma^2 - \mu_3)y^3 + \sigma^2(3\mu\sigma^2 + \mu^3 - \mu_3) + \sigma^6 = 0.$$

Setting $\mu_3' = \mu_3 - 3\mu\sigma^2 - \mu^3$, we have

$$f(y) = \sigma^2 y^4 + \mu_3' y^3 + \sigma^2 \mu_3' y - \sigma^6 = 0. \tag{47}$$

From (41), it follows that $y < 0$. Thus we are interested in negative real zeros of (47). From Descartes' Rule of Signs, and the observations that $f(0) < 0$ and $f(-\infty) > 0$, there is exactly one negative real root.

We proceed to find the root. First, observe that

$$f(y) = \sigma^2(y^2 + \sigma^2)\left(y^2 + \frac{\mu_3'}{\sigma^2} y - \sigma^2\right).$$

Hence, the unique negative root y_0 is given by

$$y_0 = \frac{-\mu_3' - (\mu_3'^2 + 4\sigma^6)^{1/2}}{2\sigma^2}$$

which provides the complete solution:

$$\lambda_1 = \frac{y_0^2}{y_0^2 + \sigma^2}, \qquad \lambda_2 = \frac{\sigma^2}{\sigma^2 + y_0^2},$$

$$x_1 = \frac{\mu_1 y_0 + \sigma^2}{y_0}, \qquad x_2 = \mu - y_0,$$

and

$$(48)\qquad F_1^*(x) = \begin{cases} 0, & x < \dfrac{\mu y_0 + \sigma^2}{y_0}, \\ \dfrac{y_0^2}{y_0^2 + \sigma^2} & \dfrac{\mu y_0 + \sigma^2}{y_0} \leqq x < \mu - y_0, \\ 1, & \mu - y_0 \leqq x. \end{cases}$$

To obtain $F_{2b}^*(x)$, the system of equations

$$\begin{aligned} \lambda_2 x_2 + \lambda_3 b &= \mu_1 \\ \lambda_2 x_2^2 + \lambda_3 b^2 &= \mu_2 \\ \lambda_2 x_2^3 + \lambda_3 b^3 &= \mu_3 \end{aligned}$$

is easily solved, yielding

$$\lambda_2 = \frac{(\mu_1 b - \mu_2)^3}{(\mu_2 b - \mu_3)(\mu_1 b^2 - 2\mu_2 b + \mu_3)}, \qquad \lambda_3 = \frac{\mu_1 \mu_3 - \mu_2^2}{b(\mu_1 b^2 - 2\mu_2 b + \mu_3)},$$

$$\lambda_1 = 1 - \lambda_2 - \lambda_3, \qquad x_2 = \frac{\mu_2 b - \mu_3}{\mu_1 b - \mu_2},$$

$$(49)\qquad F_{2b}^*(x) = \begin{cases} 0, & x < 0, \\ 1 - \lambda_2 - \lambda_3, & 0 \leqq x < \dfrac{\mu_2 b - \mu_3}{\mu_1 b - \mu_2}, \\ 1 - \lambda_3, & \dfrac{\mu_2 b - \mu_3}{\mu_1 b - \mu_2} \leqq x < b, \\ 1, & b \leqq x. \end{cases}$$

As $b \to \infty$, it is easily observed that $F_{2b}^*(x) \to F_1^*(x)$ for $k = 2$, with the addition of an infinitesimal mass at ∞.

