The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 193–199

INVARIANT INTERVAL ESTIMATION OF A LOCATION PARAMETER

BY R. S. VALAND

E. R. Squibb and Sons, Inc.

1. Summary and introduction. In the general problem of statistical estimation, it is possible to obtain from a random sample a single estimate, known as a point estimate, of a population parameter. But this estimate is not very meaningful unless associated with a measure of its reliability. One approach to this problem consists of giving a point estimate with its standard error, but this has a major drawback. In using this approach, one typically fails to make an assertion regarding the error involved in estimating the standard error.

In this paper, the problem of interval estimation is considered within the framework of decision theory, in which the cost to the statistician depends on the true value of the parameter and the interval chosen. For example, if θ is the true value of the parameter and (a, b) is the interval chosen, a typical loss function is

$$L(\theta, (a, b)) = h(a - \theta, b - \theta) \quad \text{if} \quad a < \theta < b$$
$$= h(a - \theta, b - \theta) + 1 \quad \text{if} \quad \theta < a \quad \text{or} \quad > b$$

where $h(a, b)$ is defined on $\{(a, b); a \leqq b\}$. Thus, the statistician would like to choose the interval (a, b) small to make the payment $h(a, b)$ small, and yet he wants it large enough to have a good chance of containing θ so that he will not have to pay the "extra" unit.

We consider two methods for finding optimal decision rules for the problem of interval estimation. The first one uses the Bayes principal and the second uses the invariance principle. The invariance principle is available only in decision problems which are invariant under certain transformations. Here we find a form of a best invariant interval estimate for the location parameter, and give certain conditions under which the best invariant interval estimate is minimax. In analogy of Blackwell and Girshick's suggestions, we present a loss function for which a best invariant interval estimate exists but is not minimax. (See the example in Section 4)

Finally, in Section 5, we show that a best invariant interval estimate for a scale parameter of a distribution can be found by transforming the scale parameter problem to the location parameter problem.

2. The principle of invariance. If the decision problem is symmetric, or invariant, with respect to certain operations, then it may seem reasonable to restrict the available rules to be symmetric, or invariant, with respect to those operations also. The principle of invariance involves groups of transformations over the three spaces; the parameter space Θ, the action space $\mathcal{A}$, and the sample

Received 25 1966; revised 28 August 1967.

space $\mathfrak{X}$, an n-dimensional Euclidean space. For definitions and detailed discussion on the P.I. see Lehman's *Testing Statistical Hypothesis.*

We suppose Θ to be the real line and θ to be a location parameter of the observable random variable X with density function $f_X(x, \theta)$, so that $f_X(x, \theta) = f(x - \theta)$ for some density f. In this problem, we choose $\mathcal{A}$ to be the set of all open intervals; $\mathcal{A} = \{(a, b); a \leqq b\}$ and let the loss function be

$$L(\theta, (a, b)) = h(a - \theta, b - \theta) + 1 - I_{(a,b)}(\theta) \tag{2.1}$$

where $h(a, b)$ is defined on $\{(a, b); a \leqq b\}$; for example, $h(a, b) = \int_a^b dH(u)$, where $H(u)$ is non-decreasing in u.

This decision problem is invariant under the group $\mathcal{G}$ of translations $g_{c_1}(x) = x + c_1$ with $\bar{g}_{c_1}(\theta) = \theta + c_1$ and $\tilde{g}_{c_1}(a, b) = (a + c_1, b + c_1)$. Thus the distribution of $g_{c_1}(X)$ given $\bar{g}_{c_1}(\theta)$, a location parameter is invariant. Furthermore, the loss function

$$\begin{aligned} L(\bar{g}_{c_1}(\theta), \tilde{g}_{c_1}(a, b)) &= h(a + c_1 - \theta - c_1, b + c_1 - \theta - c_1) + 1 - I_{(a+c_1, b+c_1)}(\theta + c_1) \\ &= h(a - \theta, b - \theta) + 1 - I_{(a,b)}(\theta) = L(\theta, (a, b)) \end{aligned}$$

and therefore is invariant.

