The Annals of Statistics
1980, Vol. 8, No. 4, 833–839

EXTENSIONS OF THE STOCHASTIC ORDERING PROPERTY OF LIKELIHOOD RATIOS[1]

BY GORDON SIMONS

University of North Carolina

Extensions in two directions are described of the well-known stochastic ordering property of likelihood ratios. 1. It is shown that the stochastic ordering property holds in a very general conditional sense. 2. It is shown that, for the usual setting occurring in sequential analysis, sequences of likelihood ratios possess the same stochastic ordering property in a multivariate sense. Applications to sequential analysis, and elsewhere, are described.

1. Introduction. It is well-known that a likelihood ratio $L_n = q_n(X_1, \cdots, X_n)/p_n(X_1, \cdots, X_n)$ for a sample $X_1, \cdots, X_n$ is stochastically larger under the density q_n than under the density p_n. Since this fact is equivalent to the unbiasedness of likelihood ratio tests (in the context of two simple hypotheses), it is a consequence of the Neyman-Pearson fundamental lemma. Here, we are concerned with two possible extensions. The first (Theorem 1) is concerned with the conditional behavior of a likelihood ratio. We show that, conditional on any portion of the information in $X_1, \cdots, X_n$ which one might choose to extract, likelihood ratio tests are still unbiased. The second extension (Theorem 2) is concerned with sequences of likelihood ratios. We show that, under the usual setting occurring in sequential analysis, sequences of likelihood ratios are stochastically larger, in a multivariate sense, when the (simple) alternative hypothesis is true than when the (simple) null hypothesis is correct. This stochastic ordering is an example of the rather strong ordering described in a recent paper by Kamae, Krengel and O'Brien (1977). (It is an interesting and perhaps significant fact that some of the "machinery" developed in their paper, which has potential applicability, fails for the present context.)

The stochastic ordering of likelihood ratios has been exploited in the statistical literature to show that a number of one-sided tests have a monotone power function. One needs to be working with a model that possesses the monotone likelihood ratio property. This approach has been taken, in sequential analysis settings, by Lehmann (1959; Lemma 4, page 101) and J. K. Ghosh (1960); it could have been taken by Lehmann (1959; pages 68 and 69) in a fixed sample size setting but was not because an approach based upon the Neyman-Pearson fundamental lemma gives stronger results. We return to a discussion of Ghosh's and Lehmann's approaches (which are not identical) in Section 4.

Received July 1978; revised March 1979.

[1]This research was supported by the National Science Foundation under Contract No. MCS 75-07556.

AMS 1970 *subject classifications.* Primary 62A10; secondary 62E15, 60G17.

Key words and phrases. Likelihood ratios, stochastic ordering, multivariate, Radon-Nikodym derivative, conditional probability, sample paths, partial ordering.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

Theorem 1 is the subject of Section 2. The setting for Theorem 2 is laid in Section 3, and its statement and proof are given in Section 4. Finally, Section 5 discusses the failure of Proposition 1, in Kamae, Krengel, and O'Brien (1977), to serve as a vehicle for proving Theorem 2. Its objective is to point out the need for a proof of their proposition based upon weaker assumptions.

2. Conditional stochastic ordering of likelihood ratios. In this section, we show that the *conditional* probability of rejecting a simple hypothesis in favor of a simple alternative is greater when the hypothesis is false than when it is true *no matter what information in the sample one chooses to condition upon*. This conclusion follows immediately from the following theorem.

THEOREM 1. *Suppose L is the P-Q likelihood ratio (essentially the Radon-Nikodym derivative dQ/dP) for probability measures P and Q defined on a common measurable space $(\Omega, \mathcal{F})$ and $\mathcal{E}$ is a sub-σ-field of $\mathcal{F}$ on which P and Q are equivalent. Then $Q(L > c|\mathcal{E}) \geqslant P(L > c|\mathcal{E})$* a.s. *for each real value c.*

REMARK. This result would be a trivial extension of the unconditional result were it the case, in general, that L is the $P^{\mathcal{E}}$-$Q^{\mathcal{E}}$ likelihood ratio for $\mathcal{F}$; it is not. (Here $P^{\mathcal{E}}$ and $Q^{\mathcal{E}}$ denote conditional probability measures given $\mathcal{E}$.)

