MINIMAX ESTIMATES OF THE MEAN OF A NORMAL DISTRIBUTION WITH KNOWN VARIANCE

By J. Wolfowitz[1]

Columbia University

Summary. It is proved that the classical estimation procedures for the mean of a normal distribution with known variance are minimax solutions of properly formulated problems. A result of Stein and Wald [1] is an immediate consequence. Other such optimum properties follow. Sequential and non-sequential problems can be treated in this manner. Interval and point estimation are discussed.

1. Sequential estimation by an interval of given length l. In this section we shall consider the problem of sequentially estimating the mean of a normal distribution with known variance by an interval of fixed length l. Without loss of generality we shall take the known variance to be unity. Such a sequential estimation procedure, which we shall designate generically by G, is a rule which says a) when to terminate taking random, independent observations on the normal chance variable with unknown mean $\xi(-\infty < \xi < \infty)$ and variance 1, and when this termination is to occur after the observations $x_1, \cdots, x_n$ have been obtained, gives b) the center of the estimating interval of length l as a function of $x_1, \cdots, x_n$. Let $\alpha(\xi, G)$ be the probability under G that the estimating interval will contain ξ, and let $n(\xi, G)$ be the expected number of observations when ξ is the mean and G is the estimation procedure (It is assumed that G is such that $\alpha(\xi, G)$ and $n(\xi, G)$ exist for all ξ).

Define

$$q(\xi, G) = 1 - \alpha(\xi, G),$$

and for fixed $c > 0$

$$W(\xi, G) = q(\xi, G) + cn(\xi, G). \tag{1.1}$$

Let $C(N, l)$ ($l > 0$, N a positive integer) be the classical non-sequential estimation procedure where one takes the fixed number N of observations, and estimates the mean by the interval $\left(\bar{x} - \frac{l}{2}, \bar{x} + \frac{l}{2}\right)$, where $\bar{x}$ is the sample mean. For p such that $0 < p \leq 1$, let $C(p, N, l)$ be the following estimation procedure: A chance experiment with two outcomes, N and $N + 1$, of respective probabilities p and $1 - p$, is performed. One then proceeds according to $C(i, l)$, where $i(= N, N + 1)$ is the outcome of the experiment. Finally define

$$M(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty e^{-\frac{1}{2}z^2}\, dz.$$

---

[1] Research under a contract with the Office of Naval Research.

Let us assume for a moment that the unknown ξ is itself a chance variable, normally distributed with mean zero and variance σ^2, and let us obtain a procedure G which minimizes

$$(1.2)\quad E\{q(\xi, G) + c\, n(\xi, G)\} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\{q(y, G) + cn(y, G)\}\exp\left[\frac{-y^2}{2\sigma^2}\right]dy.$$

Let $x_1, \cdots, x_m$ be m independent observations on a normal chance variable with mean ξ and variance 1. Let

$$\bar{x} = \frac{\sum_1^m x_i}{m}.$$

The a posteriori distribution of ξ, given $x_1, \cdots, x_m$, is easily verified (or see [1], eqs. (19) and (20)) to be normal with mean

$$(1.3)\qquad \bar{x}\left[1 + \frac{1}{m\sigma^2}\right]^{-1}$$

and variance

$$(1.4)\qquad \left[m + \frac{1}{\sigma^2}\right]^{-1}.$$

Thus if we stop after m observations the best procedure from the point of view of minimizing (1.2) is to put the center of the estimating interval of length l at the point (1.3). The conditional expected value of $q(\xi)$ is then

$$(1.5)\qquad Q(x_1, \cdots, x_m \mid \sigma^2) = 2M\left(\frac{l}{2}\sqrt{m + \frac{1}{\sigma^2}}\right).$$

Thus $Q(x_1, \cdots, x_m)$ is a function only of m and σ^2. Define

$$(1.6)\qquad R(m, \sigma^2) = 2M\left(\frac{l}{2}\sqrt{m + \frac{1}{\sigma^2}}\right) - 2M\left(\frac{l}{2}\sqrt{m + 1 + \frac{1}{\sigma^2}}\right).$$