6. Application of Extrema Computations to $\varphi(x)$ and $\psi(x)$. First note that $\varphi(x)$ satisfies the hypotheses of Theorem 4 for all k, in fact $\varphi(x)$ is completely monotonic. We can see this as follows:

$$\begin{aligned} \varphi^{(v)}(x) &= \sum_{i=0}^{v} \binom{v}{i} x^{-(v-i+1)}(-1)^{v-i}(v-i)!\frac{d^i}{d_x^i}[1 - e^{-(\alpha-1)x}] \\ &= x^{-(v+1)}(-1)^v v!(1 - e^{-(\alpha-1)x}) + \sum_{i=1}^{v} \frac{v!(-1)^{v+1}(\alpha-1)^i}{i!\, e^{(\alpha-1)x} x^{v-i+1}}. \end{aligned}$$

Thus

$$\frac{\varphi^{(v)}(x)(-1)^{v+1}x^{v+1}e^{(\alpha-1)x}}{v!} = \left[\sum_{i=0}^{r} \frac{x^i(\alpha-1)^i}{i!}\right] - e^{(\alpha-1)x}.$$

Since the right hand side is always negative, we have

$$\varphi^{(v)}(x) < (>)\, 0, \qquad v \text{ odd (even)}$$

for all $x \in [0, \infty)$.

Since (30) may be written

$$E(C(\alpha)) = 1 - \int_0^\infty e^{-(\alpha-1)x}\, dF(x),$$

we observe readily that $e^{-(\alpha-1)x}$ also satisfies the hypotheses of the theorem, and in fact is also completely monotonic.

When $k = 0$, $F_1^*(x)$ provides the maximum of $E\{\varphi(x)\}$ and $F_2^*(x)$ provides the infimum. From Theorem 9, we conclude that the situation is reversed for $k = 1$, i.e., $F_1^*(x)$ is a minimizing distribution and $F_2^*(x)$ provides the supremum. In general, we note this alternation every time an additional moment constraint is added. Further, the extremum which is *attained* for any value of k is not improved by the addition of the $k + 1$st moment constraint.

When $k = 0$, $F_1^*(x)$ provides the minimum of $E\{\psi(x)\}$, and $F_2^*(x)$ provides the supremum. Thus the solutions for $\varphi(x)$ and $\psi(x)$ are identical for every k upon interchanging supremum and infimum.

It is also interesting to observe the behavior of the upper and lower predictors for $d(\alpha)$ and $C(\alpha)$ as $\alpha \to \infty$. From the alternation property noted above and from Theorem 8, since the supremum of $E\{\varphi(x)\}$ has a mass point at zero for all k, the upper predictor for $d(\alpha) \to \infty$ as $\alpha \to \infty$. Similarly, we can observe, that as $\alpha \to \infty$, the lower predictor tends to a limit. This suggests the following interpretation: the upper predictor tends to infinity since, regardless of the sample size, there is no way for the experimenter to establish the nonexistence of an arbitrarily large number of classes each with negligible probability; the lower predictor tends to a limit, since there is no way for the experimenter to conclude that he will not observe all classes eventually.

Similarly, one can readily see, that the upper predictor for $C(\alpha) \to 1$ as $\alpha \to \infty$, and the lower predictor tends to a limit ξ, $0 \leqq \xi \leqq 1$. This has a similar interpretation to the corresponding result for $d(\alpha)$.

7. Historical remarks. Problems of this type have previously been investigated in papers by Corbet, Fisher, and Williams [2], Goodman [5], and Good and Toulmin [4].

The Corbet, Fisher, and Williams paper employs a parametric hypothesis as follows:

It is assumed, that for any class, the number of representatives in the sample has a Poisson distribution with mean m, where the values of m are distributed according to a Γ-type distribution. Then the expected number of classes observed is given by

$$E(d) = -\lambda \log (1 - \gamma) \tag{50}$$

and

$$E(N) = \frac{\lambda\gamma}{1 - \gamma} \tag{51}$$

where λ is independent of N, the sample size, and $\gamma/(1 - \gamma)$ is proportional to N.

Then, if the sample size is augmented to αN, or a second independent sample of size αN is observed, we have

$$\alpha N \sim \frac{\lambda \alpha \gamma}{1 - \gamma} \tag{52}$$

and

$$E\{d(\alpha)\} \sim -\lambda \log \left\{ \frac{\lambda}{\alpha N + \lambda} \right\}$$

where λ is determined from (50) and (51). Since

$$E\{d(\alpha)\} \sim d + \lambda \log \left\{ \frac{\alpha N + \lambda}{N + \lambda} \right\}, \tag{53}$$

we conclude, that for large N, the Corbet-Fisher-Williams hypothesis implies that the number of new classes to be observed when the sample is augmented by $(\alpha - 1)N$ is approximately $\lambda \log \alpha$.

