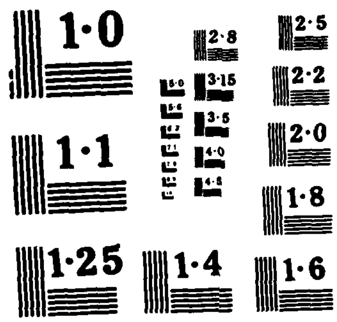
AD-A163 405 A NOTE ON CONVERGENCE OF FUNCTIONS OF RANDOM ELEMENTS (U) JOHNS HOPKINS UNIV BALTIMORE MD DEPT OF MATHEMATICAL SCIENCES R J SERFLING NOV 85 TR-459 1/1

UNCLASSIFIED N00014-79-C-0801 F/G 12/1 NL

END

FILMED

DTIC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

(13)

DEPARTMENT OF MATHEMATICAL SCIENCES
The Johns Hopkins University
Baltimore, Maryland 21218

A NOTE ON CONVERGENCE OF FUNCTIONS OF RANDOM ELEMENTS

by

Robert J. Serfling

Technical Report No. 459
ONR Technical Report No. 85-8
November, 1985

Research supported by the U.S. Department of Navy under Office of Naval Research Contract No. N00014-79-C-0801.

g (xi sub n) approaches limit of g (xi)

ABSTRACT

A NOTE ON CONVERGENCE OF FUNCTIONS OF RANDOM ELEMENTS

xi sub n xi xi sub n approaches limit of xi

- Convergence of $g(\xi_n)$ to $g(\xi)$ is considered when $\xi_n \to \xi$ in distribution or in probability, *without* the usual restriction that g be continuous a.s. under the distribution of ξ (xi). It is shown that the convergence $g(\xi_n) \to g(\xi)$ holds for arbitrary Borel-measurable g, if in addition to the assumed convergence $\xi_n \to \xi$ (xi sub n approaches limit of xi) the corresponding measures $\{P_n\}$ (P sub n) of $\{\xi_n\}$ (xi sub n) are "contiguous" to the measure P of ξ (xi) in a certain very weak sense. Some statistical applications are indicated.

AMS 1980 subject classifications: Primary 60B10, Secondary 60F05

Key words and phrases: convergence in distribution, convergence in probability, convergence of transformed random elements, contiguity

1. Introduction and preliminaries. Many statistical applications of probability theory involve a question of convergence of $g(\xi_n)$ to $g(\xi)$, $n \to \infty$, where g is a given function applied to a sequence of r.v.'s ξ_n convergent in some sense to ξ. Typically, the discontinuity set of the function g is required to have measure 0 with respect to the distribution of ξ. (See, e.g., Billingsley (1968), §5, and Serfling (1980), §1.7.) However, in some cases it is of interest to relax this assumption on g. Here it is shown that g may be an arbitrary Borel-measurable function, if in addition to the assumed convergence $\xi_n \to \xi$ the corresponding measures $\{P_n\}$ of $\{\xi_n\}$ are contiguous to the measure P of ξ in a certain very weak sense.

We will consider $\{\xi_n\}$ and ξ to be random elements of a metric space (S,ρ), and we will assume that S is separable (in order that $\rho(\xi_n,\xi)$ be a well-defined random variable and that P be regular) and locally compact (in order that Luzin's theorem be applicable). We shall treat convergence of ξ_n to ξ *in distribution*, $\xi_n \xrightarrow{d} \xi$ (by which we mean that the distributions P_n converge weakly to P, $P_n \Rightarrow P$), and *in* probability, $\xi_n \xrightarrow{p} \xi$ (by which we mean that $\{\xi_n\}$ and ξ are defined on a common probability space $(\Omega,\mathcal{A},\mathbb{P})$ and that $\mathbb{P}\{\rho(\xi_n,\xi) > \varepsilon\} \to 0$, $n \to \infty$, each $\varepsilon > 0$). As is well-known (Billingsley (1968), p. 26), $\xi_n \xrightarrow{p} \xi$ implies $\xi_n \xrightarrow{d} \xi$.

