GENERALIZATION OF THE THEOREM OF GLIVENKO-CANTELLI

By J. Wolfowitz

Cornell University and University of California at Los Angeles

Let $X_1, X_2, \cdots$ be independent chance variables with the same distribution function $F(x)$. ($F(x)$ is the probability that $X_1 < x$.) The "empiric" distribution function $F_n^*(x)$ of $X_1, \cdots, X_n$ is given by

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} \psi_x(X_i), \tag{1}$$

where

$$\begin{aligned} \psi_x(a) &= 0, \qquad x \leqq a \\ &= 1, \qquad x > a. \end{aligned}$$

Thus $F_n^*(x)$ is $1/n$ times the number of $X_1, \cdots, X_n$ which are less than x. We define the distance $\delta(G_1, G_2)$ between the two distribution functions G_1 and G_2 as

$$\delta(G_1, G_2) = \sup_x | G_1(x) - G_2(x) |. \tag{2}$$

Let $P\{ \}$ denote the probability of the relation in braces. The theorem of Glivenko-Cantelli (see, for example [1], page 260) states that

$$P\{\lim_{n\to\infty} \delta(F(x), F_n^*(x)) = 0\} = 1. \tag{3}$$

Let $Y = X_1^1, \cdots, X_1^k, X_2^1, \cdots, X_2^k, \cdots$, ad inf. be a sequence of independent chance variables such that $X_1^i, X_2^i, \cdots$, ad inf. have the same distribution function (say $F_i(x)$), $i = 1, \cdots, k$. Let q_i, $i = 1, \cdots, k$, be real parameters. We shall prove the following generalization of the theorem of Glivenko-Cantelli.

Theorem. *Let $q = (q_1, \cdots, q_k)$. Let $F(x \mid q)$ be the distribution function of $\sum_{i=1}^k q_i X_1^i$. Let $F_n^*(x \mid q)$ be the empiric distribution function of*

$$\left(\sum_{i=1}^k q_i X_j^i\right), \qquad j = 1, \cdots, n.$$

Then

$$P\{\lim_{n\to\infty} \sup_q \delta(F(x \mid q), F_n^*(x \mid q)) = 0\} = 1. \tag{4}$$

This stronger version of the Glivenko-Cantelli theorem will prove useful in mathematical statistics for the purpose of estimating unknown distribution functions. We have already made use of essentially our result in [2], [3], and [4].

For typographical simplicity we shall carry through the proof for $k = 2$, and leave to the reader the easy verification of the fact that the method is valid for

Received 9/9/53.

all k. It is easy to see that, when $k = 2$, we may, without loss of generality, take $q_2 = 1$. We write $q_1 = p$. Thus $q = (p, 1)$.

LEMMA 1. *Let Δ, η, and ϵ be positive and Q be any number. There exists a positive integer $N(\epsilon, \eta, Q)$ (which is a function only of the variables exhibited) such that*

$$P\{\delta(F(x \mid p), F_n^*(x \mid p)) < \eta + 2M(\Delta H), \quad n = N, N+1, \cdots, \text{ad inf.}, \quad |p - Q| \leqq \Delta\} > 1 - \epsilon, \tag{5}$$

where H is any positive number such that H and $-H$ are both points of continuity of $F_1(x)$,

$$F_1(H) - F_1(-H) > 1 - \frac{\eta}{6},$$

and

$$M(v) = \sup_x | F_2(x) - F_2(x - v) |.$$

PROOF. From the theorem of Glivenko-Cantelli we obtain that, for some $N_0(\epsilon, \eta, Q)$,

$$P\left\{\delta(F(x \mid Q), F_n^*(x \mid Q)) < \frac{\eta}{3}, \quad n = N_0, N_0 + 1, \cdots, \text{ad inf.}\right\} > 1 - \frac{\epsilon}{2}. \tag{6}$$

From the strong law of large numbers we have that, for some $N_1(\epsilon, \eta)$,

$$P\left\{n^{-1} \sum_{i=1}^{n} [\psi_H(X_i^1) - \psi_{-H}(X_i^1)] > 1 - \frac{\eta}{3}, \quad n = N_1, N_1 + 1, \cdots, \text{ad inf.}\right\} > 1 - \frac{\epsilon}{2}. \tag{7}$$

