NONPARAMETRIC ESTIMATION OF THE MEAN USING QUANTAL RESPONSE DATA

J. D. CHURCH AND E. BENTON COBB

(Received Feb. 9, 1970)

1. Introduction

Let Z be a random variable which takes values in [0, 1] with probability one and let $H(z)$ be the distribution function for Z. Suppose an experimenter can fix a value (level) z in [0, 1] and observe whether or not Z exceeds z, but cannot observe Z directly. For example, if Z is the minimum amount of a poison required to kill an animal, the experimenter may give the animal an amount z of poison and observe death, $Z \leqq z$, or survival, $Z > z$, where $H(z)$ is the probability of death. In this paper we discuss the use of such quantal response data; observations on Bernoulli random variables at levels $0 < z_1 \leqq z_2 \leqq \cdots \leqq z_n < 1$, to estimate $\mu = E(Z)$, or, equivalently, to estimate $\theta = 1 - \mu = \int_0^1 H(z)dz$.

The estimate considered here is $\hat{\theta} = \int_0^1 \hat{H}(z)dz$ where $\hat{H}(0) = 0$, $\hat{H}(1) = 1$, $\hat{H}$ is linear between the levels z_i, and the vector $(\hat{H}(z_1), \hat{H}(z_2), \cdots, \hat{H}(z_n))$ is obtained from the maximum likelihood estimate given by Ayers, Brunk, et al. [1], which we call the ABERS estimate.

The estimate $\hat{\theta}$ for $\theta = 1 - \mu$ is analogous to the estimate $1 - \bar{z} = \hat{\hat{\theta}} = \int_0^1 \hat{\hat{H}}(z)dz$, based on continuous data; a random sample $Z_1, Z_2, \cdots, Z_n$, where the Z_i's are observed directly, where $\hat{\hat{H}}$ is the empirical distribution function, the maximum likelihood estimate for H.

In Section 2, the ABERS estimate and an alternate way to compute it are discussed. In Section 3, properties of $\hat{\theta}$ are investigated for the particular choice of levels, $z_1 = 1/(n+1)$, $z_2 = 2/(n+1)$, $\cdots$, $z_n = n/(n+1)$. Other choices of $z_1, z_2, \cdots, z_n$ are discussed in Section 4. In Section 5, we give an application in which quantal response data is used to estimate $P(X \leqq Y)$ where X and Y are independent random variables, one with known distribution and the other unknown.

The results of this paper are clearly applicable in case the random variable Z takes values in a compact interval $[a, b]$ instead of [0, 1]. In this case let $Z'=(Z-a)/(b-a)$; then $u_{Z'}=(u_Z-a)/(b-a)$ and Z' takes values in $[0, 1]$. The levels are transformed in the same way.

In case Z takes values on an unbounded interval, the procedures given in this paper can be used to estimate the mean of a distribution H^* derived from the distribution H of the random variable Z as follows: $H^*(z)=H(z)$, $a<z<b$, $H^*(a-)=0$, and $H^*(b)=1$. That is, H is modified by placing all mass to the left of a at a and all mass to the right of b at b.

If the unbounded interval is semi-infinite, then the mean of H^* gives a bound on the mean of H (a lower bound when the interval is $[a, \infty)$ and an upper bound when the interval is $(-\infty, b]$).

However, when considering all distributions on the real line, the mean of H^* cannot be used to approximate the mean of H since given any value for the mean of H^*, the mean of H can be any real number or may not exist. Thus, additional restrictions on H are required in order that the mean of H^* be a useful approximation of the mean of H.

2. The ABERS estimate

Let $H(z)$ be an unknown distribution function on $[0, 1]$ and let $0<w_1<w_2<\cdots<w_r<1$ be r levels of experimentation. At each level w_j, $1\leqq j\leqq r$, b_j independent Bernoulli trials are performed, each with probability $H(w_j)$ of successes, where $b_j\geqq 1$. (The w_j's are related to the z_i's of Section 1 by $z_1=z_2=\cdots=z_{b_1}=w_1$, $z_{b_1+1}=\cdots=z_{b_1+b_2}=w_2$, etc.) Let a_j be the number of successes in the b_j trials at level w_j. Then the maximum likelihood estimate $(\hat{H}(w_1), \hat{H}(w_2),\cdots, \hat{H}(w_r))$ for $(H(w_1), H(w_2), \ldots, H(w_r))$ is obtained as follows [1]. If

$$\frac{a_1}{b_1}\leqq\frac{a_2}{b_2}\leqq\cdots\leqq\frac{a_r}{b_r} \tag{1}$$

then $\hat{H}(w_j)=a_j/b_j$, $1\leqq j\leqq r$. If (1) does not hold, then $a_k/b_k>a_{k+1}/b_{k+1}$ for some k, $1\leqq k<r$. The kth and $(k+1)$st experiments are then pooled; we replace the pairs (a_k, b_k) and (a_{k+1}, b_{k+1}) by the single pair $(a_k+a_{k+1}, b_k+b_{k+1})$ and replace ratios a_k/b_k and a_{k+1}/b_{k+1} by the ratio $(a_k+a_{k+1})/(b_k+b_{k+1})$. Next we check the condition

$$\frac{a_1}{b_1}\leqq\frac{a_2}{b_2}\leqq\cdots\leqq\frac{a_{k-1}}{b_{k-1}}\leqq\frac{a_k+a_{k+1}}{b_k+b_{k+1}}\leqq\frac{a_{k+2}}{b_{k+2}}\leqq\cdots\leqq\frac{a_r}{b_r}\,.$$

