ON THE GLIVENKO-CANTELLI THEOREM FOR INFINITE INVARIANT MEASURES[1]

By Eugene M. Klimko

Ohio State University

1. Introduction. Let $(\Omega, \mathcal{A}, \mu)$ be a sigma-finite measure space. Let τ be a (i) *measure preserving* (ii) *conservative* (iii) *ergodic* point-transformation on Ω. That is, we assume that: (i) $A \varepsilon \mathcal{A}$ implies $\tau^{-1}(A) \varepsilon \mathcal{A}$ and $\mu(\tau^{-1}A) = \mu(A)$; (ii) $A \varepsilon \mathcal{A}$, $A \cap \tau^{-i}A = \emptyset$ for $i = 1, 2, \cdots$ implies $\mu(A) = 0$; (iii) the *invariant sigma-field* $\mathcal{I} = \{A : \tau^{-1}A = A \varepsilon \mathcal{A}\}$ is trivial, i.e. $A \varepsilon \mathcal{I}$ implies $\mu(A) = 0$ or $\mu(\Omega - A) = 0$. In probability theory, null-recurrent Markov chains and Markov processes satisfying the Harris condition give rise to such transformations (see Harris and Robbins [4], Harris [3], Kakutani and Parry [6]).

Let X_0, Y_0 be fixed real-valued measurable functions on Ω and let $X_n = X_0 \circ \tau^n$, $Y_r = Y_0 \circ \tau^n$, $n = 1, 2, \cdots$. If s, x, t, y are extended real numbers, let

$$(1.1) \qquad F_n^{\,s}(x) = 1_{(s,x)} \circ X_n, \qquad G_n^{\,t}(y) = 1_{(t,y)} \circ Y_n, \qquad n = 0, 1, \cdots,$$

and

$$(1.2) \qquad F^s(x) = \int_\Omega F_0^{\,s}(x)\mu(d\omega), \qquad G^t(y) = \int_\Omega G_0^{\,t}(y)\mu(d\omega).$$

Our theorem asserts that the ratio $\sum_{k=0}^n F_k^{\,s}(x)/\sum_{k=0}^n G_k^{\,t}(y)$ converges almost everywhere uniformly in (x, y), which is however restricted to a set on which F^s, G^t behave with some moderation.

Theorem 1.1. *Let $s, t \varepsilon \bar{R}$ (extended real line). Let C and D be sets in $\bar{R}$ such that for some positive constants c, d*

$$(1.3) \qquad C = \{x : F^s(x) \leqq c\}, \qquad D = \{y : G^t(y) \geqq d\}.$$

Let $B = C \times D$ and

$$(1.4) \qquad \Delta_n = \sup_{(x,y)\varepsilon B} |(\textstyle\sum_{i=0}^{n-1} F_i^{\,s}(x)/\sum_{i=0}^{n-1} G_i^{\,t}(y)) - (F^s(x)/G^t(y))|.$$

Then for almost all $\omega \varepsilon \Omega$

$$(1.5) \qquad \lim_{n\to\infty} \Delta_n = 0.$$

We note that Theorem 1 implies the Glivenko-Cantelli theorem (see [9], p. 335, [7], p. 20, Tucker [10]; also Fortet and Mourier [2]). Let μ be a probability measure and let $X_0 = Y_0$. Further set $s = t = -y = -\infty$ and $c = d = 1$. Then the denominator in the first ratio in (1.4) is simply n and Theorem 1.1 asserts the uniform convergence a.e. of the experimental distribution function $n^{-1}\sum_{i=0}^{n-1} F_i^{-\infty}(x)$ of a strictly stationary ergodic process (X_n), to the distribu-

Received 25 November 1966.

[1] This paper is based on a part of a Ph.D. dissertation written at the Ohio State University under the direction of Professor L. Sucheston to whom the author gratefully acknowledges his indebtedness for advice and encouragement.

tion function $F^{-\infty}(x)$ of X_0. Indeed, a stationary process on a probability space gives rise to a measure-preserving (hence conservative) point-transformation on the sample probability space (see Doob [1], p. 452 ff.; p. 617 ff.). The uniform convergence a.e. of experimental distribution functions carries over from the second space to the first one.

Section 2 contains the proof of Theorem 1.1. In Section 3 we indicate how Theorem 1.1 extends to the non-ergodic case: the second ratio in (1.4) is then replaced by a ratio of "conditional distribution functions."