Let $(X,\mathcal{A})$ be an arbitrary measurable space. Given probability measures $\{P_n\}$ and P on $\mathcal{A}$ and a particular sequence $\{A_m\}$ in $\mathcal{A}$ such that $P(A_m) \to 0$, $m \to \infty$, we shall say that the sequence $\{P_n\}$ is *weakly contiguous to* P *at the sequence* $\{A_m\}$ if

$$\lim_{m\to\infty} \limsup_{n\to\infty} P_n(A_m) = 0. \tag{1.1}$$

This is a weak variant of the classical notion of *contiguity* due to LeCam (1960) and used by Hájek and Šidák (1967), which requires that $P_n(A_n) \to 0$ for every sequence $\{A_n\}$ such that $P(A_n) \to 0$.

REMARK 1.1. (i) Note that the above notion of weak contiguity is much weaker than, and implied by, *equicontinuity of* $\{P_n\}$ *at* $\emptyset$ (Kingman and Taylor (1966, p. 177), which requires that

$$\lim_{m\to\infty} \sup_n P_n(A_m) = 0 \tag{1.2}$$

for every sequence $\{A_m\}$ decreasing to $\emptyset$. The limit in (1.1) is less than or equal to that in (1.2), we do not require that $A_n \downarrow \emptyset$, and we permit *restriction to a single sequence* $\{A_n\}$.

(ii) Also, the above notion of weak contiguity is much weaker than, and implied by, *uniform absolute continuity* of $\{P_n\}$ with respect to P, which requires that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $P(A) < \delta$ implies $\sup_n P_n(A) < \varepsilon$.

(iii) The above weak notion of contiguity is relevant to the problem of convergence of $g(\xi_n)$ to $g(\xi)$, because our treatment will require merely that (1.1) hold for a single rather special sequence $\{B_n\}$ defined in terms of the given function g and the limit measure P through an application of Luzin's theorem, as will be seen below.

(iv) A further motivation for restricting to contiguity at a single sequence arises from certain statistical applications, to be discussed in detail in Section 2, in which the measures $\{P_n\}$ are in fact *empirical* measures based on a collection of observations each having distribution P. In such a situation, it is *not* true that the sequence $\{P_n\}$ is with probability 1 contiguous to P, or equicontinuous at $\emptyset$, or uniformly absolutely continuous with respect to P. However, it is easily seen that if the empirical measures $\{P_n\}$ (and here we are allowing consideration of empirical measures not necessarily the classical versions) satisfy a set-wise strong law (i.e., for each $A \in \mathcal{A}$, with probability 1 $\{P_n\}$ satisfies $P_n(A) \to P(A), n \to \infty$), then for any given sequence $\{A_n\}$ with $P(A_n) \to 0$, the sequence $\{P_n\}$ satisfies (1.1) with probability 1. This "set-wise strong law property" is usually easily established for reasonable notions of empirical measure. $\square$

Now take $(\mathcal{X},\mathcal{A}) = (S,\mathcal{S})$, where (S,ρ) is a separable and locally compact metric space and $\mathcal{S}$ is the Borel σ-algebra generated by the open sets. We specify as follows the sequences $\{A_m\}$ in $\mathcal{S}$ which will arise for consideration in our development. Let $g: S \to \mathbb{R}$ be $\mathcal{S}$-measurable. By separability of S the measure P on $\mathcal{S}$ is regular, so that by local compactness of S we may apply Luzin's theorem (Halmos (1950), p. 242; Cohn (1980), p. 227), which implies the existence of continuous functions $g_m: S \to \mathbb{R}$ and sets B_m such that $P(B_m) \to 0, m \to \infty$, with $S - B_m$ compact and $g = g_m$ on $S - B_m$ $(m = 1,2,\ldots)$. We shall call $\{g_m, B_m\}$ a *Luzin sequence for* g *with respect to* P.

Our main results and applications will be presented in Section 2. One application will concern Lemma 2.1 of van Zwet (1980), which in part provided the inspiration for this study. A broad class of applications is typified by the question of almost sure convergence of statistics of the form $T_n = \int g \circ F_n^{-1}(t) J_n(t)dt$, where $g(\cdot)$ and $J_n(\cdot)$, $n \geq 1$, are specified functions, $F_n(\cdot)$ is the usual empirical df based on an i.i.d. sample $X_1, \ldots, X_n$ from a df F, and $F_n^{-1}(t)$ denotes the sample t-quantile, $F_n^{-1}(t) = \inf\{x: F_n(x) \geq t\}$.

