ON THE TRANSLATION PARAMETER PROBLEM FOR DISCRETE VARIABLES[1]

By David Blackwell

Stanford University

Summary. For any chance variable $x = (x_1, \cdots, x_N)$ having known distribution, the translation parameter estimation problem is to estimate an unknown constant h, having observed $y = (x_1 + h, \cdots, x_N + h)$. Extending the work of Pitman [2], Girshick and Savage [1] have, for any loss function depending only on the error of estimate, described an estimate whose risk is a constant R independent of h, and have shown that under certain hypotheses their estimate is minimax. We investigate whether the Girshick-Savage estimate is admissible, i.e., whether it is impossible to find an estimate with risk $R(h) \leq R$ for all h and actual inequality for some h. We consider only bounded discrete variables x, and show that, if all values of x have all integer coordinates and if the loss $f(d)$ from an error d is, for instance, strictly convex and assumes its minimum value, the Girshick-Savage estimate is admissible. Two examples in which the Girshick-Savage estimate is not admissible are given.

1. Preliminaries. Let $r_1, \cdots, r_k$ be distinct points in the hyperplane $\sum_{i=1}^N x_i = 0$ in Euclidean N-space R_N, let s_{ij}, $i = 1, \cdots, k; j = 1, \cdots, m$, be real numbers with $s_{ij} \neq s_{ij'}$ whenever $j \neq j'$, and define $v_{ij} = r_i + \epsilon s_{ij}$, where $\epsilon = (1, 1, \cdots, 1)$. Let $\alpha_i > 0$, $p_{ij} \geq 0$ be numbers such that $\sum_1^k \alpha_i = 1$, $\sum_{j=1}^m p_{ij} = 1$ for each i, and let x be a chance variable such that $P\{x = v_{ij}\} = \alpha_i p_{ij}$. Clearly any N-dimensional chance variable x assuming only a finite number of values can be represented in this way. The translation parameter estimation problem is to estimate the value of an unknown constant h, having observed $y = x + \epsilon h$. An *estimate* for h is then a real valued function $t(y)$, defined for all vectors $y = r_i + \epsilon s$, $i = 1, \cdots, k$, $-\infty < s < \infty$, specifying the estimated value of h as a function of the observation y. We shall suppose that the loss to the statistician depends only on the error $d = t(x + \epsilon h) - h$, and is given by a nonnegative function $f(d)$ defined for all real d. For a given h, the risk, i.e., the expected loss, from an estimate t is

$$R(h) = \sum_{i,j} \alpha_i p_{ij} f[t(v_{ij} + \epsilon h) - h].$$

For any estimate t, the quantity $y - \epsilon t(y) = u(y)$ can be considered as an estimate of the value of x; in terms of u, the absolute value of the error is $|d| = N^{-\frac{1}{2}} |u(y) - x|$, where $|v|$ denotes the length of the vector v. In terms of u, the Girshick-Savage estimate becomes an extremely natural one. If we represent x as $r + \epsilon s$, where the sum of the components of r is 0, $-\infty < s < \infty$, the observation of y determines r, and gives certain information about s which it is

[1] Research performed under an Office of Naval Research contract.

hard to utilize unless one has a priori ideas about h. The Girshick-Savage estimate simply ignores whatever information y contains about s, and makes u a function of r only. If we are given $r = r_i$, the conditional distribution of x is $P\{x = r_i + \epsilon s_{ij}\} = p_{ij}$, and, for $u(r_i) = r_i + \epsilon w$, the conditional risk is $Q_i(w) = \sum_j p_{ij} f(s_{ij} - w)$. If $\inf_w Q_i(w) = R_i$, and W_i is the set of real numbers w with $Q_i(w) = R_i$, the Girshick-Savage estimates are the estimates $u(r)$ such that $u(r_i) = r_i + \epsilon w_i$, with $w_i \in W_i$. The risk from any Girshick-Savage estimate is $R = \sum \alpha_i R_i$ for all h.