2. Proof of Theorem 1.1. Since s and t remain fixed throughout the proof, we omit the superscripts from F^s, F_n^s, G^t, and G_n^t. We may and do assume that $c = \sup_{x \varepsilon C} F(x)$ and $d = \inf_{y \varepsilon D} G(y)$. For each positive integer M, we set $x_{MM} = \sup \{y \;\varepsilon\; C\}$, $y_{MM} = \sup \{y \;\varepsilon\; D\}$ and for $0 \leqq j < M$, we form a partition of B by letting x_{Mj} and y_{Mj} be the smallest real numbers such that

$$(2.1) \quad F(x_{Mj}) \leqq jc/M \leqq F(x_{Mj} + 0), \qquad 1/G(y_{Mj} + 0) \leqq (M - j)/dM \leqq 1/G(y_{Mj}).$$

For each pair $(x, y) \;\varepsilon\; B$, we define

$$(2.2) \qquad \delta_n(x, y) = |(\textstyle\sum_{i=0}^{n-1} F_i(x)/\sum_{i=0}^{n-1} G_i(y)) - (F(x)/G(y))|.$$

From (1.1): the definition of F_n and G_n, it follows that

$$(2.3) \qquad F_n(x) = 1_{(s,x)} \circ X_0 \circ \tau^n, \qquad G_n(y) = 1_{(t,y)} \circ Y_0 \circ \tau^n.$$

Considering F_n, G_n as functions of ω, we have for fixed x, y

$$(2.4) \qquad F_n(x) = F_0(x) \circ \tau^n, \qquad G_n(y) = G_0(y) \circ \tau^n.$$

From (1.2) and Hopf's ergodic theorem ([5], p. 49) which remains true without the assumption that τ is invertible it follows that for fixed x, y

$$(2.5) \qquad \lim_{n\to\infty} \delta_n(x, y) = 0 \qquad \text{a.e.}$$

The preceding argument is also valid with x or y in (2.5) replaced by $x + 0$ or $y + 0$, respectively. Let $(x, y) \;\varepsilon\; B$, $x \neq x_{Mo}$ and $y \neq y_{Mo}$. Since for each fixed M, the x_{Mj}'s and the y_{Mj}'s form a partition of B, there is an i and a j such that $x_{M,i-1} < x \leqq x_{Mi}$ and $y_{M,j-1} < y \leqq y_{Mj}$. The monotonicity of F, F_n, G, G_n implies that

$$(2.6) \quad F(x_{M,i-1} + 0)/G(y_{Mj}) \leqq F(x)/G(y) \leqq F(x_{Mi})/G(y_{M,j-1} + 0)$$

and

$$(2.7) \quad \begin{aligned} &\textstyle\sum_{k=0}^{n} F_k(x_{M,i-1} + 0)/\sum_{k=0}^{n} G_k(y_{Mj}) \\ &\leqq \textstyle\sum_{k=0}^{n} F_k(x)/\sum_{k=0}^{n} G_k(y) \\ &\leqq \textstyle\sum_{k=0}^{n} F_k(x_{Mi})/\sum_{k=0}^{n} G_k(y_{M,j-1} + 0); \end{aligned}$$

hence

$$\begin{aligned}(2.8)\quad &\left(\textstyle\sum_{k=0}^{n} F_k(x)/\sum_{k=0}^{n} G_k(y)\right) - (F(x)/G(y)) \\ &\leqq \left(\textstyle\sum_{k=0}^{n} F_k(x_{Mi})/\sum_{k=0}^{n} G_k(y_{M,j-1}+0)\right) \\ &\quad - (F(x_{Mi})/G(y_{M,j-1}+0)) \\ &+ (F(x_{Mi})/G(y_{M,j-1}+0)) \\ &\quad - (F(x_{M,i-1}+0)/G(y_{Mj})).\end{aligned}$$

By (2.1), the last difference in (2.8) is bounded by

$$(ic/M)\cdot((M-j+1)/dM) - ((i-1)c/M)\cdot((M-j)/dM) \leqq 2c/dM.$$

An inequality similar to (2.8), giving a *lower* bound for the left side of (2.8) is obtained, and the two inequalities together yield:

$$(2.9)\quad \delta_n(x, y) \leqq \max\,[\delta_n(x_{Mi}, y_{M,j-1}+0), \delta_n(x_{M,j-1}+0, y_{Mj})] + 2c/dM.$$

