NOTE ON ESTIMATING INFORMATION

BY
COLIN R. BLYTH

TECHNICAL REPORT NO. 17

PREPARED UNDER CONTRACT Nonr-225 (21)
(NR-042-993)
FOR
OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

JANUARY 24, 1958
NOTE ON ESTIMATING INFORMATION

BY

COLIN R. ELYTH

UNIVERSITY OF ILLINOIS AND STANFORD UNIVERSITY

TECHNICAL REPORT NO. 17

JANUARY 24, 1958

PREPARED UNDER CONTRACT Nonr-225(21)
(NR-042-993)

FOR

OFFICE OF NAVAL RESEARCH

REPRODUCTION IN WHOLE OR IN PART IS PERMITTED FOR
ANY PURPOSE OF THE UNITED STATES GOVERNMENT

APPLIED MATHEMATICS AND STATISTICS LABORATORY
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
NOTE ON ESTIMATING INFORMATION

by

Colin R. Blyth

University of Illinois and Stanford University

1. **SUMMARY.** This note is concerned with estimation of the Shannon-Wiener measure of information. Low bias estimates are obtained, and their bias and variance. These estimates are extended to the case where the number of possible values of the random variable is not known. The estimates are compared asymptotically with the maximum likelihood estimates. Minimax estimates (squared error loss function) are found for a few very special cases.

2. **INTRODUCTION.** Consider a random variable \( Y \) with finitely many distinct possible values:

\[
P(Y = a_i) = p_i, \quad i = 1, \ldots, k.
\]

A metric measure of dispersion of \( Y \) measures how spread out the distribution of \( Y \) is, in terms of distance in the space of \( Y \). If this space has no relevant distance function (e.g. \( k = 3 \), \( a_1 = \text{green}, a_2 = \text{red}, a_3 = \text{white} \)), there is no relevant metric measure of dispersion. An absolute measure of dispersion of \( Y \) measures the degree to which the total probability of \( 1 \) is broken up into pieces in the distribution of \( Y \). Such a measure is a function of \( p_1, \ldots, p_k \) only; is free from dependence on the \( a_i \)'s; is large when the probability is much broken up (e.g. \( p_1, \ldots, p_k = 1/k, \ldots, 1/k \)), small when it is not much broken up (e.g. \( p_1, \ldots, p_k = .99, .01, 0, \ldots, 0 \)).
In handling both kinds of dispersion measures the following addition property plays the same important role: (Divide the values of $Y$ into groups. Dispersion of $Y$ = between group dispersion + expected within group dispersion.). Knowing the distribution of $Y$ gives information useful in predicting $Y$. Actually observing $Y$ gives additional information - enough for perfect prediction. This additional information can be called information in $Y$ or unpredictability of $Y$ and can be measured by a measure of dispersion. In this language the addition property says that the information in observing $Y$ equals the information in observing which group $Y$ falls in plus the expected information in observing which member of that group.

For real valued $Y$ the addition property identifies variance (except for a constant multiplier) among all metric dispersion measures of the form $Ef(|Y - EY|)$ with $f$ continuous. This easily extends to weighted averages of the partial variances when $Y$ has values in a Euclidean $n$-space. Similarly the addition property identifies information or entropy $H = -\sum P_i \log_2 P_i$ (except for a constant multiplier) among all absolute dispersion measures $f_k(P_1, \ldots, P_k)$ with $f_k$ continuous and $f_k(1/k, \ldots, 1/k)$ an increasing function of $k$, as is proved in [1].

The addition property leads to very convenient mathematical simplifications. For this reason variance and information $H$ are very widely used dispersion measures. But there seems to be little intuitive necessity for the addition property. Thus in the metric case it seems quite reasonable to use measures like $E|Y - EY|$ which lack the property -- and loss functions other than squared error. Similarly in the absolute case it would seem quite reasonable
to use measures like the natural chi-square measure \((k-1) - k \sum (p_i - 1/k)^2\)
which lack the property. Essentially equivalent to this chi-square measure
is the following linear function of it, which is the terms of order up to 2
in a Taylor series for \(H\).

\[
H_2 = 1 - \frac{C}{2} \left\{ k-2 + \sum_{i=1}^{k} \left( 2p_i - 1 \right)^2 / 2 \right\}
\]

If \(f(1,0,\ldots,0) = 0\) is desired we could make the necessary subtraction from
the measure \(H_2\).

