The Annals of Statistics
1977, Vol. 5, No. 2, 370–374

UNIFORM CONVERGENCE OF THE EMPIRICAL DISTRIBUTION FUNCTION OVER CONVEX SETS[1]

BY W. F. EDDY AND J. A. HARTIGAN

Yale University

The empirical distribution function P_n converges with probability 1 to a true distribution P in R^k, uniformly over measurable convex sets, if and only if P is a countable mixture of distributions, each of which is carried by a flat and gives zero probability to the relative boundaries of convex sets included in the flat.

1. Introduction. If a density in R^k has a convex contour containing probability α, the contour may be estimated from a sample by the boundary of the convex polyhedron of minimum volume which contains a proportion α of the sample points. The consistency of this estimate is shown by establishing that the empirical distribution P_n converges with probability 1 to the true distribution P, uniformly over measurable convex sets.

The Glivenko–Cantelli theorem states that $\sup_x |F_n(x) - F(x)| \to 0$ with probability 1, where F is an arbitrary distribution and F_n is the empirical distribution function based on a sample of size n from F. Thus in one dimension, the empirical distribution converges to F with probability one, uniformly over convex sets. Ranga Rao (1962) has shown that in R^k, the empirical distribution P_n converges to P with probability one, uniformly over measurable convex sets, when the nonatomic component of P gives zero probability to the boundaries of convex sets.

More generally, Billingsley and Topsøe (1967) and Topsøe (1967), define a P-uniformity class $\mathscr{U}$ to be a class of measurable subsets of a separable metric space such that $\sup_{A\in\mathscr{U}} |P_n(A) - P(A)| \to 0$ whenever P_n converges weakly to P. (Weak convergence requires that $P_n(A) \to P(A)$ whenever $P(\partial A) = 0$ where ∂A denotes the boundary of A.) A class $\mathscr{U}$ is shown to be a P-uniformity class if and only if $\lim_{\delta\to 0} \sup_{A\in\mathscr{U}} P\{x \mid \rho(x, \partial A) < \delta\} = 0$, where $\rho(x, \partial A)$ denotes the minimum euclidean distance of x to points in ∂A. Thus the probability of a strip about the boundary of each set A approaches zero uniformly in $\mathscr{U}$ as the width of the strip approaches zero. If $\mathscr{U}$ is the class of convex sets in R^k, the condition reduces to $P(\partial A) = 0$ for $A \in \mathscr{U}$. Since the empirical distribution P_n converges weakly to P, with probability 1, it follows that with probability 1, P_n converges to P uniformly over convex sets when P gives zero probability to the boundaries of convex sets; this is Ranga Rao's condition for nonatomic P.

Received June 1974; revised July 1976.

[1] Research supported by NSF Grant DCR-75-08374.

AMS 1970 *subject classifications.* Primary 60F15; Secondary 52A20.

Key words and phrases. Empirical distribution, convex sets, Glivenko–Cantelli theorem, uniform convergence.

The concept of P-uniformity is not quite appropriate for studying uniform convergence of empirical distributions, since it is possible to have $\sup_{A \in \mathscr{U}} |P_n(A) - P(A)| \to 0$ when P_n is the empirical distribution, without having the same convergence for every P_n which converges weakly to P. This paper shows that the empirical distribution converges with probability one uniformly over measurable convex subsets in R^k if and only if P is a countable mixture of distributions, each of which is carried by a flat, and assigns zero probability to the relative boundary of every convex set contained in the flat. Thus a distribution uniform over the boundary of a triangle will not satisfy the conditions of Ranga Rao's result, and the convex sets will not be P-uniform, yet the empirical distribution converges uniformly. However a distribution uniform over the boundary of a circle will not satisfy the conditions of the theorem, nor will the empirical distribution function converge uniformly, as may be seen by taking the convex hull of the sample points which has P_n-probability 1 and P-probability zero.