Goodman considered the following problem:

A population with a known finite number of elements is partitioned into an unknown number of disjoint classes. The classes are assumed to possess no natural ordering. A random sample is drawn without replacement and we wish to estimate the number of classes in the population.

It is easily seen that this problem is identical to the problem of predicting the number of classes that will be observed in the case of an augmented sample.

Goodman has shown that in general, an unbiased estimator of $E\{d(\alpha)\}$ will not exist. However, if the maximum frequency of any class in the enlarged sample is known to be less than N, then the following estimator is unbiased.

$$\hat{d}_{G_1}(\alpha) = \sum_{i=1}^{N} A_i n_i \tag{54}$$

where

$$A_1 = \alpha,$$

$$A_i = \frac{(\alpha N)^{(i)}}{N^{(i)}} - \frac{A_1 i(\alpha N - N)^{(i-1)}}{(N-1)^{(i-1)}} - \frac{A_2 i(\alpha N - N)^{(i-2)}}{2!(N-2)^{(i-2)}} - \cdots - \frac{A_{i-1} i^{(i-1)}(\alpha N - N)}{(i-1)!(N-i+1)}, \qquad \text{for } i > 1.$$

Since $\hat{d}_{G_1}(\alpha)$ may give unreasonable answers, $\hat{d}_{G_2}(\alpha)$ was proposed by Goodman,

$$\hat{d}_{G_2}(\alpha) = \begin{cases} d, & \hat{d}_{G_1}(\alpha) < d, \\ \hat{d}_{G_1}(\alpha), & d \leqq \hat{d}_{G_1}(\alpha) \leqq \alpha N, \\ \alpha N, & \alpha N < \hat{d}_{G_1}(\alpha). \end{cases} \tag{55}$$

Goodman also proposed

$$d_{G_3}(\alpha) = \begin{cases} d, & \alpha N - \frac{\alpha(\alpha N - 1)}{N - 1} n_2 < d, \\ \alpha N - \frac{\alpha(\alpha N - 1)}{N - 1} n_2, & \text{otherwise.} \end{cases} \tag{56}$$

Good and Toulmin have obtained the following predictors.

$$\hat{d}(\alpha) = d - \sum_{i=1}^{\infty} (-1)^i(\alpha - 1)^i n_i \tag{57}$$

and

$$\hat{C}(\alpha) = 1 - \frac{1}{N} \sum_{i=1}^{\infty} (-1)^{i+1} i n_i(\alpha - 1)^{i-1}. \tag{58}$$

We present here, an alternative derivation of (57) and (58), which will exhibit their development from the moments (31).

Since $\varphi(x)$ is completely monotonic in $[0, \infty)$, it is a Laplace transform with respect to a monotonic non-decreasing function of bounded variation in $[0, \infty)$, i.e.,

$$\varphi(x) = \int_0^\infty e^{-tx}\, dG(t)$$

where $G(t)$ is non-decreasing and $G(\infty) < \infty$. Hence

$$\int_0^\infty \varphi(x)\, dF(x) = \int_0^\infty \int_0^\infty e^{-tx}\, dG(t)\, dF(x)$$

where $F(x)$ is a cumulative distribution function. The inverse Laplace transform is easily seen to be

$$G(t) = \begin{cases} t & 0 \leqq t \leqq \alpha - 1, \\ \alpha - 1, & \alpha - 1 < t. \end{cases}$$

Hence

$$\int_0^\infty \varphi(x)\, dF(x) = \int_0^\infty \int_0^{\alpha-1} e^{-tx}\, dt\, dF(x).$$