This note is concerned with estimation of information or entropy of \(Y\):

\[
H = H(p_1, \ldots, p_k) = -C \sum_{i=1}^{k} p_i \log p_i ,
\]

where \(C = \log_2 e = 1.442695\) and \(p_i \log p_i\) is taken to be 0 whenever
\(p_i = 0\). Our estimate is to be based on independent repetitions \(Y_1, \ldots, Y_n\)
of the experiment \(Y\). Then \(X_1, \ldots, X_k\), where \(X_i\) is the number of \(Y\)'s
with the value \(a_i\), is a sufficient statistic for \(p_1, \ldots, p_k\) and has the
following multinomial family of possible distributions

\[
P(X_1, \ldots, X_k = x_1, \ldots, x_k) = n! \prod_{i=1}^{k} p_i^{x_i} / x_i ! ,
\]

(1)

\[x_i = 0,1,\ldots,n ; \quad \sum_{i=1}^{k} x_i = n ; \quad 0 \leq p_i \leq 1 ; \quad \sum_{i=1}^{k} p_i = 1 .\]

We are now concerned, then, with the problem of what function \(f_k(X_1, \ldots, X_k)\)
to use as an estimate for \(H\). The maximum likelihood estimate is considered
by Miller and Madow [2] ; it is good when \( n \) is large but is likely to be poor for \( n \) small. We will try to find the best unbiased estimate, and the minimax estimate for squared error loss function.

3. **LOW BIAS ESTIMATION, \( k \) KNOWN.** Since (1) is a complete family of distributions, the problem of unbiased estimation of any function \( g(p_1, \ldots, p_k) \) is solved by Lehmann and Scheffé [3]. In fact, since \( Ef(X_1, \ldots, X_k) \) is a degree \( \leq n \) polynomial in \( p_1, \ldots, p_k \), no other functions possess unbiased estimates. And using the usual factorial notation \( x^*(v) = x(x-1) \cdots (x-v+1) \) we have \( E\{X_1 \cdots X_k\} = n \sum_{i=1}^{k} v_i \), which reduces to \( 0 = 0 \) whenever \( \sum_{i=1}^{k} v_i > n \). Thus

\[
\sum c(v_1, \ldots, v_k) x_1^{v_1} \cdots x_k^{v_k} / n \]

is the unique uniformly minimum variance (U.M.V.) unbiased estimate of \( \sum c(v_1, \ldots, v_k) p_1^{v_1} \cdots p_k^{v_k} \), where summation is over any set of \( (v_1, \ldots, v_k) \)'s with \( \sum_{i=1}^{k} v_i \leq n \) for every member. So we can write down the U.M.V. unbiased estimate for every degree \( \leq n \) polynomial, and no other function of \( p_1, \ldots, p_k \) has an unbiased estimate.

It is now clear that there is no unbiased estimate for \( H \). If low bias is what we want, the next best thing would be to use the U.M.V. unbiased estimate of the degree \( n \) polynomial which is in some sense (e.g. smallest maximum distance apart) closest to \( H \). A simpler polynomial which agrees quite closely with \( H \) is the terms of degree \( \leq n \) in the Taylor series expansion of \( H \) about the point \( (1/2, \ldots, 1/2) \). We will use the U.M.V. unbiased estimate of this polynomial as an estimate for \( H \).
Writing \( \gamma_1 = p_1 - 1/2 \) we have

\[
p_1 \log p_1 = \left( 1/2 + \gamma_1 \right) \log (1/2 + \gamma_1)
\]

\[
= - \frac{1}{2c} + (1 - \frac{1}{c}) \gamma_1 + \frac{1}{2} \left\{ \frac{(2\gamma_1)^2}{1 \cdot 2} - \frac{(2\gamma_1)^3}{2 \cdot 3} + \frac{(2\gamma_1)^4}{3 \cdot 4} + \cdots \right\}.
\]

\[
\therefore H = - C \sum_{i=1}^{k} p_i \log p_i
\]

\[
= 1 - \frac{c}{2} \left\{ \sum_{i=1}^{k} (2\gamma_i) + \sum_{i=1}^{k} \frac{(2\gamma_i)^2}{1 \cdot 2} - \sum_{i=1}^{k} \frac{(2\gamma_i)^3}{2 \cdot 3} + \cdots \right\}
\]

\[
= 1 - \frac{c}{2} \left\{ (2 - k) + \sum_{\alpha=2}^{\infty} \sum_{i=1}^{k} (-2\gamma_i)^\alpha/\alpha(2) \right\}.
\]

Here \(|2\gamma_i| \leq 1\) so all series converge absolutely and can be rearranged.

Also, \( \sum_{i=1}^{k} (2\gamma_i) = 2 - k \). For any integer \( r \leq n \) we now write

\[
H_r = 1 - \frac{c}{2} \left\{ (2 - k) + \sum_{\alpha=2}^{r} \sum_{i=1}^{k} (-2\gamma_i)^\alpha/\alpha(2) \right\}
\]

\[
= 1 - \frac{c}{2} \left\{ (2 - k) + \sum_{\alpha=2}^{r} \sum_{i=1}^{k} \frac{\alpha}{\alpha(2)} \sum_{\nu=0}^{\alpha} \frac{C\alpha}{\nu} (-2p_i)^\nu \right\}.
\]