2. Uniform convergence over convex sets. Points $x, y, z, \cdots$ will be in R^k, the set of all k-tuples of real numbers. Euclidean distance $[\sum (x_i - y_i)^2]^{\frac{1}{2}}$ between x and y will be written $\rho(x, y)$. The distance $\rho(A, B)$ between two nonempty sets A and B is $\sup [\sup_{x \in A} \inf_{y \in B} \rho(x, y), \sup_{x \in B} \inf_{y \in A} \rho(x, y)]$. This distance is a pseudometric on the class of bounded sets, and a metric on the class of compact sets.

A p-dimensional *flat* is a set consisting of all points $x_0 + \sum_{i=1}^{p} \alpha_i x_i$ where $x_0, x_1, \cdots, x_p$ are fixed and $x_1, \cdots, x_p$ are linearly independent. The *dimensionality* of a set A is the dimensionality of the smallest flat $F(A)$ containing A. A set C is convex if $x \in C, y \in C \Rightarrow \alpha x + (1 - \alpha)y \in C, 0 \leqq \alpha \leqq 1$. The *relative boundary* of C is the set of points whose every neighbourhood intersects C and $F(C) - C$; it will be denoted by $\partial_r C$.

Let P denote a probability distribution on R^k and suppose $X_1, \cdots, X_n$ denotes a random sample from P. The *empirical distribution* P_n gives probability $1/n$ to each point $X_1, \cdots, X_n$.

LEMMA. *Let $\mathscr{U}$ denote a family of measurable subsets of a space S. Let $P = \sum \alpha_i P^i$, $\alpha_i > 0$, where P, P^i are probability distributions on S, and let P_n, $P_n{}^i$ denote empirical distributions of samples of size n from P and P^i respectively.*

Then $\sup_{A \in \mathscr{U}} |P_n(A) - P(A)| \to 0$ *with probability* 1, *if and only if* $\sup_{A \in \mathscr{U}} |P_n{}^i(A) - P^i(A)| \to 0$ *with probability* 1 *for each* i.

(The quantities $\sup_{A \in \mathscr{U}} |P_n(A) - P(A)|$ may not be random variables. Convergence with probability 1 means $\sup_{A \in \mathscr{U}} |P_n(A) - P(A)| \leqq U_n$, where U_n is a random variable, $U_n \to 0$ with probability 1.) This lemma is crucial in permitting separate examination of uniform convergence on various components of P.

PROOF. First assume that uniform convergence holds for the various components P^i. A sample $X_1, \cdots, X_n$ from P is a mixture of samples from P^i, $1 \leqq i < \infty$, with k_i observations being from P^i where k_i is binomial with parameters n, α_i.

Choose $\varepsilon > 0$ and let I be such that $\sum_{i \geqq I} \alpha_i \leqq \varepsilon$. Then

$$|P_n(A) - P(A)| = \left|\sum \frac{k_i}{n} P^i_{k_i}(A) - \sum \alpha_i P^i(A)\right|$$
$$\leqq \sum_{i<I} \left|\frac{k_i}{n} - \alpha_i\right| + \sum_{i<I} |P^i_{k_i}(A) - P^i(A)| + \sum_{i \geqq I} \alpha_i + \sum_{i \geqq I} \frac{k_i}{n} .$$

With probability 1, $k_i/n \to \alpha_i$, $\sup_{A \in \mathscr{U}} |P^i_{k_i}(A) - P^i(A)| \to 0$, for $i < I$, and $\sum_{i \geqq I} k_i/n \to \sum_{i \geqq I} \alpha_i$. Thus, with probability 1,

$$\limsup_{n\to\infty} \sup_{A \in \mathscr{U}} |P_n(A) - P(A)| \leqq 2\varepsilon .$$

Since $\varepsilon > 0$ is arbitrary, $\sup_{A \in \mathscr{U}} |P_n(A) - P(A)| \to 0$ with probability 1.

It will be sufficient to prove the converse for $P = \alpha_1 P^1 + \alpha_2 P^2$. The idea is that failure of uniform convergence on P^1 means that for some $\varepsilon > 0$, $|P_n^1(A_n) - P^1(A_n)| > 2\varepsilon/\alpha_1$ for a sequence of sets A_n, $A_n \in \mathscr{U}$; some subsequence of A_n will satisfy $|P_n^2(A_n) - P^2(A_n)| < \varepsilon/\alpha_2$; thus $|P_n(A_n) - P(A_n)| > \varepsilon$ on this subsequence and P fails to converge uniformly.