Interchanging the order of integration, we have

$$\int_0^\infty \varphi(x)\, dF(x) = \int_0^{\alpha-1} M_{-x}(t)\, dt$$

where $M_{-x}(t)$ is the moment generating function of $(-X)$. Since

$$EX^r \sim \frac{(r+1)!\, n_{r+1}}{n_1}$$

we have

$$M_{-x}(t) = \sum_{r=0}^{\infty} \frac{(-1)^r(r+1)n_{r+1}\, t^r}{n_1}.$$

Upon integrating $M_{-x}(t)$ term-by-term, we get

$$n_1 \int_0^\infty \frac{1 - e^{-(\alpha-1)x}}{x}\, dF(x) = \sum_{r=0}^{\infty} (-1)^r\, n_{r+1}(\alpha - 1)^{r+1}$$

from which we obtain (57).

Similarly, we can obtain (58), by noting that

$$\int_0^\infty [1 - \psi(x)]\, dF(x) = \int_0^\infty \int_0^\infty e^{-tx}\, dH(t)\, dF(x)$$

where

$$H(t) = \begin{cases} 0, & t < \alpha - 1, \\ 1, & t \geqq \alpha - 1, \end{cases}$$

so that, by interchanging the order of integration, and introducing the moments, we have

$$\int_0^\infty [1 - \psi(x)]\, dF(x) = \sum_{r=0}^{\infty} \frac{(-1)^r (r+1) n_{r+1} (\alpha - 1)^r}{n_1}.$$

Then, from (33) we get

$$E\{C(\alpha)\} \sim 1 - \frac{1}{N} \sum_{r=0}^{\infty} (-1)^r (r+1) n_{r+1} (\alpha - 1)^r.$$

Whenever $\alpha \geqq 2$, the Good-Toulmin formulas depend heavily on the higher moments, which the experimenter knows with less precision. However, in the Good-Toulmin paper, this difficulty is largely circumvented by transforming the series, so that a sum can be obtained from any of several methods for summation of divergent series.

8. Numerical examples. Three examples have been chosen to illustrate the methodology of this paper. The first is artificially constructed. The second appears in Good and Toulmin [4], and the third appears in both Good and Toulmin and Corbet, Fisher, and Williams [2].

Example (i). One hundred observations were taken from a multinomial population with 100 equiprobable cells. The data are summarized in Table 1.

TABLE 1

r	n_r
1	41
2	19
3	7
	$d = 67$

$m_1 = 9268,$
$m_2 = 1.0244,$
$m_r = 0,$
$r \geqq 3.$

The upper and lower predictors for the number of classes, employing the first k moments, $\hat{\bar{d}}_k(\alpha)$ and $\hat{\underline{d}}_k(\alpha)$ are given in Table 2.

TABLE 2

α	$\bar{d}_0(\alpha)$	$\hat{\underline{d}}_0(\alpha)$	$\hat{\underline{d}}_1(\alpha)$	$\bar{d}_2(\alpha)$
2	108	67	93.7	94.4
3	149	67	104.3	107.9
4	190	67	108.5	116.8
5	231	67	110.2	124.2
10	436	67	111.2	157.7
∞	∞	67	111.2	∞

In this case we are unable to proceed to $\hat{\underline{d}}_3(\alpha)$, since m_3 does not satisfy the conditions of Theorem 10.

We also note that $\hat{\underline{d}}_1(\alpha)$ exceeds 100 for $\alpha > 2$. Since $E(d) = 63.4$ and the observed value of d was 67, it is clear that our predictors should give answers which are a little high in this case.

The predictors of the coverage, $\hat{\underline{C}}_k(\alpha)$ and $\bar{C}_k(\alpha)$ are given in Table 3.

