AD 608855

# THE UNIVERSITY OF WISCONSIN

madison, wisconsin



## ATHEMATICS RESEARCH CENTER

Best Available Copy

MATHEMATICS RESEARCH CENTER, UNITED STATES ARMY

THE UNIVERSITY OF WISCONSIN

Contract No.: DA-11-022-ORD-2059

# AN ASYMPTOTIC LOWER BOUND FOR THE ENTROPY OF DISCRETE POPULATIONS WITH APPLICATION TO THE ESTIMATION OF ENTROPY FOR UNIFORM POPULATIONS

E. B. Cobb and Bernard Harris

MRC Technical Summary Report #516
October 1964

Madison, Wisconsin

ABSTRACT

In this paper we obtain an asymptotic lower bound for the entropy of a multinomial population with an unknown and perhaps countably infinite number of classes. This bound is a function of the first $k + 1$ occupancy numbers of a random sample, and is a useful estimator when most of the sample information is contained in the low order occupancy numbers.

# AN ASYMPTOTIC LOWER BOUND FOR THE ENTROPY OF DISCRETE POPULATIONS WITH APPLICATION TO THE ESTIMATION OF ENTROPY FOR UNIFORM POPULATIONS

B. B. Cobb and Bernard Harris

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. That is, if $X_j$ is the $j$th observation, and $M_i$ the $i$th class, then

$$P\{X_j \in M_i\} = p_i \geq 0 \qquad i = 1, 2, \dots ; \quad j = 1, 2, \dots, N$$

and $\sum_{i=1}^{\infty} p_i = 1$. The classes are not assumed to have a natural ordering.

Let $n_r$ be the number of classes which occur exactly $r$ times in the sample. Then $\sum_{r=0}^{\infty} r\, n_r = N$.

Defining the entropy of the population by

$$(1) \qquad H(p_1, p_2, \dots) = -\sum_{i=1}^{\infty} p_i \log p_i$$

it is shown, that for the cumulative distribution function $F^*(x)$, defined by

$$(2) \qquad F^*(x) = \sum_{Np_i \leq x} Np_i e^{-Np_i} \Big/ \left(\sum_{j=1}^{\infty} Np_j e^{-Np_j}\right)$$

we have

$$(3) \qquad H(p_1, p_2, \dots) \sim \frac{1}{N} E(n_1) \int_{0}^{\infty} x^{-1} \log\left(\frac{N}{x}\right) dF^*(x) \quad .$$

---

Sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin under Contract No.: DA-11-022-ORD-2059.

In addition, in Harris [1], it is shown that the moments of $F^*(x)$, $\mu_1, \mu_2, \ldots$, are approximately given by

$$(4) \qquad \mu_r \sim \frac{(r+1)!\, E(n_{r+1})}{E(n_1)} \quad .$$

If we then replace the expected values in (4) by the observed values, defining

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

estimates of the moments of $F^*(x)$ are obtained. Then, let

$\mathfrak{F}^{[a,b]}_{(m_1, m_2, \ldots, m_k)}$ be the set of cumulative distribution functions with $F(a-0) = 0$, $F(b) = 1$, and

$$\int_{-\infty}^{\infty} x^j dF(x) = m_j \, , \quad j = 1, 2, \ldots, k \; .$$

Since $p_1, p_2, \ldots$ are all assumed to be unknown, $F^*(x)$ is unknown, and an asymptotic lower bound to (3) may be found by minimizing

$$\int_{-\infty}^{\infty} e^x \log\left(\frac{N}{x}\right) dF(x)$$

over the set $\mathfrak{F}^{[0,N]}_{(m_1, m_2, \ldots, m_k)}$. This process uses only the information contained in the first $k+1$ occupancy numbers $n_1, n_2, \ldots, n_{k+1}$, and is particularly useful, when the sample information concerning the parameters $p_1, p_2, \ldots$ is concentrated in the low order occupancy numbers. This occurs, for example, if as $N \to \infty$, $p_j \to 0$, $j = 1, 2, \ldots$, in such a way that $0 \leq Np_j < \lambda$, where $\lambda$ is approximately $k+1$.

The mimimum is explicitly computed for $k = 2$. The process employed here is compared with the maximum likelihood estimates of entropy for uniform populations with $p_j = \frac{1}{M}$, $j = 1, 2, \ldots, M$ and $M \to \infty$ as $N \to \infty$ so that $N/M \to \lambda > 0$.

