We can now establish the desired result.

THEOREM 2. *If X and Y are independent observations on the same unimodal random variable, then $X - Y$ is unimodal.*

We prove the theorem in three parts.

PART I. If X has as possible values only finitely many integers, the theorem is an immediate consequence of the preceding one. The a's are taken to be the probabilities of the successive possible values of X. Since $P(X - Y = k) = S_k$ for k a positive integer, and since $X - Y$ has a distribution symmetric about 0, the theorem follows.

PART II. Let the possible values of X now be numbers of the form $r\Delta$, where $\Delta > 0$ and r is any integer. For simplicity we may assume 0 to be a mode. For every positive integer s, define

$$X'_s = \begin{cases} X \text{ if } |X| \leqq s, \\ 0 \text{ if } |X| > s, \end{cases} \qquad Y'_s = \begin{cases} Y \text{ if } |Y| \leqq s \\ 0 \text{ if } |Y| > s. \end{cases}$$

That $X'_s - Y'_s$ has a unimodal distribution is an immediate consequence of Part I. But since $P(X'_s - Y'_s \neq X - Y) \to 0$ as $s \to \infty$, we see that $X - Y$ must also have a unimodal distribution.

PART III. Now suppose X has a density f, with mode at m. For each positive integer s, define

$$X''_s = [(X - m)\sqrt{s}]/\sqrt{s},$$

where $[u]$ denotes the greatest integer less than u. The cumulative distribution G''_s of $X''_s - Y''_s$ cannot ever differ from G by more than a quantity which tends to 0 as $s \to \infty$. However, G''_s is unimodal, by Part II. If G were not unimodal, we could find $\epsilon > 0$, $\Delta > 0$, and $u - \Delta > 0$ such that $G(u - \Delta) + G(u + \Delta) + \epsilon < 2G(u)$, which would yield a contradiction.

REFERENCE

[1] K. L. CHUNG, "Sur les lois de probabilité unimodales," *C. R. Acad. Sci. Paris*, Vol. 236 (1953), pp. 583–584.

NOTE ON A THEOREM OF LIONEL WEISS[1]

BY LUCIEN LECAM

University of California, Berkeley

1. Introduction. In a recent paper [1] it was pointed out by Lionel Weiss that the class of sequential probability ratio tests is complete in a very strong sense. The purpose of the present note is to show how this result can be derived from a

Received 2/8/54.

[1] This paper was prepared with the partial support of the Office of Naval Research.

slight extension of the usual theorems of decision theory, and to generalize this result to the case where the number of alternatives is any finite number. Similar results could also be obtained in more general cases.

2. Statement of the problem. Let $\{X_n\}$, for $n = 1, 2, \cdots$, be a sequence of mutually independent random vectors. Assume that the distribution of these vectors is given by a sequence of probability measures $\{\pi_{n,j}\}$, for $n = 1, 2, \cdots$, where j is an index taking one of the values $j = 1, 2, \cdots, k$. Suppose that the loss incurred by accepting j while i is true is a finite number W_{ij}, strictly positive if $i \neq j$ and equal to zero if $i = j$.

Let $\mathcal{W}$ be the class of matrices (W_{ij}) satisfying these conditions. If i is the true state of nature, we will assume that the cost of taking n observations is $C_i(n)$, nonnegative strictly increasing in n and such that $C_i(n)$ tends to infinity as n tends to infinity. Let $\mathcal{C}$ be the class of all k-tuples of cost functions $C = \{C_i(n)\}$ satisfying the preceding conditions. Let $\Theta = J \times \mathcal{W} \times \mathcal{C}$, where J denotes the set of integers $J = \{1, 2, \cdots, k\}$. For a particular decision function δ and a particular point $\theta = \{i, W, C\} \in \Theta$, let $R(\theta, \delta)$ denote the risk if δ is used for the state of nature i, the loss function W, and the cost function C.

