

THE INTEGRAL OF A SYMMETRIC UNIMODAL FUNCTION OVER A SYMMETRIC CONVEX SET AND SOME PROBABILITY INEQUALITIES

T. W. ANDERSON¹

1. Introduction. If one has a function $f(x)$ on the real line which is symmetric and unimodal (that is, $f(kx) \geq f(x)$, $0 \leq k \leq 1$), it is obvious that the integral of $f(x)$ over an interval of fixed length is maximized if the interval is centered at the origin; in fact, the integral is a non-increasing function of the distance of the midpoint from the origin. A direct result of this is that if a random variable X has density $f(x)$ and Y is an independent random variable, then $\Pr \{ |X| \leq a \} \geq \Pr \{ |X + Y| \leq a \}$. In this paper we generalize these results to n -space. The interval is replaced by a symmetric convex set; the condition of unimodality is expressed by the condition that the set of points for which the function is at least equal to a given value is convex. In turn we show that the probability of a random vector falling in a given symmetric convex set is at least equal to the corresponding probability for the sum of this vector and another. The results are extended to obtain inequalities on the distribution of functionals of stochastic processes, and applications to statistics are made.

2. An inequality on the integral of a symmetric unimodal function over translations of a convex symmetric set in n -space.

THEOREM 1. *Let E be a convex set in n -space, symmetric about the origin. Let $f(x) \geq 0$ be a function such that (i) $f(x) = f(-x)$, (ii) $\{x | f(x) \geq u\} = K_u$ is convex for every u ($0 < u < \infty$), and (iii) $\int_E f(x) dx < \infty$ (in the Lebesgue sense). Then*

$$(1) \quad \int_E f(x + ky) dx \geq \int_E f(x + y) dx$$

for $0 \leq k \leq 1$.

PROOF. An equivalent way of writing (1) is

$$\int_{E+ky} f(x) dx \geq \int_{E+y} f(x) dx,$$

where $E+y$ is the set E translated by the vector y . The theorem follows almost directly after we prove that for every u

Received by the editors May 10, 1954 and, in revised form, July 1, 1954.

¹ Supported by the Office of Naval Research.

$$V\{(E + ky) \cap K_u\} \geq V\{(E + y) \cap K_u\},$$

where $V\{ \}$ indicates the volume of the set. Let $\alpha[(E+y) \cap K_u] + (1-\alpha)[(E-y) \cap K_u]$ denote the set obtained by taking all linear combinations $\alpha z + (1-\alpha)w$, where $z \in (E+y) \cap K_u$ and $w \in (E-y) \cap K_u$ and $0 \leq \alpha \leq 1$. Let $\alpha = (1+k)/2$, so that $\alpha y + (1-\alpha)(-y) = ky$. Then $(E + ky) \cap K_u \supset \alpha[(E+y) \cap K_u] + (1-\alpha)[(E-y) \cap K_u]$ because K_u is convex and $E + ky \supset \alpha(E+y) + (1-\alpha)(E-y) = [\alpha E + (1-\alpha)E] + ky$. Thus

$$V\{(E + ky) \cap K_u\} \geq V\{\alpha[(E + y) \cap K_u] + (1 - \alpha)[(E - y) \cap K_u]\}.$$

$(E+y) \cap K_u$ is the mirror image through the origin of $(E-y) \cap K_u$, and therefore these two sets have the same volume. Then

$$V\{\alpha[(E + y) \cap K_u] + (1 - \alpha)[(E - y) \cap K_u]\} \geq V\{(E + y) \cap K_u\}$$

by the Brunn-Minkowski Theorem [2], which states that $V^{1/n}\{(1-\theta)E_0 + \theta E_1\} \geq (1-\theta)V^{1/n}(E_0) + \theta V^{1/n}(E_1)$ (E_0 and E_1 non-empty, $0 \leq \theta \leq 1$). Thus $V\{(E + ky) \cap K_u\} = H(u) \geq V\{(E + y) \cap K_u\} = H^*(u)$.

Definitions of the Lebesgue and Lebesgue-Stieltjes integrals show

$$\begin{aligned} (2) \quad \int_{E+ky} f(x)dx - \int_{E+y} f(x)dx &= - \int_0^\infty u dH(u) + \int_0^\infty u dH^*(u) \\ &= \int_0^\infty u d[H^*(u) - H(u)]. \end{aligned}$$

Integration by parts shows

$$\begin{aligned} (3) \quad \int_a^b u d[H^*(u) - H(u)] &= b[H^*(b) - H(b)] - a[H^*(a) - H(a)] \\ &\quad + \int_a^b [(H(u) - H^*(u))] du. \end{aligned}$$

Since $f(x)$ has a finite integral over E , $bH(b) \rightarrow 0$ as $b \rightarrow \infty$ and hence also $bH^*(b) \rightarrow 0$ as $b \rightarrow \infty$; therefore the first term on the right in (3) can be made arbitrarily small in absolute value. If $a \geq 0$, the second term above is non negative as well as the third. Thus $\int_0^\infty u d[H^*(u) - H(u)] \geq 0$.