Thus the probability of the event

$$\left\{\delta(F(x \mid Q), F_n^*(x \mid Q)) < \frac{\eta}{3}, n^{-1} \sum_{i=1}^{n} [\psi_H(X_i^1) - \psi_{-H}(X_i^1)] > 1 - \frac{\eta}{3}, \quad n = N_2, N_2 + 1, \cdots, \text{ad inf.}\right\} \tag{8}$$

exceeds $1 - \epsilon$, where $N_2 = \max(N_0, N_1)$. The event whose probability is bounded in (7) in conjunction with $|p - Q| \leqq \Delta$, implies

$$F_n^*(x - \Delta H \mid Q) - \frac{\eta}{3} \leqq F_n^*(x \mid p) \leqq F_n^*(x + \Delta H \mid Q) + \frac{\eta}{3} \tag{9}$$

for $n = N_1, N_1 + 1, \cdots$, ad inf. The event whose probability is bounded in (6), together with (9), implies

$$F(x - \Delta H \mid Q) - \frac{2\eta}{3} \leqq F_n^*(x \mid p) \leqq F(x + \Delta H \mid Q) + \frac{2\eta}{3} \tag{10}$$

for $n = N_2, N_2 + 1, \cdots$, ad inf.

From the formula for a convolution we obtain immediately that for any p and any x

$$|F(x - \Delta H | p) - F(x | p)| \leqq M(\Delta H). \tag{11}$$

Hence (10) implies

$$F(x | Q) - \frac{2\eta}{3} - M(\Delta H) \leqq F_n^*(x | p) \leqq F(x | Q) + \frac{2\eta}{3} + M(\Delta H) \tag{12}$$

for $n = N_2, N_2 + 1, \cdots$, ad inf.

Finally we consider

$$\delta(F(x | Q), F(x | p)) = \sup_x |P\{QX_1^1 + X_1^2 < x\} - P\{pX_1^1 + X_1^2 < x\}|. \tag{13}$$

We have

$$\begin{aligned} P\{pX_1^1 + X_1^2 < x\} &= P\{QX_1^1 + X_1^2 + (p - Q)X_1^1 < x\} \\ &\leqq P\{QX_1^1 + X_1^2 < x + \Delta H\} + \frac{\eta}{3} \leqq P\{QX_1^1 + X_1^2 < x\} + \frac{\eta}{3} + M(\Delta H). \end{aligned} \tag{14}$$

Similarly

$$P\{pX_1^1 + X_1^2 < x\} \geqq P\{QX_1^1 + X_1^2 < x\} - \frac{\eta}{3} - M(\Delta H) \tag{15}$$

(12), (14), and (15) imply

$$\delta(F(x | p), F_n^*(x | p)) \leqq \eta + 2M(\Delta H). \tag{16}$$

This proves Lemma 1 with $N(\epsilon, \eta, Q) = N_2$.

LEMMA 2. *Let ϵ and η be arbitrary positive numbers. If $F_1(x)$ is continuous there exist positive functions $K(\epsilon, \eta)$ and $N(\epsilon, \eta)$ such that*

$$P\{\delta(F(x | p), F_n^*(x | p)) < \eta, n = N, N + 1, \cdots, \text{ad inf.}, |p| \geqq K\} > 1 - \epsilon. \tag{17}$$

PROOF. Since $F_1(x)$ is continuous it is uniformly continuous. Let h be such that $|x_1 - x_2| \leqq h$ implies $|F_1(x_1) - F_1(x_2)| < \eta/10$. Let $K_0 > 0$ be such that $F_2(K_0) - F_2(-K_0) > 1 - \eta/10$, and K_0 and $-K_0$ are both points of continuity of $F_2(x)$. Now, if $|p| > K_0/h$, then

$$\begin{aligned} &|P\{pX_1^1 + X_1^2 < x\} - P\{pX_1^1 < x\}| \\ &\qquad \leqq \frac{\eta}{10} + \sup_x \left| F_1\left(\frac{x + K_0}{|p|}\right) - F_1\left(\frac{x}{|p|}\right)\right| < \frac{\eta}{5}. \end{aligned} \tag{18}$$

For N_1 sufficiently large we have

$$P\left\{\delta(F_1(x), F_{1n}^*(x)) < \frac{\eta}{10}, n = N_1, N_1 + 1, \cdots, \text{ad inf.}\right\} > 1 - \frac{\epsilon}{2}, \tag{19}$$

where $F_{1n}^*(x)$ is the empiric distribution function of $X_1^1, X_2^1, \cdots, X_n^1$. Obviously

$$\delta(F_1(x), F_{1n}^*(x)) = \delta\left(F_1\left(\frac{x}{p}\right), F_{1n}^*\left(\frac{x}{p}\right)\right). \tag{20}$$