We note that $R(m, \sigma^2)$ is, for fixed σ, a decreasing function of m. We conclude that a best decision as to whether or not to take another observation must be based on the value of $R(m, \sigma^2)$. If $R(m, \sigma^2) > c$ take another observation; if $R(m, \sigma^2) < c$ do not take another observation; if $R(m, \sigma^2) = c$ take either action at pleasure. Hence, if c is such that $R(N, \sigma^2) \leq c \leq R(N - 1, \sigma^2)$, a best procedure from the point of view of minimizing (1.2) is to take exactly N observations. This integer N is a function of c and σ^2, thus: $N(c, \sigma^2)$. In the next paragraph we shall show that $N(c, \sigma^2)$ can be defined for every positive c and σ^2. It is clearly a function which takes at most two values. We shall denote by $G(\sigma^2)$ the estimation procedure described above which minimizes (1.2). It consists of taking the fixed number $N(c, \sigma^2)$ of observations and putting the center of the estimating interval of length l at the point (1.3). Where $N(c, \sigma^2)$ is double-valued we may take either value at pleasure. We verify that the value of (1.2) is the same for either choice.

We now verify that $N(c, \sigma^2)$ can be defined for all positive c and σ^2. We have remarked earlier that $R(m, \sigma^2)$ is, for fixed σ^2, a monotonically decreasing function of m. We note that

$$\lim_{m=\infty} R(m, \sigma^2) = 0.$$

When $c > R(0, \sigma^2)$ we take no observations whatever and take $\bar{x} \equiv 0$. When $c = R(0, \sigma^2)$ we take zero or one observation at pleasure.

Without difficulty we compute

$$W(\xi, G(\sigma^2)) = W(\xi, \sigma^2) = cN + M\left(\sqrt{N}\,\frac{l}{2}\left[1 + \frac{1}{N\sigma^2}\right] - \frac{\xi}{\sqrt{N}\,\sigma^2}\right) + M\left(\sqrt{N}\,\frac{l}{2}\left[1 + \frac{1}{N\sigma^2}\right] + \frac{\xi}{\sqrt{N}\,\sigma^2}\right)$$

where for typographical simplicity we have written N for $N(c, \sigma^2)$. For fixed c and σ^2 the minimum of $W(\xi, \sigma^2)$ occurs at $\xi = 0$. Also $W(0, \sigma^2)$ is a monotonically increasing function of σ^2. If $N(c, \infty) > 0$ then, as $\sigma^2 \to \infty$ it approaches the limit

$$cN(c, \infty) + 2M\left(\frac{l}{2}\sqrt{N(c, \infty)}\right),$$

which is the constant value of

$$W(\xi, C(N(c, \infty), l)).$$

We therefore conclude that $C(N(c, \infty), l)$ is a minimax estimating procedure of type G, i.e.,

$$W(\xi, C(N(c, \infty), l)) = \inf_G \sup_\xi W(\xi, G)$$

for any $c > 0$. (The case $N(c, \infty) = 0$ may be verified separately. We define $\bar{x} \equiv 0$ for $C(0, l)$).

Conversely, let N_0 be a given non-negative integer. Then $C(N_0, l)$ is a minimax estimating procedure G for all $W(\xi, G)$ for which c satisfies

$$R(N_0, \infty) \le c \le R(N_0 - 1, \infty).$$

(We define $R(-1, \infty) = \infty$.) Thus we can say: For every $c > 0$ there exists a classical estimation procedure $C(N, l)$ with integral N such that

$$W(\xi, C(N, l)) = \inf_G \sup_\xi W(\xi, G).$$

For every integral N we can find at least one $c > 0$ such that the above equation holds. A method of finding N, given c, and of finding c, given N, has been described above. (We have taken the liberty of calling $C(0, l)$ a classical procedure.

Let α_0 be a given number such that

$$1 - 2M\left(\frac{l}{2}\right) \le \alpha_0 < 1.$$

Define p_0, $0 < p_0 \leq 1$, and a positive integral N_0 uniquely by

$$\alpha_0 = p_0 \left(1 - 2M\left(\sqrt{N_0}\, \frac{l}{2}\right)\right) + (1 - p_0) \left(1 - 2M\left(\sqrt{N_0 + 1}\, \frac{l}{2}\right)\right).$$