3. Characterization of invariant decision rules. Let $d(X) = (d_1(X), d_2(X))$ be an invariant decision rule; then for all X and c_1, $d_1(X + c_1) = d_1(X) + c_1$ and $d_2(X + c_1) = d_2(X) + c_1$. This implies that

$$d_1(X) = X + b_1 \qquad \text{and} \qquad d_2(X) = X + b_2$$

where $b_1 = d_1(0)$ and $b_2 = d_2(0)$. Therefore, every invariant decision rule has

$$d_b(X) = (X + b_1, X + b_2) \tag{3.1}$$

for some $b = (b_1, b_2)$. Since the risk function for $d_b(X)$ is independent of θ (Blackwell and Girshick)

$$\begin{aligned} R(\theta, d_b) &= E_0\{h(X + b_1, X + b_2) + 1 - P_0(-b_2 < X < -b_1)\} \\ &= \int_{X \in \mathfrak{X}} h(X + b_1, X + b_2) f(X/0)\, dX + 1 - \int_{-b_2}^{-b_1} f(X/0)\, dX. \end{aligned} \tag{3.2}$$

If $R(0, d_b)$ exists for all θ, then any estimate of the form (3.1) is an invariant decision rule.

Among all invariant rules, the rule which minimizes equation (3.2) is obviously best. Suppose there exists (b_1^0, b_2^0) such that

$$L(0, (X + b_1^0, X + b_2^0)) = \inf_{(b_1, b_2)} E_0 L(0, (X + b_1, X + b_2))$$

where the infimum is taken over all (b_1, b_2) for which $E_0 L(0, (X + b_1, X + b_2))$ exists and E_0 is the expectation when $\theta = 0$; then, the rule

$$d_{b^0}(X) = (X + b_1^0, X + b_2^0) \tag{3.3}$$

is a best invariant rule. Thus, we can state the following theorem.

THEOREM 1. *In the problem of estimating a location parameter by interval estimation, with loss function as defined in equation* (2.1), *if* $E_0L(0, (X + b_1, X + b_2))$ *exists for some* (b_1, b_2), *and if there exists a* (b_1^0, b_2^0) *such that*

$$(3.4)\quad E_0L(0, (X + b_1^0, X + b_2^0)) = \inf_{(b_1,b_2)} E_0L(0, (X + b_1, X + b_2))$$

where the infimum is taken over all (b_1, b_2) *for which* $E_0L(0, (X + b_1, X + b_2))$ *exists, then* $d_{b^0}(X) = (X + b_1^0, X + b_2^0)$ *is a best invariant rule, and has a constant risk equal to* $E_0L(0, (X + b_1^0, X + b_2^0))$.

4. The minimax character of the best invariant interval estimates. One usually expects best invariant decision rules to be minimax. The theorem and lemmas below give conditions under which a best invariant interval estimate of a location parameter is minimax. H. Kudō has found such conditions in a slightly different problem of invariant set estimation (H. Kudō, Theorem 2.7) using a more general group of transformations, rather than a group of translations. However, he restricted himself to a loss function (specialized to a group of translations) of the form

$$L(\theta, A) = c(A) + 1 - I_A(\theta)$$

where A is a set (not just an open interval), and $c > 0$. His minimax invariant decision function ϕ is

$$\phi(A_{\text{opt}}(x), x) = 1, \qquad A_{\text{opt}}(x) = \{\theta \varepsilon \Theta; p(\theta/x) > c\},$$

where $p(\theta/x)$ is the conditional density of θ, given X, and with probability one for all $\theta \varepsilon \Theta$, $A_{\text{opt}}(x)$ is given by $p(\theta/x) > c$.

Here we are interested in proving that these estimates are minimax for more general loss functions similar to that found in Blackwell and Girshick for point estimation of a location parameter. (Blackwell and Girshick, Section 11.3, Theorem 11.3.1).

THEOREM 2. *In the problem of interval estimation of a location parameter with a loss function*

$$(4.1)\quad L(\theta, (a, b)) = h(a - \theta, b - \theta) + 1 - I_{(a,b)}(\theta)$$

where $h(a, b) \geqq 0$, *if for every* $\epsilon > 0$, *there exists an* N *such that*

$$(4.2)\quad \int_{-N}^{N} L(0, (X + b_1, X + b_2))\, dF(X/0) \geqq R_0 - \epsilon$$

for all b_1 *and* b_2 $(b_1 \leqq b_2)$, *where*

$$R_0 = \text{Inf}_{(b^1,b^2)} E_0\{L(0, (X + b_1, X + b_2))\}$$

then the best invariant decision rule is minimax.