(18), (19), and (20) imply that

$$P\left\{\delta\left(F(x \mid p), F_{1n}^*\left(\frac{x}{p}\right)\right) < \frac{3\eta}{10}, n = N_1, N_1 + 1, \cdots, \text{ad inf.}\right\} > 1 - \frac{\epsilon}{2}. \tag{21}$$

From the strong law of large numbers it follows that, for N_2 sufficiently large,

$$P\left\{n^{-1} \sum_{i=1}^{n} [\psi_{K_0}(X_i^2) - \psi_{-K_0}(X_i^2)] > 1 - \frac{\eta}{5}, \right. \\ \left. n = N_2, N_2 + 1, \cdots, \text{ad inf.}\right\} > 1 - \frac{\epsilon}{2}. \tag{22}$$

Now $p > K_0/h$ together with the event whose probability is bounded in (22) implies the event

$$\left\{F_{1n}^*\left(\frac{x - K_0}{p}\right) - \frac{\eta}{5} \leqq F_n^*(x \mid p) \leqq F_{1n}^*\left(\frac{x + K_0}{p}\right) + \frac{\eta}{5}, \right. \\ \left. n = N_2, N_2 + 1, \cdots, \text{ad inf.}\right\} \tag{23}$$

From (19), (20), (22), (23), and the definition of h we obtain, with $N_3 = \max(N_1, N_2)$, that

$$P\left\{F_{1n}^*\left(\frac{x}{p}\right) - \frac{\eta}{2} \leqq F_n^*(x \mid p) \leqq F_{1n}^*\left(\frac{x}{p}\right) + \frac{\eta}{2}, \right. \\ \left. n = N_3, N_3 + 1, \cdots, \text{ad inf.}, p > \frac{K_0}{h}\right\} > 1 - \epsilon. \tag{24}$$

The same result obviously holds for $p < -K_0/h$. From (24) and (21) we obtain the desired result with $K(\epsilon, \eta) = K_0/h$, and $N(\epsilon, \eta) = N_3$.

LEMMA 3. *Lemma 2 holds even when $F_1(x)$ is not continuous.*

PROOF. Let $d_1, d_2, \cdots$ be the (necessarily denumerable) points of discontinuity of $F_1(x)$ and let t_i be the saltus of $F_1(x)$ at d_i, $i = 1, 2, \cdots$, ad inf. Let r be such that

$$\sum_{r+1}^{\infty} t_i < \frac{\eta}{10}. \tag{25}$$

Write $F_1(x) = F_1'(x) + F_2'(x)$, where $F_1'(x)$ is continuous and nondecreasing, $F_2'(x)$ is a nondecreasing step-function with saltuses of size t_i at the points d_i,

$i = 1, 2, \cdots$, ad inf., and $F_1'(-\infty) = F_2'(-\infty) = 0$. Define

$$t_0^* = 1 - \textstyle\sum_{i=1}^{\infty} t_i$$

$$t_i^* = t_i \qquad (i = 1, \cdots, r).$$

$$t_{r+1}^* = 1 - \textstyle\sum_{i=0}^{r} t_i^*$$

We shall assume that $p > 0$ and $t_0^* > 0$. The modifications needed in the proof below when $p < 0$ and/or $t_0^* = 0$ will be obvious. Let W_{jn}, $j = 0, 1, \cdots, r + 1$; $n = 1, 2, \cdots$, ad inf., be chance variables distributed independently of each other and of the elements of Y, with distributions given by the following (for all n):

$$P\{W_{0n} < x\} = \frac{1}{t_0^*} F_1'(x).$$

$$P\{W_{in} = d_i\} = 1 \qquad (i = 1, \cdots, r).$$

$$P\{W_{(r+1)n} = 0\} = 1$$

Let Z_n, $n = 1, 2, \cdots$, ad inf., be (independently distributed) chance variables defined for all n by the following:

$$P\{Z_n = W_{in} \mid Z_1, \cdots, Z_{n-1}\} = t_i^* \qquad (i = 0, 1, \cdots, r + 1).$$

For all positive p and all positive integral n we define chance variables Z_{pn} by $Z_{pn} = pZ_n + X_n^2$. Write $V(x \mid p)$ for the distribution function of Z_{pn}. We have immediately from (25) that

$$\delta(F(x \mid p), V(x \mid p)) < \frac{\eta}{10}. \tag{26}$$

Let $n\gamma_i(n)$, $(i = 0, 1, \cdots, r + 1)$ be the number of indices j for which $Z_j = W_{ij}$, $j = 1, \cdots, n$. From the strong law of large numbers it follows that, for any positive ϵ and η and for some $N'(\epsilon, \eta)$ large enough,

$$P\left\{\mid \gamma_i(n) - t_i^* \mid < \frac{\eta}{20(r+1)}, i = 0, 1, \cdots, r + 1, \quad n = N', N' + 1, \cdots, \text{ad inf.}\right\} > 1 - \frac{\epsilon}{2}. \tag{27}$$