If this condition is satisfied the pooling is complete; if not, pooling is continued in the same way until a sequence of non-decreasing ratios

$$\frac{A_1}{B_1}\leqq\frac{A_2}{B_2}\leqq\cdots\leqq\frac{A_s}{B_s}\,,$$

where $s \leqq r$, is obtained. Then $\hat{H}(w_j)=A_m/B_m$ if the level w_j experiment is in the pool associated with the mth ratio. It is shown in [1] that the order of pooling is immaterial.

Below we give another procedure for computing the ABERS estimate which we find more adaptable to a computer program. Let a_j and b_j be as previously defined and let

$$N_k^{(1)}=\sum_{j=1}^{k} a_j\ , \qquad D_k^{(1)}=\sum_{j=1}^{k} b_j\ , \qquad \text{for } 1 \leqq k \leqq r\ .$$

Then define $c_1=\min_{1\leqq k\leqq r}\{N_k^{(1)}/D_k^{(1)}\}$, and let $k_0=0$ and $k_1=\max\{k : N_k^{(1)}/D_k^{(1)}=c_1\}$. Set $\hat{H}(w_1)=\hat{H}(w_2)=\cdots=\hat{H}(w_{k_1})=c_1$. If $k_1<r$ repeat the procedure starting with the (k_1+1)st level; define $N_k^{(2)}=\sum_{j=k_1+1}^{k} a_j$, and $D_k^{(2)}=\sum_{j=k_1+1}^{k} b_j$, for $k_1+1\leqq k\leqq r$, let $c_2=\min_{k_1<k\leqq r}\{N_k^{(2)}/D_k^{(2)}\}$, let $k_2=\max\{k : k>k_1 \text{ and } N_k^{(2)}/D_k^{(2)}=c_2\}$, and set $\hat{H}(w_{k_1+1})=\cdots=\hat{H}(w_{k_2})=c_2$. This procedure is repeated until $\hat{H}(w_r)$ is computed. Te see that $c_2>c_1$ (when $k_1<r$), observe that for $e\geqq 0$, and positive f, g, h, $(e+g)/(f+h)>e/f$ implies $e/f<g/h$ and we have

$$\frac{N_{k_1}^{(1)}}{D_{k_1}^{(1)}}<\frac{N_{k_2}^{(1)}}{D_{k_2}^{(1)}}=\frac{N_{k_1}^{(1)}+N_{k_2}^{(2)}}{D_{k_1}^{(1)}+D_{k_2}^{(2)}}$$

which implies

$$c_1=\frac{N_{k_1}^{(1)}}{D_{k_1}^{(1)}}<\frac{N_{k_2}^{(2)}}{D_{k_2}^{(2)}}=c_2\ .$$

This can be extended to show that for the $\hat{H}(w_j)$ obtained by the above procedure, we have $\hat{H}(w_1)\leqq\hat{H}(w_2)\leqq\cdots\leqq\hat{H}(w_r)$.

We call this procedure clumping and refer to the pools so obtained as clumps (clumps of levels, not observations). We now show that clumping yields the same estimates as the ABERS method.

First, consider restricting the ABERS pooling to the tth clump of levels $\{w_{k_{t-1}+1}, w_{k_{t-1}+2}, \cdots, w_{k_t}\}$. Note that, if the tth clump is subdivided into right and left subclumps of levels, the ratio of successes to trials in the right subclump is less than or equal that ratio for the left subclump, since for $1\leqq l<k_t-k_{t-1}$,

$$\frac{N_{k_t}^{(t)}}{D_{k_t}^{(t)}}=\frac{N_{k_{t-1}+l}^{(t)}+(N_{k_t}^{(t)}-N_{k_{t-1}+l}^{(t)})}{D_{k_{t-1}+l}^{(t)}+(D_{k_t}^{(t)}-D_{k_{t-1}+l}^{(t)})}\leqq\frac{N_{k_{t-1}+l}^{(t)}}{D_{k_{t-1}+l}^{(t)}}$$

implies

$$\frac{N_{k_{t-1}+l}^{(t)}}{D_{k_{t-1}+l}^{(t)}}\geqq\frac{N_{k_t}^{(t)}-N_{k_{t-1}+l}^{(t)}}{D_{k_t}^{(t)}-D_{k_{t-1}+l}^{(t)}}$$

where l is the number of levels in the left subclump and the ratios associated with the right and left subclumps are given by the corresponding sides of the last inequality. This can be generalized to m subclumps of the tth clump, to obtain a non-increasing sequence of ratios. Now if the ABERS pooling is completed within clump t, we have, in general, subclumps of clump t, and the ratios associated with the subclumps must be constant and equal to c_t since the ABERS procedure gives a non-decreasing sequence of ratios and we have shown they must also be non-increasing.