Any estimate $u(y)$ is specified by k real functions $z_1(s), \cdots, z_k(s)$: when $y = r_i + \epsilon s$, $u(y) = r_i + \epsilon z_i(s)$; and conversely every set of k functions determines an estimate. The corresponding estimate of h is $t(y) = s - z_i(s)$. The risk is

$$R(h) = \sum_{i=1}^{k} \alpha_i \sum_{j=1}^{m} p_{ij} f[s_{ij} - z_i(s_{ij} + h)].$$

Thus formulated, the N-dimensional estimation problem is simply a collection of k one-dimensional problems, with the particular one-dimensional problem to be faced by the statistician selected according to the probabilities $\alpha_1, \cdots, \alpha_k$. This fact enables us to restrict attention largely to one-dimensional problems.

2. The main result.

We have seen that the risk from a Girshick-Savage estimate is a constant R independent of h. A question raised by Girshick and Savage is whether their estimate is admissible, i.e., whether it is impossible to find another estimate with $R(h) \leq R$ with actual inequality for some h. The theorem of this section gives some conditions under which the Girshick-Savage estimate is admissible; essentially the result is that, for bounded variables x for which all s_{ij} are integers, and strictly convex loss functions $f(d)$ with $f(d) \to \infty$ as $d \to \pm\infty$, the estimate is admissible. Some cases in which the estimate is not admissible are described in the next section. The main result is a consequence of the following lemma, an analogue of which has been obtained by Lehmann [oral communication] for normally distributed x's.

LEMMA. *If all s_{ij} are integers and if $f(d)$ is continuous and such that* (a) *each $Q_i(w) = \sum_j p_{ij} f(s_{ij} - w)$ assumes its minimum R_i at a unique point w_i, i.e., the Girshick-Savage estimate exists and is unique, and* (b) *$Q_i(d_n) \to R_i$ as $n \to \infty$ implies $d_n \to w_i$, then for any estimate $z_1(s), \cdots, z_k(s)$ we have*

$$\lim_{\substack{A \to -\infty \\ B \to -\infty}} \sum_{A}^{B} [R(h) - R] = \sum_{i=1}^{k} \alpha_i \sum_{s=-\infty}^{\infty} [Q_i[z_i(s)] - R_i],$$

where h, s assume only integer values.

PROOF. Since the hypothesis for the N-dimensional problem implies the hypothesis for each of its one-dimensional components, and the conclusion for each component implies the conclusion for the entire problem, it is sufficient to prove the theorem in the one-dimensional case. Suppose, then, that x is one-dimensional, integer-valued, and bounded, say $P\{x = j\} = p_j$, $\sum_{-m}^{m} p_j = 1$, $\min Q(w) = Q(w_1) = R$, where $Q(w) = \sum_{-m}^{m} p_j f(j - w)$, and let z_h be any estimate defined for all integers h. We have

$$R(h) = \sum_{-m}^{m} p_j f(j - z_{j+h}).$$

For any integers A, B with $A \le B$, we have

$$\sum_{A}^{B} R(h) = \sum_{i=A-m}^{B+m} \sum_{j=\max(-m,i-B)}^{\min(m,i-A)} p_j f[j - z_i].$$

For $B - A \ge 2m$,

$$(1) \quad \sum_{A}^{B} R(h) = \sum_{i=A-m}^{A+m-1} \sum_{j=-m}^{i-A} + \sum_{i=B-m+1}^{B+m} \sum_{j=i-B}^{m} + \sum_{i=A+m}^{B-m} \sum_{j=-m}^{m} p_i f[j - z_i].$$

For any set of $2m$ numbers $(u_{-m}, \cdots, u_{m-1}) = u$, define

$$g(u) = \sum_{i=-m}^{m-1} \sum_{j=-m}^{i} p_j f[j - u_i],$$

$$G(u) = \sum_{i=-m}^{m-1} \sum_{j=i+1}^{m} p_j f[j - u_i],$$

so that $g(u) + G(u) = \sum_{i=-m}^{m-1} Q(u_i)$. Then

$$(2) \quad \sum_{A}^{B} R(h) = g(P_{A-m}) + G(P_{B-m+1}) + \sum_{A+m}^{B-m} Q(z_i),$$

where, for any integer a, $P_a = (z_a, \cdots, z_{a+2m-1})$. (2) may also be written as

$$(3) \quad \sum_{A}^{B} R(h) = g(P_{A-m}) - g(P_{B-m+1}) + \sum_{A+m}^{B+m} Q(z_i)$$

or

$$(4) \quad \sum_{A}^{B} R(h) = -G(P_{A-m}) - g(P_{B-m+1}) + \sum_{A-m}^{B+m} Q(z_i).$$