TABLE 3

α	$\bar{C}_0(\alpha)$	$\hat{\underline{C}}_0(\alpha)$	$\hat{C}_1(\alpha)$	$\hat{\underline{C}}_2(\alpha)$
2	1.00	.59	.838	.820
3	1.00	59	.936	.896
4	1.00	.59	.975	.921
5	1.00	59	.990	.930
10	1.00	.59	1.000	.934
∞	1.00	.59	1.000	.934

$$\hat{C}(1) = .59.$$

Example (ii). This example is due to Good and Toulmin [4]. 1000 words from "Our Mutual Friend" by Charles Dickens were tabulated and the results are given in Table 4. It is to be noted that the method of sampling, i.e., choosing the last words of lines on pages congruent to 5 modulo 25, is not random sampling; however, the data will nevertheless suffice to illustrate the method.

TABLE 4

r	n_r	r	n_r
1	404	6	3
2	57	7	0
3	24	8	3
4	16	≥ 9	15
5	6		
			$d = 528$

$m_1 = .2822,$
$m_2 = .3564,$
$m_3 = .9505.$

We obtain as predictors for the number of classes the results shown in Table 5.

TABLE 5

α	$\bar{\hat{d}}_0(\alpha)$	$\underline{\hat{d}}_0(\alpha)$	$\underline{\hat{d}}_1(\alpha)$	$\bar{\hat{d}}_2(\alpha)$	$\underline{\hat{d}}_3(\alpha)$
2	932	528	880	893	889
3	1336	528	1145	1221	1188
4	1740	528	1345	1539	1438
5	2144	528	1497	1854	1648
10	4164	528	1846	3423	2286
∞	∞	528	1960	∞	2737

Good and Toulmin have computed $\hat{d}(5)$ using their predictor and get 1683. The upper and lower predictors employing three moments are 1854 and 1643. The predictors of the coverage are given in Table 6.

TABLE 6

α	$\bar{\hat{C}}_0(\alpha)$	$\underline{\hat{C}}_0(\alpha)$	$\bar{\hat{C}}_1(\alpha)$	$\underline{\hat{C}}_2(\alpha)$	$\bar{\hat{C}}_3(\alpha)$
2	1.000	.596	.695	.661	.674
3	1.000	.596	.770	.679	.727
4	1.000	.596	.827	.684	.771
5	1.000	.596	.869	.686	.808
10	1.000	.596	968	.686	920
∞	1.000	.596	1.000	.686	1.000

$$\hat{C}(1) = .596.$$

Example (iii). This example is due to Corbet, Fisher, and Williams [2]. 15,609 Macrolepidoptera were caught in a light trap at Rothamsted and classified by species. The data are summarized in Table 7.

TABLE 7

r	n_r	r	n_r
1	35	6	11
2	11	7	5
3	15	≥ 8	139
4	14		
5	10		$d = 240$

$$m_1 = .6285, \quad m_2 = 2.5714, \quad m_3 = 9.6000.$$

The upper and lower predictors for the number of classes employing the first k moments, $\bar{d}_k(\alpha)$ and $\underline{\hat{d}}_k(\alpha)$ are given in Table 8.

TABLE 8

α	$\bar{d}_0(\alpha)$	$\underline{\hat{d}}_0(\alpha)$	$\underline{\hat{d}}_1(\alpha)$	$\bar{d}_2(\alpha)$
2	275	240	266.0	270.9
3	310	240	279.8	300.6
4	345	240	287.2	330.2
5	380	240	291.2	359.8
10	555	240	295.5	507.9
∞	∞	240	295.7	∞

Note that here, as in Example (i), we are unable to proceed to $\underline{\hat{d}}_3(\alpha)$, since m_3 does not satisfy the condition of Theorem 10 and hence only m_1 and m_2 are realizable as moments of a cumulative distribution function on $[0, \infty)$.

The corresponding predictors of coverage are given in Table 9.