2. The computation of the lower bound for entropy. In Harris [1], it was shown that for $r^2 = O(N)$ as $N \to \infty$,

(5) $$E(n_r) \sim \frac{1}{r!} \Sigma_{j=1}^{\infty} (Np_j)^r e^{-Np_j} ,$$

where the approximation is valid, in the sense that, either both sides are negligible, or the ratio of the two sides approaches unity.

In particular,

(6) $$E(n_1) \sim \Sigma_{j=1}^{\infty} Np_j e^{-Np_j} ;$$

hence

$$\frac{1}{N} E(n_1) \int_{-\infty}^{\infty} e^x \log (\frac{N}{x}) \, dF^*(x)$$

$$\sim \frac{1}{N} \Sigma_{j=1}^{\infty} e^{Np_j} \log (\frac{1}{pj}) Np_j e^{-Np_j}$$

$$= H(p_1, p_2, \ldots) .$$

Let $h(x) = e^x \log \frac{N}{x}$. Then we wish to determine $F_0(x) \in \mathfrak{F}_{(m_1, m_2)}^{[0, N]}$ such that

(7) $$\min_{F(x) \in \mathfrak{F}_{(m_1, m_2)}^{[0, N]}} \int_{-\infty}^{\infty} h(x) \, dF(x) = \int_{-\infty}^{\infty} h(x) \, dF_0(x) .$$

Since $h(0)$ does not exist, we consider instead $\pi^{[\epsilon, N]}_{(m_1, m_2)}$, where $\epsilon > 0$, is arbitrary. Then $h(x)$ is bounded on $[\epsilon, N]$ for every $\epsilon > 0$ and it is well-known [1] that $F_\epsilon(x)$ defined by

$$(8) \qquad \min_{F(x) \epsilon \pi^{[\epsilon, N]}_{(m_1, m_2)}} \int_{-\infty}^{\infty} h(x)\, dF(x) = \int_{-\infty}^{\infty} h(x)\, dF_\epsilon(x) ,$$

is obtainable as a discrete cumulative distribution function with at most three jumps, say at $x_1, x_2, x_3$, $\epsilon \leq x_1 < x_2 < x_3 \leq N$. Hence, there exists $\lambda_1, \lambda_2, \lambda_3 \geq 0$, $\Sigma_{i=1}^{3} \lambda_i = 1$, with

$$(9) \qquad \begin{cases} \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 = m_1 \\ \\ \lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2 = m_2 , \end{cases}$$

such that

$$(10) \qquad F_\epsilon(x) = \begin{cases} 0 \ , & x < x_1 \\ \lambda_1 \ , & x_1 \leq x < x_2 \\ \lambda_1 + \lambda_2 , & x_2 \leq x < x_3 \\ 1 \ , & x \geq x_3 \end{cases}$$

whenever $m_2 \geq m_1^2$, a condition which we will assume throughout the remainder of this discussion. With no loss in generality, we may assume that $m_2 > m_1^2$, since otherwise $F_\epsilon(x)$ is a cumulative distribution function with exactly one jump, and (8) has a trivial solution.

It can be shown that $\lambda_i \geq 0$, $i = 1, 2, 3$, if and only if

$$(11) \qquad (-1)^{i+j-1}(x_i x_j - m_1(x_i+x_j)+m_2) \geq 0, \qquad 1 \leq i < j \leq 3 .$$

In addition, from Harris [1], there exist real numbers $\alpha_0, \alpha_1, \alpha_2$ such that $x_1, x_2$, and $x_3$ are roots of

$$(12) \qquad g(x) = \Sigma_{i=0}^{2} \alpha_i x^i - h(x) = 0,$$

and

$$(13) \qquad \Sigma_{i=0}^{2} \alpha_i x^i - h(x) \leq 0, \qquad \epsilon \leq x \leq N .$$

From (11) and (12), we also have that for $\epsilon < x_i < N$, $i = 1, 2, 3$;

$$(14) \qquad g'(x_i) = \alpha_1 + 2\alpha_2 x_i - h'(x_i) = 0 .$$

To solve (9), (12), (13) and (14), observe that there exist numbers $\delta_1, \delta_2, \delta_3$, $0 < \delta_1 < \delta_2 < \delta_3 < N$, such that

$$h'(x) \begin{cases} < 0, & 0 < x < \delta_1, \\ > 0, & \delta_1 < x < \delta_3, \\ < 0, & \delta_3 < x \leq N, \end{cases}$$

and

$$h''(x) \begin{cases} > 0, & 0 < x < \delta_2 \\ < 0, & \delta_2 < x \leq N \end{cases}$$

with

$$\begin{aligned} &\delta_1 \to 0 , && N \to \infty \\ &\delta_2 = (N-2)+O(\tfrac{1}{N}), && N \to \infty \\ &\delta_3 = (N-1)+O(\tfrac{1}{N}), && N \to \infty \end{aligned}$$

and $h''(x)$ is strictly decreasing on $(0, \delta_1)$ and $(\delta_2, N)$. We now establish the following