The U.M.V. unbiased estimate of \( H_r \) is

\[
Z_r = 1 - \frac{c}{2} \left\{ (2 - k) + \sum_{\alpha=2}^{r} \sum_{i=1}^{k} \sum_{\nu=0}^{\alpha} \frac{\alpha}{\alpha(2)} \frac{1}{\alpha(2)} (-2)^\nu \frac{X_i^{(\nu)}}{n^{(\nu)}} \right\}.
\]
The bias of $Z_r$ as an estimate for $H$ is

$$B_r = E[Z_r] - H = H_r - H$$

$$= \frac{C}{2} \sum_{\alpha=r+1}^{\infty} \sum_{i=1}^{k} (-2\gamma_i)^{\alpha}/\alpha(2).$$

Now $u^s$ is convex for $s$ an even integer, and $-u^s$ is convex on $u \leq 0$ for $s$ an odd integer. From this it is easily shown that if $\sum_{i=1}^{k} u_i = 2k$ and $|u_i| \leq 1$, then

$$(k-1+(-1)^s)(1-\frac{2}{k})^s \leq (-1)^s \sum_{i=1}^{k} u_i^s \leq k-1+(-1)^s.$$

These lower, upper bounds are achieved by the choices $(u_1, \ldots, u_k) = (2/k-1, \ldots, 2/k-1)$ and $(1, -1, \ldots, -1)$ respectively. Applying this to the series for $B_r$ gives

$$\frac{C}{2} \sum_{\alpha=r+1}^{\infty} \frac{(k-1+(-1)\alpha)(1-2/k)^\alpha}{\alpha(2)} \leq B_r \leq \frac{C}{2} \left\{ \frac{k-1-(-1)^r}{r} - 2 \sum_{\alpha=r+1}^{\infty} \frac{(-1)^\alpha}{\alpha} \right\}.$$

This lower bound is achieved when the $p_i$'s are all $1/k$, and the upper bound is achieved when some $p_i = 1$. If $k > 2$ this lower bound is positive and we will use the estimate

$$Z'_r = Z_r - \frac{C}{2} \sum_{\alpha=r+1}^{\infty} \frac{(k-1+(-1)\alpha)(1-2/k)^\alpha}{\alpha(2)}$$

instead of $Z_r$ for $H$ because $Z'_r$ has the same variance as $Z_r$ and
uniformly smaller bias. For a fixed set \( \gamma_1, \ldots, \gamma_k \) we have

\[
B_r \leq \frac{C_k}{2^r} \left( \max_i \lvert 2\gamma_i \rvert \right)^{r+1}
\]

and the corresponding result for the bias of the improved estimate \( Z'_r \).

To compute the variance of \( Z_r \), we now use the fact

\[
\begin{align*}
(v_1)_{X_1} &= \sum_{t=0}^{v_1} \frac{v_1(t) v_2}{t!} (v_1 - v_2) (v_1 - t) \\
v_1(t)_t &= \sum_{t=0}^{\min(v_1, v_2)} \frac{v_1(t) v_2}{t!} \frac{v_1 + v_2 - t}{x_i} .
\end{align*}
\]

which gives

\[
\begin{align*}
\frac{(v_1)}{X_1} \frac{(v_2)}{X_1} &= \sum_{t=0}^{\min(v_1, v_2)} \frac{v_1(t) v_2}{t!} \frac{v_1 + v_2 - t}{x_i} .
\end{align*}
\]

Further routine calculations now give

\[
\begin{align*}
E \left( \frac{(v_1)}{X_1} - \frac{v_1}{p_1} \right) \left( \frac{(v_2)}{X_1} - \frac{v_2}{p_1} \right) &= \sum_{t=1}^{\min(v_1, v_2)} \frac{v_1(t) v_2}{t!} \frac{v_1 + v_2 - t}{n(t)} (1 - p_i)^t ,
\end{align*}
\]

\[
\begin{align*}
E \left( \frac{(v_1)}{X_1} - \frac{v_1}{p_1} \right) \left( \frac{(v_2)}{X_j} - \frac{v_2}{p_1} \right) &= \left( \frac{(v_1 + v_2)}{n(v_1 v_2)} - 1 \right) \frac{v_1}{p_1} \frac{v_2}{p_j} , \quad i \neq j .
\end{align*}
\]

From these, the variance of \( Z_r \) is seen to be
\[ \text{var } Z_r = E(Z_r - H_r)^2 \]

\[ = \frac{C^2}{n} E \left( \sum_{\alpha=2}^r \sum_{v=0}^{(\alpha-2)/2} \left( \frac{\alpha}{\alpha(2)} \sum_{i=1}^k \left( \frac{x_i^{(v)}}{n^{(v)}} - p_i^{(v)} \right) \right)^2 \right) \]

\[ = \frac{C^2}{n} \sum_{\alpha_1=2}^r \sum_{\alpha_2=2}^r \frac{\alpha_1 \alpha_2}{v_1=0} \frac{v_1^{v_1+v_2}}{v_2=0} \left( \frac{v_1^{v_1+v_2}}{v_2} \right) \left[ \sum_{i_1 \neq i_2=1}^k p_{i_1} p_{i_2} \frac{n \left( v_1 \right) \left( v_2 \right)}{n} - 1 \right] \]

\[ + \sum_{i=1}^k \frac{\min(v_1^{v_2})}{v_1^{v_2}} \frac{v_1^{v_2}}{v_2} \frac{p_1^{v_1+v_2-t}}{v_1^{v_2}} \frac{(1-p_1)^t}{n^{(v_2)}} \right) \}