In case $x = x_{Mo}$, a computation similar to the previous one shows that (2.9) holds with x_{Mi} and $x_{M,i-1}$ both replaced by x_{Mo}; similarly if $y = y_{Mo}$. Hence, if $y_{Mo} \not\varepsilon D$, an upper bound for $\Delta_n = \sup_{(x,y)\varepsilon B} \delta_n(x, y)$, is given by:

$$(2.10)\quad \max_{i=1,2}\,[\Delta_{n,M}^{(i)}] + 2c/dM$$

where

$$\begin{aligned}(2.11)\quad \Delta_{n,M}^{(1)} &\overset{\text{def}}{=}{}_{\text{def}} \max_{0\leqq i\leqq M, 0\leqq j<M} \delta_n(x_{Mi}, y_{Mj}+0) \\ \Delta_{n,M}^{(2)} &\overset{\text{def}}{=}{}_{\text{def}} \max_{0\leqq i<M, 0<j\leqq M} \delta_n(x_{Mi}+0, y_{Mj}).\end{aligned}$$

If $y_{M0}\ \varepsilon\ D$, we allow $j = 0$ in the definition of $\Delta_{n,M}^{(2)}$. It follows from (2.5) that for $i = 1, 2$, each M and almost every $\omega\ \varepsilon\ \Omega$, $\lim_n \Delta_{n,M}^{(i)} = 0$, and therefore $\limsup \Delta_n \leqq 2c/dM$. Since M is arbitrary, it follows that $\lim \Delta_n = 0$ almost everywhere, which completes the proof of the theorem.

3. The non-ergodic case. In this section we show that Theorem 1.1, with suitable modifications, remains true even though the invariant sigma-field $\mathcal{I}$ is not trivial. We assume that μ is sigma-finite on $\mathcal{I}$. For any integrable function f, $E(f \mid \mathcal{I})$ has its usual meaning of a Radon-Nikodým derivative of a finite measure with respect to a sigma-finite measure; i.e., μ restricted to $\mathcal{I}$. The limit in the Hopf ergodic Theorem [5] is now the ratio $E(f \mid \mathcal{I})/E(g \mid \mathcal{I})$; this identification is easily seen to be equivalent with the one made in [5]. (Even when μ is not sigma-finite on $\mathcal{I}$, it is still possible to compute the limit as a ratio of conditional expectations with respect to an equivalent finite measure.) For each $s, x, t, y\ \varepsilon\ \bar{R}$, we define

$$(3.1)\quad F^s(x \mid \mathcal{I}) = E(1_{(s,x)} \circ X_0 \mid \mathcal{I}), \qquad G^t(y \mid \mathcal{I}) = E(1_{(t,y)} \circ Y_0 \mid \mathcal{I}).$$

Using the method of regularization as in the case of conditional probability distributions, we may and do assume that for *every* $\omega\ \varepsilon\ \Omega$, $F^s(x \mid \mathcal{I})$ and $G^t(y \mid \mathcal{I})$

are (i) nondecreasing in $x(y)$ (ii) left-continuous and (iii) $F^s(s \mid \mathcal{I}) = G^t(t \mid \mathcal{I}) = 0$. In the sequel, $F^s(x)$ and $G^t(y)$ are assumed to be replaced in Δ_n, $\delta_n(x, y)$ etc. by $F^s(x \mid \mathcal{I})$ and $G^t(y \mid \mathcal{I})$ respectively. The proof of the next theorem uses an idea of Tucker [10] and is an extension of his result.

THEOREM 3.1. *Let $s, t \in \bar{R}$ and let C and D be sets in $\bar{R}$ such that for some positive* a.e. *finite-valued $\mathcal{I}$ measurable functions $c(\omega)$ and $d(\omega)$,*

$$C = \{x: F^s(x \mid \mathcal{I}) \leqq c(\omega)\} \qquad D = \{y: G^t(y \mid \mathcal{I}) \geqq d(\omega)\}, \tag{3.2}$$

the inequalities holding for all ω outside of a null set N independent of x, y. Let $B = C \times D$. Then for almost all $\omega \in \Omega$,