**Lemma.** If $\epsilon < x_1 < x_2 < N$ $(0 < \epsilon < \delta_1)$, the following conditions cannot be satisfied simultaneously

(15) $$\Sigma_{i=0}^{2} \alpha_i x^i \leq h(x), \qquad \epsilon \leq x \leq N$$

(16) $$\Sigma_{i=0}^{2} \alpha_i x_j^i = h(x_j), \qquad j = 1, 2 .$$

**Proof.** Assume (15) and (16) hold. Let $p(x) = \Sigma_{i=0}^{2} \alpha_i x^i$ . Then

(17) $$h'(x_j) = p'(x_j), \qquad j = 1, 2 .$$

Let $I_1 = (\epsilon, \delta_1]$, $I_2 = (\delta_1, \delta_2]$, $I_3 = (\delta_2, N)$. Assume $\alpha_2 > 0$. Then if $x_2 \in I_3$, since $p(x)$ is strictly convex and $h(x)$ is strictly concave in $I_3$, by (16) and (17), we have $p(x_0) > h(x_0)$ for some $x_0 \in I_3$, contradicting (15). If $x_2 \in I_2$, then $p'(x_2) > 0$, hence $p(N) > p(x_2) > 0 = h(N)$, contradicting (15). If $x_2 \in I_1$, then $\epsilon < x_1 < x_2 \leq \delta_1$, and by (16) and Rolle's Theorem, there exist $\xi_1, \xi_2$, $x_1 < \xi_1 < \xi_2 < x_2$ such that $g''(\xi_j) = 0$, $j = 1, 2$. This, however, implies that $h''(\xi_j) = 2\alpha_2$, $j = 1, 2$, contradicting the monotonicity of $h''(x)$.

If $\alpha_2 < 0$, the argument is similar. The case $\alpha_2 = 0$ is trivial.

We now obtain $F_0(x)$ .

**Theorem 1.** There exists a unique cumulative distribution function $F_0(x) \in \mathcal{F}_{(m_1, m_2)}^{[0, N]}$ such that

$$\int_{-\infty}^{\infty} h(x)\, dF_0(x) = \min_{F(x) \in \mathcal{F}_{(m_1, m_2)}^{[0, N]}} \int_{-\infty}^{\infty} h(x)\, dF(x)$$

given by

$$
(18) \qquad F_0(x) = \begin{cases} 0 & , \quad x < \dfrac{Nm_1 - m_2}{N - m_1} \\[2ex] \dfrac{(N-m_1)^2}{(N-m_1)^2 + (m_2 - m_1^2)} & , \quad \dfrac{Nm_1 - m_2}{N - m_1} \le x < N \\[2ex] 1 & , \quad x \ge N \end{cases}
$$

Proof. By the above lemma, we have $x_1 = \epsilon$, $\epsilon < x_2 < N$, $x_3 = N$. From (11), we have

$$
(19) \qquad \frac{Nm_1 - m_2}{N - m_1} \le x_2 \le \frac{m_2 - m_1 \epsilon}{m_1 - \epsilon} .
$$

Thus, by (9), we have

$$
\lambda_1(x_2, \epsilon) = \frac{Nx_2 - m_1(N + x_2) + m_2}{(x_2 - \epsilon)(N - \epsilon)}
$$

$$
\lambda_2(x_2, \epsilon) = \frac{-(\epsilon N - m_1(N + \epsilon) + m_2)}{(x_2 - \epsilon)(N - x_2)} ,
$$

and

$$
\lim_{\epsilon \to 0} \lambda_1(x_2, \epsilon) = \frac{N x_2 - m_1(N + x_2) + m_2)}{x_2 N} ,
$$

$$
\lim_{\epsilon \to 0} \lambda_2(x_2, \epsilon) = \frac{N m_1 - m_2)}{x_2(N - x_2)} .
$$