Grouping together terms of like order in \( n \), the asymptotic variance as \( n \to \infty \) of the sequence \( \{Z_n\} \) of estimates is seen to be

\[ \text{var}(Z_n) \sim \frac{C^2}{n} \sum_{i=1}^k p_i \left( \log p_i - \sum_{j=1}^k p_j \log p_j \right)^2 \]

provided the non-zero \( p_i \)'s are not all equal; and

\[ \text{var}(Z_n) \sim \frac{C^2}{n} \cdot \frac{k^* - 1}{2(n-1)} \]

if the non-zero \( p_i \)'s are all equal, where \( k^* \) is the number of non-zero \( p_i \)'s. And for bias of this sequence of estimates we have asymptotically as \( n \to \infty \)

\[ \frac{kC}{2n(n+1)} \left( 1 - \frac{2}{k} \right)^{n+2} \leq B_n \leq \frac{kC}{2n(n+1)} \left( \max \left| 2\gamma \right| \right)^{n+1} \]
Comparison with asymptotic results obtained by Miller and Madow for the maximum likelihood estimates \( \{H'_n\} \) and the estimates
\( \{H''_n\} = \{H'_n + C(k-1)/2n\} \) shows the following. No asymptotic differences in variance. Asymptotic differences in expected squared error only in the special cases where this has order \( 1/n^2 \) or smaller. Bias is asymptotically much smaller for \( \{Z_n\} \) than for \( \{H'_n\} \) or \( \{H''_n\} \) except for the special case when some \( p_i = 1 \).

For small \( n \), numerical checking shows the following. \( Z_n \) has a smaller bias than \( H'_n \) or \( H''_n \) over most of the range of \( (p_1, \ldots, p_k) \). Comparing expected squared error as a whole, \( H'_n \) is quite poor and there is little to choose between \( Z_n \) and \( H''_n \) : sometimes one seems better, sometimes the other. In these comparisons \( Z_n \) and \( H''_n \) are modified by substituting \( C \log k \) for any value exceeding \( C \log k \), since this uniformly reduces expected squared error. For example in the following table for \( k = 2 \), every value exceeding \( C \log 2 = 1 \) would be replaced by 1. This table gives \( Z_n(x_1) \) in the upper part of each column and \( H''_n(x_1) \) in the lower part, for \( n = 2, 3, \ldots, 7 \) and all possible \( x_1 \). Values not tabulated are obtained from \( Z_n(n-x_1) = Z_n(x_1) \) and \( H''_n(n-x_1) = H''_n(x_1) \).
4. LOW BIAS ESTIMATION, k UNKNOWN. When k is unknown we shall consider the estimates obtained by acting as though k were equal to the observed number of different Y values and using the estimates of the preceding section. Now we have

\[ Z_r = 1 - \frac{C}{2} \left( W_1 + \cdots + W_k \right) \]

where

\[ W_i = \left( 2 \frac{X_1}{n} - 1 \right) + \sum_{\alpha=1}^{r} \sum_{v=0}^{\alpha} \binom{\alpha}{v} \frac{(-2)^v X_1^{(v)}}{\binom{\alpha}{2} \left( \frac{n}{v} \right)} . \]

The modification of \( Z_r \) just suggested for use in the case k unknown is

\[ Z_r^* = 1 - \frac{C}{2} \left( W_1^* + \cdots + W_k^* \right) \]
where

\[ W_i^* = W_i \quad \text{if} \quad X_i \neq 0 , \]
\[ = 0 \quad \text{if} \quad X_i = 0 . \]

Since \( W_i = -1/r \) when \( X_i = 0 \) we have

\[ P(W_i^* = W_i) = 1 - (1-p_i)^n , \]
\[ P(W_i^* = W_i + 1/r) = (1-p_i)^n . \]

\[ \therefore \quad E W_i^* = E W_i + \frac{1}{r} (1-p_i)^n , \]
\[ E Z_r^* = 1 - \frac{C}{2r} (E W_1 + \cdots + E W_k + \frac{1}{r} \sum_{i=1}^{k} (1-p_i)^n) \]
\[ = E Z_r - \frac{C}{2r} \sum_{i=1}^{k} (1-p_i)^n = H_r - \frac{C}{2r} \sum_{i=1}^{k} (1-p_i)^n . \]