Let P^0 be the distribution $P^0(1) = \alpha_1$, $P^0(2) = \alpha_2$. The probability space $\{S^i, B^i, Q^i\}$ is the space of the infinite sequence of observations $X_1^i, X_2^i, \cdots, X_n^i, \cdots$ from P^i. A sample point in S^i will be denoted by $w^i = \{x_1^i, x_2^i, \cdots, x_n^i, \cdots\}$. The product space of $\{S^i, B^i, Q^i\}$ for $i = 0, 1, 2$ is denoted by $\{S^*, B^*, Q^*\}$. A sequence of observations $X_1, \cdots, X_n, \cdots$ from P is generated on this product space by mixing observations from P^1 and P^2 according to the mixing probabilities P^0; explicitly

$$X_n(w^0, w^1, w^2) = (2 - x_n^0)x^1_{j_n} + (x_n^0 - 1)x^2_{n-j_n}$$

where j_n is the number of times 1 occurs in $x_1^0, \cdots, x_n^0$. Thus for example $X_1 = x_1^1$ if $x_1^0 = 1$ and $X_1 = x_1^2$ if $x_1^0 = 2$, and so X_1 is sampled from $P = \alpha_1 P^1 + \alpha_2 P^2$.

Let Ω^1 be the set of points w^1 such that for some sequence of sets $A_m \in \mathscr{U}$, $|P_m^1(A_m) - P^1(A_m)| > \varepsilon$ infinitely often. In terms of the observations $X_1, \cdots, X_n$ from P, for $w^1 \in \Omega^1$, the event $|P^1_{j_n}(A_{j_n}) - P^1(A_{j_n})| > \varepsilon$ occurs infinitely often. Let $\Omega^0 = \{w^0 \mid |j_n/n - \alpha_1| < \frac{1}{2}\varepsilon\alpha_1^2 \text{ for } n > N\}$ and choose $N(w_0)$ so that $Q^0(\Omega^0) = 1$. Let $\Omega^2(w^0, w^1)$ be a set of points w^2 such that $|P^2_{n-j_n}(A_{j_n}) - P^2(A_{j_n})| < \frac{1}{2}\varepsilon\alpha_1$ on a subsequence of the sets A_{j_n} where $|P^1_{j_n}(A_{j_n}) - P^1(A_{j_n})| > \varepsilon$. For a particular j_n, $|P^2_{n-j_n}(A_{j_n}) - P^2(A_{j_n})| < \frac{1}{2}\varepsilon\alpha_1$ with probability at least $1 - 1/[\varepsilon^2\alpha_1^2(n - j_n)]$ by Chebyshev, so by suitable selection of a subsequence of the A_{j_n}, $Q^2[\Omega^2(w^0, w^1)] = 1$.

For $w^1 \in \Omega^1$, $w^0 \in \Omega^0$, $w^2 \in \Omega^2(w^0, w^1)$, there is a subsequence of $\{n\}$, depending on w^0, w^1, w^2, such that

$$|P_n(A_{j_n}) - P(A_{j_n})|$$
$$= \left|\frac{j_n}{n} P^1_{j_n}(A_{j_n}) - \alpha_1 P(A_{j_n}) + \left(1 - \frac{j_n}{n}\right) P^2_{n-j_n}(A_{j_n}) - \alpha_2 P^2(A_{j_n})\right|$$
$$\geqq \alpha_1 |P^1_{j_n}(A_{j_n}) - P(A_{j_n})| - \alpha_2 |P^2_{n-j_n}(A_{j_n}) - P(A_{j_n})| - \left|\frac{j_n}{n} - \alpha_1\right|$$
$$\geqq \alpha_1\varepsilon - (1 - \alpha_1)\tfrac{1}{2}\varepsilon\alpha_1 - \tfrac{1}{2}\varepsilon\alpha_1^2 = \tfrac{1}{2}\varepsilon\alpha_1 .$$