PROOF. Let $(d_1(y), d_2(y))$ be any decision rule for a location parameter of the observable random variable Y with density function $f(y - \theta)$. We shall, now, exhibit a sequence τ_n of *a priori* distributions and show that $\lim_{n\to\infty} \inf_d r(\tau_n, d) \geqq R_0$; where $d = (d_1(y), d_2(y))$. In other words, we shall show that d is an extended Bayes rule. If a rule with constant risk is an extended Bayes rule, then it is minimax (Blackwell and Girshick [2]). We shall consider the distribution of the parameter to be the uniform distribution over the interval $(-M, M)$, call it τ_M.

Let $\epsilon > 0$, and find N to satisfy condition (4.2). Let $M > N$. Then, since the integrals may be interchanged due to the non-negativity of the integrand, we have

$$(4.3)\quad r(\tau_M, d) = (2M)^{-1} \int_{-M}^{M} \int [h(d_1(y) - \theta, d_2(y) - \theta) + 1 - I_{(d_1(y), d_2(y))}(\theta)]\, dF(y - \theta)\, d\theta.$$

Let $y - \theta = x$,

$$r(\tau_M, d) = (2M)^{-1} \int \int_{-M}^{M} [h(d_1(x + \theta) - \theta, d_2(x + \theta) - \theta) + 1 - - I_{(d_1(x+\theta), d_2(x+\theta))}(\theta)]\, d\theta\, dF(x).$$

Let $\theta = z - x$, then

$$\begin{aligned} r(\tau_M, d) &= (2M)^{-1} \int \int_{x-M}^{x+M} [h(d_1(z) - z + x, d_2(z) - z + x) + 1 \\ &\qquad - I_{(d_1(z), d_2(z))}(z - x)]\, dz\, dF(x) \\ &= (2M)^{-1} \int \int_{z-M}^{z+M} [h(d_1(z) - z + x, d_2(z) - z + x) + 1 \\ (4.4) \qquad &\qquad - I_{(z-d_2(z), z-d_1(z))}(x)]\, dF(x)\, dz \\ &\geqq (2M)^{-1} \int_{(M-N)}^{M-N} \int_{z-M}^{z+M} [L(0, (x + d_1(z) - z, x + d_2(z) - z))] \\ &\qquad dF(x/0)dz \\ &\geqq (2M)^{-1} \int_{(M-N)}^{M-N} (R_0 - \epsilon)\, dz \\ &\geqq (R_0 - \epsilon) 2(M - N)(2M)^{-1}. \end{aligned}$$

The second inequality in equation (4.4) follows, since for $z \varepsilon (-(M - N), (M - N))$, the inside integral is not less than the integral in equation (4.2). The inequality in (4.4) is true for all (including randomized) decision rules. Therefore,

$$(4.5)\quad \liminf_{M\to\infty} \inf_d r(\tau_M, d) \geqq \liminf_{M\to\infty} (R_0 - \epsilon)(2M - 2N)(2M)^{-1} = R_0 - \epsilon$$

for all $\epsilon > 0$, completing the proof.

The following two lemmas give simple sufficient conditions in order that (4.2) hold.

Lemma 1. *If $h(a, b)$ is bounded, then condition* (4.2) *is satisfied.*

Proof. Let $B \geqq h(a - \theta, b - \theta)$ and find N so that

$$P\{|X| > N\} \leqq \epsilon/(B + 1), \quad \text{i.e.,} \quad 1 - \int_{-N}^{N} dF(x) \leqq \epsilon/(B + 1).$$

Then,

$$\int_{-N}^{N} L(0, (x + b_1, x + b_2))\, dF(x/0) = \int_{-\infty}^{+\infty} L(0, (x + b_1, x + b_2))\, dF(x/0)$$
$$- \int_{-\infty}^{-N} L(0, (x + b_1, x + b_2))\, dF(x/0)$$
$$- \int_{N}^{\infty} L(0, (x + b_1, x + b_2))\, dF(x/0)$$
$$\geqq R_0 - (B + 1)P_0\{|X| > N\} \tag{4.6}$$
$$\geqq R_0 - \epsilon. \qquad \text{Q.E.D.}$$

Lemma 2. *If* (1) $h(a, b)$ *and* F *are continous,* (2) $h(a, b) \to \infty$ *as* $b \to \infty$ *or* $a \to -\infty$, *and* (3) $h(a, b)$ *is non-decreasing in* b *for fixed* a, *and* $h(a, b)$ *is non-increasing in* a *for fixed* b, *then condition* (4.2) *is satisfied.*