2. A convergence theorem and some applications. Without ado, we give

THEOREM 2.1. *Let (S,ρ) be a separable and locally compact metric space and $\mathcal{S}$ the corresponding Borel σ-algebra, and let $\{P_n\}$ and P be probability measures on $\mathcal{S}$. Let $g: S \to \mathbb{R}$ be $\mathcal{S}$-measurable and suppose that $\{P_n\}$ is weakly contiguous to P at $\{B_m\}$, where $\{g_m, B_m\}$ is a Luzin sequence for g with respect to P.*

(i) *Suppose that $P_n \Rightarrow P$. Then*

$$P_n g^{-1} \Rightarrow P g^{-1}. \tag{2.1}$$

(ii) *Let $\{\xi_n\}$ and ξ be measurable mappings from a probability space $(\Omega, A, \mathbb{P})$ into S with corresponding measures $\{P_n\}$ and P, and suppose that $\xi_n \xrightarrow{p} \xi$. Then*

$$g(\xi_n) \xrightarrow{p} g(\xi). \tag{2.2}$$

PROOF. We first prove (ii), which then will be utilized in the proof of (i). Let $\varepsilon > 0$ be given, choose an integer $m \geq 1$, and write

$$\mathbb{P}\{|g(\xi_n) - g(\xi)| > \varepsilon\} \leq \alpha_{n,m} + \beta_{n,m} + \gamma_m,$$

where $\alpha_{n,m} = \mathbb{P}\{\xi_n \notin B_m, \xi \notin B_m, |g_m(\xi_n) - g_m(\xi)| > \varepsilon\}$, $\beta_{n,m} = \mathbb{P}\{\xi_n \in B_m\} = P_n(B_m)$, and $\gamma_m = \mathbb{P}\{\xi \in B_m\} = P(B_m)$. (This is possible by the definitions and discussion of Section 1.) By the uniform continuity of g_m on B_m, $\exists\ \delta_{m,\varepsilon}$ such that $\rho(x,y) < \delta_{m,\varepsilon}$ implies $|g_m(x) - g_m(y)| < \varepsilon$ for $x,y \in S-B_m$. Hence $\alpha_{n,m} \leq \mathbb{P}\{\rho(\xi_n,\xi) > \delta_{m,\varepsilon}\}$, and by $\xi_n \xrightarrow{P} \xi$ we thus have $\alpha_{n,m} \to 0$ as $n \to \infty$. Therefore,

$$(2.3) \qquad \limsup_{n\to\infty} \mathbb{P}\{|g(\xi_n) - g(\xi)| > \varepsilon\} \leq \limsup_{n\to\infty} \beta_{n,m} + \gamma_m.$$

We now let $m \to \infty$, applying (1.1) and the fact that $\gamma_m \to 0$, and obtain (2.2). Thus (ii) is proved.

To prove (i), we apply a representation of Skorokhod (1956) as extended by Dudley (1968), to introduce a probability space $(\Omega_0, \mathcal{B}_0, \mu_0)$ and *$\mathcal{B}_0 - \mathcal{S}$-measurable S-valued functions $\{\eta_n\}$ and η defined on Ω_0* such that the distribution of η_n is $P_n (n = 1,2,...)$ and that of η is P, and $\eta_n \to \eta, n \to \infty$, a.s. $[\mu_0]$. Here $\Omega_0 = \Omega \times \Omega_*$, where $\Omega_* = \Pi_{n=1}^{\infty} \Omega_n$, with $\Omega_n = S_n \times I_n, S_n$ a copy of S, and I_n a copy of the unit interval $[0,1]$. Thus Ω_0 is separable and locally compact, by virtue of these properties for the spaces S and $[0,1]$, and taking $\mathcal{B}_0$ to be the product σ-algebra on Ω_0. Also, Ω_0 is metrizable. Therefore, noting that $\eta_n \to \eta$ in μ_0-measure, it is easily seen that we may apply part (ii) already proved, to assert that $g(\eta_n) \to g(\eta)$ in μ_0-measure, which implies (2.1). $\square$

REMARK 2.1. (i) It is possible to relax the assumption of separability in Theorem 2.1, at the expense of greatly complicating the formulation and tools. This entails replacing the Borel σ-algebra $\mathcal{B}$ with a suitable sub-σ-algebra, considering non-Borel measures, and utilizing a further extension of Skorokhod's theorem by Wichura (1970) to arbitrary metric spaces. For discussion of convergence in distribution and probability in this setting, see Gaenssler (1983), Chapter 3.