PROOF. For simplicity, we shall assume that P and Q are equivalent on $\mathcal{F}$ so that

$$Q(F) = \int_F L\, dP, \qquad F \in \mathcal{F}. \tag{1}$$

This implies that P and Q are equivalent on $\mathcal{E}$ and that, for any bounded $\mathcal{E}$-measurable random variable Y,

$$\int_E Y\, dP = \int_E Y L_0^{-1} dQ, \qquad E \in \mathcal{E}, \tag{2}$$

where L_0 denotes the P-Q likelihood ratio for $\mathcal{E}$. What we wish to show is equivalent to

$$Q(L > c|\mathcal{E})P(L \leqslant c|\mathcal{E}) \geqslant P(L > c|\mathcal{E})Q(L \leqslant c|\mathcal{E}) \quad \text{a.s.,}$$

which follows from the two inequalities:

$$Q(L > c|\mathcal{E}) \geqslant cP(L > c|\mathcal{E})L_0^{-1} \quad \text{a.s.,} \tag{3}$$

and

$$Q(L \leqslant c|\mathcal{E}) \leqslant cP(L \leqslant c|\mathcal{E})L_0^{-1} \quad \text{a.s.} \tag{4}$$

Since the right-hand side of (3) is $\mathcal{E}$-measurable, (3) can be verified by showing that

$$\int_E Q(L > c|\mathcal{E})dQ \geqslant \int_E cP(L > c|\mathcal{E})L_0^{-1} dQ, \qquad E \in \mathcal{E},$$

which, as a consequence of (2), is equivalent to

$$Q(L > c, E) \geqslant cP(L > c, E), \qquad E \in \mathcal{E},$$

and which, in turn, is immediate from (1). Thus (3) is justified, and, in a like manner, (4) can be justified. □

A multivariate generalization of Theorem 1 is described in Section 4. As far as we can judge, this generalization does not have a statistical interpretation, but it does generalize a lemma in Eisenberg, B.K. Ghosh and Simons (1976) which has some statistical content.

3 Types of ordering. In this section, we introduce the multivariate generalization of stochastic ordering, appropriate for our purpose, by focusing on the special case of two dimensions. The salient new features are present in $\mathbb{R}^2$. (Readers who wish to consider this topic in the abstract setting of a partially ordered Polish space should read the first section of Kamae, Krengel and O'Brien (1977).)

Let P and Q be probability measures on a common measurable space. A random variable U is stochastically larger under Q than under P if

$$Q(U > u) \geqslant P(U > u), \qquad -\infty < u < \infty. \tag{5}$$

One possible generalization of (5) to two random variables U and V is the condition

$$Q(U > u, V > v) \geqslant P(U > u, V > v), \qquad -\infty < u, v < \infty. \tag{6}$$

However, neither this condition nor any variation of it is as satisfactory, for our purposes, as the condition

$$Q((U, V) \in I) \geqslant P((U, V) \in I), \qquad I \in \mathcal{I}, \tag{7}$$

where $\mathcal{I}$ denotes the class of Borel measurable increasing sets in $\mathbb{R}^2$ (sets I such that $(u_1, v_1) \in I$, $u_1 \leqslant u_2$, $v_1 \leqslant v_2 \Rightarrow (u_2, v_2) \in I$). Notice that (7) is stronger than (6) but is equivalent to

$$Q((U, V) \in D) \leqslant P((U, V) \in D), \qquad D \in \mathcal{D},$$

where $\mathcal{D}$ is the class of Borel measurable "decreasing sets" in $\mathbb{R}^2$ (with an obvious meaning), due to the fact that decreasing sets are complements of increasing sets. There are obvious analogs of (7) for random vectors of every dimension.

Now suppose there exists a probability space which admits random vectors (U_1, V_1) and (U_2, V_2) whose distributions agree with those of (U, V) under P and Q, respectively. Further, suppose $U_1 \leqslant U_2$ and $V_1 \leqslant V_2$. Then, clearly, condition (7) must hold for (U, V). More importantly, Strassen (1965), has proven an equivalence. Namely, condition (7) holds, if and only if, there exists a probability space with random vectors (U_1, V_1) and (U_2, V_2) having the properties described. His result generalizes to random vectors of any (countable) dimension (and further). Additional equivalences are described by Kamae, Krengel, and O'Brien (1977).

