AN INEQUALITY SUGGESTED BY THE THEORY OF STATISTICAL INFERENCE

Dedicated to Hans Rademacher on the occasion of his seventieth birthday

BY

HENRY B. MANN[1]

We consider Borel regions W in the n-dimensional Euclidean space R_n and two functions F_1, F_2 such that $\int_W dF_1$, $\int_W dF_2$ are nonnegative. The integral $\int_W dF_1 = P_1(W)$ is called the size of W. The integral $\int_W dF_2 = P_2(W)$ will be called the power of W.

A family L of regions in R_n will be called an additive family if sums, intersections, and differences of L are again in L.

In the following definitions and theorems all regions considered will be regions of an additive family L. To avoid cumbersome language we shall however just speak of regions. A most powerful region W will mean a region of the family L such that $P_2(W) \geqq P_2(W')$ for all $W' \in L$ for which $P_1(W) = P_1(W')$. We shall also assume that our regions satisfy the following condition.

(i) *If W is any region of size α, and if $0 \leqq \beta < \alpha$, then W has a subregion of size β.*

Condition (i) obviously implies

(i′) *If W is any region of size α, and if $\alpha = \sum_{i=1}^{n} \beta_i$, then $W = \sum W_i$, where W_i has size $\beta_i \geqq 0$.*

LEMMA 1. *Let W be a most powerful region of size α, and W_1 any subregion of W of size $\beta \leqq \alpha$. Let K be any region of size β and $K \cap W$ empty. Then $P_2(K) \leqq P_2(W_1)$.*

Otherwise the region $W - W_1 + K$ would have size α and higher power than W.

LEMMA 2. *Let W be a most powerful region of size α, and W^* any subregion of W of size $\beta \leqq \alpha$. Let K be any region of size $k\beta$ such that $K \cap W$ is empty; then*

$$kP_2(W^*) \geqq P_2(K). \tag{1}$$

Proof. If $P_2(K) = \infty$, then K must have a subregion K_1 of size $\leqq \beta$ such that $P_2(K_1) = \infty$. By Lemma 1 this implies $P_2(W^*) = \infty$, and (1) holds.

Received April 3, 1961.

[1] This paper grew out of work done by the author in the summer of 1960 at the Mathematics Research Center, U. S. Army, University of Wisconsin, Madison, Wisconsin.

If $P_2(K)$ is finite, choose $\varepsilon > 0$ and arbitrary. Put $n = [1/\varepsilon] + 1$. Divide K into n regions of size $k\beta/n$. In at least one of these regions, say in D_1, we must have $P_2(D_1) \leqq \varepsilon\, P_2(K)$. Hence if $0 \leqq \delta \leqq k\beta/n$, there is a region $D \subset K$ such that $P_1(D) = \delta, P_2(D) \leqq \varepsilon\, P_2(K)$. Now let $\beta = m\delta, m$ integral, $k\beta = km\delta = [km]\delta + \delta_1$ where $\delta_1 < \delta$. Let $W^* = \sum_1^m W_i$, where W_i has size δ. Let $K = \sum K_i + D$, where $P_1(K_i) = \delta$, $P_1(D) = \delta_1$, $P_2(D) \leqq \varepsilon\, P_2(K)$. Let $p = \min P_2(W_i)$. By Lemma 1 we have $p \geqq P_2(K_i)$, whence

$$P_2(K) \leqq [km]p + \varepsilon\, P_2(K) \leqq kP_2(W^*) + \varepsilon\, P_2(K).$$

Since ε was arbitrary, (1) follows.

THEOREM 1. *Let W be a most powerful region of size α. Let $W_1, \cdots, W_t$ be t regions, $P_1(W_i) = \alpha_i$, and let $p_1, \cdots, p_t$ be t nonnegative numbers such that*

$$\sum p_i\, \alpha_i = \alpha, \qquad \sum p_i = 1.$$

Then $\sum_i p_i\, P_2(W_i) \leqq P_2(W)$.