Let

$$c_0 = R(N_0, \infty).$$

For $c = c_0$ we verify readily that both $C(N_0, l)$ and $C(N_0 + 1, l)$ are minimax estimating procedures G, so that

$$\begin{aligned} W(\xi, C(N_0, l)) &= W(\xi, C(N_0 + 1, l)) \\ &= p_0\, W(\xi, C(N_0, l)) + (1 - p_0)\, W(\xi, C(N_0 + 1, l)) \\ &= (1 - \alpha_0) + c_0[p_0 N_0 + (1 - p_0)(N_0 + 1)] \\ &= (1 - \alpha_0) + c_0[N_0 + (1 - p_0)]. \end{aligned}$$

Therefore, for any G whatever,

$$\begin{aligned} (1 - \alpha_0) + c_0[N_0 + (1 - p_0)] &\leq \sup_\xi \{q(\xi, G) + c_0\, n(\xi, G)\} \\ &\leq \sup_\xi q(\xi, G) + c_0 \sup_\xi n(\xi, G). \end{aligned}$$

Hence

$$\sup_\xi q(\xi, G) \leq 1 - \alpha_0$$

implies

$$\sup_\xi n(\xi, G) \geq N_0 + (1 - p_0),$$

a result first proved by Stein and Wald [1].
Also

$$\sup_\xi n(\xi, G) \leq N_0 + (1 - p_0)$$

implies

$$\sup_\xi q(\xi, G) \geq 1 - \alpha_0,$$

a result also proved in [1].

2. A sequential upper bound for the mean. The fact that in the last section l was a constant made matters simpler, as we see when we begin to consider the problem of a sequential upper bound for $\xi(-\infty < \xi < \infty)$. This of course means that we wish to use as estimating interval the interval $(-\infty, L(x_1, \cdots, x_n))$ where L is a function of the observations $x_1, \cdots, x_n$, and n (a chance variable) is the number of observations before the process of taking observations is terminated. What is wanted now is a suitable definition of the "length" of this interval. Also we shall admit the possibility that it might be in some sense advantageous to have intervals of varying length; this poses the problem of optimum choice of the function $L(x_1, \cdots, x_n)$.

As before, let ξ be the mean of a normal distribution with unit variance. Let T be the generic estimation procedure which consists of a rule for terminating the taking of observations, and of a function $L_T(x_1, \cdots, x_n)$ which is used to estimate ξ by the interval $(-\infty, L_T)$. Define

$$q(\xi, T) = P\{L_T \leq \xi\},$$
$$\lambda(\xi, T) = E(L_T - \xi)^2,$$

and

$$W(\xi, T) = q(\xi, T) + k\lambda(\xi, T) + cn(\xi, T), \tag{2.1}$$

where c and k are positive constants. (We admit only such T for which the quantities q, λ, and n are defined for all real ξ.) As before, let us temporarily assume that ξ is normally distributed with mean zero and variance σ^2, and set ourselves the task of minimizing

$$\frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} W(y, T)\, e^{-(y^2/2\sigma^2)}\, dy = W^*(T, \sigma^2) \tag{2.2}$$

with respect to T. In the next paragraph we digress for a moment to derive a needed elementary inequality.

Let us prove that, if h, h_1, and h_2 are non-negative, and

$$h^2 = p\, h_1^2 + (1 - p)\, h_2^2, \tag{2.3}$$

where $0 < p < 1$, then

$$M(h) \leq p\, M(h_1) + (1 - p)\, M(h_2). \tag{2.4}$$

Hold h and p fixed. The desired result is obviously true when $h_1 = h_2 = h$. Let h_1 and h_2 vary, subject to (2.3). Then

$$\frac{dh_2}{dh_1} = \frac{-ph_1}{(1 - p)h_2}.$$

Also

$$\frac{pdM(h_1)}{dh_1} = \frac{-p}{\sqrt{2\pi}} e^{-\frac{1}{2}h_1^2}$$

and

$$(1 - p)\frac{dM(h_2)}{dh_1} = (1 - p)\frac{dM(h_2)}{dh_2}\frac{dh_2}{dh_1} = \frac{ph_1}{\sqrt{2\pi}h_2} e^{-\frac{1}{2}h_2^2}.$$

Thus the derivative of the right member of (2.4) with respect to h_1 is 0 when $h_1 = h$, positive when $h_1 > h$, and negative when $h_1 < h$. From this we get (2.4).

