When $u = 1$, $\theta = \theta'$ and

$$E_{\theta'}(n) = \frac{ab}{E_{\theta'}[A(x) + cB(x)]^2}. \tag{2.18}$$

The author wishes to express his thanks to L. J. Savage for his interest in this problem and his many useful suggestions.

REFERENCES

[1] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, Inc., New York, 1954.
[2] A. Wald, *Sequential Analysis*, John Wiley and Sons, Inc., New York, 1947.
[3] L. J. Savage, "When different pairs of hypotheses have the same family of likelihood-ratio," this issue of these *Annals*.
[4] M. A. Girshick, Contributions to the Theory of Sequential Analysis, II, III, *Ann. Math. Stat.* Vol. 17 (1946), pp. 282–298.

WHEN DIFFERENT PAIRS OF HYPOTHESES HAVE THE SAME FAMILY OF LIKELIHOOD-RATIO TEST REGIONS[1]

By Leonard J. Savage

University of Chicago

Blasbalg [1], in this issue of these *Annals*, shows that certain families of distributions are especially simple, or degenerate, from the point of view of sequential tests. The main object of this note is to show briefly that these are (at least practically) the only families thus degenerate; some preliminary and related conclusions are also demonstrated.

Let F and G be a pair of probability measures on a space X with elements x, and let ℓ be the logarithm of the likelihood ratio of F with respect to G. ℓ is of course defined only mod $(F + G)$, that is, only up to sets simultaneously of F and G measure 0. If x_i is a sequence of values of x, then a likelihood-ratio critical region in X^n is defined by

$$R(A, n) = \left\{(x_1, \cdots, x_n) : \sum_1^n \ell(x_i) \leqq A\right\}. \tag{1}$$

The innocuous ambiguity of ℓ of course induces corresponding ambiguity in R.

This family of sets R is simplest to study when the distribution of ℓ is non-

Received May 3, 1957.

[1] Research was carried out at the Statistical Research Center, University of Chicago, under sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

I wish to thank H. Blasbalg for his help in coordinating his paper [1] and this note, and I thank W. H. Kruskal for many suggestions. Before writing this, I had the privilege of seeing a related manuscript based on an idea of M. A. Girshick.

atomic under both F and G and when the essential range of ℓ is $(-\infty, \infty)$ under both F and G, less rigorously, when ℓ has no plateaus and takes on almost all values. It is henceforth assumed that F and G are such that ℓ has these properties. Consequently, if $R(A, n) = R(B, n)$, for some $n \geqq 1$, then $A = B$.

Now introduce a second pair of distributions F' and G' also on X, together with their ℓ' and their critical regions R', and subject F' and G' to the same conditions as F and G.

When can some R' be the same as some R? The answers for $n = 1$ and 2 are easy but uninteresting and will be deferred.

THEOREM 1. *If (for some A and A', some $n \geqq 3$, and some representation ℓ and ℓ' of the logarithms of the likelihood ratios) $R(A, n) = R'(A', n)$, then $\ell' = \alpha\ell + \beta$ for some $\alpha > 0$ and for some β, and $A' = \alpha A + n\beta$. Conversely, if $\ell' = \alpha\ell + \beta$, for some $\alpha > 0$, then $R(A, n) = R'(\alpha A + n\beta, n)$ for all A and n, $n \geqq 1$.*

PROOF. The second part of the theorem is obvious.[2] The full proof of the first part will be clear from the proof for $n = 3$.

If $\ell(u) < \ell(v)$, then $\ell'(u) < \ell'(v)$. Indeed, if $\ell(u) < \ell(v)$, there clearly are x and y such that

$$\ell(x) + \ell(y) + \ell(u) \leqq A, \qquad \ell(x) + \ell(y) + \ell(v) > A. \tag{2}$$

This implies the corresponding relations for ℓ' and A', which in turn imply that $\ell'(u) < \ell'(v)$. It follows that $\ell'(x)$ is a strictly increasing function, $t[\ell(x)]$, of $\ell(x)$. From the density of the range of t, it follows that t has a continuous and one-to-one extension to all the real numbers. (Note that this line of argument applies even if $n = 2$.)

Now, working with this extension of t, note that

$$a + b + c = A \quad \text{if and only if} \quad t(a) + t(b) + t(c) = A', \tag{3}$$

whence (making essential use of the fact that $n > 2$)

$$t(a + b) + t(0) + t(c) = A' \quad \text{if and only if} \quad t(a) + t(b) + t(c) = A', \tag{4}$$

and therefore

$$[t(a + b) - t(0)] = [t(a) - t(0)] + [t(b) - t(0)]. \tag{5}$$

This shows that the strictly increasing and continuous function t is of the form $\alpha\ell + \beta$ with $\alpha > 0$. Finally, in view of (3), $A' = \alpha A + 3\beta$, and the proof for $n = 3$ is complete.