Proof. Suppose first that $W \not\subseteq W_1$, $W_1 \not\subseteq W$. Let

$$P_1(W - W \cap W_1) = \beta_1, \qquad P_1(W_1 - W \cap W_1) = \beta_2 .$$

Suppose $\beta_1 \geqq \beta_2$. Then $P_2(W_1)$ is not decreased if W_1 is replaced by the region $W_1 \cap W + W^*$ where W^* is a subregion of $(W - W \cap W_1)$ of size β_2. If $\beta_1 < \beta_2$, choose $W^* \subseteq W_1 - W \cap W_1$ of size β_1, and replace it by $W - W_1 \cap W$.

Without loss of generality we may therefore assume that either $W_i \supseteq W$ or $W_i \subseteq W$.

Now suppose $W_1 \supseteq W$, $W_2 \subseteq W$. Let $P_1(W_1 - W) = \beta_1$, $P_1(W - W_2) = \beta_2$. If $p_1\beta_1 \geqq p_2\beta_2$, replace W_2 by W, and subtract from $W_1 - W$ a region W^* of size $(p_2/p_1)\beta_2$. Let the two new regions be W_1', W_2'. Then

$$p_1 P_1(W_1') + p_2 P_1(W_2') = p_1 P_1(W_1) + p_2 P_1(W_2),$$

and (Lemma 2)

$$p_1 P_2(W_1') + p_2 P_2(W_2') = p_1 P_2(W_1) + p_2 P_2(W_2) - p_1 P_2(W^*)$$
$$+ p_2 P_2(W - W_2) \geqq p_1 P_2(W_1) + p_2 P_2(W_2).$$

If $p_1\beta_1 < p_2\beta_2$, replace W_1 by W and add a region $W^* \subseteq W - W_2$ of size $p_1\beta_1/p_2$ to W_2.

We can continue this process until either $W_i \supseteq W$ for all W_i or $W \subseteq W_i$ for all W_i, but then the equation $\sum p_i P_1(W_i) = \alpha$ shows that all W_i are of size α, and we can replace them by W. This proves Theorem 1.

COROLLARY TO THEOREM 1. *Let $\{W_i\}$ be an infinite sequence of regions, and $\{p_i\}$ a sequence of nonnegative numbers such that $\sum p_i P_1(W_i) = P_1(W)$ and*

$\sum p_i = 1$, *where* W *is a most powerful region. Then*

$$\sum p_i P_2(W_i) \leqq P_2(W).$$

If $P_2(W) = \infty$, then the corollary is obvious. If $P_2(W) < \infty$, then by Theorem 1 we have for every N

$$\sum_1^N p_i P_2(W_i) \leqq P_2(W);$$

hence $\sum_1^N p_i P_2(W_i)$ converges and does not exceed $P_2(W)$.

THEOREM 2. *Let* S *be a space, and* Q *a probability measure defined over* S. *For every* $z \epsilon S$ *let* $W(z)$ *be a region in* R_n. *Let further*

$$\int P_1(W(z))\,dQ = \alpha,$$

and assume that $\int P_2(W(z))\,dQ$ *exists. If* W *is a most powerful region of size* α, *then*

$$\int P_2(W(z))\,dQ \leqq P_2(W). \tag{2}$$

We may assume that $P_2(W) < \infty$.

If $\alpha = 0$, then $P_1(W(z)) = 0$ for a set of z of Q-measure 1. Hence $P_2(W(z)) \leqq P_2(W)$ except in a set of Q-measure 0, and (2) follows. We may therefore assume that $\alpha > 0$.