Let T be any estimation procedure and $L_T(x_1, \cdots, x_n)$ its associated function. Write

$$l_T(x_1, \cdots, x_n) = L_T(x_1, \cdots, x_n) - \bar{x}\left[1 + \frac{1}{n\sigma^2}\right]^{-1}.$$

If $n = m$ and $x_1, \cdots, x_m$ is the sample obtained, we have that the conditional expected value of $W^*(T, \sigma^2)$ is

$$M\left(l_T(x_1, \cdots, x_m)\sqrt{m + \frac{1}{\sigma^2}}\right) + cm + kE(U_m^* + l_T(x_1, \cdots, x_m))^2, \tag{2.5}$$

where U_m^* is a normally distributed chance variable with mean zero and variance $\left(m + \frac{1}{\sigma^2}\right)^{-1}$. The last term in (2.5) is therefore

$$k\left[\left(m + \frac{1}{\sigma^2}\right)^{-1} + l_T^2(x_1, \cdots, x_m)\right].$$

This is an even function of l_T, while the first term of (2.5) is a monotonically decreasing function of l_T. Thus (2.5) and hence $W^*(T, \sigma^2)$ will be minimized by taking l_T non-negative. Now take the expected value of (2.5) over the set of samples where $n = m$. Application of the result of the preceding paragraph to the finite sums which approximate the integral gives the result that $W^*(T, \sigma^2)$ is minimized when $l_T(x_1, \cdots, x_m)$ is a function only of m. Hence we may restrict ourselves to consideration of procedures T for which (2.5) takes the value

$$M\left(\sqrt{m + \frac{1}{\sigma^2}}\, l_T(m)\right) + cm + k\left[\left(m + \frac{1}{\sigma^2}\right)^{-1} + \{l_T(m)\}^2\right]. \tag{2.6}$$

For any such procedure T, since k and c are fixed positive numbers (and σ^2 is held fixed for the present), the expression (2.6) takes its minimum for some value of m. Thus, in our quest for a procedure T which will minimize $W^*(T, \sigma^2)$ we may restrict ourselves to procedures of *fixed* sample size. This fixed sample size and the (constant) value of l_T are functions of k, c, and σ^2. For fixed m,

$$M\left(\sqrt{m + \frac{1}{\sigma^2}}\, l^0\right) + k(l^0)^2$$

has an absolute minimum at l_m, say, since it is a continuous function of $l^0 (l^0 \geq 0)$ which approaches ∞ with l^0. The case $m = 0$ must be considered. (In this event $\bar{x} \equiv 0$.) Now consider the sequence

$$\left\{M\left(\sqrt{m + \frac{1}{\sigma^2}}\, l_m\right) + cm + k\left[\left(m + \frac{1}{\sigma^2}\right)^{-1} + l_m^2\right]\right\}$$

for $m = 0, 1, 2, \cdots$ ad inf. This sequence condenses only at ∞. Hence there exists a value $N(k, c, \sigma^2)$ of m for which the elements of this sequence have a minimum value. We may choose $N(k, c, \sigma^2)$ so that $\lim_{\sigma^2=\infty} N(k, c, \sigma^2)$ exists. (We verify easily that this is always possible.) Designate this limit by $N(k, c, \infty)$, and the associated l by $l(k, c, \infty)$. The l associated with $N(k, c, \sigma^2)$ will be designated by $l(k, c, \sigma^2)$. Thus a best procedure for minimizing $W^*(T, \sigma^2)$ is to take the fixed number $N(k, c, \sigma^2)$ observations, and to use, as upper bound for ξ, the quantity

$$\bar{x}\left[1 + \frac{1}{\sigma^2 N(k, c, \sigma^2)}\right]^{-1} + l(k, c, \sigma^2).$$

We see readily that

$$l(k, c, \infty) = \lim_{\sigma^2 = \infty} l(k, c, \sigma^2)$$

and that

$$M(\sqrt{N(k, c, \infty)}\; l(k, c, \infty)) = \lim_{\sigma^2 = \infty} M\left(\sqrt{N(k, c, \sigma^2) + \frac{1}{\sigma^2}}\; l(k, c, \sigma^2)\right).$$

Let $T(\sigma^2)$ be the procedure described above which is a best procedure T in the sense of minimizing $W^*(T, \sigma^2)$ when σ^2 is the variance of ξ.