The integral $\phi(y) = \int_E f(x+y)dx$ is a symmetric function and is unimodal in the sense that along a given ray through the origin the integral is a nondecreasing function of the distance from the origin. However, $\phi(y)$ does not necessarily satisfy the condition of unimodality imposed on $f(x)$; that is, $\{y | \phi(y) \geq u\}$ is not necessarily convex.

It will be noticed that we obtain strict inequality in (1) if and only if for at least one u , $H(u) > H^*(u)$ (because $H(u)$ is continuous on the left). For $H(u) = H^*(u)$ we need equality in the Brunn-Minkowski Theorem which implies $(E-y) \cap K_u = (E+y) \cap K_u + z$ for some z ; then the linear combination set is $(E+y) \cap K_u + (1-\alpha)z$. This must be $(E+ky) \cap K_u$ also in order that $H(u) = H^*(u)$. Since the latter set must include the former set z can be taken as $-2y$, and thus the condition is $(E+y) \cap K_u = (E+ky) \cap K_u + (1-k)y$. Since $(E+ky) \cap K_u \supseteq k[(E+y) \cap K_u] + (1-k)E \cap K_u$, the condition (for $k < 1$) becomes $(E+y) \cap K_u = E \cap K_u + y$.

COROLLARY 1. *In Theorem 1 the equality in (1) holds for $k < 1$ if and only if, for every u , $(E+y) \cap K_u = E \cap K_u + y$.*

3. Some probability inequalities. In this section we derive some probability inequalities, and in the next section we make some applications of these to statistics. If $f(x)$ is a probability density, Theorem 1 and Corollary 1 can be stated in probability terms as follows:

COROLLARY 2. *Let X be a random vector with density $f(x)$ such that (i) $f(x) = f(-x)$ and (ii) $\{x | f(x) \geq u\}$ is convex for every u ($0 \leq u < \infty$). If E is a convex set, symmetric about the origin, $\Pr \{X+ky \in E\} \geq \Pr \{X+y \in E\}$ for $0 \leq k \leq 1$. If $h(x)$ is a symmetric function such that $\{x | h(x) \leq v\}$ is convex, then $\Pr \{h(X+ky) \leq v\} \geq \Pr \{h(X+y) \leq v\}$. For $k < 1$, the equality holds if and only if, for every u , $(E+y) \cap K_u = E \cap K_u + y$.*

The second assertion of the corollary says that the cumulative distribution of $h(X+y)$ is bounded by that of $h(X+ky)$ and in particular by that of $h(X)$. A particular density satisfying the conditions is the normal one, $f(x) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \{-\frac{1}{2}x'\Sigma^{-1}x\}$, where Σ is a positive definite matrix.

THEOREM 2. *Let X be a random vector with density $f(x)$ such that (i) $f(x) = f(-x)$ and (ii) $\{x | f(x) \geq u\}$ is convex for every u ($0 \leq u < \infty$). Let Y be independently distributed. If E is a convex set, symmetric about the origin, then*

$$(4) \quad \Pr \{(X+kY) \in E\} \geq \Pr \{X+Y \in E\},$$

for $0 \leq k \leq 1$. If $h(x)$ is a symmetric function such that $\{x | h(x) \leq v\}$ is convex, then $\Pr \{h(X+kY) \leq v\} \geq \Pr \{h(X+Y) \leq v\}$ for $0 \leq k \leq 1$. For $k < 1$, equality holds if and only if $(E+Y) \cap K_u = E \cap K_u + Y$ with probability one for every u .

PROOF. Let the cumulative distribution of Y be $G(y)$. Then the

density of $X+kY$ is $\int_R f(z-ky)dG(y)$, where R denotes the entire n -space. Thus

$$\begin{aligned} \Pr \{X+kY \in E\} &= \int_E \int_R f(z-ky)dG(y)dz \\ (5) \qquad \qquad &= \int_R \int_E f(z-ky)dzdG(y) \\ &= \int_R \int_{E-ky} f(w)dw dG(y). \end{aligned}$$

Then Theorem 1 and Corollary 1 imply the desired results.

Theorem 2 shows that in a certain sense the distribution of $X+Y$ is more spread out than the distribution of $X+kY$ ($0 \leq k \leq 1$).