Now, if we ABERS procedure is applied to all the data, but the order of pooling is such that pooling is done first inside clumps, a non-decreasing sequence of ratios is obtained just from pooling inside clumps, the ABERS procedure is completed, and the resulting estimate $(\hat{H}(w_1), \cdots, \hat{H}(w_r))$ is the same as that for clumping.

3. Estimation of the mean of $H(z)$

For levels $0<z_1\leqq z_2\leqq\cdots\leqq z_n<1$, let $\hat{H}(z_i)=\hat{H}(w_j)$, the ABERS estimate for $H(w_j)$, when $z_i=w_j$. Define $\hat{H}(0)=0$, $\hat{H}(1)=1$, and for $z_i<z<z_{i+1}$, let

$$\hat{H}(z)=\hat{H}(z_i)+\left(\frac{z-z_i}{z_{i+1}-z_i}\right)[\hat{H}(z_{i+1})-\hat{H}(z_i)]$$

where $z_0=0$ and $z_{n+1}=1$. Since

$$\mu=\int_0^1 z dH(z)=1-\int_0^1 H(z)dz ,$$

estimation of μ is equivalent to estimation of $\theta=\int_0^1 H(z)dz$, and a natural estimate for θ is $\hat{\theta}=\int_0^1 \hat{H}(z)dz$.

We shall consider this estimate in detail for the case $z_i=i/(n+1)$, $1\leqq i\leqq n$, (equally spaced levels) and one trial at each level. Properties of the estimate for various experimental designs will be considered in Section 4.

For the case at hand, we have $b_i=1$ and $a_i=0$ or 1 for $1\leqq i\leqq n$ (using the notation of Section 2). Suppose clumping yields ratios $A_1/B_1<A_2/B_2<\cdots<A_s/B_s$, where

$$A_i=\sum_{j=k_{i-1}+1}^{k_i} a_j \quad\text{and}\quad B_i=\sum_{j=k_{i-1}+1}^{k_i} b_j=(k_i-k_{i-1}) \qquad \text{for } 1\leqq i\leqq s .$$

Then $\hat{H}(z_j)=A_i/B_i$ if $k_{i-1}<j\leqq k_i$ and

$$\int_0^1 \hat{H}(z)dz = \frac{1}{2(n+1)} \sum_{j=0}^{n} [\hat{H}(z_{j+1}) + \hat{H}(z_j)]$$
$$= \frac{1}{n+1} \sum_{j=1}^{n} \hat{H}(z_j) + \frac{1}{2(n+1)}$$
$$= \frac{1}{n+1} \left[\sum_{i=1}^{s} A_i + \frac{1}{2} \right] .$$

Since $\sum_{i=1}^{s} A_i$ is the total number of successes in the n trials, if we let $x = \sum_{i=1}^{s} A_i = \sum_{j=1}^{n} a_j$, we have

$$\hat{\theta} = \int_0^1 \hat{H}(z)dz = \frac{x+1/2}{n+1} .$$

Thus for equally spaced levels, the estimate $\hat{\theta}$ may be computed without computing the ABERS estimate.

It should be noted that the formula $(x+1/2)/(n+1)$ does not characterize $\hat{\theta}$, since the distribution of $\hat{\theta}$ for given H depends on the design $\tilde{z} = (z_1, z_2, \cdots, z_n)$. As is pointed out in the next section, there are other designs for which $\hat{\theta}$ is of the form $(x+1/2)/(n+1)$. We assume here that the design is fixed at $\tilde{z} = (1/(n+1), 2/(n+1), \cdots, n/(n+1))$ and hence the distribution of $\hat{\theta} = (x+1/2)/(n+1)$ is determined for each H.

