NOTES

This section is devoted to brief research and expository articles and other short items.

ON A THEOREM OF LYAPUNOV

By David Blackwell

Howard University

The purpose of this note is to point out two extensions of the following theorem of Lyapunov[1], and to note an interesting statistical consequence of each.[2]

Lyapunov's Theorem: *Let $u_1, \cdots, u_n$ be non-atomic[3] measures on a Borel field $\mathcal{B}$ of subsets of a space X. The set R of vectors $[u_1(E), \cdots, u_n(E)]$, $E \in \mathcal{B}$, is convex, i.e., if $r_1, r_2 \in R$, so does $tr_1 + (1 - t)r_2$ for $0 \leq t \leq 1$.*

Extension 1. *Let $u_1, \cdots, u_n$ be non-atomic measures on a Borel field of subsets of a space X and let A be any subset of n-dimensional Euclidean space. Let $f = a(x) = [a_1(x), \cdots, a_n(x)]$ be any $\mathcal{B}$-measurable function defined on X with values in A, and define $v(f) = [\int a_1(x)\, du_1, \cdots, \int a_n(x)\, du_n]$. The set of vectors $v(f)$ is convex.*

Lyapunov's theorem is the special case in which A consists of two points $(0, \cdots, 0)$ and $(1, \cdots, 1)$.

Proof. Let $v(f_i) = v_i$, $f_i = [a_{i1}(x), \cdots, a_{in}(x)]$, $i = 1, 2$, and consider the $2n$-dimensional measure

$$w(E) = \int_E a_{11}(x)\, du_1 \cdots, \int_E a_{1n}(x)\, du_n, \int_E a_{21}(x)\, du_1, \cdots, \int_E a_{2n}(x)\, du_n.$$

Since $w(N) = (0, \cdots, 0)$ where N is the null set, $w(X) = (v_1, v_2)$, for any t, $0 < t < 1$, there is, by Lyapunov's theorem, a set $E \in \mathcal{B}$ with $w(E) = (tv_1, tv_2)$,

---

[1] "Sur les fonctions-vecteurs complètement additives," *Bull. Acad. Sci. URSS. Sér. Math.* Vol. 4 (1940), pp. 465–478. For a simplified proof of Lyapunov's results, see Halmos, "The range of a vector measure," *Bull. Amer. Math. Soc.*, Vol. 54 (1948), pp. 416–421.

[2] Since this note was submitted, results obtained earlier by Dvoretzky, Wald, and Wolfowitz have appeared in the April 1950 *Proceedings of the National Academy of Sciences*. Their results are closely related to those presented here, and anticipate the general conclusion reached here: that in dealing with non-atomic distributions, mixed strategies are unnecessary. Their principal tool is also an extension of Lyapunov's theorem; their extension does not appear to contain or be contained in either of the extensions given here. The situation considered here is more general in that an infinite number of possible terminal actions are possible, but more restricted in that only mixtures of a finite number of pure strategies are considered here.

[3] A measure u is non-atomic if every set of non-zero measure has a subset of different non-zero measure.

so that $w(CE) = [(1 - t)v_1, (1 - t)v_2]$. Define $f = f_1$ on E, $f = f_2$ on CE. Then $v(f) = tv_1 + (1 - t)v_2$. This completes the proof.

This extension may be reformulated using statistical language, in the special case where $u_1, \cdots, u_n$ are probability measures, as follows: *In a statistical decision problem in which there are only a finite number of possible distributions, each of which is non-atomic, mixed strategies on the part of the statistician are unnecessary: anything which can be achieved with mixed strategies can already be achieved with pure strategies.*

In amplification, $u_1, \cdots, u_n$ are probability distributions, and x is an observation chosen according to one of them. Having observed x, the statistician must choose an action d from a set D of possible actions. His loss in choosing an action d is $a(1, d), \cdots, a(n, d)$ when the true distribution of x is $u_1, \cdots, u_n$, respectively. Thus the choice of d may be described as choosing a point $a \in A$, the subset of n-dimensional space consisting of the set of loss vectors

$$[a(1, d), \cdots, a(n, d)], d \in D.$$

Of course several points d may lead to the same a. From our point of view, two d's with the same a may be identified, so that it is no loss of generality to consider A itself as the set of possible actions.

A strategy for the statistician is then a function $f = a(x)$ from X into A, specifying the action to be taken (i.e., the loss vector to be chosen) when x is observed. We shall consider only $\mathfrak{B}$-measurable strategies f. The expected loss vector from a strategy f is $v[f] = \int a_1(x)\, du_1, \cdots, \int a_n(x)\, du_n$; the i-th component is the expected loss from f when the true distribution is u_i. Thus the range R of $v(f)$ is the set of expected loss vectors attainable with pure strategies f. By mixed strategies, i.e., using strategies $f_1, \cdots, f_k$ with probabilities

$$p_1, \cdots, p_k, p_i \geq 1, \Sigma p_i = 1,$$

the statistician can attain all vectors in the convex set determined by R, and only those. Thus if R is already convex, nothing is gained by the use of mixed strategies.[4]

Sequential sampling. The above discussion applies directly only to the action to be taken after a sample point x has been obtained, sequentially or otherwise, and asserts that, in the non-atomic case, nothing is gained by mixing actions. It is still possible that a mixture of sampling plans, for instance tossing a coin to decide whether to take another observation, might, even with non-atomic distributions, achieve an expected loss vector not attainable with any one sampling plan. It turns out, however, that nothing is gained by mixing sampling plans, provided all sampling plans provide for at least one observation, and that the distributions of this observation are non-atomic. Formally, we have the

[4] It has been shown by the author in a paper submitted to the *Proceedings of the American Mathematical Society* that if A is closed, R is closed. Closure of R implies that a minimax strategy for the statistician exists.