The bias of \( Z_r^* \) as an estimate for \( H \) is

\[ E^*_r = E Z_r^* - H = H_r - H - \frac{C}{2r} \sum_{i=1}^{k} (1-p_i)^n \]
\[ = B_r - \frac{C}{2r} \sum_{i=1}^{k} (1-p_i)^n . \]

The variance of \( Z_r^* \) is found as follows:

\[ \text{var} \; Z_r^* = \frac{C^2}{4} \left( \sum_{i=1}^{k} \text{var} \; W_i^* + \sum_{i \neq j=1}^{k} \text{cov} \; W_i^* W_j^* \right) . \]
\[
\text{var } W_i^* = \text{E}(W_i^* - \text{E } W_i^*)^2 \\
= [1-(1-p_1)^n] \text{E}((W_i - \text{E } W_i) - \frac{1}{r} (1-p_1)^n)^2 \\
+ (1-p_1)^n \text{E}((W_i - \text{E } W_i) + \frac{1}{r} - \frac{1}{r} (1-p_1)^n)^2 \\
= \text{E}(W_i - \text{E } W_i)^2 + \frac{1}{r^2} \{(1-p_1)^n - (1-p_1)^{2n}\} \\
= \text{var } W_i + \frac{1}{r^2} \{(1-p_1)^n - (1-p_1)^{2n}\} .
\]

Furthermore, we have for \(i \neq j\)

\[
P(W_i^*, W_j^* = W_i, W_j) = 1-(1-p_1)^n - (1-p_j)^n + (1-p_1-p_j)^n ,
\]

\[
P(W_i^*, W_j^* = W_i + \frac{1}{r}, W_j) = (1-p_1)^n - (1-p_1-p_j)^n ,
\]

\[
P(W_i^*, W_j^* = W_i, W_j + \frac{1}{r}) = (1-p_j)^n - (1-p_1-p_j)^n ,
\]

\[
P(W_i^*, W_j^* = W_i + \frac{1}{r}, W_j + \frac{1}{r}) = (1-p_1-p_j)^n .
\]

So the covariance of \(W_i^*, W_j^*, i \neq j\), is

\[
\text{cov } W_i^*, W_j^* = \text{E}(W_i^* - \text{E } W_i^*)(W_j^* - \text{E } W_j^*) \\
= [1-(1-p_1)^n-(1-p_j)^n-(1-p_1-p_j)^n] \text{E}\{\left\{(W_i - \text{E } W_i) - \frac{(1-p_1)^n}{r}\right\}\left\{(W_j - \text{E } W_j) - \frac{(1-p_j)^n}{r}\right\}\} \\
+ [(1-p_1)^n-(1-p_1-p_j)^n] \text{E}\{\left\{(W_i - \text{E } W_i) + \frac{1-(1-p_1)^n}{r}\right\}\left\{(W_j - \text{E } W_j) - \frac{(1-p_j)^n}{r}\right\}\} \\
+ [(1-p_1)^n-(1-p_1-p_j)^n] \text{E}\{\left\{(W_i - \text{E } W_i) - \frac{(1-p_1)^n}{r}\right\}\left\{(W_j - \text{E } W_j) + \frac{1-(1-p_j)^n}{r}\right\}\} \\
+ [(1-p_j)^n-(1-p_1-p_j)^n] \text{E}\{\left\{(W_i - \text{E } W_i) + \frac{1-(1-p_1)^n}{r}\right\}\left\{(W_j - \text{E } W_j) + \frac{1-(1-p_j)^n}{r}\right\}\} .
\]
\[
+ (1-p_i-p_j)^n \mathbb{E}\left\{ (W_i - EW_i) + \frac{1-(1-p_i)^n}{r} \right\} \left\{ (W_j - EW_j) + \frac{1-(1-p_j)^n}{r} \right\} \\
= \text{cov} \ W_i, W_j + \frac{1}{r^2} \left\{ (1-p_i-p_j)^n - (1-p_i)^n(1-p_j)^n \right\}.
\]

\[\therefore \text{var} \ Z^*_r = \text{var} \ Z_r + \frac{c^2}{4r^2} \left\{ \sum_{i=1}^{k} [(1-p_i)^n - (1-p_i)^{2n}] \right\} + \sum_{i \neq j=1}^{k} [(1-p_i-p_j)^n - (1-p_i)^n(1-p_j)^n] \]

\[= \text{var} \ Z_r + \frac{c^2}{4r^2} \left\{ \sum_{i=1}^{k} (1-p_i)^n - [\sum_{i=1}^{k} (1-p_i)^n]^2 + \sum_{i \neq j=1}^{k} (1-p_i-p_j)^n \right\}.
\]