If D is any subset of the set $\mathfrak{D}$ of all measurable decision procedures such that

(1) D contains all $\delta \varepsilon \mathfrak{D}$ which minimize linear combinations of the form

$$K(\mu, \delta) = \sum_{\xi=1}^{m} \mu_\xi K(\theta_\xi, \delta) \qquad \mu_\xi > 0; \quad \sum_{\xi=1}^{m} \mu_\xi = 1;$$

(2) D is compact in the sense defined in [2];

Then it follows from a modification ([2] Theorem (5)) of a theorem of Wald ([3] Theorem 3.18)) that D is essentially complete. This means that whatever $\delta_0 \varepsilon \mathfrak{D}$, there exists $\delta_1 \varepsilon D$ such that $R(\theta, \delta_1) \leqq R(\theta, \delta_0)$ for every $\theta \varepsilon \Theta$. Such an essentially complete class is described below.

3. Description of the complete class. Let Δ be the set of probability distributions on J. Let Z_n be the vector $Z_n = \{X_1 \cdots X_n\}$ and let $q(Z_{n,p}) = \{q_j(Z_{n,p})\}$ be the vector representing the a posteriori distribution of $i \varepsilon J$ given Z_n, when the a priori distribution of $i \varepsilon J$ is p. Let p be fixed. Consider the class D of all decision functions defined in the following way. For each $n \geqq 0$ choose k closed convex sets $\{S_{n,j}\}$ with $j = 1, 2, \cdots, k$, each contained in Δ, with disjoint interiors and such that $S_{n,j}$ contains q if $q = \{q_i\}$ with $q_j = 1$.

The decision function δ consists of the following rule: if $q(Z_{n,p}) \varepsilon S_{n,j}$, then stop and accept j; if $q(Z_{n,p})$ is not a member of $\bigcup_j S_{n,j}$, then take one more observation; if $q(Z_{n,p})$ is a limit point of one or many $S_{n,j}$, randomize appropriately.

It is clear that the preceding description uses characteristics not depending explicitly on p, so that p may be fixed and, for instance, taken equal to the uniform distribution on J. For this class D the following theorem holds.

THEOREM. *There exists on $\mathfrak{D}$ a topology $\mathfrak{J}$ for which* (1) *$R(\theta, \delta)$ is lower semicontinuous in δ for each $\theta \varepsilon \Theta$ and* (2) *$\mathfrak{D}$ is compact and D is a closed, compact, subset of $\mathfrak{D}$.*

If δ_0 is any measurable decision function $\delta \varepsilon \mathfrak{D}$, there exists a $\delta_1 \varepsilon D$ such that $R(\theta, \delta_0) \leqq R(\theta, \delta_1)$, whatever may be $\theta \varepsilon Z$.

PROOF. A topology $\mathfrak{I}$ having the desired properties has been defined more generally [2] by a process analogous to the one used by Wald [3] for the definition of regular convergence. A classical theorem (see [4], Vol. 1, p. 246; Vol. 2, p. 21; or [5]) states that the space of closed subsets of a compact metric space is compact for the usual definition of distance between sets. It then follows from the relationship between compactness in this sense and compactness in the sense of $\mathfrak{I}$ (or of regular convergence in [3]) that D is compact. This proves the first part of the theorem.

The second part is an immediate consequence of Theorem (5) in [2], provided we can show that Bayes' solutions belong to D. To show this, let $P_{ij}(\delta)$ be the probability of accepting j if i is true and δ is used, and let $Q_i(n, \delta)$ be the probability of taking at least n observations if i is true and δ is used. Let $\theta = \{i, W, C\}$. Then

$$R_i(W, C, \delta) = R(\theta, \delta) = \sum_j W_{ij}P_{ij}(\delta) + \sum_{n=0}^{\infty} C_i(n)[Q_i(n, \delta) - Q_i(n + 1, \delta)].$$

Consequently,

$$\sum_{\xi=1}^{m} \mu_\xi R(\theta_\xi, \delta) = \sum_{i=1}^{k} p_i R_i(W, C, \delta)$$

for suitable values of W, C, and $\{p_i\}$. Therefore the Bayes' solutions for our problem have the same structure as the Bayes' solutions for the now classical problem in which W and C are fixed. A very slight modification of the argument given by Arrow, Blackwell, and Girshick [6] yields the desired result. This completes the proof of the theorem.