(ii) It would be of interest also to develop an analogous theorem for convergence of $g(\xi_n)$ to $g(\xi)$ almost surely, given $\xi_n \to \xi$ almost surely $[\mathbb{P}]$. However, a preliminary investigation suggests the need for a refinement of Luzin's theorem. This will be pursued elsewhere.

From Remark 1.1 (i), (ii) we immediately have

COROLLARY 2.1. *Let* (S,ρ) *be a separable and locally compact metric space, with corresponding Borel* σ*-algebra* $\mathcal{S}$. *Let* $\{P_n\}$ *and* P *be probability measures on* $\mathcal{S}$ *such that either* $\{P_n\}$ *is equicontinuous at* $\emptyset$ *or* $\{P_n\}$ *is uniformly absolutely continuous with respect to* P. *Then, for every* $\mathcal{S}$*-measurable function* $g\colon S \to \mathbb{R}$, *we have:*

(i) $P_n g^{-1} \Rightarrow P_g^{-1}$

and

(ii) $g(\xi_n) \xrightarrow{P} g(\xi)$, *whenever* $\{\xi_n\}$ *and* ξ *are measurable mappings of some* $(\Omega,\mathcal{A},\mathbb{P})$ *into* S *with* $\xi_n \xrightarrow{P} \xi$.

The next result is of special interest in connection with problems such as the statistical application discussed at the end of Section 1. We denote by λ the Lebesgue measure on $([0,1],\mathcal{B})$. For any df G on $\mathbb{R}$, put $G^{-1}(t) = \inf\{x\colon G(x) \geq t\}, t \in (0,1)$.

COROLLARY 2.2. *Let* $g\colon \mathbb{R} \to \mathbb{R}$ *be Borel-measurable. Let* $\{G_n\}$ *and* G *be df's on* $\mathbb{R}$ *such that* $G_n \Rightarrow G$ *and* $\{G_n\}$ *is weakly contiguous to* G *at a Luzin sequence for* g *at* G. *Then* $g \circ G_n^{-1}$ *converges to* $g \circ G^{-1}$ *in Lebesgue measure, i.e., for every* $\varepsilon > 0$,

$$(2.4) \qquad \lim_{n\to\infty} \lambda\{t: |g \circ G_n^{-1}(t) - g \circ G^{-1}(t)| > \varepsilon\} = 0.$$

PROOF. We apply Theorem 2.1 with $S = \mathbb{R}$ and $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}, \lambda)$. Now, by Lemma 1.5.6 of Serfling (1980), the convergence $G_n \Rightarrow G$ implies that the set $\{t: G_n^{-1}(t) \not\to G^{-1}(t), n \to \infty\}$ has at most countably many elements. That is, the sequence of r.v.'s $\xi_n(t) = G_n^{-1}(t)$ defined on $([0,1], \mathcal{B})$ converges a.s. $[\lambda]$ and hence in λ-measure to the r.v. $\xi(t) = G^{-1}(t)$. And, of course, the df's of $\{\xi_n\}$ and ξ are simply $\{G_n\}$ and G, respectively. Thus (2.4) follows by part (ii) of Theorem 2.1. $\square$