Proof. The function $R_N(b_1, b_2)$ defined for $b_1 \leqq b_2$ as

$$R(b_1, b_2) = \int_{-N}^{N} [L(0, (x + b_1, x + b_2))]\, dF(x/0)$$
$$= \int_{-N}^{N} [h(x + b_1, x + b_2) + 1 - I_{(x+b_1, x+b_2)}(0)]\, dF(x/0)$$

is (i) non-decreasing in N for each fixed (b_1, b_2), (ii) continuous in (b_1, b_2) since h and F are continuous, and (iii) for sufficiently large N, $R_N(b_1, b_2) \to \infty$ as $b_1 \to -\infty$ or $b_2 \to \infty$. From (iii), there is an N_0 and compact set A in two dimensions such that for (b_1, b_2) outside A, $R_{N_0}(b_1, b_2) \geqq R_0 + 1$. From (i) this inequality is valid for all $N \geqq N_0$. Since $\inf_{(b_1, b_2)} R_N(b_1, b_2) \leqq R_0$, (ii) implies that the infimum of $R_N(b_1, b_2)$ is assumed at some point $(b_1^N, b_2^N) \varepsilon A$ when $N \geqq N_0$. Hence, there exists a limit point (b_1^1, b_2^1). Since $n > N > N_0$ implies

$$R_N(b_1^n, b_2^n) \leqq R_n(b_1^n, b_2^n) \leqq R_0, \tag{4.7}$$

we have $R_N(b_1^1, b_2^1) \leqq R_0$ for all $N > N_0$, but as $N \to \infty$, $R_N(b_1^1, b_2^1)$ converges to something at least R_0. Therefore, from (4.7) $R_n(b_1^n, b_2^n)$ converges to R_0; thus, there exists an N such that $R_N(b_1^N, b_2^N) > R_0 - \epsilon$.

Counter Example. Blackwell and Girshick point out that for some loss functions, best invariant estimates exist and are not minimax. In analogy to their suggestions, we present a loss function for which a best invariant interval estimate exists but is not minimax.

Suppose the random variable X has a distribution function

$$f(X \mid \theta) = 1/(X - \theta)(X - \theta + 1), \qquad X = \theta + 1, \theta + 2, \cdots.$$

Let Θ be the real line, and $\mathfrak{A}$ be the half plane $\{(a, b); a \leqq b\}$, where $a > -\infty$. Let the loss function be

$$L(\theta, (a, b)) = \max\left(\tfrac{1}{2}(b + a) - \theta, 0\right) + 1 - I_{(a,b)}(\theta). \tag{4.8}$$

For this problem, the risk of an invariant rule $(X + b_1, X + b_2)$ is

$$E_0 L(0, (X + b_1, X + b_2)) \tag{4.9}$$
$$\geqq \sum_{X + \frac{1}{2}(b_1 + b_2) > 0} (X + \tfrac{1}{2}(b_1 + b_2))(X(X + 1))^{-1}.$$

For all b_1 and b_2 this series diverges to $+\infty$. Now, consider a non-invariant rule of the following form

$$d(x) = (X - C_a|X|, X - C_b|X|), \qquad \text{where } \tfrac{1}{2}(C_a + C_b) > 1.$$

The risk function for this rule is

(4.10) $R(\theta, d) < \log 2[(\tfrac{1}{2}(C_a + C_b) + 1)/(\tfrac{1}{2}(C_a + C_b) - 1)]$.

Thus, the best invariant rule is not minimax. Moreover, this example does not depend entirely on the fact that all invariant rules have infinite risk.

5. Invariant interval estimation of a scale parameter. We shall find the best invariant rule for a scale parameter of the observable random variable, call it X, for some density, call it f, by transforming the scale parameter problem to a location parameter problem. Let Θ be a real line, and θ be a scale parameter of the observable random variable X. We choose $\mathcal{A}$ to be the set of all open intervals;

$$\mathcal{A} = \{(a, b); a \leqq b\}.$$

Let the loss function be

(5.1) $$L(\theta, (a, b)) = h(a/\theta, b/\theta) + 1 - I_{(a,b)}(\theta)$$

where $h(a, b)$ is defined on $\{(a, b); a \leqq b\}$.