Since $\sup_{A \in \mathscr{A}} |P_n(A) - P(A)| \to 0$ with probability 1, $Q^*[w^1 \in \Omega^1, w^0 \in \Omega^0, w^2 \in \Omega^2(w^0, w^1)] = 0$. The set $\Omega^2(w^0, w^1)$ has $Q^2[w^2 \in \Omega^2(w^0, w^1)] = 1$ for each (w^0, w^1). Thus by Fubini's theorem, the set $\{(w^0, w^1, w^2) \mid w^2 \in \Omega^2(w^0, w^1)\}$ has $Q^*[w^2 \in \Omega^2(w^0, w^1)] = 1$. Also $Q^*(\Omega^0) = Q^0(\Omega^0) = 1$. Thus $Q^*(\Omega^1) = Q^1(\Omega^1) = 0$, showing uniform convergence of $P_n^{\,1}$ to P^1 as required.

THEOREM. *Let $\mathscr{C}$ denote the family of measurable convex subsets of R^k. Then $\lim_{n\to\infty} \sup_{C \in \mathscr{C}} |P_n(C) - P(C)| = 0$ if and only if P is a countable mixture of distributions, each of which is carried by a flat, and assigns zero probability to the relative boundary of every convex set contained in the flat.*

PROOF. First assume that relative boundaries of convex sets have zero probability. Choose $\varepsilon > 0$ and find a closed ball S such that $P(S) > 1 - \varepsilon$.

For each convex set C included in S, define $N_\delta(C) = \{A \mid A \text{ convex}, \rho(A, C) < \delta\}$. If $A \in N_\delta(C)$, and C has dimension k, then $\partial_r C = \partial C$ and $A \triangle C \subset \{x \mid \rho(x, \partial C) < \delta\} = B_\delta(C)$, say. Since $\lim_{\delta\downarrow 0} B_\delta(C) = \partial_r C$, $\lim_{\delta \downarrow 0} P[B_\delta(C)] = 0$. Choose $\delta(C)$ for each C so that $B(C) = B_\delta(C)$ satisfies $P(B(C)) < \varepsilon$. If C has dimension less than k, its closure C^* is a subset of a relative boundary and $P(C^*) = 0$. Also $A \triangle C \subset \{x \mid \rho(x, C) < \delta\} = B_\delta(C)$ where $\lim_{\delta\downarrow 0} B_\delta(C) = C^*$. Thus choose δ small so that $B(C) = B_\delta(C)$ satisfies $P(B(C)) < \varepsilon$.

For each sample sequence $w = x_1, \cdots, x_n, \cdots$ choose $M(C, w)$ so that $|P_n(C) - P(C)| < \varepsilon$ and $P_n(B(C)) < 2\varepsilon$ for all $n > M(C, w)$. The strong law of large numbers guarantees $M(C, w) < \infty$ except for $w \in \Omega$, $P(\Omega) = 0$.

The Blaschke selection theorem (Eggleston, 1958, page 59) states that the class of convex sets included in S is compact under the pseudometric ρ. The family of open sets $N_{\delta(C)}(C)$ includes all convex sets inside S, and so all convex sets are included in a finite subfamily $N_{\delta(C_i)}(C_i) = N_i$, $1 \leqq i \leqq p$. For $C \in N_i$,

$$\begin{aligned} |P_n(C) - P(C)| &\leqq |P_n(C_i) - P(C_i)| + P_n(C \triangle C_i) + P(C \triangle C_i) \\ &\leqq |P_n(C_i) - P(C_i)| + P_n[B(C_i)] + P[B(C_i)] \\ &\leqq 4\varepsilon \qquad \text{for} \quad n > M(C_i, w)\,. \end{aligned}$$

Thus $\sup_{C \subset S} |P_n(C) - P(C)| \leqq 4\varepsilon$ for $n > \max_i M(C_i, w)$, where $M_1(w) = \max_i M(C_i, w)$ is finite except on a set of measure zero.

