ON STOCHASTIC LIMIT AND ORDER RELATIONSHIPS

By H. B. Mann and A. Wald[1]

Columbia University

1. Introduction. The concept of a stochastic limit is frequently used in statistical literature. Writers of papers on problems in statistics and probability usually prove only those special cases of more general theorems which are necessary for the solution of their particular problems. Thus readers of statistical papers are confronted with the necessity of laboriously ploughing through details, a task which is made more difficult by the fact that no uniform notation has as yet been introduced. It is therefore the purpose of the present paper to outline a systematic theory of stochastic limit and order relationships and at the same time to propose a convenient notation analogous to the notation of ordinary limit and order relationships. The theorems derived in this paper are of a more general nature and seem to contain to the authors' knowledge all previous results in the literature. For instance the so-called δ-method for the derivation of asymptotic standard deviations and limit distributions, also two lemmas by J. L. Doob [1] on products, sums and quotients of random variables and a theorem derived by W. G. Madow [2] are special cases of our results. It is hoped that such a general theory together with a convenient notation will considerably facilitate the derivation of theorems concerning stochastic limits and limit distributions. In section 2 we define the notion of convergence in probability and that of stochastic order and derive 5 theorems of a very general nature. Section 2 contains 2 corollaries of these general theorems which have so far been most important in applications.

We shall frequently need the concept of a vector. A vector $a = (a^1, \cdots, a^r)$ is an ordered set of r numbers $a^1, \cdots, a^r$. The numbers $a^1, \cdots, a^r$ are called the components of a. If the components are random variables then the vector is called a random vector.

We shall generally denote by a, b constant vectors by x, y random vectors and by $a^1, \cdots, a^r, x^1, \cdots, x^r$ their components. Differing from the usual practice we shall put $|a| = (|a^1|, \cdots, |a^r|)$ and we shall write $a < b$ or $a \leq b$ if $a^i < b^i$ or $a^i \leq b^i$ for every i. This notation saves a great amount of writing, since all our theorems except theorem 4 are valid for sequences of any number of jointly distributed variates.

We shall review here the ordinary order notation. In all that follows let $f(N)$ be a positive function defined for all positive integers N.

[1] Research under a grant-in-aid from the Carnegie Corporation of New York.

We write

$a_N = o[f(N)]$ if $\lim_{N\to\infty} a_N/f(N) = 0$.

$a_N = O[f(N)]$ if $|a_N| \leq Mf(N)$ for all N and a fixed $M > 0$.

$a_N = \Omega[f(N)]$ if $0 < M'f(N) \leq |a_N| \leq Mf(N)$ for almost all N and for two fixed numbers $M > M' > 0$.

$a_N = \omega[f(N)]$ if $0 < Mf(N) \leq |a_N|$ for almost all N and a fixed $M > 0$.

For instance, $\log N = o(N^\epsilon)$ for every $\epsilon > 0$, or $\sin N/N = O(1/N)$, $3 + 4\cdot N/(4 + 8\sqrt{N}) = \Omega(\sqrt{N})$ $5/\sin N = \omega(1)$.

For any statement V we shall denote by $P(V)$ the probability that V holds.

2. General theorems on stochastic limit and order relationships.

DEFINITION 1. *We write* $\operatorname{plim}_{N\to\infty} x_N = 0$. (*In words x_N converges in probability to 0 with increasing N*) *if for every* $\epsilon > 0$ $\lim P(|x_N| \leq \epsilon) = 1$. *Further* $\operatorname{plim}_{N\to\infty} x_N = x$ *if* $\operatorname{plim}_{N\to\infty} (x_N - x) = 0$.

DEFINITION 2. *We write* $x_N = o_p[f(N)]$ (*x_N is of probability order $o[f(N)]$*) *if* $\operatorname{plim}_{N\to\infty} x_N/f(N) = 0$.