Write $F_p''(x)$ for the convolution of $t_0^{*-1} F_1'\left(\frac{x}{p}\right)$ with $F_2(x)$. Let $H_i(x \mid p, n\gamma_i(n))$, $i = 0, 1, \cdots, r + 1$, be the empiric distribution function of those Z_{pj}, $j = 1, \cdots, n$, for which the corresponding Z_j equals W_{ij}, $j = 1, \cdots, n$. (The saltuses of H_i are integral multiples of $(n\gamma_i(n))^{-1}$.) From (27), Lemma 2, and the theorem of Glivenko-Cantelli we conclude that there exist $N''(\epsilon, \eta)$ and $K(\epsilon, \eta)$ such that the probability exceeds $1 - \epsilon$ that the following events will all occur for every $n \geqq N''(\epsilon, \eta)$ and every $p > K(\epsilon, \eta)$:

$$(28)\qquad | \gamma_i(n) - t_i^* | < \frac{\eta}{20(r+1)}, \qquad i = 0, 1, \cdots, r+1$$

$$(29)\qquad \delta(H_0(x \mid p, n\gamma_0(n)), F_p''(x)) < \frac{\eta}{10(r+1)}$$

and

$$(30)\qquad \delta(H_i(x \mid p, n\gamma_i(n)), F_2(x - pd_i)) < \frac{\eta}{10(r+1)} \qquad i = 1, \cdots, r.$$

(We get (29) from Lemma 2, because $F_1'(x)$ is continuous. We get (30) from the Glivenko-Cantelli theorem applied to $F_2(x)$, because the role of the W_{in} for $i = 1, \cdots, r$, is to supply an additive constant which merely translates the distribution functions.) When (28) holds we have from (25)

$$(31)\qquad \sup_x \left| F_n^*(x \mid p) - \sum_{i=0}^{r} \gamma_i(n) H_i(x \mid p, n\gamma_i(n)) \right| < \frac{\eta}{5}.$$

Also (29), (30), and (31) imply

$$(32)\qquad \sup_x \left| F_n^*(x \mid p) - \sum_{i=1}^{r} \gamma_i(n) F_2(x - pd_i) - \gamma_0(n) F_p''(x) \right| < \frac{3\eta}{10}.$$

From (28) and (32) we obtain

$$(33)\qquad \sup_x \left| F_n^*(x \mid p) - \sum_{i=1}^{r} t_i^* F_2(x - pd_i) - t_0^* F_p''(x) \right| < \frac{7\eta}{20}.$$

From (33) we obtain

$$(34)\qquad \delta(F_n^*(x \mid p), V(x \mid p)) < \frac{9\eta}{20}.$$

Finally (26) and (34) yield

$$(35)\qquad \delta(F(x \mid p), F_n^*(x \mid p) < \eta.$$

This completes the proof of Lemma 3.

Proof of the Theorem. First suppose $F_2(x)$ is continuous. For given η and ϵ let $K(\epsilon/2, \eta)$ and $N_0(\epsilon/2, \eta)$ be functions for which Lemmas 2 and 3 hold for η and $\epsilon/2$. Choose H as in Lemma 1, and choose $\Delta > 0$ sufficiently small so that $2M(\Delta H) < \eta/2$. (The latter can be done when $F_2(x)$ is continuous.) Define $[\Delta^{-1}K(\epsilon/2, \eta)]$ as the smallest integer $\geqq \Delta^{-1}K(\epsilon/2, \eta)$. Let Q_i be defined by

$$(36)\qquad Q_i = K\left(\frac{\epsilon}{2}, \eta\right) - (2i - 1)\Delta.$$

As in Lemma 1 choose N_i, $i = 1, \cdots, [\Delta^{-1}K(\epsilon/2, \eta)]$, so large that

$$(37)\qquad P\{\delta(F(x \mid p), F_n^*(x \mid p)) < \eta, | p - Q_i | < \Delta, n = N_i, N_i + 1, \cdots, \text{ad inf.}\} > 1 - \frac{\epsilon}{2}\left[\frac{K(\epsilon/2, \eta)}{\Delta}\right]^{-1}.$$