When only one value of \( Y \) is observed, which happens with probability \( \sum_{i=1}^{k} p_i^n \), the value of \( Z^*_r \) is \( (C/2)((-1)^r/r - 2 \sum_{\alpha=r}^{\infty} (-1)^{\alpha}/\alpha) \). Since \( u_m > 0 \), \( u_m \to 0 \), \( u_m > u_{m+1} \) and \( u_{m+1} - u_m > u_{m+2} - u_{m+1} \) together imply convergence of \( \sum_{m=1}^{\infty} (-1)^{m-1} u_m \) to a value \( > u_1/2 \), this value of \( Z^*_r \) has the sign of \( (-1)^{r-1} \). In the case \( r \) even, this negative value should be replaced by \( 0 \); bias and variance of the resulting modification of \( Z^*_r \) are easily found. This point does not arise in Section 3 because if \( k = 1 \) is known, \( H = 0 \) is known and estimation is not needed.

A similar discussion can be given for the estimates \( H^*_r = H_r' + C(k^*-1)/2r \). The maximum likelihood estimates \( H^*_r \) do not require knowledge of \( k \) so can be used unchanged in the case \( k \) unknown.
5. **MINIMAX ESTIMATION**, \( k - 2 \). Using squared error loss function, the risk function of the estimate \( Z = z(x_1, \ldots, x_n) \) for \( H \) is

\[
R_Z(p_1, \ldots, p_k) = E(Z - H)^2
\]

\[
= n! \Sigma [z(x_1, \ldots, x_k) - H]^2 \prod_{i=1}^{k} p_i^i / x_i!
\]

In searching for a minimax \( Z \) we need consider only those satisfying

\( 0 \leq z(x_1, \ldots, x_k) \leq C \log k \) since no other estimates are admissible.

Also, \( H(p_1, \ldots, p_k) \) being symmetric we need consider only symmetric

\( z(x_1, \ldots, x_k) \) since it is easily seen \([4]\) that no other estimates can

be minimax, or Bayes for an a priori distribution symmetric in \( p_1, \ldots, p_k \).

Thus in the case \( k = 2 \) where \( p_1, p_2 = p, q \) and \( x_1, x_2 = X, n-X \) with

\( X \) binomial \((n, p)\), we need consider only estimates \( z(X) \) satisfying

\( 0 \leq z(x) \leq 1 \) and \( z(x) = z(n-x) \).

When \( k=2 \) we can find minimax estimates for \( H \) by the usual method

of guessing a least favorable a priori distribution \( \lambda \) for \( p \) and finding

the corresponding Bayes estimate \( Z_\lambda \) for \( H \). If \( R_{Z_\lambda}(p) \) assumes its

maximum value with \( \lambda \)-probability 1, then \( \lambda \) is least favorable and \( Z_\lambda \)

is minimax (also admissible if \( Z_\lambda \) was unique Bayes).

**Case** \( n = 1 \). If \( z(0) = z(1) = a \), \( R_Z(p) = (a-H)^2 \). Minimax \( Z \)

is clearly given by the choice \( a = 1/2 \) and its maximum risk is \( 1/4 = .25 \).

**Case** \( n = 2 \). If \( z(0) = z(2) = a \), \( z(1) = b \) we have

\[
R_Z(p) = (1-2pq)(a-H)^2 + 2pq(b-H)^2
\]
If for a priori distribution \( \lambda \) we take

\[
P(p = 0) = P(p = 1) = \alpha/2, \quad P(p = 1/2) = 1 - \alpha
\]

then the expected risk is

\[
E_{\lambda} R_z(p) = \alpha a^2 + \frac{1 - \alpha}{2} [(a-1)^2 + (b-1)^2]
\]

which is minimized by the choice \( Z_\lambda \):

\[
a_\lambda = \frac{1 - \alpha}{1 + \alpha}, \quad b_\lambda = 1.
\]

Now choose \( \alpha \) so as to make

\[
[R_{Z_{\lambda}}(0) = R_{Z_{\lambda}}(1)] = R_{Z_{\lambda}}(1/2)
\]

which gives \( \alpha^* = \sqrt{2} - 1 \), \( a^*_{\lambda^*} = \sqrt{2} - 1 \).

The estimate \( Z_{\lambda^*} \): \( z(0) = z(2) = \sqrt{2} - 1 \), \( z(1) = 1 \), Bayes for \( \lambda^* \), has the risk \( a^2_{\lambda^*} = 3 - 2\sqrt{2} = .1716 \) with \( \lambda^* \)-probability 1. This estimate is now seen to be minimax by checking that \( R_{Z_{\lambda^*}}(p) \leq 3 - 2\sqrt{2} \) for all \( p \).