Since x is the sum of independent Bernoulli random variables, we have

$$E(\hat{\theta}) = \frac{\sum_{i=1}^{n} \theta_i + \frac{1}{2}}{n+1} \qquad \text{and} \qquad \sigma_{\hat{\theta}}^2 = \sum_{i=1}^{n} \frac{\theta_i(1-\theta_i)}{(n+1)^2}$$

where $\theta_i = H(z_i)$, $0 \leqq i \leqq n+1$. It is also easy to obtain a bound on the bias of $\hat{\theta}$. Since $\theta_i \leqq H(z) \leqq \theta_{i+1}$ for $z_i \leqq z \leqq z_{i+1}$, $0 \leqq i \leqq n$, we have

$$\frac{\sum_{i=1}^{n} \theta_i}{n+1} \leqq \theta = \int_0^1 H(z)dz \leqq \frac{\sum_{i=1}^{n} \theta_i + 1}{n+1} .$$

Thus

$$|\text{bias}| = |E(\hat{\theta}) - \theta| = \left| \frac{\sum_{i=1}^{n} \theta_i + \frac{1}{2}}{n+1} - \theta \right| \leqq \frac{1}{2(n+1)} .$$

This bound is seen to be sharp by considering H degenerate at $1/(n+1)$.

The mean square error (MSE) is given by

$$\text{MSE} = E(\theta - \hat{\theta})^2 = \sigma_{\hat{\theta}}^2 + (\text{bias})^2$$

$$=\frac{1}{(n+1)^2}\sum_{i=1}^{n}\theta_i(1-\theta_i)+\left(\frac{\sum_{i=1}^{n}\theta_i+\frac{1}{2}}{n+1}-\theta\right)^2.$$

If we let $\bar{\theta}=\frac{1}{n}\sum_{i=1}^{n}\theta_i$, then

$$\text{MSE}\leqq\frac{n\bar{\theta}(1-\bar{\theta})}{(n+1)^2}+\frac{1}{4(n+1)^2}$$

since the maximum of $\sum_{i=1}^{n}\theta_i(1-\theta_i)$ subject to the side condition $\sum_{i=1}^{n}\theta_i=n\theta$ is given by $n\bar{\theta}(1-\bar{\theta})$ and since $(\text{bias})^2\leqq[1/2(n+1)]^2$. Further, since $\bar{\theta}(1-\bar{\theta})\leqq 1/4$, we have the bound

$$\text{MSE}\leqq\frac{n/4}{(n+1)^2}+\frac{1}{4(n+1)^2}=\frac{1}{4(n+1)}$$

independent of any of the θ_i or θ. To see that this bound cannot be improved, consider, for $0<\epsilon<1/(n+1)$, H_ϵ, where

$$H_\epsilon(z)=\begin{cases}0, & z<0\\ \frac{1}{2}, & 0\leqq z<\frac{n}{n+1}+\epsilon\\ 1, & z\geqq\frac{n}{n+1}+\epsilon\end{cases}$$

and verify, for each such ϵ, that $\text{MSE}>(1/4(n+1))-\epsilon$.

Next, consider the distribution of $\hat{\theta}$ for large n. We will show that, if $\hat{\theta}$ is not degenerate, then $\sqrt{n}(\hat{\theta}-\theta)$ has an approximate normal distribution with mean zero and variance given by $\int_0^1 H(z)(1-H(z))dz$. By Lindeberg's Theorem (see Feller [2]), if $x=\sum_{i=1}^{n}x_i$ is a sum of independent Bernoulli random variables where x_i has parameter θ_i, and if $\sum_{i=1}^{n}\sigma_{x_i}^2=\sum_{i=1}^{n}\theta_i(1-\theta_i)\to\infty$ as $n\to\infty$, then the distribution of $\sum_{i=1}^{n}(x_i-\theta_i)\Big/\left[\sum_{i=1}^{n}\theta_i(1-\theta_i)\right]^{1/2}$ tends to a standard normal distribution. Clearly, if H is not degenerate, then

$$\sum_{i=1}^{n}\theta_i(1-\theta_i)=\sum_{i=1}^{n}H\left(\frac{i}{n+1}\right)\left[1-H\left(\frac{i}{n+1}\right)\right]\to\infty$$

as $n\to\infty$. Further,

$$\sqrt{n}(\hat{\theta}-\theta)=\sqrt{n}\left(\frac{x-\sum_{i=1}^{n}\theta_i}{n+1}\right)+\sqrt{n}\left(\frac{\sum_{i=1}^{n}\theta_i+\frac{1}{2}}{n+1}-\theta\right)$$

$$=\sqrt{n}\left(\frac{x-\sum_{i=1}^{n}\theta_i}{n+1}\right)+\sqrt{n}(\text{bias}) .$$

But $\sqrt{n}\,|\text{bias}|\leqq(\sqrt{n}/2(n+1))\rightarrow 0$ as $n\rightarrow\infty$, and hence $\sqrt{n}(\hat{\theta}-\theta)$ behaves asymptotically as $\sqrt{n}\left(\left(x-\sum_{i=1}^{n}\theta_i\right)\Big/(n+1)\right)$ which is approximately normal with mean zero and variance $\frac{1}{n}\sum_{i=1}^{n}\theta_i(1-\theta_i)$ for large n. Recalling that $\theta_i=H(i/(n+1))$, $1\leqq i\leqq n$, it is clear that $\frac{1}{n}\sum_{i=1}^{n}\theta_i(1-\theta_i)\rightarrow\int_0^1 H(z)\cdot(1-H(z))dz$ as $n\rightarrow\infty$, and the asserted asymptotic normality of $\hat{\theta}$ is established.