Since $g(P)$ and $G(P)$ are nonnegative for all P,

$$\sum_{A+m}^{B-m} Q(z_i) \le \sum_{A}^{B} R(h) \le \sum_{A-m}^{B+m} Q(z_i),$$

so that

$$\sum_{A+m}^{B-m} (Q(z_i) - R) - 2mR \le \sum_{A}^{B} (R(h) - R) \le \sum_{A-m}^{B+m} (Q(z_i) - R) + 2mR.$$

Now $Q(z_i) \ge R$ for all i. If $\sum_{-\infty}^{\infty} [Q(z_i) - R]$ diverges, then it follows that $\lim_{\substack{A \to -\infty \\ B \to \infty}} \sum_{A}^{B} [R(h) - R] = \infty$. If $\sum_{-\infty}^{\infty} [Q(z_i) - R]$ converges, then, as $i \to \pm\infty$, $Q(z_i)$

$\to R$, $z_i \to w_1$, and $P_i \to (w_1, w_1, \cdots, w_1) = P^*$. Since g is continuous, as $A \to -\infty$, $B \to \infty$, we have $g(P_{A-m}) \to g(P^*)$, $g(P_{B-m+1}) \to g(P^*)$, so that (3) yields

$$\lim_{\substack{A\to-\infty \\ B\to\infty}} \sum_A^B (R(h) - R) = \sum_{-\infty}^{\infty} [Q(z_i) - R].$$

THEOREM 1. *Under the same hypotheses as those of the Lemma, if $z_1(s), \cdots, z_k(s)$ is any estimate with $R(h) \le R$ for all integers h, then $R(h) = R$ for all h, and $z_i(s) = w_i$ for all integers s and all $i = 1, \cdots, k$: the estimate is the Girshick-Savage estimate for integers.*

PROOF. According to the Lemma, we have

$$\sum_{-\infty}^{\infty} [R(h) - R] = \sum_{i=1}^{k} \alpha_i \sum_{s=-\infty}^{\infty} [Q_i(z_i(s)) - R_i],$$

so that both sides are zero, and $R(h) = R$ for all h, $z_i(s) = w_i$ for all s.

When the s_{ij} are restricted to be integral, h may as well also be so restricted, since the statistician can, by considering $y^* = r_i + \epsilon[s]$ when $y = r_i + \epsilon s$, reduce the problem to one where h is replaced by $[h]$. For completeness, however, we prove

THEOREM 2. *Under the same hypotheses as those of the Lemma, the Girshick-Savage estimate is admissible.*

PROOF. Let $z_1(s), \cdots, z_k(s)$ be any estimate for which $R(h) \le R$ for all h. (s, h now assume all real values.) For any h_0, consider the estimate $z_i^*(s) = z_i(h_0 + s)$. Then $R^*(h) = R(h_0 + h) \le R$ for all h. In particular, $R^*(h) \le R$ for all integral h, so that, by Theorem 1, $R^*(h) = R$, $z_i^*(s) = w_i$ for all integers h, s, and all i. Choosing $h = 0$ and $s = 0$ yields $R(h_0) = R$, $z_i(h_0) = w_i$ for all $i: z_1(s), \cdots, z_k(s)$ is the Girshick-Savage estimate.