TABLE 9

α	$\bar{C}_0(\alpha)$	$\underline{\hat{C}}_0(\alpha)$	$\bar{C}_1(\alpha)$	$\underline{\hat{C}}_2(\alpha)$
2	1.0000	.9978	.9988	.9981
3	1.0000	.9978	.9994	.9981
4	1.0000	.9978	.9997	.9981
5	1.0000	.9978	.9998	.9981
10	1.0000	.9978	1.0000	.9981
∞	1.0000	.9978	1.0000	.9981

$\hat{C}(1) = .9978.$

Employing the parametric hypotheses of Corbet, Fisher, and Williams, we have $Ed(2) \sim 267.9$.

Good and Toulmin obtain $\hat{d}(2) = 261.9$.

Williams (in Corbet et al) noted that doubling the sample size would approximately halve the proportion of the population not represented in the sample. In Good and Toulmin, $\hat{C}(2)$ is given as .9991.

APPENDIX A

In Section 2, the following approximations are employed

$$(1) \qquad \sum_{j=1}^{\infty} [1 - (1 - p_j)^N] \backsim \sum_{j=1}^{\infty} [1 - e^{-Np_j}]$$

$$(2) \qquad \sum_{j=1}^{\infty} \binom{N}{r} p_j^r (1 - p_j)^{N-r} \backsim \frac{1}{r!} \sum_{j=1}^{\infty} (Np_j)^r e^{-Np_j}$$

$$(3) \qquad 1 - \sum_{j=1}^{\infty} p_j (1 - p_j)^N \backsim 1 - \sum_{j=1}^{\infty} p_j e^{-Np_j}.$$

We will show that the approximations are satisfactory for large N. We first establish (1).

Consider

$$\left| \frac{\sum_{j=1}^{\infty} [1 - (1 - p_j)^N] - \sum_{j=1}^{\infty} [1 - e^{-Np_j}]}{\sum_{j=1}^{\infty} [1 - e^{-Np_j}]} \right|$$

where $0 < p_j < 1$ and $\sum_j p_j = 1$.

It is easy to see that if $a_i, b_i > 0$, $i = 1, 2, \cdots$; and $a_0/b_0 = \sup_i a_i/b_i$, then

$$\frac{a_0}{b_0} \geqq \frac{\sum_i a_i}{\sum_i b_i}.$$

Thus

$$\left| \frac{\sum_{j=1}^{\infty} [1 - (1 - p_j)^N] - \sum_{j=1}^{\infty} [1 - e^{-Np_j}]}{\sum_{j=1}^{\infty} [1 - e^{-Np_j}]} \right| \leqq \sup_p \frac{e^{-Np} - (1 - p)^N}{1 - e^{-Np}}.$$

Since

$$\begin{aligned}(1 - p)^N &= e^{N \log(1-p)} \\ &= \exp\left[- Np - \frac{Np^2}{2(1 - \xi)^2} \right] \qquad 0 \leqq \xi \leqq p\end{aligned}$$

we have

$$\begin{aligned}\frac{e^{-Np} - (1 - p)^N}{1 - e^{-Np}} &= \frac{e^{-Np}\left(1 - \exp\left[- \frac{Np^2}{2(1 - \xi)^2} \right]\right)}{1 - e^{-Np}} \\ &\leqq \frac{e^{-Np}\left(1 - \exp\left[- \frac{Np^2}{2(1 - p)^2} \right]\right)}{1 - e^{-Np}}.\end{aligned}$$

If $p \geqq 1/\sqrt{N}$, then

$$\frac{e^{-Np} - (1 - p)^N}{1 - e^{-Np}} \leqq \frac{e^{-Np}}{1 - e^{-Np}} \leqq \frac{e^{-\sqrt{N}}}{1 - e^{-\sqrt{N}}}$$

which clearly tends to zero as $N \to \infty$. For $p < 1/\sqrt{N}$,

$$\frac{e^{-Np} - (1 - p)^N}{1 - e^{-Np}} \leqq \frac{e^{-Np}(1 - e^{-\alpha(Np^2/2)})}{1 - e^{-Np}} = h_N(p)$$

where $\alpha = N/(\sqrt{N} - 1)^2$.