DEFINITION 3. *We write* $x_N = O_p[f(N)]$ (*x_N is of probability order $O[f(N)]$*) *if for each* $\epsilon > 0$ *there exists an* $A_\epsilon > 0$ *such that* $P(|x_N| \leq A_\epsilon f(N)) \geq 1 - \epsilon$ *for all values of N.*

DEFINITION 4. $x_N = \Omega_p[f(N)]$ *if for each* $\epsilon > 0$ *there exist two numbers* $A_\epsilon > 0$ and $B_\epsilon > 0$ *and an integer* N_ϵ *such that* $P[A_\epsilon f(N) \leq |x_N| \leq B_\epsilon f(N)] \geq 1 - \epsilon$ *for all* $N \geq N_\epsilon$.

DEFINITION 5. $x_N = \omega_p[f(N)]$ *if for every* $\epsilon > 0$ *there exists an* $A_\epsilon > 0$ *and an integer* N_ϵ *such that* $P[A_\epsilon f(N) < |x_N|] \geq 1 - \epsilon$ *for all* $N \geq N_\epsilon$.

Let E denote a vector space. For any subset E' of E the symbol $a \subset E'$ will mean that a is an element of E'.

Since $P(x \subset E_1 \,\&\, x \subset E_2) \geq P(x \subset E_1) - P(x \not\subset E_2)$ we evidently have

LEMMA 1. *If* $P(x \subset E_1) \geq 1 - \epsilon$, $P(x \subset E_2) \geq 1 - \epsilon'$, *then* $P(x \subset E_1;\ x \subset E_2) \geq 1 - \epsilon - \epsilon'$.

We now put $O^1 = o$, $O^2 = O$, $O^3 = \Omega$, $O^4 = \omega$.

THEOREM 1. *For every* $\epsilon > 0$ *let* $\{R_N(\epsilon)\}$ *be a sequence of subsets of the r-dimensional Cartesian space such that* $P(x_N \subset R_N(\epsilon)) \geq 1 - \epsilon$ *for all N greater than a certain integer* N_ϵ. *Let* $\{g_N(x)\}$ *be a sequence of functions of* $x = (x^1, x^2, \cdots x^r)$ *such that* $g_N(a_N) = O^i[f(N)]$ *for any* $\epsilon > 0$ *and for any sequence* $\{a_N\}$ *for which* $a_N \subset R_N(\epsilon)$. *Then we have* $g_N(x_N) = O_p^i[f(N)]$.

PROOF: For $i = 1, 2, 3$, there exists a positive integer $\bar{N}_\epsilon$ such that $|g_N(a)|$ is a bounded function of a in $R_N(\epsilon)$ for $N > \bar{N}_\epsilon$. For otherwise we could construct a sequence $\{a_N\}$ with a_N in $R_N(\epsilon)$ such that $|g_N(a_N)| > Mf(N)$ for any M and for infinitely many values of N which contradicts the hypothesis of our theorem. Hence there exists an $\bar{N}_\epsilon$ such that for $N > \bar{N}_\epsilon$ the function $|g_N(a)|$ is bounded in $R_N(\epsilon)$. Let $M_N(\epsilon)$ be the l.u.b. of $|g_N(a)|/f(N)$ in $R_N(\epsilon)$. We can construct a sequence $\{a_N\}$ with $a_N \subset R_N(\epsilon)$ such that $|g_N(a_N)|/f(N) \geq M_N(\epsilon)/2$ for all $N > \bar{N}_\epsilon$. Hence for $i = 2, 3$ the sequence $M_N(\epsilon)$ must be bounded and for

$i = 1$ we must have $\lim_{N\to\infty} M_N(\epsilon) = 0$. Let $M(\epsilon)$ be the l.u.b. of $M_N(\epsilon)$. For $i = 3, 4$ one shows in exactly the same manner the existence of a g.l.b. $\bar{M}(\epsilon)$ of $|g_N(a)|/f(N)$ if $a \subset R_N(\epsilon)$ and for $N > N'_\epsilon$. Hence for sufficiently large N we have

$$P[|g_N(x_N)| \leq M_N(\epsilon)f(N)] \geq 1 - \epsilon \text{ with } \lim_{N\to\infty} M_N(\epsilon) = 0 \text{ for } i = 1,$$

$$P[|g_N(x_N)| \leq M(\epsilon)f(N)] > 1 - \epsilon \qquad \text{for } i = 2,$$

$$P[\bar{M}(\epsilon)f(N) \leq |g_N(x_N)| \leq M(\epsilon)f(N)] \geq 1 - \epsilon \qquad \text{for } i = 3,$$

$$P[\bar{M}(\epsilon)f(N) \leq |g_N(x_N)|] \geq 1 - \epsilon \qquad \text{for } i = 4.$$