Choose $M_2(w)$ so that $P_n(S) > 1 - 2\varepsilon$.

Then $\sup_C |P_n(C) - P(C)| \leqq 7\varepsilon$ for $n > M_1(w), M_2(w)$.

Since this is true for every $\varepsilon > 0$, $\sup_C |P_n(C) - P(C)| \to 0$ with probability 1, as has already been proven by Ranga Rao (1962). If P satisfies the conditions of the theorem, the above convergence occurs for every component of the mixture, and so it occurs for P by the lemma.

Now consider P not satisfying the conditions of the theorem. Let $\{P_{i1}\}$ denote the one point distributions carried by the atoms of P, and choose α_{i1} so that $P - \sum \alpha_{i1} P_{i1}$ has no atoms. Let $\{P_{i2}\}$ denote the distributions carried by 1-dimensional flats to which $P - \sum \alpha_{i1} P_{i1}$ gives positive measure, and choose α_{i2} so that $P - \sum \alpha_{i1} P_{i1} - \sum \alpha_{i2} P_{i2}$ gives zero measure to every 1-dimensional flat.

Continuing in this way, $P_{1k+1} = P - \sum_{j=1}^{k} \sum \alpha_{ij} P_{ij}$ gives zero measure to all $(k-1)$-dimensional flats, and P_{ij} is a distribution carried by a $(j-1)$-dimensional flat which gives zero probability to all $(j-2)$-dimensional flats. Since P does not satisfy the conditions of the theorem, some P_{ij+1} gives positive probability to the relative boundary $\partial_r C$ of a convex set C of dimension j, lying in the flat F carrying P_{ij+1}. Necessarily $j > 0$, because convex sets of dimension 0 are points with null relative boundaries. Let T be the set of points in $\partial_r C$ which belong to no P-positive flat of less than j dimensions. Since P_{ij+1} gives zero probability to the excluded flats, $P_{ij+1}(T) > 0$. Thus T is a subset of the relative boundary of C, of positive probability α, such that every $(j-1)$-dimensional subset of T has zero probability. It may be assumed that $T \subset C$. Set $P = \alpha P^1 + (1-\alpha)P^2$ where P^1 is carried by T. Let $X_1, \cdots, X_n$ be observations from P^1 and let C_n be their convex hull; then $C_n \subset C$. Since C_n contains $X_1, \cdots, X_n$, $P_n^1(C_n) = 1$.

If C_n has dimensionality less than j, $P^1(C_n) = 0$ by assumption. If C_n has dimensionality j, then F must be the minimal flat containing C_n. If $x \in C_n - \partial_r C_n$, then x has a neighbourhood which does not intersect $F - C_n$ and so does not intersect $F - C \subset F - C_n$. Thus $C_n - \partial_r C_n \subset C - \partial_r C$, $P^1(C_n - \partial_n C_n) \leqq P^1(C - \partial_r C) = 0$. Also $\partial_r C_n$ is the union of a finite number of $(j-1)$ dimensional sets, so that $P^1(\partial_r C_n) = 0$. It follows that $P^1(C_n) = 0$ whether C_n has dimensionality equal to j or less than j.

Thus $P_n^1(C)$ does not converge with probability one to $P^1(C)$ uniformly over convex sets C, and from the lemma, $P_n(C)$ does not converge with probability one uniformly to $P(C)$.

3. Acknowledgment. We are indebted to a referee who pointed out an error in the original statement and proof of the theorem, and to David Pollard for helping with the lemma.

REFERENCES

[1] Billingsley, P. and Topsøe, F. (1967). Uniformity in weak convergence. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **7** 1–16.

[2] Eggleston, H. G. (1958). *Convexity*. University Press, Cambridge.

[3] Rao, Ranga (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.

[4] Topsøe, F. (1967). On the connection between P-continuity and P-uniformity in weak convergence. *Theor. Probability Appl.* **12** 241–250.

Department of Statistics
Yale University
Box 2179
Yale Station
New Haven, Connecticut 06520