Differentiating $h_N(p)$ and equating to zero, we have

$$h_N'(p) = \frac{(1 - e^{-\alpha Np^2/2})(-Ne^{-Np}) + (N\alpha pe^{-Np-(Np^2\alpha/2)})(1 - e^{-Np})}{(1 - e^{-Np})^2} = 0$$

whence, we have

$$-(1 - e^{-\alpha Np^2/2}) + \alpha p(e^{-N\alpha p^2/2})(1 - e^{-Np}) = 0.$$

Thus

$$-1 + e^{-\alpha Np^2/2}(1 + \alpha p[1 - e^{-Np}]) = 0$$

or, since $p < 1/\sqrt{N}$

$$\log (1 + \alpha p[1 - e^{-Np}]) \sim \alpha p[1 - e^{-Np}]$$

we have

$$-\frac{N\alpha p^2}{2} + \alpha p[1 - e^{-Np}] \sim 0 \qquad \text{and} \qquad 1 - e^{-Np} \sim \frac{Np}{2}.$$

Thus, we can establish that the maximum of $h_N(p)$ occurs when $p \sim 1.6/N$. However,

$$h_N\left(\frac{1.6}{N}\right) \sim \frac{e^{1.6}}{1 - e^{-1.6}} (1 - e^{-(2.56\alpha/2N)})$$

which clearly tends to zero as $N \to \infty$. If $p_j = 1$ for some j, then all other terms in both sums in (1) = 0, and the approximation holds trivially. If $p_j = 0$, for some j, neither sum is increased, and hence we have established (1).

In considering (2), we suppose r^2 to be very small compared to N. We shall show that the approximation is satisfactory in the following sense; either both sides of (2) are negligible for sufficiently large N; or the ratio of the error to the expected value of the number of cells with r elements is small.

Thus, since

$$\binom{N}{r} \sim \frac{N^r}{r!} \exp\left[-\frac{r(r-1)}{2N}\right]$$

and

$$(1 - p)^{N-r} \sim \exp\left[-(N - r)p - \frac{(N - r)p^2}{2}\right], \qquad \text{for } p < 1,$$

we have

$$\sum_{j=0}^{\infty} \frac{(Np_j)^r e^{-Np_j}}{r!} - \sum_{j=0}^{\infty} \binom{N}{r} p_j^r(1 - p_j)^{N-r} \sim \sum_{j=0}^{\infty} \frac{(Np_j)^r}{r!} e^{-Np_j}$$
$$\left\{1 - \exp\left[rp_j - \frac{r(r-1)}{2N} - \frac{(N-r)}{2} p_j^2 - \cdots\right]\right\}.$$

First consider those indices j for which $p_j \geqq 1/N^{2/3}$. Then

$$\sum_{p_j \geqq (1/N^{2/3})} \frac{(Np_j)^r}{r!} e^{-Np_j} \left\{1 - \exp\left[rp_j - \frac{r(r-1)}{2N} - \frac{(N-r)}{2} p_j^2 - \cdots\right]\right\} \leqq \sum_{p_j \geqq (1/N^{2/3})} \frac{(Np_j)^r}{r!} e^{-Np_j}.$$

For N sufficiently large,

$$\sum_{p_j \geqq (1/N^{2/3})} \frac{(Np_j)^r e^{-Np_j}}{r!} \leqq \frac{N^{(r+2)/3}}{r!} e^{-N^{1/3}},$$

which is negligible for sufficiently large N. The approximation holds trivially for the case $p_j = 1$ for some j, and 0 for all other indices j.