We now compute $W(\xi, T(\sigma^2))$ and obtain

$$W(\xi, T(\sigma^2)) = cN + k\left[\frac{N\sigma^4}{(1 + N\sigma^2)^2} + \left(l - \frac{\xi}{1 + N\sigma^2}\right)^2\right] + M\left(\frac{1 + N\sigma^2}{\sqrt{N}\sigma^2}\left[l - \frac{\xi}{1 + N\sigma^2}\right]\right), \tag{2.7}$$

where for brevity we have written N and l for $N(k, c, \sigma^2)$ and $l(k, c, \sigma^2)$. Let

$$l - \frac{\xi}{1 + N\sigma^2} = x, \qquad \frac{1 + N\sigma^2}{\sqrt{N}\sigma^2} = \sqrt{N} + \epsilon.$$

Then

$$W = cN + k\left[\frac{1}{(\sqrt{N} + \epsilon)^2} + x^2\right] + M([\sqrt{N} + \epsilon]\, x), \tag{2.8}$$

$$\frac{\partial W}{\partial x} = 2kx - \frac{(\sqrt{N} + \epsilon)}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\{(\sqrt{N} + \epsilon)^2 x^2\}\right]. \tag{2.9}$$

The second term above is always of the same sign and the exponential decreases as $|x|$ increases. Thus $\partial W/\partial x = 0$ has the unique positive root x^*. Put x^* for x in W (in 2.8) and call the result W^*. W is a continuous function of x and approaches ∞ as $|x| \to \infty$. Since the root x^* is unique it follows that W^* is the minimum value of W with respect to x. Now $N(k, c, \sigma^2)$ is constant for σ^2 sufficiently large. Hence, for such σ^2, we have

$$\begin{aligned}\frac{\partial W^*}{\partial \epsilon} &= \frac{-2k}{(\sqrt{N} + \epsilon)^3} + 2kx^*\frac{dx^*}{d\epsilon} - \frac{dx^*}{d\epsilon}\frac{(\sqrt{N} + \epsilon)}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\{(\sqrt{N} + \epsilon)^2 x^{*2}\}\right] \\ &\quad - \frac{x^*}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\{(\sqrt{N} + \epsilon)^2 x^{*2}\}\right] \\ &= \frac{-2k}{(\sqrt{N} + \epsilon)^3} - \frac{x^*}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\{(\sqrt{N} + \epsilon)^2 x^{*2}\}\right]\end{aligned} \tag{2.10}$$

since x^* is the root of $\partial W/\partial x = 0$. Also ϵ is positive and, for σ^2 sufficiently large, approaches zero monotonically as σ^2 approaches ∞. For $\epsilon > 0$ we have that $\partial W^*/\partial\epsilon < 0$, since $x^* > 0$. We conclude: For σ^2 sufficiently large,

$$\min_{\xi} W(\xi, T(\sigma^2))$$

increases monotonically with σ^2 and approaches

$$cN + k\left[\frac{1}{N} + \{x_N(k)\}^2\right] + M(\sqrt{N}x_N(k)),$$

where N is short for $N(k, c, \infty)$ and $x_N(k)$ is the unique positive root of the equation in x

$$2kx = \frac{\sqrt{N}}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}Nx^2\right].$$

Going back to the definition of $l(k, c, \infty)$ we see that the latter satisfies the equation in l:

$$\frac{d}{dl}\{M(\sqrt{N}\, l) + kl^2\} = 0.$$

Hence

$$x_N(k) = l(k, c, \infty).$$

Thus the classical estimation procedure C_0 where one takes the fixed number $N(k, c, \infty)$ of observations and uses as upper bound for the mean $\bar{x} + l(k, c, \infty)$ is a minimax procedure T, i.e.,

$$W(\xi, C_0) = \inf_T \sup_\xi W(\xi, T).$$

For fixed N, $x_N(k)$ decreases monotonically from $+\infty$ to 0 as k increases from 0 to $+\infty$. Hence, for given positive integral N_0 and $l^* > 0$, there is a unique positive value k_0 such that $x_{N_0}(k_0) = l^*$. Consider the expression

$$B(m) = M(\sqrt{m}\, x_m(k_0)) + cm + k_0\left[\frac{1}{m} + \{x_m(k_0)\}^2\right], \tag{2.11}$$

where m is a positive, continuous variable. We have

$$\frac{dB(m)}{dm} = c - \frac{k_0}{m^2} + \frac{dx_m(k_0)}{dm}\frac{\partial}{\partial x_m(k_0)}\left\{M(\sqrt{m}\, x_m(k_0)) + k_0[x_m(k_0)]^2\right\} + \frac{\partial M(\sqrt{m}\, x_m(k_0))}{\partial m}. \tag{2.12}$$