Remark. The above results are closely related to

THEOREM 3. *Let S be any closed bounded strictly convex subset of N-space which is tangent to the line $x_1 = \cdots = x_N$ at the point $(w, \cdots, w) = P^*$. The only sequence of numbers $\{z_n\}$, $-\infty < n < \infty$, for which each point $P_n = (z_{n+1}, \cdots, z_{n+N}) \in S$ is $z_n \equiv w$.*

Thus, if $f(d)$ is strictly convex and $f(d) \to \infty$ as $d \to \pm\infty$, and $p_i > 0$, $|i| \le m$, the set $\sum_{-m}^{m} p_i f(u_i - i) \le \min_w \sum p_i f(w - i)$ is a closed bounded strictly convex subset of R_{2m+1}, tangent to the line $u_{-m} = \cdots = u_m$ at the point $(w_0, w_0, \cdots, w_0)$, where $\min_w \sum p_i f(w - i)$ occurs at w_0. The theorem then asserts that the only estimate $\{z_n\}$ with $R(h) \le R$ for all h is the Girshick-Savage estimate. The proof of this theorem follows the pattern of the proof of the lemma but is simpler in detail, as follows.

Let $L(x) = \sum_1^N a_i x_i = 0$ be a tangent plane to S at P^* which contains the line $x_1 = \cdots = x_N$; say $L(x) \le 0$ for $x \in S$. For $B - A \ge N$,

$$\sum_A^B L(P_h) = \sum_{A+1}^{B+N} z_i \sum_{\max(1,i-B)}^{\min(N,i-A)} a_j \sum_1^{N-1} b_i z_{A+i} - \sum_1^{N-1} b_i z_{B+i+1}$$
$$= M(P_A) - M(P_{B+1}),$$

where $b_i = \sum_1^i a_j$ and $M(x) = \sum_1^{N-1} b_i x_i$, using the fact that $\sum_1^N a_i x_i = 0$ contains the point $(1, \cdots, 1)$, so that $\sum_1^N a_i = 0$. If all points $P_h \varepsilon S$, then $L(P_h) \geq 0$ for all h. Since M is bounded on S, $\sum_{-\infty}^{\infty} L(P_h)$ converges and, as $h \to \pm\infty$, $L(P_h) \to 0$ and $P_h \to P^*$. Then

$$\sum_{-\infty}^{\infty} L(P_h) = \lim_{\substack{A \to -\infty \\ B \to \infty}} \{M(P_A) - M(P_B)\} = 0, \qquad L(P_h) = 0$$

for all h, so that $P_h = P^*$ for all h.

3. Examples. We present here two one-dimensional examples in which the Girshick-Savage estimate fails to be admissible.

EXAMPLE 1. $x = \pm 1$, each with probability $\frac{1}{2}$, $f(d) = |d|$ for $|d| \geq 1$, $f(d) = 1$ for $|d| > 1$. We have $Q(w) = \frac{1}{2}f(w - 1) + \frac{1}{2}f(w + 1)$, $\min Q(w) = \frac{1}{2} = Q(-1) = Q(1)$. Thus there are two Girshick-Savage estimates: $z = -1$ and $z = 1$, each yielding the constant risk $R = \frac{1}{2}$. (The corresponding estimates for h when $y = n$ is observed are $n + 1$ and $n - 1$.) The estimate $z_n = -1$ for $n < 0$, $z_n = 1$ for $n \geq 0$ (i.e., for h, estimate $n + 1$ if $n < 0$, and $n - 1$ if $n \geq 0$), which is not a Girshick-Savage estimate, has $R(h) = \frac{1}{2}$ for $h \neq -1, 0$, $R(-1) = R(0) = 0$. One can even be frivolous at a single point, setting $z_n = -1$ for $n < 0$, $z_n = 1$ for $n > 0$, $z_0 = 7$, and still obtain $R(h) = \frac{1}{2}$ for $h \neq 0$, $R(0) = 0$, an improvement over the constant risk $\frac{1}{2}$. The extension of either estimate to all h can be made, for instance by defining $z(y) = z_{[y]}$, where $[y]$ denotes the largest integer not exceeding y.

EXAMPLE 2. In Example 1, the pathology occurred in the loss function. We now set $f(d) = d^2$, so that the expected loss is simply the mean square error, and exhibit an x for which the Girshick-Savage estimate is not admissible. Since, for any x, $f(d) = d^2$ will satisfy the hypotheses of Theorem 1, we must go beyond bounded, integer-valued variables.