Now consider,

$$\frac{\displaystyle\sum_{p_j < (1/N^{2/3})} \frac{(Np_j)^r e^{-Np_j}}{r!} \left\{1 - \exp\left[rp_j - \frac{r(r-1)}{2N} - \frac{(N-r)}{2} p_j^2 - \cdots\right]\right\}}{\displaystyle\sum_{p_j < (1/N^{2/3})} \frac{(Np_j)^r e^{-Np_j}}{r!}}$$

$$\leqq \sup_{p < (1/N^{2/3})} \frac{\displaystyle\frac{(Np)^r e^{-Np}}{r!} \left\{1 - \exp\left[rp - \frac{r(r-1)}{2N} - \frac{(N-r)}{2} p^2 - \cdots\right]\right\}}{\displaystyle\frac{(Np)^r e^{-Np}}{r!}}$$

$$= \sup_{p < (1/N^{2/3})} 1 - \exp\left[rp - \frac{r(r-1)}{2N} - \frac{(N-r)}{2} p^2 - \cdots\right]$$

$$= 1 - e^{0(1/N^{1/3})}.$$

The discussion of (3) is almost identical with that of (2) and will not be produced here.

APPENDIX B

Asymptotic mean square error of $\hat{C}(1)$. In Section 3, we established $\hat{C}(1) = 1 - (n_1/N)$. It is clear that $E[C - \hat{C}(1)]^2 = E(D - (n_1/N))^2$ where $D = 1 - C$. From (14) and (15) in Section 2, we have

$$ED^2 = \sum_{i,j} p_i p_j (1 - p_i - p_j)^N + \sum_j p_j^2[(1 - p_j)^N - (1 - 2p_j)^N].$$

From (11) in Section 2, we have

$$E\frac{n_1^2}{N^2} = \frac{1}{N^2} En_1 + \frac{1}{N^2} \sum_{i,j} \frac{N!}{(N-2)!} p_i p_j (1 - p_i - p_j)^{N-2} - \frac{1}{N^2} \sum_{j=1}^{\infty} \frac{N!}{(N-2)!} p_j^2 (1 - 2p_j)^{N-2}.$$

Then, to find $E(Dn_1)$, we introduce two random variables

$$X_j = \begin{cases} 1, & \text{if } j\text{th cell does not occur in sample,} \\ 0, & \text{otherwise,} \end{cases}$$

$$Y_j = \begin{cases} 1, & \text{if } j\text{th cell occurs once in sample,} \\ 0, & \text{otherwise.} \end{cases}$$

Then $D = \sum_j p_j X_j$, $n_1 = \sum_j Y_j$. Thus

$$\begin{aligned} E(Dn_1) &= \sum_{i,j} E(p_i X_i Y_j) \\ &= \sum_{i \neq j} N p_i p_j (1 - p_i - p_j)^{N-1} \\ &= \sum_{i,j} N p_i p_j (1 - p_i - p_j)^{N-1} - \sum_j N p_j^2 (1 - 2p_j)^{N-1}. \end{aligned}$$

Hence, upon introducing the exponential approximations, we have

$$E\left(D - \frac{n_1}{N}\right)^2 \sim \frac{2En_2}{N^2} + \frac{En_1}{N^2}.$$

REFERENCES

[1] H. Chernoff and S. Reiter, "Selection of a distribution function to minimize an expectation subject to side conditions," *Stanford University Technical Report No.* 23, 1954.

[2] A. S. Corbet, R. A. Fisher, and C. B. Williams, "The relation between the number of species and the number of individuals in a random sample of an animal population," *J. Anim. Ecol.*, Vol. 12 (1943), pp. 42–58.

[3] I. J. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, Vol. 40 (1953), pp. 237–264.

[4] I. J. Good and G. H. Toulmin, "The number of new species and the increase in population coverage when a sample is increased," *Biometrika*, Vol. 43 (1956), pp. 45–63.

[5] L. A. Goodman, "On the estimation of the number of classes in a population," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 572–579.

[6] S. Karlin and L. S. Shapley, "The geometry of moment spaces," *Memoirs of the Amer. Math. Soc.* No. 12, 1953.

[7] J. S. Rustagi, "On minimizing and maximizing a certain integral with statistical applications," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 308–328.

[8] A. Wald, "Limits of a distribution function determined by absolute moments and inequalities satisfied by absolute moments," *Trans. Amer. Math. Soc.* Vol. 46 (1939), pp. 280–306.