The third term of the right member is identically zero because

$$2k_0 x_m(k_0) = \frac{\sqrt{m}}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}m[x_m(k_0)]^2\right\}. \tag{2.13}$$

Further we have

$$\frac{d^2B(m)}{dm^2} = \frac{2k_0}{m^3} - \frac{d}{dm}\left\{\frac{m^{-\frac{1}{2}}x_m(k_0)}{2\sqrt{2\pi}}\, e^{-\frac{1}{2}mx_m^2(k_0)}\right\}$$
(2.14)
$$= \frac{2k_0}{m^3} - \frac{k_0\, d\{m^{-1}(x_m(k_0))^2\}}{dm}.$$

For typographic simplicity we shall use y for $x_m(k_0)$ in the computations of the next few lines. From (2.13) we obtain

$$\log 2k_0 + \log y = -\log\sqrt{2\pi} + \tfrac{1}{2}\log m - \tfrac{1}{2}\, m\, y^2,$$

$$\frac{1}{y}\frac{dy}{dm} = \frac{1}{2m} - \frac{y^2}{2} - my\frac{dy}{dm},$$

$$\frac{dy}{dm} = \frac{y(1 - my^2)}{2m(1 + my^2)}.$$

Hence

$$\frac{d^2B(m)}{dm^2} = 2k_0 m^{-3} + k_0 m^{-2}\, y^2 - 2k_0 m^{-1}\, y\frac{dy}{dm}$$
(2.15)
$$= 2k_0 m^{-3} + k_0 m^{-2}\, y^2 - \frac{k_0 y^2(1 - my^2)}{m^2(1 + my^2)}$$
$$= 2k_0 m^{-3} + \frac{2y^4k_0}{m(1 + my^2)} > 0.$$

Since $c > 0$, we have

$$\lim_{m=0} B(m) = \lim_{m=\infty} B(m) = +\infty.$$

Hence there exists a value of m for which $B(m)$ takes its minimum value. If in $d\,B\,(m)/dm$ we put $m = N_0$ and set the resulting expression equal to zero, we obtain an equation in c whose unique solution c_0, if it is positive, assures us that, when $c = c_0$ and $k = k_0$, $B\,(m)$ takes its minimum at $m = N_0$. A simple computation gives

$$c_0 = \frac{k_0}{N_0^2} + \frac{l^* \exp\{-\frac{1}{2}N_0 l^{*2}\}}{2\sqrt{2\pi N_0}} > 0. \tag{2.16}$$

Actually we are interested in considering $B\,(m)$ only for positive integral values of m. We see readily that the minimum of $B\,(m)$ occurs then at $m = N_0$ when c is such that

$$c_1\,(N_0, k_0) \leq c \leq c_2\,(N_0, k_0), \tag{2.17}$$

with c_1 and c_2 roots of the following equations in c:

$$B\,(N_0) = B\,(N_0 + 1),$$
$$B\,(N_0) = B\,(N_0 - 1).$$

(If $N_0 = 1$, then $c_2 = \infty$.)

Let C_0 (N_0, l^*) be the classical (non-sequential) procedure where one takes N_0 observations and uses $\bar{x} + l^*$ as upper bound for the mean. Choose $k = k_0$ and c such that (2.17) is satisfied. Then

$$W(\xi, C_0(N_0, l^*)) = cN_0 + k_0\left(\frac{1}{N_0} + l^{*2}\right) + M(\sqrt{N_0}\, l^*)$$

identically in ξ. $C_0(N_0, l^*)$ is a procedure T such that

$$W(\xi, C_0) = \inf_T \sup_\xi W(\xi, T). \tag{2.18}$$

Whenever c and k are given, the N and l of the minimax solution may be obtained as follows: First we obtain an integer N such that

$$c_1(N, k) \leq c \leq c_2(N, k).$$

Knowing N and k we can then solve for l.

The results of this section may be summarized as follows: For every positive c and k there exists a classical estimation procedure $C_0(N, l)$ with positive integral N and $l > 0$ such that (2.18) holds. Conversely, for every such pair (N, l) there exists a positive pair (c, k) so that (2.18) holds. A method of finding one member of the pair of couples (c, k) and (N, l) when the other is given, has been indicated above.