4. Multivariate stochastic ordering of likelihood ratios. Let P and Q be probability measures on a common measurable space $(\Omega, \mathcal{F})$, and let $\{\mathcal{F}_n\}$ denote a nondecreasing sequence of sub-σ-fields of $\mathcal{F}$. Further, let L_n denote the P-Q likelihood ratio for $\mathcal{F}_n$, $n \geqslant 1$. For simplicity, we shall assume that P and Q are equivalent on each $\mathcal{F}_n$ so that, for each n,

$$Q(E) = \int_E L_n \, dP, \qquad E \in \mathcal{F}_n. \tag{8}$$

Finally, let $\mathcal{I}(\mathbb{R}^n)$ and $\mathcal{I}(\mathbb{R}^\infty)$ denote the measurable increasing sets of $\mathbb{R}^n$ and $\mathbb{R}^\infty$ respectively.

THEOREM 2.

$$Q((L_1, \cdots, L_n) \in I) \geqslant P((L_1, \cdots, L_n) \in I), \qquad I \in \mathcal{I}(\mathbb{R}^n), n \geqslant 1, \tag{9}$$

and

$$Q((L_1, L_2, \cdots) \in I) \geqslant P((L_1, L_2, \cdots) \in I), \qquad I \in \mathcal{I}(\mathbb{R}^\infty). \tag{10}$$

PROOF. We shall use induction to prove (9). According to Proposition 2 of Kamae, Krengel, and O'Brien (1977), (9) implies (10). Suppose $I \in \mathcal{I}(\mathbb{R}^n)$ for some $n \geqslant 1$. Observe that

$$(L_1, \cdots, L_n) \in I \quad \text{and} \quad (L_1, \cdots, L_{n-1}, 1) \notin I \Rightarrow L_n > 1.$$

Thus, it follows from (8) that

$$\begin{aligned} Q((L_1, \cdots, L_n) \in I, (L_1, \cdots, L_{n-1}, 1) \notin I) \\ \geqslant P((L_1, \cdots, L_n) \in I, (L_1, \cdots, L_{n-1}, 1) \notin I). \end{aligned} \tag{11}$$

Likewise, it follows from (8) that

$$\begin{aligned} Q((L_1, \cdots, L_n) \notin I, (L_1, \cdots, L_{n-1}, 1) \in I) \\ \leqslant P((L_1, \cdots, L_n) \notin I, (L_1, \cdots, L_{n-1}, 1) \in I). \end{aligned} \tag{12}$$

Next, we shall need an induction hypothesis to conclude that

$$Q((L_1, \cdots, L_{n-1}, 1) \in I) \geqslant P((L_1, \cdots, L_{n-1}, 1) \in I). \tag{13}$$

The initial case is $n - 1 = 0$, for which (13) is a triviality. Inequality (13) can be established for $n - 1 \geqslant 1$ by making the induction hypothesis that the inequality in (9) holds with n replaced by $n - 1$. This is because the event $[(L_1, \cdots, L_{n-1}, 1) \in I)]$ can be expressed as $[(L_1, \cdots, L_{n-1}) \in I']$ for some increasing set of I' in $\mathcal{I}(\mathbb{R}^{n-1})$. Finally, inequalities (11), (12) and (13) combine to yield the inequality in (9). □

The following corollary is really a theorem, due to J.K. Ghosh (1960), about generalized sequential probability ratio tests. Its proof simply depends upon the observation that the event in question is expressible in the form $[(L_1, L_2, \cdots) \in I]$ for some $I \in \mathcal{I}(\mathbb{R}^\infty)$.

COROLLARY. *Suppose $\{A_n\}$ and $\{B_n\}$ are sequences of constants with $0 \leqslant A_n \leqslant B_n \leqslant \infty$ for $n \geqslant 1$, and that $N = \inf\{n : L_n \notin (A_n, B_n)\}$. Then $Q(N < \infty$ and $L_N \geqslant B_N) \geqslant P(N < \infty$ and $L_N \geqslant B_N)$.*

REMARKS. 1. If $P(N < \infty) = Q(N < \infty) = 1$, this corollary says that the generalized sequential probability ratio test, which chooses P or Q as $L_N \leqslant A_N$ or $L_N \geqslant B_N$, is unbiased. Ghosh uses this result to obtain a more interesting result concerning the monotocity of certain one-sided sequential tests.