COROLLARY 1. *If the hypothesis of the first part of Theorem 1 holds with fixed A and A' for two different values of n (of which only one need be as great as 3), then $\ell = \alpha\ell' \bmod (F + G + F' + G')$, with $\alpha > 0$, for all representations of ℓ and ℓ'.*

If $\ell = \alpha\ell' \bmod (F + G + F' + G')$ for some $\alpha > 0$, call (F, G) parallel to (F', G'). Parallelism is obviously an equivalence relation.

---

[2] No condition on the distribution of ℓ and ℓ' is needed for such "sufficiency" conclusion as this.

The techniques used in proving Theorem 1 lead easily to the following conclusions for $n = 1$ and 2. $R(A, 1) = R'(A', 1)$ if and only if $\ell(x) \leqq A$ where and only where $\ell'(x) \leqq A'$. $R(A, 2) = R'(A', 2)$ if and only if $\ell'(x) \equiv t[\ell(x)]$ with t continuous, strictly increasing and subject to the identity $t(\ell) + t(A - \ell) = A'$. Another way to describe these functions is to prescribe that $t(\ell) = q(\ell - \frac{1}{2}A) + \frac{1}{2}A'$, where q is strictly monotone, continuous, and antisymmetric.

A sequential likelihood-ratio test for F and G (together with a determination of ℓ) defines three critical sets S, T, U in each X^n, which can be described formally thus:

$$\ell_m = \sum_1^m \ell(x_j), \tag{6}$$

$$S(A, B; n) = \{(x_1, \cdots, x_n): B \leqq \ell_m \leqq A, m < n; B > \ell_n\}, \tag{7}$$

$$T(A, B; n) = \{(x_1, \cdots, x_n): B \leqq \ell_m \leqq A, m \leqq n\}, \tag{8}$$

$$U(A, B; n) = \{(x_1, \cdots, x_n): B \leqq \ell_m \leqq A, m < n; \ell_n > A\}. \tag{9}$$

It can happen that, for each pair A, B with $B < A$, there exist A' and B' such that $S(A, B; n) = S'(A', B'; n)$, $T(A, B; n) = T'(A', B'; n)$, and $U(A, B; n) = U'(A', B'; n)$ for all n. It is clearly sufficient for this that (F, G) be parallel to (F', G'). Parallelism is also necessary (even if n is confined to the range 1 and 2) as the next two paragraphs prove.

Studying $n = 1$, you see that A' is determined by A alone and B' by B. Continuing with $n = 1$, if $\ell(u) < \ell(v)$ it follows that $\ell'(u) < \ell'(v)$. Therefore ℓ' is a strictly increasing function t of ℓ, and, in view of the density of the ranges of ℓ and ℓ', t is extendible to a continuous, strictly increasing, function. Also $A' = t(A)$, and $B' = t(B)$.

Now turn to $n = 2$. Consider two real numbers c and d with $d \geqq 0$. Letting $c + d = A$, you see that, since $c \leqq A$, $t(c) + t(d) \leqq A' = t(c + d)$. But, in view of the continuity and strict monotony of t, equality actually obtains. A dual argument leads to the same conclusion if $d < 0$. Therefore, t is linear, homogeneous and increasing, so ℓ and ℓ' are indeed proportional.

The possibility that S, T, U equals S', T', U' only for some *one* quadruple of parameters A, B, A', B' and all n may be of interest, though the answer is a little complicated. The following conditions are almost obviously sufficient: $(A' - B') = \alpha(A - B)$, $\ell' = \alpha\ell + A' - \alpha A$ for $\ell \in [B, A]$, $\ell' = \alpha\ell$ for $\ell \in [B - A, A - B]$, $\alpha > 0$. These conditions are necessary (even if n is confined to the range 1, 2, 3) as the next paragraph proves. Note that, if $[B, A]$ and $[B - A, A - B]$ intersect, as they do in the usual configuration $B < 0 < A$, these conditions simplify to: $A' = \alpha A$, $B' = \alpha B$, and $\ell' = \alpha\ell$ for $\ell \in [B, A] \cup [B - A, A - B]$.

The following facts are easily checked successively. For $\ell \in [B, A]$, $\ell' = r(\ell)$ is an invertible function connecting the values of $\ell \in [B, A]$ and $\ell' \in [B', A']$. r is strictly increasing and has a continuous extension, and $r(A) = A'$, $r(B) = B'$.

Similarly, for $\ell \varepsilon [B - A, A - B]$, $\ell' = t(\ell)$, with t strictly increasing with a continuous extension. t is linear and homogeneous. Finally, r is of the form $\alpha\ell + A' - \alpha A$ *and* $\alpha\ell + B' - \alpha B$.

Now let $F(s)$ be a one-parameter family of probability measures on X parametrized by s and defined by probability densities $f(s)$ with respect to a fixed (σ-finite) measure μ, and let $\ln f(s) = h(s)$. Since the introduction of sequential likelihood-ratio tests, it has often been remarked that certain important families $F(s)$ are especially simple with respect to sequential tests, as has recently been emphasized in [1]. Indeed, it is the special feature of these families that every pair (s, s') is imbedded in a one-parameter family of pairs, say $[s(\theta), s'(\theta)]$ such that each $[F(s(\theta)), F(s'(\theta))]$ resembles $[F(s), F(s')]$ and is in fact parallel to it in the technical sense of this paper. What does this simplicity imply about the structure of the family $F(s)$?