Case \( n = 3 \). If \( z(0) = z(3) = a \), \( z(1) = z(2) = b \), we have

\[
R_z(p) = (1 - 3pq)(a - H)^2 + 3pq(b - H)^2.
\]
If for a priori distribution \( \lambda \) we take

\[
P(p = 0) = P(p = 1) = \frac{\alpha}{2}, \quad P(p = p_0) = P(p = 1 - p_0) = \frac{1 - \alpha}{2}
\]

then the expected risk is

\[
E_{\lambda} R_z(p) = \alpha R_z(0) + (1 - \alpha)R_z(p_0)
\]

\[
= \alpha a^2 + (1 - \alpha)\{(1 - 3 p_0 q_0) [a - H(p_0)]^2 + 3 p_0 q_0 [b - H(p_0)]^2\}
\]

which is minimized by the choice

\[
a_{\lambda} = \frac{H(p_0)}{1 + \alpha/(1 - \alpha) \sqrt{1 - 3 p_0 q_0}}, \quad b_{\lambda} = H(p_0)
\]

For this estimate to be minimax we must have

\[
R_z(0) = R_z(p_0)
\]

\[
R_z'(p_0) = 0.
\]

This pair of equations is easily solved numerically, giving \( p_0 = .361596 \), \( \alpha = .356705 \). Corresponding to this \( \lambda \) we have \( a_{\lambda} = .33673 \), \( b_{\lambda} = .94400 \). This estimate is then proved minimax by checking that for it we have

\[
R_{z,\lambda}^2(0) \leq a_{\lambda}^2 = .11339.
\]

**General** \( n \). In general we would try for \( \lambda \) a symmetric distribution about \( p = 1/2 \) on \( n + 1 \) points of which one is 0, with \( P(p = p_1) = \alpha_1 \). We can write down the corresponding Bayes estimate for \( H \). If this is to
he minimax we must have

\[ R(0) = R(p_1) = \cdots = R(p_{n/2}) \]

\[ R'(p_1) = \cdots = R'(p_{n/2}) = 0 \]

We would have to solve these \(2(n/2)\) equations numerically for the \(2(n/2)\) unknowns \(p_1, \alpha_i\) and then check that the resulting estimate has a risk function which never rises above its value at \(p = 0\).

The same method can be used to find minimax estimates of \(pq\), from which we can also write down minimax estimates of any linear function of \(pq\) such as the chi square dispersion measure \(E_2\). We can restrict attention to estimates \(z(x)\) for \(pq\) satisfying \(0 \leq z(x) \leq 1/4\) and \(z(x) = z(n-x)\). The following table gives minimax estimates of \(pq\) for \(n = 1, 2, 3, 4, 5\). For \(n = 3, 4, 5\) the minimax estimate has constant risk. If this is true for larger \(n\), we could find minimax estimates by finding what estimate has constant risk and then finding an a priori distribution (symmetric and discrete) with respect to which this estimate is Bayes.

<table>
<thead>
<tr>
<th>(n\times)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Max. Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.125000</td>
<td>1/8</td>
<td></td>
<td>.015625</td>
</tr>
<tr>
<td>2</td>
<td>.103553</td>
<td>.250000</td>
<td>((\sqrt{2}-1)/4)</td>
<td>.010724</td>
</tr>
<tr>
<td>3</td>
<td>.083333</td>
<td>.250000</td>
<td>1/4</td>
<td>.006944</td>
</tr>
<tr>
<td>4</td>
<td>.071584</td>
<td>.202278</td>
<td>.245843</td>
<td>.005124</td>
</tr>
<tr>
<td>5</td>
<td>.065076</td>
<td>.177968</td>
<td>.228414</td>
<td>.004235</td>
</tr>
</tbody>
</table>
REFERENCES


<table>
<thead>
<tr>
<th>Name and Address</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chief, Bureau of Ships, Dept. of the Navy, Washington 25, D.C., Attn: H. Weingarten</td>
<td>1</td>
</tr>
<tr>
<td>U. S. Naval Avionics Facility, Indianapolis 18, Indiana, Attn: Library</td>
<td>2</td>
</tr>
<tr>
<td>Commanding Officer, U. S. Naval Powder Factory, Indianhead, Md., Attn: F. Freshman</td>
<td>2</td>
</tr>
<tr>
<td>Scientific Section, Office of Naval Research, Dept. of the Navy, San Francisco</td>
<td>3</td>
</tr>
<tr>
<td>Director, Naval Research Lab., Washington 25, D.C., Attn: Technical Information</td>
<td>4</td>
</tr>
<tr>
<td>Office of Naval Research, Dept. of the Navy, Washington 25, D.C., Attn:</td>
<td>5</td>
</tr>
<tr>
<td>Dept. of Mathematical Statistics, University of North Carolina, Chapel Hill, N. C.</td>
<td>6</td>
</tr>
<tr>
<td>Office of the Asst. Naval Attache for Research, Naval Attache, American</td>
<td>7</td>
</tr>
<tr>
<td>Office of Technical Services, Dept. of Commerce, Washington 25, D.C.</td>
<td>8</td>
</tr>
</tbody>
</table>
Dr. Julius R. Blum
Dept. of Mathematics
Indiana University
Bloomington, Indiana

Dr. Paul Blunk
Box 532
Fair Oaks, Calif.