Let T_1 be any procedure for giving an upper bound for ξ. We shall say that T_1 is optimum if for any other procedure T_2 such that

$$\sup_\xi q(\xi, T_2) \leq \sup_\xi q(\xi, T_1),$$

$$\sup_\xi \lambda(\xi, T_2) \leq \sup_\xi \lambda(\xi, T_1),$$

we have

$$\sup_\xi n(\xi, T_2) \geq \sup_\xi n(\xi, T_1).$$

It is easy to prove that the classical procedure C_0 with any positive l and positive integral N is optimum by using the results of the last paragraph. For let $1 - \alpha = M\,(l\sqrt{N})$ and let k and c be the corresponding parameters. We have then

$$\sup_\xi q(\xi, T_2) + k \sup_\xi \lambda(\xi, T_2) + c \sup_\xi n(\xi, T_2) \geq \sup_\xi \{q(\xi, T_2)$$

$$+ k\,\lambda(\xi, T_2) + cn(\xi, T_2)\} \geq (1 - \alpha) + k\left(\frac{1}{N} + l^2\right) + cN.$$

Since $\sup q(\xi, T_2) \leq (1 - \alpha)$ and $\sup \lambda(\xi, T_2) \leq 1/N + l^2$, we must have

$$\sup_\xi n(\xi, T_2) \geq N,$$

which is the desired result.

In a general unprecise way we may say that an estimation procedure is the better the smaller the three quantities

$$\beta_1(T) = \sup_\xi q(\xi, T), \qquad \beta_2(T) = \sup_\xi \lambda(\xi, T), \qquad \beta_3(T) = \sup_\xi n(\xi, T).$$

We can now assert the following: No sequential procedure T can be superior to the classical fixed sample procedure C in the sense that

$$\beta_i(T) \leq \beta_i(C) \qquad \text{for } i = 1, 2, 3$$

and the inequality sign holds for at least one i.

In concluding this section we may remark that the case $\alpha \leq \frac{1}{2}$, i.e., $l \leq 0$, may be handled in the same manner as above except that we use $M(-l\sqrt{m})$ in place of $M(l\sqrt{m})$.

3. Miscellaneous results; point estimation. Without going into the necessarily involved details, we content ourselves with pointing out that the problem of estimating sequentially the mean of a normal distribution by a finite interval of length not specified in advance, can be solved in similar fashion. As before let ξ be the unknown mean of a normal distribution with unit variance, where ξ may be any real value. We want to estimate by an interval

$$(L_1(x_1, \cdots, x_n), \qquad L_2(x_1, \cdots, x_n)).$$

Let c, k_1, and k_2 be positive constants and consider the problem of minimizing the supremum with respect to ξ of

$$1 - P\{L_1 < \xi < L_2 \mid G^1\} + cn(\xi, G^1) + k_1 E[(L_1 - \xi)^2 \mid G^1] + k_2 E[(L_2 - \xi)^2 \mid G^1],$$

where G^1 is the generic designation of the estimation procedure. As before, employ an a priori normal distribution of ξ with mean zero and variance σ^2, and let $\sigma^2 \to \infty$. A fixed sample size procedure will be a minimax solution. It will possess optimum properties similar to those described in the preceding sections. The problem of minimizing the supremum with respect to ξ of

$$1 - P\{L_1 < \xi < L_2 \mid G^1\} + cn(\xi, G^1) + kE\{(L_2 - L_1)^2 \mid \xi, G^1\}$$

can be treated similarly.

Suppose the sample size is fixed in advance. The problem of finding an estimate which will minimize

$$\sup_{\xi}[1 - P\{L_1 < \xi < L_2 \mid G^1\} + k_1 E\{(L_1 - \xi)^2 \mid G^1\} + k_2 E\{(L_2 - \xi)^2 \mid G^1\}]$$

or

$$\sup_{\xi}[1 - P\{L_1 < \xi < L_2 \mid G^1\} + kE\{(L_2 - L_1)^2 \mid \xi, G^1\}]$$

can be treated by the method of the preceding sections.

The problem of estimating (sequentially or with fixed sample size) the means of a multivariate normal distribution with known covariance matrix can be treated in similar fashion.