COROLLARY 3. *Let X be normally distributed with mean 0 and covariance matrix Σ ; let Z be normally distributed with mean 0 and covariance matrix Ψ , where $\Psi-\Sigma$ is positive semi-definite. If E is a convex set, symmetric about the origin, then $\Pr \{X \in E\} \geq \Pr \{Z \in E\}$. If $h(x)$ is a symmetric function such that $\{x | h(x) \leq v\}$ is convex, then $\Pr \{h(X) \leq v\} \geq \Pr \{h(Z) \leq v\}$. If E is a bounded set and $\Psi-\Sigma \neq 0$, then strict inequality holds.*

PROOF. Let Y be normally distributed with mean 0 and covariance matrix $\Psi-\Sigma$. Then Z has the same distribution as $X+Y$, and Corollary 3 follows from Theorem 2.

These inequalities can be useful in statistics. The inequality in Theorem 2 indicates that the cumulative distribution of $h(X+Y)$ is bounded from above by the distribution of $h(X)$. This is useful if it is possible to obtain the distribution of $h(X+Y)$, but not that of $h(X)$. Some examples² of functions that satisfy the conditions are $\max |x_i|$ and $\sum x_i^2$.

These results can be applied to certain functionals of Gaussian stochastic processes.

COROLLARY 4. *Let $X_i(t)$ ($i=1, 2; 0 \leq t \leq T$) be a Gaussian stochastic process with mean 0 and covariance function $EX_i(t)X_i(s) = \sigma_i(t, s)$, jointly continuous ($0 \leq t, s \leq T$). Suppose $\sigma_2(t, s) - \sigma_1(t, s)$ is a positive function. Then*

$$(6) \qquad \Pr \left\{ \int_0^T X_1^2(t)dt \leq v \right\} \geq \Pr \left\{ \int_0^T X_2^2(t)dt \leq v \right\}.$$

² The author wishes to acknowledge Murray Rosenblatt's stimulation of this paper by suggesting this problem for $\max |x_i|$ under conditions of Corollary 3.

If $X_i(t)$ is defined to be a separable process, then

$$(7) \quad \Pr \left\{ \sup_{0 \leq t \leq T} |X_1(t)| \leq v \right\} \geq \Pr \left\{ \sup_{0 \leq t \leq T} |X_2(t)| \leq v \right\}.$$

PROOF. Let $0 = t_0 < t_1 < \dots < t_n = T$ be a partition of $(0, T)$. Then

$$\Pr \left\{ \sum_{j=1}^n X_1^2(t_j)(t_j - t_{j-1}) \leq v \right\} - \Pr \left\{ \sum_{j=1}^n X_2^2(t_j)(t_j - t_{j-1}) \leq v \right\} \geq 0$$

by Corollary 3 (for $\{x | \sum_1^n x_j^2(t_j - t_{j-1}) \leq v\}$ is convex). This implies (6) because $\sum_1^n X_i^2(t_j)(t_j - t_{j-1})$ approaches $\int_0^T X_i^2(t)dt$ (defined as the limit in quadratic mean) in distribution as we take a sequence of partitions becoming dense in $(0, T)$.

In order for the second assertion of Corollary 4 to have meaning, the stochastic processes must be defined in such a way that the probabilities of the sets $\{\sup |X_i(t)| \leq v\}$ exist. The requirement that $X_i(t)$ be Gaussian with mean 0 and covariance function $\sigma_i(t, s)$ defines a class of probability measures on the function space $X(t)$ ($0 \leq t \leq T$); in this class we wish to consider a measure such that the set $\{\sup |X(t)| \leq v\}$ is measurable. Doob [4] has proposed a criterion of separability, which insures that the set is measurable (and that its measure can be approximated by the measure of $\{\max |X(t_j)| \leq v\}$). To prove (7) we note

$$(8) \quad \Pr \left\{ \max_{0 \leq j \leq n} |X_1(t_j)| \leq v \right\} - \Pr \left\{ \max_{0 \leq j \leq n} |X_2(t_j)| \leq v \right\} \geq 0$$

by Corollary 3. Since $\sigma_i(t, s)$ is jointly continuous,

$$\lim_{n \rightarrow \infty} \max_{0 \leq j \leq n} |X_i(t_j)| = \sup_{0 \leq t \leq T} |X_i(t)|$$

with probability one (Theorem 2.2 of Chapter II of [4]). Then (8) implies (7).

COROLLARY 5. Let $X_i(t)$ ($i = 1, 2; 0 \leq t \leq T$) be a Gaussian stochastic process with covariance function $E[X_i(t) - EX_i(t)][X_i(s) - EX_i(s)] = \sigma(t, s)$, jointly continuous ($0 \leq t, s \leq T$). Let $EX_1(t) = kEX_2(t)$ ($0 \leq k \leq 1$). Then (6) and (7) hold when $X_i(t)$ is defined to be a separable process.