For $i = 2$ the existence of an $M'(\epsilon)$ such that $P[|g_N(x_N)| \leq M'(\epsilon)f(N)] \geq 1 - \epsilon$ for all N follows easily from this result. Hence our theorem is proved.

COROLLARY 1. *If $x_N^j = O_p^{i_j}[f_j(N)]$ for $j = 1, 2, \cdots, r$ and $\{\bar{R}_N(\epsilon)\}$ is a sequence of subsets of the k-dimensional space $y^1, y^2, \cdots y^k$ such that $P[y_N \subset \bar{R}_N(\epsilon)] \geq 1 - \epsilon$ for sufficiently large N, and if $\{g_N(x^1, x^2, \cdots x^r, y^1, y^2, \cdots y^k)\}$ is a sequence of functions of $x^1, x^2, \cdots x^r, y^1, y^2, \cdots y^k$ such that for any $\epsilon > 0$ we have $g_N(a_N, b_N) = O^i[f(N)]$ for every sequence $\{a_N, b_N\}$ with $a_N^j = O^{i_j}[f_j(N)]$ $(j = 1, 2, \cdots, r)$ and $b_N \subset \bar{R}_N(\epsilon)$, then $g_N(x_N, y_N) = O_p^i[f(N)]$.*

PROOF: It follows from Lemma 1, the definition of the relation $x_N^j = O_p^{i_j}[f_j(N)]$ and the hypothesis of our corollary that for any $\epsilon > 0$ there exists a sequence of subsets $\{R_N(\epsilon)\}$ of the space $x^1, \cdots, x^r, y^1, \cdots, y^k$ which satisfies the conditions of Theorem 1 with respect to the sequence of functions $\{g_N\}$. Hence Corollary 1 is an immediate consequence of Theorem 1.

Corollary 1 implies *inter alia* that all operational rules for the ordinary order and limit relations are also applicable to stochastic limit and order relations. For instance $o[f(N)]/\Omega[g(N)] = o[f(N)/g(N)]$. Hence also $o_p[f(N)]/\Omega_p[g(N)] = o_p[f(N)/g(N)]$.

DEFINITION 6. *For any N let R_N be a region, $f_N(a)$ a function defined on R_N. The sequence $\{f_N(a)\}$ will be said to be uniformly continuous with respect to $\{R_N\}$ if the following condition is fulfilled. For every $\epsilon > 0$ there exists a vector $\delta > 0$ such that for almost all N*

$$|f_N(a + \bar{\delta}) - f_N(a)| \leq \epsilon \qquad \textit{for any } |\bar{\delta}| < \delta, \textit{ and for any } a \subset R_N$$

THEOREM 2. *Let $\operatorname{plim}_{N\to\infty} (x_N - y_N) = 0$. For every $\epsilon > 0$ let $\{R_N(\epsilon)\}$ be a sequence of subsets of the r-dimensional vector space such that for almost all N we have $P[y_N \subset R_N(\epsilon)] \geq 1 - \epsilon$. If the sequence of functions $\{f_N(a)\}$ is uniformly continuous with respect to $\{R_N(\epsilon)\}$ for every $\epsilon > 0$, then $\operatorname{plim}_{N\to\infty} [f_N(x_N) - f_N(y_N)] = 0$.*

PROOF: We have $f_N(x_N) - f_N(y_N) = f_N(y_N + z_N) - f_N(y_N)$ where $z_N^j = o(1)$ for $j = 1, \cdots, r$. Because of the uniform continuity of $f_N(a)$ with respect to $R_N(\epsilon)$ we see that for every sequence $\{a_N, b_N\}$ with $a_N \subset R_N(\epsilon)$ and $b_N^j = o(1)$ $(j = 1, 2, \cdots, r)$.

$$f_N(a_N + b_N) - f_N(a_N) = o(1).$$

Hence Theorem 2 follows from Corollary 1.