Suppose it is desired to estimate sequentially the mean ξ $(-\infty < \xi < \infty)$ of a normal distribution with unit variance by means of a chance point

$\hat{\xi}\ (x_1, \cdots, x_n)$. Let $R(\xi, \xi^1)$ be the Wald risk function (cf. [2]), a non-negative function which measures the loss incurred in using the particular value ξ^1 as an estimate when ξ is the actual value. The functions $\hat{\xi}\ (x_1, \cdots, x_n)$ and $R(\xi, \xi^1)$ must have suitable measurability properties for which we refer the reader to [2]. Let us seek a procedure ξ^* such that

$$\sup_{\xi}[E\{R(\xi, \xi^*)\} + cn(\xi, \xi^*)] = \inf_{\hat{\xi}} \sup_{\xi}[E\{R(\xi, \hat{\xi})\} + c\ n(\xi, \hat{\xi})].$$

Here $n(\xi, \hat{\xi})$ is the average number of observations under $\hat{\xi}$ when ξ is the "true" mean. The procedure ξ^* will be called a minimax solution. We shall assume that $R(a, b)$ is a monotonically non-decreasing function of $| a - b |$, and that there exists a positive number g such that

$$\int_0^{\infty} R(0, x) \exp\left\{\frac{-x^2}{2g}\right\} dx < \infty .$$

As examples of functions with these properties we may cite

$$R(a, b) = | a - b |,$$

$$R(a, b) = (a - b)^2.$$

As before, assume temporarily that ξ is normally distributed with mean zero and variance σ^2. We verify without difficulty that a solution $\hat{\xi} = \xi_0$ which minimizes

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} [E\{R(\xi, \hat{\xi})\} + cn(\xi, \hat{\xi})] \exp\left\{-\tfrac{1}{2} \frac{\xi^2}{\sigma^2}\right\} d\xi$$

is the following: n is identically a suitable constant, say N, and ξ_0 is $\bar{x}(1 + 1/ N\sigma^2)^{-1} = \bar{x}h$ say, so that $h < 1$. For this solution we have

$$E\{R(\xi, \xi_0)\} + cn(\xi, \xi_0) = cN + \frac{\sqrt{N}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R(\xi, \bar{x}h) \exp\left\{-\frac{N}{2} (\bar{x} - \xi)^2\right\} d\bar{x} .$$

Write $u = \bar{x} - \xi$. Then

$$R(\xi, \bar{x}h) = R(\xi, h [\xi + u]) = R(0, hu - [1 - h]\xi),$$

$$\int_{-\infty}^{\infty} R(\xi, \bar{x}h) \exp\left\{-\frac{N}{2} (\bar{x} - \xi)^2\right\} d\bar{x}$$

$$= \int_{-\infty}^{\infty} R(0, hu - [1 - h]\xi) \exp\left\{-\frac{Nu^2}{2}\right\} du$$

$$= \int_{-\infty}^{\infty} R(0, v) \exp\left\{-\frac{N}{2h^2} (v + [1 - h]\xi)^2\right\} \frac{1}{h}\, dv.$$

Because of the assumptions on the function R the last expression is a minimum when $\xi = 0$. We may always choose N such that, for large enough σ^2, the integer N is a constant, say N_0. Also $h \to 1$ as $\sigma^2 \to \infty$. Thus we conclude that the following is a minimax solution: $n = N_0$ and $\hat{\xi} = \xi^* = \bar{x}$. If any estimation procedure $\hat{\xi}$ is such that $\sup_{\xi} n(\xi, \hat{\xi}) \leq N_0$ then

$$\sup_{\xi} E\{R(\xi, \hat{\xi})\} \geq E\{R(\xi, \xi^*)\}.$$

If $\hat{\xi}$ is such that

$$\sup_{\xi} E\{R(\xi, \hat{\xi})\} \leq E\{R(\xi, \xi^*)\},$$

then

$$\sup_{\xi} n(\xi, \hat{\xi}) \geq N_0 .$$

If the restrictions imposed above on R are satisfied and if the sample must always be of given size N, the above argument still holds when $1/N \leq g$, and shows that the estimate $\bar{x}$ minimizes

$$\sup_{\xi} E\{R(\xi, \hat{\xi})\}$$

with respect to $\hat{\xi}$.

REFERENCES

[1] C. Stein and A. Wald, "Sequential confidence intervals for the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 427–433.

[2] A. Wald, "Statistical decision functions," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 165–205.