Let $e_0 = -1$, $e_i > 0$ for $i = 1, \cdots, k$ be $k + 1$ rationally incommensurable numbers (i.e., the only integers n_i such that $\sum_0^k n_i e_i = 0$ are $n_i = 0, i = 0, \cdots, k$), and let $P\{x = -e_i\} = p_i$, with the p_i chosen so that $\sum_0^k p_i = 1$, $p_i e_i =$ constant for $i = 1, \cdots, k$, and $\sum_0^k p_i e_i = 0$. For given e_i, these requirements determine $p_0, \cdots, p_k$ uniquely, and $p_i > 0$ for $i = 0, \cdots, k$. Let S be the additive group determined by $e_0, \cdots, e_k$, i.e., the set of all numbers representable as $\sum_0^k n_i e_i$, where $n_0, \cdots, n_k$ are integers. Then for $h \varepsilon S$, all values of $x + h \varepsilon$ S. We shall define an estimate $z(s)$ for $s \varepsilon S$ for which $R(h) < R$ for all $h \varepsilon S$. The extension of z to all real numbers will be straightforward. For any $z(s)$, the inequality $R(h) \leq R$ becomes

$$\sum_{i=0}^{k} p_i[e_i + z(h - e_i)]^2 \leq \sum_{i=0}^{k} p_i e_i^2,$$

that is,

$$\frac{1}{2} \sum_{i=0}^{k} p_i z^2(h - e_i) \leq -\sum_{i=0}^{k} p_i e_i z(h - e_i),$$

or

$$\frac{1}{2}\left[z^2(h - e_0) + \frac{1}{k}\sum_1^k \frac{z^2(h - e_i)}{e_i}\right] \le z(h - e_0) - \frac{1}{k}\sum_1^k z(h - e_i).$$

If $s = \sum_{i=0}^k n_i e_i$, we define $z(s) = 0$ unless $\sum_0^k n_i = 0$. If $\sum_0^k n_i = 0$, we represent s by the vector $v = (n_1, \cdots, n_k)$. Let $y_1, y_2, \cdots$ be independent vector chance variables, with $P(y = \delta_j) = \frac{1}{k}, j = 1, \cdots, k; i = 1, 2, \cdots$ (where δ_j is the vector with k components of which the jth is 1 and the others are all 0's). Let $z_0(v)$ be the probability that $y_1 + \cdots + y_N = v$, where $N = \sum_1^k n_i$, $z_0(0) = 1$. Then $z_0(v) = 0$ if any $n_i < 0$, and $z_0(v) = N!/(n_1! \cdots n_k! k^N)$ if all $n_i \ge 0$. We shall define $z(v) = a_N z_0(v)$, and choose nonnegative numbers a_N so as to satisfy $R(h) \le R$. This inequality becomes, for $v \ne 0$,

$$\frac{1}{2}\left[a_{N+1}^2 z_0^2(v) + \frac{a_N^2}{k}\sum_1^k \frac{z_0^2(v - \delta_i)}{e_i}\right] \le (a_{N+1} - a_N)z_0(v), \tag{5}$$