Dr. Charles Bell
General Electric Co.
735 State St.
Santa Barbara, Calif.

Professor W. G. Cochran
Biostatistics Dept.
The Johns Hopkins University
Baltimore 5, Md.

Professor Lee Cronbach
Bureau of Education Research
1007 S. Wright
Champaign, Illinois

Dr. Joseph Daly
Bureau of the Census
Washington 25, D.C.

Dr. Francis Dresch
Stanford Research Institute
1915 University Ave.
Palo Alto, Calif.

Professor Meyer Dwass
Dept. of Mathematics
Northwestern University
Evanston, Ill.

Mr. Murray A. Geisler
Logistics Section
The RAND Corporation
1700 Main St.
Santa Monica, Calif.

Mr. Geoffrey Gregory
4 Osborne Grove
Gatley, Cheadle
Cheshire, England
c/o Office of Naval Research
London Branch
Navy No. 100, Fleet Post Office
New York, N. Y.

Professor E. J. Gumbel
Industrial Engr. Dept.
409 Engineering Bldg.
Columbia University
New York 27, N. Y.

Dr. Theodore E. Harris
The RAND Corporation
1700 Main St.
Santa Monica, Calif.

Professor Leonid Hurwicz
School of Business Administration
University of Minnesota
Minneapolis 14, Minn.

Professor Stanley Isaacson
4715 Pleasant Street
Des Moines, Iowa

Professor Leo Katz
Statistics Dept.
Michigan State University
East Lansing, Mich.

Professor Tosio Kitagawa
Mathematical Institute
Faculty of Science
Kyusyu University
Fukuoka, Japan

Dr. Dennis V. Lindley
Statistical Lab.
University of Cambridge
Cambridge, England
c/o Office of Naval Research
London Branch
Navy No. 100, Fleet Post Office
New York, N. Y.

Dr. Eugene Lukaas
Dept. of Mathematics
Catholic University
Washington 17, D. C.

Mr. Monroe Norden
Research Division
Engineering Statistics Group
College of Engineering
New York University
401 W. 205 St.

New York 54, N. Y.
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution/Department</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor Stanley Reiter</td>
<td>Dept. of Economics, Purdue University, Lafayette, Indiana</td>
<td></td>
</tr>
<tr>
<td>Dr. A. R. Roy</td>
<td>Statistical Wing, Indian Council of Agricultural Research, Linlithgow Ave., New Delhi, India</td>
<td></td>
</tr>
<tr>
<td>Professor Herman Rubin</td>
<td>Dept. of Mathematics, University of Oregon, Eugene, Oregon</td>
<td></td>
</tr>
<tr>
<td>Professor Seymour Sherman</td>
<td>Moore Sch. of Electrical Engineering, University of Pennsylvania, Philadelphia 4, Pa.</td>
<td></td>
</tr>
<tr>
<td>Dr. Milton Sobel</td>
<td>Bell Telephone Labs., 555 Union Blvd., Allentown, Pa.</td>
<td></td>
</tr>
<tr>
<td>Professor Herbert Solomon</td>
<td>Teachers' College, Columbia University, New York 27, N. Y.</td>
<td></td>
</tr>
<tr>
<td>Professor Donald Truax</td>
<td>Mathematics Dept., University of Kansas, Lawrence, Kansas</td>
<td></td>
</tr>
<tr>
<td>Mr. Cesareo Villegas</td>
<td>Institute de Mathematica y Estadistica, Av. J. Herrera y Reissig 565, Montevideo, Uruguay</td>
<td></td>
</tr>
<tr>
<td>Professor W. Allen Wallis</td>
<td>Committee on Statistics, University of Chicago, Chicago 37, Ill.</td>
<td></td>
</tr>
<tr>
<td>Professor Irving Weiss</td>
<td>Bell Telephone Labs., 1600 Osgood St., North Andover, Mass.</td>
<td></td>
</tr>
<tr>
<td>Dr. Oscar Wesler</td>
<td>Dept. of Mathematics, University of Michigan, Ann Arbor, Mich.</td>
<td></td>
</tr>
<tr>
<td>Dr. John D. Wilkes</td>
<td>Office of Naval Research, Code 200, Washington 25, D.C.</td>
<td></td>
</tr>
<tr>
<td>Professor S. S. Wilks</td>
<td>Room 110, Fine Hall, Box 708, Princeton, N. J.</td>
<td></td>
</tr>
<tr>
<td>Professor J. Wolfowitz</td>
<td>Mathematics Dept., Cornell University, Ithaca, N. Y.</td>
<td></td>
</tr>
<tr>
<td>Professor M. A. Woodbury</td>
<td>Dept. of Mathematics, New York University, New York 53, N. Y.</td>
<td></td>
</tr>
</tbody>
</table>

Additional copies for project leader and assistants and reserve for future requirements.