2. The proof of this corollary depends upon (10). A more elementary argument, in the sense that it does not depend implicitly upon Proposition 2 of Kamae, Krengel and O'Brien (1977) (whose proof relies upon the theory of weak convergence), can be given by just using (9). All one needs to do is to prove that $Q(N \leqslant n, \lambda_N \geqslant B_N) \geqslant P(N \leqslant n, \lambda_N \geqslant B_N)$ for $n \geqslant 1$, which is easily done. These same inequalities were validated by Ghosh, who used an induction argument that is somewhat more intricate than our proof of (9).

3. A proof of this corollary, in a very restricted setting, appears in Lehmann's book (1959), Lemma 4, page 101). His approach is to show that there exist two sequences of random variables $\{L_{1n}\}$ and $\{L_{2n}\}$, defined on some probability space, with $L_{1n} \leqslant L_{2n}$, $n \geqslant 1$, and with the distributions of these sequences the same as $\{L_n\}$ under P and Q, respectively. By combining (10) with Strassen's (1965) result (described in Section 3), one sees that Lehmann's approach is possible in full generality. Of course, there is no guarantee that the construction of the sequences $\{L_{1n}\}$ and $\{L_{2n}\}$ is sufficiently easy, in general, to make this an appealing approach.

4. Theorem 2 applies to many situations that have nothing to do with generalized sequential probability ratio tests. For instance, it immediately follows that

$$Q(L_n > c_n \text{ i.o.}) \geqslant P(L_n > c_n \text{ i.o.}),$$

regardless of the sequence c_n. (This is well-known in the elementary "i.i.d. context", in which L_n takes the form $\prod_{i=1}^{n} q(X_i)/p(X_i)$, $n \geqslant 1$, with $X_1, X_2 \cdots$ i.i.d. with common density p or q. For such situations, $Q(L_n \to \infty) = P(L_n \to 0) = 1$ except in the trivial instance that $L_1 = 1$ a.s.) Likewise, it immediately follows that the expected number of times L_n exceeds c_n is at least as large under Q as under P.

5. A conditional version of Theorem 2 along the lines of Theorem 1 is possible. Specifically, if $\mathcal{E}$ is a sub-σ-field of $\mathcal{F}_1$ and P and Q are equivalent on $\mathcal{E}$, then

$$Q((L_1, \cdots, L_n) \in I|\mathcal{E}) \geqslant P((L_1, \cdots, L_n) \in I|\mathcal{E}) \text{ a.s.}, \quad I \in \mathcal{I}(\mathbb{R}^n), n \geqslant 1, \tag{14}$$

and

$$Q((L_1, L_2, \cdots) \in I|\mathcal{E}) \geqslant P((L_1, L_2, \cdots) \in I|\mathcal{E}) \text{ a.s.}, \quad I \in \mathcal{I}(\mathbb{R}^\infty). \tag{15}$$

These results generalize Lemma 2 of Eisenberg, B.K. Ghosh and Simons (1976). From the conditional version of Theorem 2, one can easily obtain conclusions such as

$$Q(L_n > c_n \text{ i.o.}|(L_1, \cdots, L_k) \in A) \geqslant P(L_n > c_n \text{ i.o.}|(L_1, \cdots, L_k) \in A),$$

where c_n is an arbitrary sequence, and A is an arbitrary Borel set in $\mathbb{R}^k$. It is even possible to replace the condition $(L_1, \cdots, L_k) \in A$ by the more general condition $(L_1, L_2, \cdots) \in B$, where B is a Borel set in $\mathbb{R}^\infty$. Stated more precisely and more generally, if $\mathcal{E}$ is any sub-σ-field of $\bigvee_{n=1}^{\infty}\mathcal{F}_n$, on which P and Q are equivalent, and if I is an increasing set in $\mathcal{G}(\mathbb{R}^\infty)$ such that the event $[(L_1, L_2 \cdots) \in I]$ belongs to the tail σ-field of $\{L_n\}$, then

$$Q((L_1, L_2, \cdots) \in I|\mathcal{E}) \geqslant P((L_1, L_2, \cdots) \in I|\mathcal{E}) \quad \text{a.s.}$$

5. Discussion. The nesting of the σ-fields $\mathcal{F}_n$ is crucial for the validity of Theorem 2, as the following counter-example shows:

COUNTER-EXAMPLE 1. Let $P(X_1 = X_2 = i) = \frac{1}{2}$ for $i = 0$ and 1 and $Q(X_1 = 0, X_2 = 1) = Q(X_1 = 1, X_2 = 0) = Q(X_1 = X_2 = 1) = \frac{1}{3}$. Further, let $\mathcal{F}_1 = \sigma(X_1)$ and $\mathcal{F}_2 = \sigma(X_2)$. Then $L_i = \frac{2}{3}$ or $\frac{4}{3}$ as $X_i = 0$ or 1, $i = 1, 2$, and

$$\begin{aligned} Q(L_1 > 1, L_2 > 1) &= Q(X_1 = X_2 = 1) < P(X_1 = X_2 = 1) \\ &= P(L_1 > 1, L_2 > 1). \end{aligned}$$

This means that (9) would not hold for $n = 2$ (in fact, for no $n \geqslant 2$). □

If the assumptions of Proposition 1 of Kamae, Krengel and O'Brien were weaker, it would provide an alternative route for verifying (9) and, hence, Theorem 2. The assumption which causes the problem, when translated to our context, is the following: For $n \geqslant 1$, $s^{(n)} = (s_1, \cdots, s_n)$, $t^{(n)} = (t_1, \cdots, t_n)$ and $s^{(n)} \leqslant t^{(n)}$ (component-wise),

$$(16) \quad Q(L_{n+1} > c|(L_1, \cdots, L_n) = t^{(n)}) \geqslant P(L_{n+1} > c|(L_1, \cdots, L_n) = s^{(n)}), \quad 0 < c < \infty.$$

But (16) is to strong:

COUNTER-EXAMPLE 2. Let the joint distributions of X_1 and X_2 under P and Q be defined by the following table:

(i,j)	$P(X_1 = i, X_2 = j)$	$Q(X_1 = i, X_2 = 1)$
(0, 0)	7/12	1/12
(0, 1)	1/12	1/4
(1, 1)	1/3	2/3

Further, let $\mathcal{F}_1 = \sigma(X_1)$ and $\mathcal{F}_2 = \sigma(X_1\ X_2)$. Then $L_1 = 1/2$ or 2 as $X_1 = 0$ or 1; $L_2 = 1/7$, 3 or 2 as $(X_1, X_2) = (0, 0), (0, 1)$ or $(1, 1)$; and

$$Q(L_2 > 2|L_1 = 2) = 0 < \tfrac{1}{8} = P(L_2 > 2|L_1 = \tfrac{1}{2}),$$

which contradicts (16) for realizable values of $s^{(1)}$ and $t^{(1)}$, $s^{(1)} < t^{(1)}$.□

Note that (16) *does* hold for $s^{(n)} = t^{(n)}$; i.e., for $n \geqslant 1$,

$$(17) \quad Q(L_{n+1} > c|L_1, \cdots, L_n) \geqslant P(L_{n+1} > c|L_1, \cdots, L_n), \qquad 0 < c < \infty,$$

on account of Theorem 1. Unfortunately, (17) (for $n \geqslant 1$) and the other assumption required for Proposition 1, namely that

$$(18) \qquad Q(L_1 > c) \geqslant P(L_1 > c), \qquad 0 < c < \infty,$$

are together, not strong enough to prove (9):

COUNTER-EXAMPLE 3. Let U and V be Bernoulli random variables whose joint distributions under P and Q are implicitly specified by: $P(U = 0) = Q(U = 1) = Q(V = 1|U = 0) = P(V = 1|U = 0) = Q(V = 0|U = 1) = P(V = 0|U = 1) = \frac{2}{3}$. Identify U with L_1, V with L_2, and check that (17) and (18) hold when $n = 1$. Consider the increasing set $I = \{(u, v): v \geqslant 1\}$ in $\mathcal{I}(\mathbb{R}^2)$ and check that $Q((U, V) \in I) = \frac{4}{9}$, which is less than $P((U, V) \in I) = \frac{5}{9}$. Thus (17) and (18) can not imply (9).

REFERENCES

[1] EISENBERG, B., GHOSH, B. K. and SIMONS, G. (1976). Properties of generalized sequential probability ratio tests. *Ann. Statist.* **4** 237-251.

[2] GHOSH, J. K. (1960). On the monotonicity of the OC of a class of sequential probability ratio tests. *Calcutta Statist. Assoc. Bull.* **35** 139-144.

[3] KAMAE, T., KRENGEL, U. and O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probability* **6** 899-912.

[5] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[6] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423-439.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
321 PHILLIPS HALL 039A
CHAPEL HILL, NORTH CAROLINA 27514