This gives a parametric family of cumulative distribution functions $F_{0, x_2}(x)$. Since $\lim_{x \to 0+} h(x) = \infty$, we must have $\lambda_1(x_2, \epsilon) = O(\frac{1}{h(\epsilon)})$, $\epsilon \to 0$, since otherwise $F_\epsilon(x)$ would not satisfy (8). Hence $\lim_{\epsilon \to 0} \lambda_1(x_2, \epsilon) = 0$ and

$x_2 \to \dfrac{Nm_1 - m_2}{N - m_1}$ as $\epsilon \to 0$. Since $\lambda_1(x_2, \epsilon)\, h(\epsilon) \geq 0$ for every $\epsilon > 0$, it follows that $\lambda_1(x_2, \epsilon) = o(\frac{1}{h(\epsilon)})$ as $\epsilon \to 0$, establishing the theorem.

Finally we have:

<u>Theorem 2.</u> The required lower bound for the entropy is

$$\frac{n_1}{N} \int_{-\infty}^{\infty} h(x)\, dF_0(x) =$$

$$\frac{n_1}{N} \frac{(N-m_1)^2}{(N-m_1)^2 + (m_2 - m_1^2)}\, e^{\frac{Nm_1 - m_2}{N - m_1}} \log \frac{N(N-m_1)}{Nm_1 - m_2} .$$

<u>Remark.</u> Krein [2] has studied minimization problems similar to (8). However, Krein's methods require that $1, x, x^2, h(x)$ form a Tschebycheffian system of functions on $[\epsilon, N]$. A necessary condition for the above (see Pólya and Szegö [3]) is that the Wronskians

$$W(x) = \begin{vmatrix} 1 & x & x^2 & h(x) \\ 0 & 1 & 2x & h'(x) \\ 0 & 0 & 2 & h''(x) \\ 0 & 0 & 0 & h'''(x) \end{vmatrix}, \quad \epsilon \leq x \leq N ,$$

be non-negative (non-positive) on $[\epsilon, N]$. This condition is clearly not satisfied in this case and Krein's methods are therefore inapplicable.

## 3. <u>The Estimation of the Entropy of Uniform Populations</u>.

Let

$$p_j = \begin{cases} \dfrac{1}{M} & j = 1, 2, \ldots, M \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$F^*(x) \begin{cases} 0 & x < N/M \\ 1 & x \geq N/M \end{cases}$$

where

$$N \to \infty, \; M \to \infty \quad \text{so that} \quad N/M \to \lambda > 0 .$$

Then,

$$E(m_r) \sim \frac{M}{r!} \lambda^r e^{-\lambda} \qquad r = 1, 2, \ldots$$

and

$$\mu_r = \lambda^r \qquad r = 1, 2, \ldots$$

In this case,

$$\frac{1}{N} E(m_1) \int_{-\infty}^{\infty} h(x)\, dF^*(x) = e^{-\lambda} h(\lambda) - \log M$$

as required.

In addition, the class $\mathcal{F}\left\{ {0, N \atop (\mu, \mu_2)} \right\}$ contains only $F^*(x)$, so that the solution of (7) provides an estimation of $H(p_1, p_2, \ldots)$ rather than a lower bound.

In the replacement of $\mu_1, \mu_2$ by the sample quantities $m_1, m_2$, it may happen that $m_2 < m_1^2$. This, of course suggests that $F^*(x)$ is degenerate, and in such cases, we take $m_2 = m_1^2$.

By way of contrast, the maximum likelihood estimate $\hat{H}$ is poor under the limiting process employed there, since

$$E(\hat{H}) = \Sigma_{i=1}^{N} E(n_i)\frac{i}{N}\log\left(\frac{i}{N}\right)$$

and for $M = 1000$, $N = 100$, we have $E(n_1) = 90.48$, $E(n_2) = 4.52$, $E(n_3) = .15$ obtaining

$$E(\hat{H}) = 4.271$$

and log $M = 6.908$.