$$\lim_{n \to \infty} \Delta_n = 0. \tag{3.3}$$

PROOF. The proof is similar to that of Theorem 1.1 and we merely sketch it, indicating the essential changes. We may and do assume that for every ω, $c(\omega) = \sup_{x \in C} F(x \mid \mathcal{I})(\omega)$ and $d(\omega) = \inf_{y \in D} G(y \mid \mathcal{I})(\omega)$. Let M and j be integers with $0 \leqq j < M$. Set $X_{MM} = \sup \{x \in C\}$ and $Y_{MM} = \sup \{y \in D\}$. We define $\mathcal{I}$ measurable functions X_{Mj} and Y_{Mj} by letting for each fixed ω, X_{Mj} and Y_{Mj} be the smallest real numbers for which

$$\begin{gathered} F(X_{Mj} \mid \mathcal{I}) \leqq jc(\omega)/M \leqq F(X_{Mj} + 0 \mid \mathcal{I}), \\ 1/G(Y_{Mj} + 0 \mid \mathcal{I}) \leqq (M - j)/M\, d(\omega) \leqq 1/G(Y_{Mj} \mid \mathcal{I}). \end{gathered} \tag{3.4}$$

$\mathcal{I}$ measurable functions are shift invariant; since X_{Mj}, Y_{Mj}, are $\mathcal{I}$ measurable,

$$\begin{gathered} \tau^{-1}[s < X_n < X_{Mj}] = [s < X_{n+1} < X_{Mj}] \\ \tau^{-1}[t < Y_n < Y_{Mj}] = [t < Y_{n+1} < Y_{Mj}], \end{gathered} \tag{3.5}$$

and therefore we can write

$$F_n(X_{Mj}) = F_0(X_{Mj}) \circ \tau^n, \qquad G_n(Y_{Mj}) = G_0(Y_{Mj}) \circ \tau^n. \tag{3.6}$$

From (3.6) we can conclude by Hopf's ergodic theorem that

$$\lim_{n \to \infty} \delta_n(X_{Mi}, Y_{Mj}) = 0 \qquad \text{a.e.} \tag{3.7}$$

Applying the argument of Section 2 for each fixed $\omega \in \Omega$, we obtain

$$\Delta_n \leqq \max_{i=1,2} [\Delta_{n,M}^{(i)}] + 2c(\omega)/M\, d(\omega). \tag{3.8}$$

The theorem then follows by noting that $c(\omega)/d(\omega)$ is a.e. finite valued and M is arbitrary.

4. Concluding remarks. One may ask whether the stationarity of the sequence X_n is essential. In the case when τ is onto and invertible and $\mu(A) = 0$ implies $\mu(\tau A) = \mu(\tau^{-1}A) = 0$, we may drop the assumption that τ is measure preserving provided that F_n^s and G_n^t are suitably weighted. Let φ_n be the Radon-Nikodým derivative of $\mu \circ \tau^n$ with respect to μ. Theorems 1.1 and 3.1 remain valid if F_n^s and G_n^t in (1.4) are multiplied by φ_n. Indeed, in this case the role of

Hopf's ergodic theorem may be played by the ergodic theorem of Hurewicz-Halmos (see [8]).

REFERENCES

[1] Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.
[2] Fortet, R. and Mourier, E. (1953). Convergence de la répartition empirique vers la répartition théorique. *Ann. École Normale Sup.* **70** 267–285.
[3] Harris, T. E. (1955). The existence of stationary measures for certain Markov processes. *Proc. Third Berk. Symp. Math. Statist. Prob.* **2** 113–124.
[4] Harris, T. E. and Robbins, H. (1953). Ergodic theory of Markov chains admitting an infinite invariant measure. *Proc. Nat. Acad. Sci. U.S.A.* **39** 860–864.
[5] Hopf, E. (1937). *Ergodentheorie*. J. Springer, Berlin. (reprinted Chelsea Publishing Co., New York, 1948).
[6] Kakutani, S. and Parry, W. (1963). Infinite measure preserving transformations with "mixing". *Bull. Amer. Math. Soc.* **69** 752–756.
[7] Loève, M. (1955). *Probability Theory*, D. van Nostrand, New York.
[8] Oxtoby, J. C. (1948). On the ergodic theorem of Hurewicz. *Ann. Math.* **49** 872–884.
[9] Rényi, A. (1962). *Wahrschinlichkeitsrechnung*. Veb Deutsher Verlag. Berlin.
[10] Tucker, H. G. (1959). A generalization of the Glivenko-Cantelli theorem. *Ann. Math. Statist.* **30** 828–830.