THEOREM: *Let $x = (x_1, x_2, \cdots)$ be a sequence of chance variables whose joint distribution is one of n probability distributions $u_1, \cdots, u_n$. Let $S_1, \cdots, S_N$ be N sequential decision functions, each requiring the observation of x_1, and suppose the distributions of x_1 under $u_1, \cdots, u_n$ are non-atomic. Then any expected loss vector attainable from a mixture of $S_1, \cdots, S_N$ is also attainable from a single decision function S.*

PROOF. Let $d_{ij}(x)$ be the loss from S_j when the distribution of x is u_i. (The loss is a function of x as well as i, j, since the cost of observations may vary with x.) Then $a_j = (Ed_{ij}, \cdots, Ed_{nj})$ is the expected loss vector from S_j. Since $S_1, \cdots, S_N$ all involve observing x_1, the statistician need not make up his mind about which decision procedure to use until after x_1 is observed, i.e., a possible decision procedure is a division $\mathfrak{D}$ of sample space into N mutually exclusive x_1-sets $D_1, \cdots, D_N$, and to use decision procedure S_j if $x_1 \epsilon D_j$. The expected loss vector from $\mathfrak{D}$ is

$$v(\mathfrak{D}) = \left(\sum_{j=1}^{N} \int_{D_j} \phi_{1j}(x_1)\, du_1(x_1), \cdots, \sum_{j=1}^{N} \int_{D_j} \phi_{nj}(x_1)\, du_n(x_1)\right),$$

where $\phi_{ij}(x_1)$ is the conditional expectation of d_{ij} with respect to x_1. If $\mathfrak{D}$ is the decision procedure with D_j = space X, D_i = null set for $i \neq j$, then $v(\mathfrak{D}) = a_j$. Thus it is sufficient to show that the range of $v(\mathfrak{D})$ is convex.

The convexity of the range of $v(\mathfrak{D})$ is the special case where $u_1, \cdots, u_n$ are probability measures of

EXTENSION 2. *Let $u_1, \cdots, u_n$ be non-atomic measures on a Borel field $\mathfrak{B}$ of subsets of a space X, let $\phi_{ij}(x)$, $i = 1, \cdots, n$, $j = 1, \cdots, N$,* be *$\mathfrak{B}$-measurable functions of x such that ϕ_{ij} is u_i-integrable over X, let $\mathfrak{D} = (D_1, \cdots, D_N)$ be a decompositon of X into N disjoint subsets, and define*

$$v(\mathfrak{D}) = \left(\sum_{j=1}^{N} \int_{D_j} \phi_{1j}\, du_1, \cdots, \sum_{j=1}^{N} \int_{D_j} \phi_{nj}\, du_n\right).$$

The range of $v(\mathfrak{D})$ is convex.

PROOF. Let $\mathfrak{D}_k = (D_{k1}, \cdots, D_{kN})$, $k = 1, 2$ be two decompositions. We must show that for any t, $0 \leq t \leq 1$, there is a $\mathfrak{D}$ with $v(\mathfrak{D}) = tv(\mathfrak{D}_1) + (1 - t)v(\mathfrak{D}_2)$. Write $m_{ij}(B) = \int_B \phi_{ij}\, du_i$, and consider the $2nN$-dimensional measure $w(B) = m_{ij}(BD_{kj})$, $i = 1, \cdots, n$, $j = 1, \cdots, N$, $k = 1, 2$. Since $w(B)$ is non-atomic, Lyapunov's theorem asserts there is a B with $w(B) = tw(x)$, i.e., $m_{ij}(BD_{kj}) = tm_{ij}(D_{kj})$. Then $m_{ij}(C(B)D_{kj}) = (1 - t)m_{ij}(D_{kj})$. Define $D_j = BD_{1j} + C(B)D_{2j}$, $j = 1, \cdots, N$, $\mathfrak{D} = (D_1, \cdots, D_N)$. Then

$$\begin{aligned} v(\mathfrak{D}) &= \sum_{j=1}^{N} [m_{1j}(D_j), \cdots, m_{nj}(D_j)] \\ &= t \sum_{j=1}^{N} [m_{1j}(D_{1j}), \cdots, m_{nj}(D_{1j})] + (1 - t) \sum_{j=1}^{N} [m_{1j}(D_{2j}), \cdots, m_{nj}(D_{2j})] \\ &= tv(\mathfrak{D}_1) + (1 - t)v(\mathfrak{D}_2). \end{aligned}$$