One graphic way to couch the answer is to remark that, for each s, $h(s)$ is a vector in a function space. The family $h(s)$ is a curve in this space. And the technical condition of parallelism is simply that the chord $[h(s) - h(s')]$ be parallel in the ordinary geometrical sense to the chords $[h(s(\theta)) - h(s'(\theta))]$. Thus $h(s)$ needs to be a curve of which every chord is parallel to many others—I am purposely a little vague in order to admit more than one possibly equally interesting interpretation of "many others." This condition is obviously met in a wide sense if $h(s)$ is any plane curve, that is, if $h(s)$ can be represented as

$$h_0 + \eta_1(s)h_1 + \eta_2(s)h_2, \tag{10}$$

where h_0, h_1, h_2 are fixed vectors (that is, real-valued functions of x) and $\eta_1(s)$, $\eta_2(s)$ are real-valued functions of s. For narrower senses, regularity conditions might be imposed on $\eta_1(s)$, $\eta_2(s)$.

Moreover, it is presumably only a plane curve $h(s)$ that can satisfy the condition in any way that would be considered natural. By an unnatural way, I here mean resort to a space filling curve or the like. By a natural way, I mean one with enough regularity to justify something like the following proof that $h(s)$ must be a plane curve.

$$\begin{aligned}[h(s) - h(0)] &= \lambda(s, \Delta s)[h(s + \Delta s) - h(\phi(s, \Delta s))] \\ &= [h(s) - h(0)] \\ &\quad + \Delta s\left\{\frac{\partial\lambda(s, 0)}{\partial\Delta s}[h(s) - h(0)] + \left[\dot{h}(s) - \dot{h}(0)\frac{\partial\phi(s, 0)}{\partial\Delta s}\right]\right\} + o(\Delta s), \\ &\qquad\text{for } s > 0 \text{ but sufficiently small,}\end{aligned} \tag{11}$$

where the dot indicates the derivative with respect to s. Therefore,

$$\frac{d}{ds}[h(s) - h(0)] = -\lambda'(s)[h(s) - h(0)] + \phi'(s)\dot{h}(0), \quad \text{for } s > 0 \text{ but sufficiently small,} \tag{12}$$

using evident abbreviations.

Multiplying through by an arbitrary linear functional, you see that, for given λ' and ϕ', (12) is in effect a collection of first order, linear, ordinary differential equations that can be treated separately. In particular, the whole curve $h(s)$ (for the range where (12) is satisfied) is determined through (12) by the value of h at any one value of s, say for symbolic simplicity, at $s = 1$. But it is easily verified that (12) is solved by an $h(s)$ of the form

$$h(0) + \alpha(s)[h(1) - h(0)] + \beta(s)h^{\cdot}(0), \quad \text{with} \quad \alpha(1) = 1, \qquad \beta(1) = 0.$$

In fact, α and β are obviously determined by the calculation

$$(13) \quad \alpha^{\cdot}(s)[h(1) - h(0)] + \beta^{\cdot}(s)h^{\cdot}(0) = -\lambda'(s)\{\alpha(s)[h(1) - h(0)] + \beta(s)h^{\cdot}(0)]\} + \phi'(s)h^{\cdot}(0),$$

$$(14) \quad \alpha^{\cdot}(s) = -\lambda'(s)\alpha(s),$$

$$(15) \quad \beta^{\cdot}(s) = -\lambda'(s)\beta(s) + \phi'(s).$$

Thus the initial segment of h is a plane curve, and by piecing, this conclusion can be extended to the whole curve.

To summarize, the logarithmic densities of a family that is, so to speak, degenerate with respect to sequential tests can with more than sufficient generality be represented by (10). The corresponding probability densities are of the form

$$(16) \qquad f(x, s) = f_0(x)f_1(x)^{\eta_1(s)}f_2(x)^{\eta_2(s)}.$$

Such families exist in great abundance. The choice of f_0, f_1, and f_2 is nearly arbitrary, subject only to $f_0 \geqq 0$, $f_1, f_2 > 0$ (at least where $f_0 > 0$) and mild integrability conditions. Typically, η_1 and η_2 will, for given f_0, f_1, f_2, be constrained to lie on a convex curve in order that f shall be normalized, that is, integrate to 1. This curve can be parametrized arbitrarily, to complete the construction. See [1] for important examples.

The form (16) has a natural extension to families with two or more parameters. For example, the two-parameter, bivariate, normal family corresponding to two random variables with means 0, equal variances σ^2, and correlation coefficient ρ is of the form

$$f_0(x)f_1(x)^{\eta_1(\rho,\sigma)}f_2(x)^{\eta_2(\rho,\sigma)}f_3(x)^{\eta_3(\rho,\sigma)}.$$

REFERENCE

[1] H. Blasbalg, "Transformation of the fundamental relationships in sequential analysis," this issue of these *Annals*.