Example. Three random samples were chosen with $N = 1000$, $M = 1000$. The data are summarized below.

| | Sample #1 | Sample #2 | Sample #3 |
|---|---|---|---|
| $n_1$ | 373 | 341 | 377 |
| $n_2$ | 199 | 179 | 169 |
| $n_3$ | 62 | 70 | 60 |
| $n_4$ | 8 | 17 | 25 |
| $n_5$ | 1 | 2 | 1 |
| $n_6$ | 1 | 1 | 0 |
| $n_7$ | 0 | 1 | 0 |
| $m_1$ | 1.067 | 1.050 | .897 |
| $m_2$ | .997 | 1.232 | .955 |
| $\frac{n_1}{N}\int h(x)\,dF_0(x)$ | ..... | 6.683 | 6.486 |
| $H(p_1,\cdots,p_M)$ | 6.908 | 6.908 | 6.908 |
| $\hat{H}$ | 6.364 | 6.294 | 6.329 |

In sample #1, $m_2 < m_1^2$, then supposing $F^*_{(x)}$ to be degenerate with a jump of 1 at $m_1$, we get, using $m_2 = m_1^2$, $\frac{n_1}{N}\int h(x)\,dF_0(x) = 7.419$.

## REFERENCES

[1] Harris, Bernard (1959). Determining bounds on integrals with applications to cataloging problems, Ann. Math. Stat. **30**, 521-548.

[2] Krein, M. G. (1951). The ideas of P. L. Čebyšev and A. A. Markov in the theory of limiting values of integrals and their further development. (In Russian). Uspehi Mat. Nauk (N. S.) **6** No. 4(44) 3-120; in English: Amer. Math. Soc. Translations Ser. 2 **12** (1959), 1-121.

[3] Pólya, G. and Szegö, G. (1925). Aufgaben und Lehrsätze aus der Analysis. Bd. II, Springer, Berlin.

Mathematics Research Center, U.S. Army — UNCLASSIFIED

AN ASYMPTOTIC LOWER BOUND FOR THE ENTROPY OF DISCRETE POPULATIONS WITH APPLICATION TO THE ESTIMATION OF ENTROPY FOR UNIFORM POPULATIONS

E. B. Cobb and Bernard Harris

MRC Report No. 516 AD________

Contract No.: DA-11-022-ORD-2059

Statistics
Multinomial populations
Estimation of entropy

In this paper we obtain an asymptotic lower bound for the entropy of a multinomial population with an unknown and perhaps countably infinite number of classes. This bound is a function of the first $k + 1$ occupancy numbers of a random sample, and is a useful estimator when most of the sample information is contained in the low order occupancy numbers. pp. 11

Mathematics Research Center, U.S. Army — UNCLASSIFIED

AN ASYMPTOTIC LOWER BOUND FOR THE ENTROPY OF DISCRETE POPULATIONS WITH APPLICATION TO THE ESTIMATION OF ENTROPY FOR UNIFORM POPULATIONS

E. B. Cobb and Bernard Harris

MRC Report No. 516 AD________

Contract No.: DA-11-022-ORD-2059

Statistics
Multinomial populations
Estimation of entropy

In this paper we obtain an asymptotic lower bound for the entropy of a multinomial population with an unknown and perhaps countably infinite number of classes. This bound is a function of the first $k + 1$ occupancy numbers of a random sample, and is a useful estimator when most of the sample information is contained in the low order occupancy numbers. pp. 11

Mathematics Research Center, U.S. Army — UNCLASSIFIED

AN ASYMPTOTIC LOWER BOUND FOR THE ENTROPY OF DISCRETE POPULATIONS WITH APPLICATION TO THE ESTIMATION OF ENTROPY FOR UNIFORM POPULATIONS

E. B. Cobb and Bernard Harris

MRC Report No. 516 AD________

Contract No.: DA-11-022-ORD-2059

Statistics
Multinomial populations
Estimation of entropy

In this paper we obtain an asymptotic lower bound for the entropy of a multinomial population with an unknown and perhaps countably infinite number of classes. This bound is a function of the first $k + 1$ occupancy numbers of a random sample, and is a useful estimator when most of the sample information is contained in the low order occupancy numbers. pp. 11

Mathematics Research Center, U.S. Army — UNCLASSIFIED

AN ASYMPTOTIC LOWER BOUND FOR THE ENTROPY OF DISCRETE POPULATIONS WITH APPLICATION TO THE ESTIMATION OF ENTROPY FOR UNIFORM POPULATIONS

E. B. Cobb and Bernard Harris

MRC Report No. 516 AD________

Contract No.: DA-11-022-ORD-2059

Statistics
Multinomial populations
Estimation of entropy

In this paper we obtain an asymptotic lower bound for the entropy of a multinomial population with an unknown and perhaps countably infinite number of classes. This bound is a function of the first $k + 1$ occupancy numbers of a random sample, and is a useful estimator when most of the sample information is contained in the low order occupancy numbers. pp. 11