As a particular case, if $\delta_0 \varepsilon \mathfrak{D}$ is such that $\lim_{n\to\infty} Q_i(n, \delta_0) = 0$ for $i \varepsilon J$, the preceding theorem implies that there exists $\delta_1 \varepsilon D$ satisfying:

$$P_{ij}(\delta_0) \geqq P_{ij}(\delta_1), \qquad \text{for every } i, j \varepsilon J; \quad i \neq j;$$

$$Q_i(n, \delta_0) \geqq Q_i(n, \delta_1), \qquad \text{for every } i \varepsilon J; \quad n \geqq 0.$$

If, furthermore, the probabilities $\{\pi_{n,j}\}$ satisfy the condition imposed by Weiss [1], the boundaries of the S_{nj} have measure zero and randomization is unncessary.

4. Remarks.

(1) the technique of enlarging the space of strategies of nature, say Ω, to a product $\Omega \times S$ with S finite has been used systematically by Weiss [7], [8] and Lindley [9]. A more general type of extension is implicitly contained in the assumptions of [2]. The standard form of Bayes' solutions given by Theorem 4.7 of [3], or its generalization, remains usually valid under such modifications of Ω.

(2) The proofs of the optimum character of the sequential probability ratio test given in [6] or [10] also make use of classes of Bayes' solutions obtained by varying W and C. However, in these proofs C remains proportional to a given C_0. This is not sufficient for the present purpose.

REFERENCES

[1] LIONEL WEISS, "Testing one simple hypothesis against another," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 273-281.
[2] L. LECAM, "An extension of Wald's theory of statistical decision functions," unpublished.
[3] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, 1950.
[4] C. KURATOWSKI, "Topologie I," 2nd ed., Warsaw, 1948. "Topologie II," Warsaw, 1950.
[5] G. CHOQUET, "Convergences," Annales de l'université de Grenoble, N.S., Vol. 23 (1947-1948), pp. 57-112.
[6] K. J. ARROW, D. BLACKWELL AND M. A. GIRSHICK, "Bayes and minimax solutions of sequential decision problems," *Econometrica*, Vol. 17 (1949), pp. 213-244.
[7] L. WEISS, "A higher order complete class theorem," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 677-680.
[8] L. WEISS, "Sequential procedures that control the individual probabilities of coming to the various decisions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 779-784.
[9] D. V. LINDLEY, "Statistical inference," *J. Roy. Stat. Soc. Series B*, Vol. 15, No. 1 (1953), pp. 30-76.
[10] A. WALD AND J. WOLFOWITZ, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 326-339.

THE DISTRIBUTION OF DISTANCE IN A HYPERSPHERE

BY R. D. LORD

The Royal Technical College, Glasgow

1. In a note with the above title, Hammersley [2] has used ad hoc methods to deal with the distribution of the distance AB, when A and B are points uniformly distributed in a sphere of radius a in s dimensions. I show here how this question may be treated by general methods which I have developed elsewhere [3] for random vectors with spherical distributions. A random vector $\mathbf{r}$ will be said to have a *spherical distribution* if its probability function is a function of $|\mathbf{r}|$ only.

I start with the observation that the problem is in fact one of the addition of independent random vectors with spherical distributions. We require the distribution of $\mathbf{r}_1 - \mathbf{r}_2$ where $\mathbf{r}_1$ and $\mathbf{r}_2$ are random vectors with the same uniform spherical distribution. But on account of the spherical symmetry, $-\mathbf{r}_2$ has the same distribution as $\mathbf{r}_2$, so that the problem is equivalent to finding the distribution of $\mathbf{r}_1 + \mathbf{r}_2$. It will be dealt with in this form in what follows.

2. The first method uses the polar form of the characteristic function. For any spherical distribution in s dimensions let

$$P(r)\,dr = Pr\{r < |\mathbf{r}| < r + dr\}.$$

Received 10/5/53, revised 4/12/54.