The admissibility of $\hat{\theta}=(x+1/2)/(n+1)$ together with the design $\tilde{z}=(1/(n+1), 2/(n+1),\cdots, n/(n+1))$ is established in the next section. We show here that any design for which $\hat{\theta}$ is of the form $\hat{\theta}=(x+1/2)/(n+1)$ is Bayes against a prior distribution on the class of distribution functions on $[0, 1]$. Let $\mathscr{H}_c$ denote the class of distribution functions which are constant on $(0, 1)$, and let $H_p\in\mathscr{H}_c$ have value p on $(0, 1)$. Then $\theta_p=\int_0^1 H_p(z)dz=p$, and given that $H\in\mathscr{H}_c$, the estimation problem reduces to estimating θ when x is binomial with parameters n and θ. It is well known that $\phi(x)=(x+\alpha)/(n+\alpha+\beta)$ is the essentially unique Bayes rule against a beta (α, β) prior for squared error loss. (Ferguson [3]). Then $\hat{\theta}=(x+1/2)/(n+1)$ is Bayes against a beta prior (with $\alpha=\beta=1/2$) on the parameter space $\Theta=\{\theta: 0<\theta<1\}$, which is equivalent to a prior over $\mathscr{H}_c$.

4. The design problem

In this section we consider the choice of an experimental design, (choice of $\tilde{z}=(z_1, z_2,\cdots, z_n)$) given that the number of observations, n, is fixed, and given that the estimator will be $\hat{\theta}=\int_0^1\hat{H}(z)dz$.

We first consider some components of the mean squared error (MSE) for a given design, $\tilde{z}$, and any distribution function H. The estimate $\hat{\theta}=\frac{1}{2}\sum_{i=1}^{n}(z_{i+1}-z_{i-1})\hat{H}(z_i)+\frac{1}{2}(1-z_n)$ may be considered an estimate for the "true" value of the quadrature

$$Q(\tilde{z}, H)=\frac{1}{2}\sum_{i=1}^{n}(z_{i+1}-z_{i-1})H(z_i)+\frac{1}{2}(1-z_n) .$$

The quadrature error is $QE=Q(\tilde{z}, H)-\int_0^1 H(z)dz=Q(\tilde{z}, H)-\theta$. The bias is $E(\hat{\theta}-\theta)=QE+QB$ where $QB=E(\hat{\theta})-Q(\tilde{z}, H)$, the bias of $\hat{\theta}$ as an estimator of the quadrature. Then $\text{MSE}=\sigma_{\hat{\theta}}^2+(QB+QE)^2$. If $\hat{\theta}$ is unbiased for $Q(\tilde{z}, H)$, then $QB=0$, and bias$=QE$.

Given $H(z_i)$, $0\leqq i\leqq n$, a bound on the quadrature error is $1/2\cdot \sum_{i=0}^{n}(z_{i+1}-z_i)[H(z_{i+1})-H(z_i)]$. If the $H(z_i)$ are not given, then the maximum absolute quadrature error occurs for H_0 which is degenerate at one end of the interval $[z_j, z_{j+1}]$, where $z_{j+1}-z_j=\max_{0\leqq i\leqq n}(z_{i+1}-z_i)$, and for H_0 the absolute quadrature error is $(z_{j+1}-z_j)/2$. In this case $\hat{\theta}=Q(\tilde{z}, H_0)$ with probability one and the MSE at H_0 is given by $[(z_{j+1}-z_j)/2]^2$. Then maximum MSE is small only for designs for which $\max_{0\leqq i\leqq n}(z_{i+1}-z_i)$ is small. In particular, if $\max_{0\leqq i\leqq n}(z_{i+1}-z_i)>1/\sqrt{n+1}$, then the maximum MSE$>$ $1/4(n+1)$ which is the maximum MSE for the design $\tilde{z}=(1/(n+1), 2/(n+1),\cdots, n/(n+1))$.

The design of the last section, with one observation at each of n equally spaced levels (hereafter, the eq. sp. design) seems a natural choice in ignorance of H. As was shown, with eq. sp., for squared error loss, $\hat{\theta}$ is Bayes against a prior on the set of distribution functions. With eq. sp., the estimate $\hat{\theta}$ is easily computed; the clumping procedure can be avoided. Also, eq. sp. minimizes the maximum absolute bias, since for eq. sp. $\hat{\theta}$ is unbiased for $Q(\tilde{z}, H)$, clearly eq. sp. uniquely minimizes the maximum absolute quadrature error, and the maximum absolute quadrature error may always be achieved with either sign.