For every Q-measurable set S_i let $Q(S_i)$ be the Q-measure of S_i. For every $\varepsilon_1 > 0$, $1 > \varepsilon > 0$ we can find a covering $\{S_i\}$ of S such that

$$\int P_2(W(z))\,dQ = \sum P_2(W(\xi_i))Q(S_i) - \eta_1,$$

$0 \leqq \eta_1 \leqq \varepsilon_1$, $\xi_i \epsilon S_i$, and

$$\alpha = \int P_1(W(z))\,dQ = \sum P_1(W(\xi_i))Q(S_i) + \eta, \qquad |\eta| \leqq \varepsilon\alpha.$$

If η is positive, apply the Corollary to Theorem 1 to a most powerful subregion of W of size $\alpha - \eta$. If η is negative, choose a most powerful region W^* of size $\alpha + \eta$ containing W. Then

$$\int P_2(W(z))\,dQ \leqq \sum P_2(W(\xi_i))Q(S_i) \leqq P_2(W^*) = P_2(W) + P_2(W^* - W)$$
$$\leqq P_2(W) + \varepsilon P_2(W).$$

This proves (2).

In a perfectly analogous manner we can also define a least powerful region and obtain

THEOREM 2a. *Let S be a space, and Q a probability measure defined over S. For every $z \in S$, let $W(z)$ be a region in R_n. Let further*

$$\int P_1(W(z))\, dQ = \alpha,$$

and assume that $\int P_2(W(z))\, dQ$ exists. If W is a least powerful region of size α, then

$$\int P_2(W(z))\, dQ \geqq P_2(W). \tag{2a}$$

The proof of Theorem 2a can be obtained from that of Theorem 2 by minor and fairly obvious changes and may be left to the reader.

Applications

The results we shall derive in this section are well known, but the proofs given in the literature [1] are rather cumbersome. Here we obtain them by simple substitutions in Theorem 2.

THEOREM 3. *Let $F(x)$ be a function with a nonincreasing and nonnegative derivative, defined for $x \geqq a$. Let Q be a distribution over $x \geqq a$, and let E denote mathematical expectation under Q. Then*

$$E(F(x)) \leqq F(E(x)). \tag{3}$$

Proof. In Theorem 2, put $F_1(x) = x - a$, $F_2(x) = F(x) - F(a)$. Let L be the additive class of regions generated by the intervals in $[a, \infty)$. A moment's reflection will show that the region (a, x) is a most powerful region of size $x - a$. Now put $W(z) = (a, z)$. Then Theorem 2 gives: If $E(z - a) = x - a$, then $E(F(z) - F(a)) \leqq F(x) - F(a)$; and this is (3).

If in (3) we put $F(x) = \log x$, we get

$$E(\log x) \leqq \log E(x), \tag{4}$$

or the geometric mean is less than or equal to the arithmetic mean.

If in (3) we put $F_1(x) = x$, $F_2(x) = x^k$, $k < 1$, $a = 0$, then we obtain

$$E(x^k) \leqq (E(x))^k, \tag{5}$$

where Q is any distribution over $[0, \infty)$. If in (5) we put $x = z^v$, $k = \rho/v$, $0 < \rho < v$, we get

$$(E(z^\rho))^{1/\rho} \leqq (E(z^v))^{1/v}, \tag{6}$$

a well-known inequality. If in (6) we put $\rho = 1$ and let Q be a discrete distribution where z takes the value x_i with probability p_i, we get

$$\sum p_i x_i \leqq \left(\sum p_i x_i^v\right)^{1/v}.$$

Putting now $1/v + 1/v' = 1$, $p_i = t_i^{v'} / \sum t_i^{v'}$, $x_i = z_i/t_i^{v'/v}$, we obtain

$$\sum z_i t_i \leqq \left(\sum z_i^v\right)^{1/v} \left(\sum t_i^{v'}\right)^{1/v'}, \tag{7}$$

the Hoelder inequality for $v > 1$, $1/v + 1/v' = 1$. Similarly if in (6) we put $v = 1$, then $\rho < 1$, and an analogous substitution leads to

$$\sum z_i t_i \geqq \left(\sum t_i^{\rho}\right)^{1/\rho}\left(\sum z_i^{\rho'}\right)^{1/\rho'} \tag{7'}$$

for $\rho < 1$, $1/\rho + 1/\rho' = 1$, the Hoelder inequality for $\rho < 1$.