PROOF. The proof applies Corollary 1 in the same way that the proof of Corollary 4 applied Corollary 3.

4. Statistical applications. We give one application of Corollary 4 to the asymptotic theory of certain statistical tests of goodness of

fit. Let x_1, \dots, x_N be (scalar) observations from some distribution. To test the hypothesis that this distribution has a specified continuous cumulative distribution function $F(x)$, one has available the criteria $U_N = N^{1/2} \sup |F_N(x) - F(x)|$ and $W_N = \int [F_N(x) - F(x)]^2 dF(x)$ where $F_N(x) = (\text{number of } \{x_i\} \leq x) / N$. It has been shown that the limiting distributions of these criteria when the observations actually are distributed according to $F(x)$ are those of

$$U = \sup_{0 \leq t \leq 1} |Z(t)| \quad \text{and} \quad W = \int_0^1 Z^2(t) dt,$$

where $Z(t)$ is a Gaussian process with mean 0 and covariance function $EZ(t)Z(s) = \min(t, s) - ts = r(t, s)$, say. The distributions of U and W have been tabulated ([5] and [1], respectively). Now consider the problem of testing the hypothesis that x_1, \dots, x_N comes from some member of a one-parameter family of continuous distributions, $F(x, \theta)$. We can use the criteria $U_N^* = N^{1/2} \sup |F_N(x) - F(x, \theta^*)|$ and $W_N^* = \int [F_N(x) - F(x, \theta^*)]^2 dF(x, \theta^*)$, where θ^* is the maximum likelihood estimate of θ . Darling [3] has shown under certain conditions that the limiting distributions of U_N^* and W_N^* are those of $U^* = \sup_{0 \leq t \leq 1} |X(t)|$ and $W^* = \int_0^1 X^2(t) dt$, respectively, where $X(t)$ is a Gaussian process with mean 0 and covariance function $EX(t)X(s) = r(s, t) - \phi(s)\phi(t)$, where

$$\phi(F(x, \theta)) = \frac{\partial}{\partial \theta} F(x, \theta) (\lim E(\theta^* - \theta)^2)^{1/2}.$$

It then follows from Corollary 4 that $\Pr \{U^* \leq a\} \geq \Pr \{U \leq a\}$ and $\Pr \{W^* \leq a\} \geq \Pr \{W \leq a\}$. Thus for U^* , say, we can use the significance point of U and know that under the null hypothesis the probability of rejection is not greater than the assumed significance level.

We can apply Corollary 2 to a problem in the theory of testing hypotheses. Suppose we have a density $f(x - \theta)$, where θ is a vector of (location) parameters, and suppose $f(x) = f(-x)$ and $\{x | f(x) \geq u\}$ is convex for every u . We wish to test the null hypothesis that $\theta = 0$ on the basis of one observation on the random vector X . Let E be a convex set, symmetric about the origin such that

$$(9) \quad \int_E f(x) dx = 1 - \alpha,$$

where α ($0 < \alpha < 1$) is the desired significance level of the test. The test procedure is to reject the hypothesis if the observation falls outside

of E . Then by Corollary 2, this test is unbiased; that is, the probability of rejection under any alternative hypothesis ($\theta \neq 0$) is at least equal to that under the null hypothesis ($\theta = 0$). The test is strictly unbiased if for every $\theta \neq 0$ there is at least one u such that $(E + \theta) \cap K_u \neq E \cap K_u + \theta$.

A particular example of this result is the case where we are sampling from a multivariate normal distribution with mean θ and known covariance matrix. Then the mean of a sample has a normal distribution with mean θ and a known covariance matrix. Then if E is a region for which (9) holds, the corresponding test is unbiased. If E is bounded, the test is strictly unbiased. An extension of the argument shows that the generalized T^2 test (when the covariance matrix is unknown) is strictly unbiased.

REFERENCES

1. T. W. Anderson and D. A. Darling, *Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes*, Ann. Math. Statist. vol. 23 (1952) pp. 193-212.
2. T. Bonnesen and W. Fenchel, *Theorie der konvexen Körper*, New York, Chelsea, 1948.
3. D. A. Darling, *The Cramér-Smirnov test in the parametric case*, Ann. Math. Statist. vol. 26 (1955), in press.
4. J. L. Doob, *Stochastic processes*, New York, Wiley, 1953.
5. N. V. Smirnov, *Table for estimating goodness of fit of empirical distributions*, Ann. Math. Statist. vol. 19 (1948) pp. 279-281.

COLUMBIA UNIVERSITY AND
STANFORD UNIVERSITY