The admissibility of eq. sp. can be established as follows. As shown above, for any design $\tilde{z}$, there is a degenerate distribution H_0 against which $\text{MSE}=\left[\max_{0\leqq i\leqq n}(z_{i+1}-z_i)\Big/2\right]^2$. Clearly $\left[\max_{0\leqq i\leqq n}(z_{i+1}-z_i)\Big/2\right]^2\geqq 1/4(n+1)^2$ with equality if and only if $\tilde{z}$ is the eq. sp. design. Against this H_0, for eq. sp., $\text{MSE}\leqq 1/4(n+1)^2$. Thus eq. sp. cannot be weakly dominated by a different design, and hence is admissible.

We will show that eq. sp. is minimax among the designs for which $\hat{\theta}$ depends only on x, the number of successes observed. These designs have many of the desirable properties of eq. sp.; we call them equal weight (eq. wt.) designs.

It can be verified that eq. wt. designs are characterized by r distinct levels w_j, $1\leqq j\leqq r$, and b_j observations at each level w_j, where $\sum_{j=1}^{r}b_j=n$

and $(w_{j+1}-w_{j-1})/b_j$ is constant for $1\leqq j\leqq r$. The z_i's and w_j's are related as in Section 2. If $\hat{H}(w_j)$ is the common value of $\hat{H}(z_i)$, $\sum_{r=1}^{j-1} b_r < i \leqq \sum_{r=1}^{j} b_r$ and $w_0=0$, $w_{r+1}=1$, we have

$$\begin{aligned}\hat{\theta}&=\frac{1}{2}\sum_{i=1}^{n}(z_{i+1}-z_{i-1})\hat{H}(z_i)+\frac{1}{2}(1-z_n)\\&=\frac{1}{2}\sum_{j=1}^{r}(w_{j+1}-w_{j-1})\hat{H}(w_j)+\frac{1}{2}(1-w_r)\\&=\frac{c}{2}\sum_{j=1}^{r}b_j\hat{H}(w_j)+\frac{1}{2}(1-w_r)\\&=\frac{c}{2}x+\frac{1}{2}(1-w_r)\ .\end{aligned}$$

where c is the constant value of $(w_{j+1}-w_{j-1})/b_j$ and x is the total number of successes observed. Since $\sum_{j=1}^{r}(w_{j+1}-w_{j-1})=1+w_r-w_1$ and also $\sum_{j=1}^{r}(w_{j+1}-w_{j-1})=\sum_{j=1}^{r}cb_j=cn$, we have $c=(1+w_r-w_1)/n$ and thus

$$\begin{aligned}\hat{\theta}&=\frac{1+w_r-w_1}{2n}x+\frac{1}{2}(1-w_r)\\&=\frac{x+\dfrac{n(1-w_r)}{1+w_r-w_1}}{n+\dfrac{n(1-w_r)}{1+w_r-w_1}+\dfrac{nw_1}{1+w_r-w_1}}\\&=\frac{x+\alpha(\tilde{z})}{n+\alpha(\tilde{z})+\beta(\tilde{z})}\ .\end{aligned}$$

As in the previous section, it follows that for each eq. wt. design, $\hat{\theta}$ is Bayes against a beta prior on the "flat" H's.

If $1-w_r=w_1=1/(n+1)$, we have $\alpha(\tilde{z})=\beta(\tilde{z})=1/2$, and $\hat{\theta}$ has the form $(x+1/2)/(n+1)$. To see that this may occur without using the eq. sp. design, consider the following example. Let $r=3$; $n=7$; $w_1=1/8$, $w_2=1/2$, $w_3=7/8$ and $b_1=2$, $b_2=3$, $b_3=2$. Then $(w_{j+1}-w_{j-1})/b_j=1/4$ for $1\leqq j\leqq 3$, and $1/(n+1)=1/8=w_1=1-w_3$.

It can be shown that $\hat{\theta}$ is unbiased for $Q(\tilde{z}, H)$ if and only if $\tilde{z}$ is an eq. wt. design. For any eq. wt. design

$$Q(\tilde{z}, H)=\frac{\sum_{i=1}^{n}\theta_i+\alpha(\tilde{z})}{n+\alpha(\tilde{z})+\beta(\tilde{z})}$$

and since x is unbiased for $\sum_{i=1}^{n}\theta_i$, $\hat{\theta}$ is unbiased for $Q(\tilde{z}, H)$. If $\tilde{z}$ is not an eq. wt. design then there is an $i\leqq r-1$ such that $(w_{i+1}-w_{i-1})/b_i\neq$

$(w_{i+2}-w_i)/b_{i+1}$. If we consider H_1 with $H_1(w_j)=0$ for $1 \leqq j \leqq i-1$, $H_1(w_i)=H_1(w_{i+1})=1/2$, and $H_1(w_j)=1$ for $i+2 \leqq j \leqq r+1$, it is easily verified that $E(\hat{\theta}) \neq Q(\tilde{z}, H_1)$.