(In the notation of Lemma 1, one can take $N_i = N(\epsilon/2[K(\epsilon/2, \eta)/\Delta]^{-1}, \eta/2, Q_i)$, since $(\eta/2) + 2M(\Delta H) < \eta$). Let $N_0^* = \max \{N_i\}, i = 1, \cdots, [K(\epsilon/2, \eta)/\Delta]$. Therefore, for

$$N^* \geqq \max \{N_0(\epsilon/2, \eta), N_0^*\} \tag{38}$$

we have

$$P\{\delta(F(x \mid p), F_n^*(x \mid p)) < \eta, -\infty < p < \infty, \quad n = N^*, N^* + 1, \cdots, \text{ad inf.}\} > 1 - \epsilon. \tag{39}$$

This proves the theorem when $F_2(x)$ is continuous.

To prove the theorem for the case when $F_2(x)$ has discontinuities, proceed as in Lemma 3. Except for a probability sufficiently small so that it can be ignored, $F_2(x)$ consists of a continuous part and a step-function with a finite number of saltuses. We have already proved the theorem for the continuous portion. When the X_i^2, $i = 1, \cdots, n$, assume one of the values at which a saltus occurs, the effect is simply to translate both the distribution function and the empiric distribution function. In this case the Glivenko-Cantelli theorem already gives the desired result. Thus the theorem is proved when $F_2(x)$ is discontinuous and our proof is complete.

The underlying ideas of the above proof are the following:

A) When $|p|$ is a large number and $F_1(x)$ is continuous the variables $\{X_j^2\}$ play a small role in determining $\delta(F(x \mid q), F_n^*(x \mid q))$ (Lemma 2). This is made plausible by the following fact. Let $J(x \mid q)$ be the distribution function of $X_1^1 + p^{-1}X_1^2$, and $J_n^*(x \mid q)$ be the empiric distribution function of $\{X_j^1 + p^{-1}X_j^2\}$, $j = 1, \cdots, n$. Then

$$\delta(F(x \mid q), F_n^*(x \mid q)) = \delta(J(x \mid q), J_n^*(x \mid q)).$$

B) The discontinuities in $F_1(x)$ act essentially to displace the distributions laterally and the distance is left invariant (Lemma 3, especially equation (30)). Hence, when $|p|$ is large, say greater than a suitable number L^*, the variables $\{X_j^2\}$ play a small role in determining $\delta(F(x \mid q), F_n^*(x \mid q))$, whether or not $F_1(x)$ is continuous.

C) The theorem is true when p varies in a small interval (Lemma 1), essentially because of the Glivenko-Cantelli theorem.

D) The theorem is therefore true in general, because the interval $-L^* \leqq p \leqq L^*$ can be subdivided into a finite number of small intervals, for each of which C) holds, and the case $|p| > L^*$ is taken care of by B).

These considerations show that our theorem holds with essentially the same proof under hypotheses much weaker than those we have stated. We shall content ourselves with indicating just a few possible generalizations:

a) The chance variables $\{X_j^i\}$ (i fixed, $j = 1, 2, \cdots$, ad inf.) need not be independent of each other. If, for example, for each i, $\{X_j^i\}$ is a metrically transitive stationary sequence of chance variables, the Glivenko-Cantelli theorem will hold and so will our generalization of it. (As an example, see [4],[1] equation (6.3).)

b) X_1^1 and X_1^2 need not be independent, provided the dependence does not prevent B) and C) from holding. (As examples, see [2], Lemma 1, [4], equation (5.11), and [4], equation (6.10).)

c) The chance variables may be vectors and need not be scalars. (As examples, see [4], equations (5.11) and (6.10).)

REFERENCES

[1] M. Fréchet, "Recherches théoriques modernes sur la theorie des probabilités," Gauthier-Villars, Paris, 1937.

[2] J. Wolfowitz, "Consistent estimators of the parameters of a linear structural relation," *Skand. Aktuarietids.*, Vol. 35 (1952), pp. 132–151.

[3] J. Wolfowitz, "Estimation by the minimum distance method," *Ann. Inst. Stat. Math.*, Tokyo, Vol. 5 (1953), pp 9–23.

[4] J. Wolfowitz, "Estimation by the minimum distance method in nonparametric stochastic difference equations," to appear in *Ann. Math. Stat.*, Vol. 25, No. 2 (1954).

[1] The reference here is to a paper which it had been hoped to publish in the present issue of the *Annals* but which will appear in the next issue.