Let $F(x)$ be a monotonic, continuous function in $[0, \infty)$ such that

$$F(x + y) \leqq F(x) + F(y);$$

then $F(x + y) - F(x) \leqq F(y)$, so that, for $F_1(x) = x$, $F_2(x) = F(x)$, the region $[0, y]$ is a most powerful region provided $F(0) = 0$. Hence we get from Theorem 2 for any positive random variable x

$$E(F(x)) \leqq F(E(x)). \tag{8}$$

A similar argument shows, by using Theorem 2a, that $F(0) = 0$, $F(x + y) \geqq F(x) + F(y)$ implies

$$E(F(x)) \geqq F(E(x)). \tag{8'}$$

If we call a function convex if it satisfies (8′) for every positive random variable, and concave if it satisfies (8), we may state

THEOREM 4. *Let $F(x)$ be monotone, continuous, and defined in $x \geqq 0$. Let further $F(0) = 0$. Then $F(x)$ is concave if and only if*

$$F(x + y) \leqq F(x) + F(y). \tag{9}$$

$F(x)$ is convex if and only if

$$F(x + y) \geqq F(x) + F(y). \tag{9'}$$

That the conditions are sufficient has been proved above. To prove their necessity consider a random variable taking two values 0, x, with probabilities p, q respectively. Then (8) gives $q\,F(x) \leqq F(qx)$ for $q \leqq 1$, and so $q\,F(x) \geqq F(qx)$ for $q \geqq 1$. If $y = qx$, $q < 1$, we get

$$F(x + y) = F((1 + q)x) \leqq (1 + q)F(x) = F(x) + q\,F(x)$$
$$\leqq F(x) + F(qx) = F(x) + F(y).$$

The inequality (9′) follows from (8′) in the same way.

Applying the inequality (3) to the function GF^{-1}, where G and F are monotonic functions and GF^{-1} has a decreasing derivative, we get

$$E(GF^{-1}(x)) \leqq GF^{-1}(E(x)),$$

and replacing x by $F(x)$ we get

$$G^{-1}E(G(x)) \leqq F^{-1}(EF(x)). \tag{10}$$

It may be amusing to mention the motivation for the inequality (2).

If we interpret $F_1(x)$, $F_2(x)$ as distribution functions, then W becomes a most powerful critical region for testing the hypothesis that the random variable ξ has the distribution function $F_1(x)$, when the only possible alternative is the distribution $F_2(x)$. The variable z represents a random variable which is unrelated to the hypothesis tested. If we consider a procedure by which critical regions are determined by a random observation on z in such a way that the first equation of Theorem 2 is satisfied, then (2) expresses the fact that the power of our critical region cannot be increased by random experiments which are not related to the hypothesis tested. Thus we have derived Hoelder's inequality from the fact that we cannot increase our knowledge on the milk yield of cows by flipping a coin or by measuring the weight of herrings. However, lest we trust our intuition too much in these matters, we shall show by an example that the condition (i) is indeed necessary for Theorems 1 and 2.

Let ξ be a random variable taking the values 0, 1, 2. Let

$$\begin{aligned} F_1(x) &= 0 \quad \text{for} \quad x < 0, \\ &= \tfrac{1}{4} \quad \text{for} \quad 0 \leqq x < 1, \\ &= \tfrac{3}{4} \quad \text{for} \quad 1 \leqq x < 2, \\ &= 1 \quad \text{for} \quad 2 \leqq x; \end{aligned} \qquad \begin{aligned} F_2(x) &= 0 \quad \text{for} \quad x < 1, \\ &= 1 \quad \text{for} \quad x \geqq 1. \end{aligned}$$

All available regions of size $\frac{1}{4}$ have power 0. We can however get size $\frac{1}{4}$ and power $\frac{1}{2}$ if we flip a coin and choose for W the point 1 whenever a head appears and the point 3 whenever a tail appears.

Reference

1. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, 1952.

Ohio State University
Columbus, Ohio