This problem is invariant under the group $\mathcal{G}$ of translations $g_{c_1}(X) = c_1X$, where $c_1 > 0$, with $\bar{g}_{c_1}(\theta) = c_1\theta$ and $\tilde{g}_{c_1}(a, b) = (c_1a, c_1b)$, the distribution of $g_{c_1}(X)$ given $\bar{g}_{c_1}(\theta)$ is the same as the distribution of X given θ, since θ is a scale parameter. Furthermore, the loss function

$$L(\bar{g}_{c_1}(\theta), \tilde{g}_{c_1}(a, b)) = h(c_1a/c_1\theta, c_1b/c_1\theta) + 1 - I_{(c_1a, c_1b)}(c_1\theta)$$
$$= h(a/\theta, b/\theta) + 1 - I_{(a,b)}(\theta)$$

is invariant.

The above problem is identical to the problem of estimating a location parameter $\theta' = \log \theta$ for the distribution of $X' = \log X$ $(X > 0)$ and of $X' = \log(-X)$ $(X < 0)$, and the loss function for the transformed problem is

$$L'(\theta', (a', b')) = L(e^{\theta'}, (e^{a'}, e^{b'})) = L(1, (e^{a'-\theta'}, e^{b'-\theta'})).$$

Let us define $L'(x, y) = L(e^x, e^y)$, then

(5.2) $$L'(\theta', (a', b')) = L'(0, (a' - \theta', b' - \theta'))$$

where $a' = \log a$; $b' = \log b$. By using equation (3.3), the best invariant rule for θ' is

(5.3) $$(X' + b_1'^0, X' + b_2'^0)$$

where $(b_1'^0, b_2'^0)$ is such that

(5.4) $E_0L'(0, (X' + b_1^0, X' + b_2^0)) = \inf_{(b_1', b_2')} E_0L'(0, (X' + b_1', X' + b_2'))$

for all b_1' and $b_2'(b_1' < b_2')$, and E_0 is the expectation when $\theta' = 0$. Hence, the best invariant estimate of log θ is

$$(5.5) \qquad (\log X + \log b_1^0, \log X + \log b_1^0)$$

where (b_1^0, b_2^0) minimizes

$$E(L(1, (e^{X'+b_1'}, e^{X'+b_2'}))/\theta' = 0) = E(L(1, (b_1X, b_2X))/\theta = 1).$$

Thus, the best invariant estimate of θ is (b_1^0X, b_2^0X) where (b_1^0, b_2^0) is that value of (b_1, b_2) which minimizes $E(L(1, (b_1X, b_2X))/\theta = 1)$. Therefore, we can state the following generalization to the above theory:

In the problem of estimating a scale parameter with loss function

$$L(\theta, (a, b)) = h(a/\theta, b/\theta) + 1 - I_{(a,b)}(\theta)$$

if $E_1L(1, (b_1X, b_2X))$ exists for some (b_1, b_2), and if there exists a (b_1^0, b_2^0) such that

$$E_1L(1, (b_1^0X, b_2^0X)) = \inf_{(b_1,b_2)} E_1L(1, (b_1X, b_2X)),$$

where the infimum is taken over all (b_1, b_2) for which $E_1L(1, (b_1X, b_2X))$ exists, and E_1 is the expectation when $\theta = 1$, then $d_b^0(X) = (b_1^0X, b_2^0X)$ is a best invariant rule, and has a constant risk equal

$$E_1L(1, (b_1^0X, b_2^0X)) = E_1\{h(b_1^0X, b_2^0X) + 1 - I_{(b_1^0X, b_2^0X)}(1)\}.$$

Acknowledgment. I should like to thank Professor Thomas S. Ferguson, for his continued guidance and encouragement and Professor F. J. Massey for many helpful suggestions, during the preparation of the dissertation, on which the present paper is based. I should also like to thank the referee for valuable suggestions.

REFERENCES

[1] Blackwell, D. (1951). On the translation parameter problem for discrete variables. *Ann. Math. Statist.* **22** 393–407.

[2] Blackwell, David and M. A. Girshick. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

[3] Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.* **28** 573–601.

[4] Konijn, H. S. (1963). Minimax interval estimates with a shortness criteria. A new formulation. *Ann. Inst. Statist. Math.* **15** 79–81.

[5] Kudō, H. (1955). On minimax invariant estimates of the transformation parameter. *Nat. Sci. Rep. Ochanomizu Univ.* **6** 31–73.

[6] Lehmann, H. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[7] Neyman, J. (1952). *Lectures and conferences on mathematical statistics and probability.* U. S. Department of Agriculture, Washington.

[8] Weiss, Lionel. (1961). *Statistical Decision Theory.* McGraw Hill, New York.