Finally, to show eq. sp. is minimax among eq. wt. designs, consider an eq. wt. design against H_2, where $H_2(z)=0$, $0 \leqq z < w_1$, and $H_2(z)=1/2$ for $w_1 \leqq z < 1$. Then $\sigma_{\hat{\theta}}^2=((1+w_r-w_1)/2)^2/4n$, $QE=(1-w_r+w_1)/4$, and

$$\mathrm{MSE}=\left(\frac{1+w_r-w_1}{2}\right)^2 \frac{1}{4n}+\left(\frac{1-(w_r-w_1)}{4}\right)^2 .$$

Letting $\delta=w_r-w_1$, we can write

$$\mathrm{MSE}=\left(\frac{1}{4}\right)^2\left[\frac{1}{n}(1+\delta)^2+(1-\delta)^2\right]$$

and it is easily seen that $\mathrm{MSE} \geqq 1/4(n+1)$. Thus eq. sp. is minimax, since for eq. sp., the maximum $\mathrm{MSE}=1/4(n+1)$.

The eq. sp. design and many of the eq. wt. designs are contained in a larger class of designs called symmetric (sym.) designs, characterized by $z_i=1-z_{n+1-i}$ for all $1 \leqq i \leqq n$. We are led to restrict our attention to sym. designs by imposing a reasonable invariance requirement on our estimate for θ. For continuous distribution functions H, using squared error loss, we require that $\hat{\theta}$ be invariant with respect to the two-element group of transformations generated by T where $T(H)$ is given by $T(H)(x)=1-H(1-x)$ for $0 \leqq x \leqq 1$. Clearly $T^2(H)=H$. If H has a density $h(x)$, then $T(H)$ has density $h(1-x)$; T redistributes the mass on $[0, 1]$ to positions symmetric with respect to the midpoint of the interval. $\hat{\theta}$ has this invariant property if and only if a sym. design is used. The long proof of this is omitted.

Using continuity and compactness it can be shown that among the sym. designs there is a minimax design. By a theorem in Ferguson [3], a procedure which is minimax among procedures invariant with respect to a finite group is minimax in general. Thus the class of sym. designs contains a minimax design.

Computations in special cases suggest that the eq. sp. design is "nearly minimax" but that the actual minimax design is a sym. design for which $z_{i+1}-z_{i-1}$ is a non-increasing, somewhere decreasing function of i, for $1 \leqq i \leqq k$ where k is the greatest integer with $k \leqq (n+1)/2$. However, in ignorance of the form of the unknown distribution function H, we prefer to use the eq. sp. design. We believe that for no other design does the estimate $\hat{\theta}$ behave much better over the set of all distribution functions on $[0, 1]$.

5. Application to the estimation of $P(X \leqq Y)$

In this section we consider the problem of estimating $P(X \leqq Y)$ under the conditions:

(i) X and Y are independent random variables with distribution functions F and G respectively.

(ii) G is known and F is unknown.

(iii) The statistician can select n levels, say $y_1 \leqq y_2 \leqq \cdots \leqq y_n$, for experimentation. A random sample $X_1, X_2, \cdots, X_n$ is drawn and the statistician can observe whether or not $X_i \leqq y_i$, $1 \leqq i \leqq n$, but he cannot observe the X_i's directly.

This problem arises in applications in which a component with random strength is subjected to a random stress, and fails when stress exceeds strength. The reliability of the population of components is the probability that strength exceeds stress. Estimation of the reliability when one of the distributions is known is equivalent to the problem described above.

Under the three conditions, we will show that the estimation of $P(X \leqq Y)$ is equivalent to estimation of $\theta = \int_0^1 H(z)dz$, where H is a distribution function on $[0, 1]$, under the conditions described in Section 1. To establish this we first need the following:

LEMMA. *Let ϕ be a Borel measurable function on the reals, let G be a distribution function, and define $\tilde{G}(z) = \sup\{y : G(y) \leqq z\}$ for $0 < z < 1$. Then, $\int_{-\infty}^{\infty} \phi(y)dG(y) = \int_0^1 \phi(\tilde{G}(z))dz$.*

PROOF. Since G is non-decreasing and continuous on the right, it follows that for all $\epsilon > 0$,

$$G[\tilde{G}(z) - \epsilon] \leqq z \leqq G[\tilde{G}(z)] \; . \tag{1}$$

Note that if G has an inverse function G^{-1}, then $\tilde{G} = G^{-1}$. Now, if μ is Lebesque measure and $\mu\tilde{G}^{-1}$ is the measure induced by $\tilde{G}$ (i.e., for any Borel subset A, $[\mu\tilde{G}^{-1}](A) = \mu[\tilde{G}^{-1}(A)]$), then it is true that

$$\int_{-\infty}^{\infty} \phi(y)d[\mu\tilde{G}^{-1}](y) = \int_0^1 \phi[\tilde{G}(z)]d\mu(z) \; ,$$