where $h - e_0 = \sum_0^k n_i e_i$, $v = (n_1, \cdots, n_k)$, $\sum_1^k n_i = N + 1$, using the fact that $z_0(v) = \frac{1}{k}\sum_{i=1}^k z_0(v - \delta_i)$ for $v \ne 0$. For $v = 0$, the requirement is $\frac{1}{2}a_0^2 \le a_0$, i.e., $a_0 \le 2$. Since $z_0(v) = 0$ when $\sum_1^k n_i \le 0, v \ne 0$, (5) is satisfied for $N < 0$, $v \ne 0$. For $N \ge 0$, let $w_N = \max_{\sum_1^k n_i = N} z_0(v)$. This maximum occurs for the choice of $n_1, \cdots, n_k$, unique except for order, for which $|n_i - n_j| \le 1$ for all i, j, and Stirling's formula yields $\frac{w_N}{N^{\frac{1}{2}(1-k)}} \to c_1$ as $N \to \infty$, where c_1 is a positive constant. Then there is a $c > 0$ with $w_N < c\, N^{\frac{1}{2}(1-k)}$ for $N = 1, 2, \cdots$. Since $z_0(v) = \frac{1}{k}\sum_1^k z_0(v - \delta_i)$ and $z_0 \ge 0$ for all v, we have $z_0(x - \delta_i) \le k z_0(v)$. Thus for nondecreasing a_N, the left member of (5) is less than $d a_{N+1}^2 z_0^2(v)$, where d is a positive constant (for fixed k and e_i). Thus (5) is satisfied if $d a_{N+1}^2 z_0^2(v) \leqq (a_{N+1} - a_N)z_0(v)$, i.e., if $d a_{N+1}^2 z_0(v) \le a_{N+1} - a_N$. Since $z_0(v) < c\,(N + 1)^{\frac{1}{2}(1-k)}$, it is sufficient to choose a_N such that

$$a_{N+1}^2 \le b(a_{N+1} - a_N)(N + 1)^{\frac{1}{2}(1-k)}, \tag{6}$$

where b is a positive constant. If $a_N = N^\epsilon$, $0 < \epsilon < \frac{1}{2}$, and $k > 3$, (6) will be satisfied for sufficiently large N, say $N \ge' N_0$. Setting $a_N = 0$ for $N \le N_0$ and $a_N = N^\epsilon$ for $N > N_0$ satisfies (6) for all N, with actual inequality for sufficiently large N.

Thus, for $k > 3$, we have defined an estimate $z(s)$ for $s \varepsilon S$ with $R(h) \le R$ for all $h \varepsilon S$ and $R(h) < R$ for at least one $h \varepsilon S$, say h_0, where

$$R(h) = \sum_{i=0}^k p_i[e_i + z(h - e_i)]^2.$$

For any $h_1 \varepsilon S$, let $z_1(s) = z(s + h_1)$. Then $R_1(h) = R(h + h_1)$, so that $R_1(h) \leq R$ for all h and $R_1(h_0 + h_1) < R$. Thus for each $h^* \varepsilon S$ there is an estimate $z_{h^*}(s)$ with $R_{h^*}(h) \leq R$ for all h, $R_{h^*}(h^*) < R$. Let a_h be a set of positive numbers with $\sum_{h \varepsilon S} a_h = 1$, and define $z^*(s) = \sum_{h \varepsilon S} a_h z_h(s)$; since the original z is bounded, the series converges. Since $R(h)$ is a convex function of z, $R^*(s) \leq \sum_{h \varepsilon S} a_h R_h(s)$ for $s \varepsilon S$, and $z^*(s)$ is the required estimate.

To extend z^* to all real numbers, divide all real numbers into classes s_α, with y_1 in the same class as y_2 if and only if $y_1 - y_2 \varepsilon S$, and choose a representative t_α of each S_α. Then every y has a unique representation $y = t_\alpha + s$, $s \varepsilon S$; define $z(y) = z^*(s)$. For any $h = t_\alpha + s$,

$$R(h) = \sum_{i=0}^{k} p_i[e_i + z^*(s - e_i)]^2 = R^*(s) < R.$$

Notice that the extension $z(y)$ of z^* is a nonmeasurable (Lebesgue) function of y. It can be shown that any $z(y)$ with $R(h) < R$ for all h is necessarily nonmeasurable; a variation of the method of proof of Theorem 3, evaluating $\int_A^B L(P_h)\, dh$ instead of $\sum_A^B L(P_h)$ over integral h, shows that for any Lebesgue measurable $z(y)$ with $R(h) \leq R$ for all h, the set $R(h) < R$ has Lebesgue measure zero and that, for almost all y, the estimate $z(y) = 0$, agreeing with the Girshick-Savage estimate.

REFERENCES

[1] M. A. Girshick and L. J. Savage, "Bayes and minimax estimates arising from quadratic risk functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951.

[2] E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1939), pp. 391–421.