Statistical applications of Corollary 2.2 are as follows. Let $X_1, X_2, \ldots$ be independent r.v.'s on $\mathbb{R}$ with common df F, and define $F_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}\{X_i \le x\}$, $x \in \mathbb{R}$. By the Glivenko-Cantelli theorem it follows that with probability 1 $\{F_n(\cdot)\}$ is a sequence of df's satisfying $F_n \Rightarrow F$. With this fact and Remark 1.1 (iv), it follows that with probability 1 $\{F_n(\cdot)\}$ is a sequence of df's satisfying the requirements of Corollary 2.2 with $G_n = F_n$ and $G = F$. Thus, for any Borel-measurable $g\colon \mathbb{R} \to \mathbb{R}$, we have: *with probability* 1, $g \circ F_n^{-1}$ *converges to* $g \circ F^{-1}$ *in Lebesgue measure, i.e.,* $\lambda\{t: |g \circ F_n^{-1}(t) - g \circ F^{-1}(t)| > \varepsilon\} \to 0$, *each* $\varepsilon > 0$. This result is precisely (with different notation) Lemma 2.1 of van Zwet (1980), which he employs instrumentally in establishing strong convergence of statistics of the form $T_n = \int_0^1 g \circ F_n^{-1}(t) J_n(t) dt$, under suitable conditions on $\{J_n\}$ and F. Extensions of this work to a much broader class of statistics are made possible by the generalization of van Zwet's lemma given by Theorem 2.1 (ii); see Helmers, Janssen and Serfling (1985).

Acknowledgment. I am greatly indebted to my colleague Alan F. Karr for extensive and very helpful discussions on the issues underlying this development.

REFERENCES

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

Cohn, D.L. (1980), *Measure Theory*, Birkhäuser, Boston.

Dudley, R.M. (1968), "Distances of probability measures and random variables," *Ann. Math. Statist.*, 39, 1563-1572.

Gaenssler, P. (1983), *Empirical Processes*, Institute of Mathematical Statistics Monograph Series, Vol. 3, Hayward, California.

Hájek, J. and Šidák, Z. (1967), *Theory of Rank Tests*, Academic Press, New York.

Halmos, P.R. (1950), *Measure Theory*, Van Nostrand, Princeton, New Jersey.

Helmers, R., Janssen, P. and Serfling, R. (1985), "Glivenko-Cantelli properties of some generalized empirical df's and strong convergence of generalized L-statistics," Technical Report No. 460, Department of Mathematical Sciences, Johns Hopkins University, Baltimore, Maryland.

Kingman, J.F.C. and Taylor, S.J. (1966), *Introduction to Measure and Probability*, Cambridge University Press, Cambridge.

LeCam, L. (1960), "Locally asymptotically normal families of distributions," *Univ. of Calif. Public. in Statist.* 3, 37-98.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Skorokhod, A.V. (1956), "Limit theorems for stochastic processes," *Th. Prob. Applic.*, 1, 261-290.

van Zwet, W.R. (1980), "A strong law for linear functions of order statistics," *Ann. Prob.*, 8, 986-990.

Wichura, M.J. (1970), "On the construction of almost uniformly convergent random variables with given weakly convergent image laws," *Ann. Math. Statist.*, 41, 284-291.

Robert J. Serfling
Department of Mathematical Sciences
Johns Hopkins University
Baltimore, Maryland 21218 USA

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER ONR No. 85-8	2. GOVT ACCESSION NO. AD-A163405	3. RECIPIENT CATALOG NUMBER
4. TITLE A note on convergence of functions of random elements		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORGANIZATION REPORT NO. Technical Report No. 459
7. AUTHOR(s) Robert J. Serfling		8. CONTRACT OR GRANT NUMBER(s) ONR No. N00014-79-C-0801
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematical Sciences The Johns Hopkins University Baltimore, Maryland 21218		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics and Probability Program Arlington, Virginia 22217		12. REPORT DATE November, 1985 13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

convergence in distribution, convergence in probability, convergence of *transformed random elements, contiguity*

20. ABSTRACT

Convergence of $g(\xi_n)$ to $g(\xi)$ is considered when $\xi_n \to \xi$ in distribution or in probability, *without* the usual restriction that g be continuous a.s. under the distribution of ξ. It is shown that the convergence $g(\xi_n) \to g(\xi)$ holds for arbitrary Borel-measurable g, if in addition to the assumed convergence $\xi_n \to \xi$ the corresponding measures $\{P_n\}$ of $\{\xi_n\}$ are "contiguous" to the measure P of ξ in a certain very weak sense. Some statistical applications are indicated.

END

FILMED

2-86

DTIC