(Halmos, [4]). The proof will be completed if we show that for any Borel subset A, $P_G(A) = [\mu\tilde{G}^{-1}](A)$, where $P_G(A)$ is the probability assigned to A by the distribution function G. It is sufficient to show equality for $A = (-\infty, a]$. In this case, $\tilde{G}^{-1}(A) = \{z : \tilde{G}(z) \leqq a\}$. But if

$\tilde{G}(z) \leqq a$, then $z \leqq G[\tilde{G}(z)] \leqq G(a)$, where the first inequality follows from (1). Conversely, if $\tilde{G}(z) > a$, then for sufficiently small $\epsilon > 0$,

$$G(a) \leqq G[\tilde{G}(z) - \epsilon] \leqq z .$$

Thus $\tilde{G}^{-1}(A)$ is either the interval $(0, G(a))$ or $(0, G(a)]$; and in either case $[\mu\tilde{G}^{-1}](A) = G(a) = P_G(A)$, and the lemma is established.

The following theorem gives the desired representation of $P(X \leqq Y)$.

THEOREM. *Let X, Y be independent random variables with distribution functions F, G respectively. Then $P(X \leqq Y) = \int_0^1 H(z)dz$, where $H(z)$ is a distribution function on* $[0, 1]$, *and $H(z) = F[\tilde{G}(z)]$ for $0 < z < 1$.*

PROOF. Since $P(X \leqq Y) = \int_{-\infty}^{\infty} F(y)dG(y)$, we have by the previous lemma, that for $H(z) = F[\tilde{G}(z)]$, $0 < z < 1$, $P(X \leqq Y) = \int_0^1 H(z)dz$. We need only show that H can be a distribution function. It can be verified that $\tilde{G}$ is continuous on the right; the details of this are omitted. Since both F and $\tilde{G}$ are non-decreasing and continuous on the right, it follows that H has these properties also, and if we define $H(1) = 1$ and $H(0) = F[\tilde{G}(0+)]$, the theorem is established.

The theorem makes clear that estimation of $P(X \leqq Y)$ is equivalent to estimation of $\int_0^1 H(z)dz$, as in Section 1. Consider now how the estimate of Section 3 is used. First n levels, $0 < z_1 \leqq z_2 \leqq \cdots \leqq z_n < 1$ are selected. The corresponding Y levels are $y_i = \tilde{G}(z_i)$, $1 \leqq i \leqq n$. Note that $P(X_i \leqq y_i) = F[\tilde{G}(z_i)] = H(z_i)$. The experimenter observes whether or not $X_i \leqq y_i$ for $1 \leqq i \leqq n$, and the outcomes of these Bernoulli trials are used to obtain the maximum likelihood estimate $(\hat{H}(z_1), \hat{H}(z_2), \cdots, \hat{H}(z_n))$ and $\hat{H}(z)$ is further defined as in Section 3. Then the estimate for $P(X \leqq Y)$ is $\hat{P} = \int_0^1 \hat{H}(z)dz$.

If the distribution function H is arbitrary, then $(\hat{H}(z_1), \hat{H}(z_2), \cdots, \hat{H}(z_n))$ is the ABERS estimate. However, since G is not necessarily continuous, then we may have $\tilde{G}(z_i) = \tilde{G}(z_{i+1})$ for $z_i < z_{i+1}$. In this case $H(z_i) = H(z_{i+1})$, which is additional information about H. For arbitrary levels $0 < z_1 \leqq z_2 \leqq \cdots \leqq z_n < 1$, the maximum likelihood estimate must be modified; if $H(z_i) = H(z_{i+1})$ for $z_i < z_{i+1}$, then additional pooling may have to be performed to obtain a common estimate of $H(z_i) = H(z_{i+1})$. This additional pooling may be done after clumping is completed. If G is continuous, then H is an arbitrary distribution function on $[0, 1]$, since F

is arbitrary, and no additional pooling is appropriate.

If an eq. wt. design is used, the additional information about where H is constant becomes irrelevant, since for eq. wt. designs, $\hat{\theta}$ depends only on the number of successes observed and $\hat{\theta}$ is unchanged by additional pooling.

Acknowledgement

This research is an offspring of earlier work, by one of the authors and Bernard Harris, of the Mathematics Research Center, University of Wisconsin, on a parametric version of the problem of Section 5. Professor Harris contributed much toward formulation of the problem treated here.

This work was supported by a grant from The General Research Fund, University of Kansas.

THE UNIVERSITY OF KANSAS

REFERENCES

[1] Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information, *Ann. Math. Statist.*, **26**, 641-647.

[2] Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, **II**, New York, Wiley.

[3] Ferguson, T. (1967). *Mathematical Statistics: A Decision Theoretic Approach*, New York, Academic.

[4] Halmos, P. R. (1950). *Measure Theory*, Princeton, Van Nostrand.