In the following we shall abbreviate "cumulative distribution function" by d.f.

DEFINITION 7. *Let* $\{x_N\}$ *be a sequence of random variables. Let* F_N *be the d.f. of* x_N. *Let* x *have the distribution* F. *We shall write* $d\infty(x_N) = d(x)$ *if* $\lim_{N\to\infty} F_N = F$ *in every continuity point of* F.

THEOREM 3. *Let* $\operatorname{plim}_{N\to\infty} (x_N - y_N) = 0$ *and* $d\infty(y_N) = d(y)$; *then* $d\infty(x_N) = d(y)$.

PROOF: Let G_N, F_N be the d.f.'s of x_N, y_N resp. For any $\delta > 0$ we have

$$P(y_N \leq a + \delta) \geq P(x_N \leq a; y_N \leq a + \delta) \geq P(x_N \leq a; | y_N - x_N | \leq \delta)$$
$$\geq P(x_N \leq a) - P(| y_N - x_N | > \delta),$$
$$P(x_N \leq a) \geq P(x_N \leq a; y_N \leq a - \delta) \geq P(y_N \leq a - \delta)$$
$$- P(| x_N - y_N | > \delta).$$

Hence since $P(y_N \leq a) = F_N(a)$, $P(x_N \leq a) = G_N(a)$, $\lim_{N\to\infty} P(| x_N - y_N | > \delta) = 0$ we have lim. sup. $F_N(a + \delta) \geq$ lim. sup. $G_N(a) \geq$ lim. inf. $G_N(a) \geq$ lim. inf. $F_N(a - \delta)$.

If $a + \delta$ and $a - \delta$ are continuity points of F we have

$$F(a + \delta) \geq \text{lim. sup. } G_N(a) \geq \text{lim. inf. } G_N(a) \geq F(a - \delta) .$$

For any $\delta_0 > 0$ there exists a positive $\delta < \delta_0$ such that $a - \delta$ and $a + \delta$ are continuity points of F. Hence we can choose δ arbitrarily small and if a is a continuity point of F we must have

$$\text{lim. } G_N(a) = F(a).$$

THEOREM 4. *Let* x_N, y_N *be two sequences of one-dimensional vectors and let* $\operatorname{plim}_{N\to\infty} (x_N - y_N) = 0$. *Let* F_N, G_N *be the cumulative distribution functions of* x_N *and* y_N *respectively. Let* $R_N(\epsilon)$ *be the set of points* a *for which* $| F_N(a) - G_N(a) | > \epsilon$. *Let* $M_N(\epsilon)$ *be the Lebesgue measure of this set. Then* $\lim_{N\to\infty} M_N(\epsilon) = 0$ *for every* $\epsilon > 0$.

We first prove the following lemma.

LEMMA 2. *Let* δ, ϵ *be any arbitrary positive numbers and let* f *be a distribution function. The set of points* a *for which* $f(a + \delta) - f(a) \geq \epsilon$ *has at most the Lebesgue measure* δ/ϵ.

PROOF: The points a for which $f(a + \delta) - f(a) \geq \epsilon$ must have a lower bound $\bar{a}$. Otherwise we could find infinitely many such points whose distance from each other is more than δ. But this contradicts the requirement that $f(\infty) = 1$. Let a_1 be the g.l.b. of the a's. Then for any $\eta > 0$ in the interval $(a_1 \leq x \leq a_1 + \delta + \eta)$ the value of F increases at least by the amount ϵ. Let now a_2 be the g.l.b. of the a's outside of this interval. We continue our construction by constructing the interval $(a_2 \leq x \leq a_2 + \delta + \eta)$ and so forth. But after at most

$1/\epsilon$ such steps the construction must stop. Hence all points a for which $f(a + \delta) - f(a) \geq \epsilon$ are contained in at most $1/\epsilon$ intervals of length $\delta + \eta$. Hence since η was arbitrary the Lebesgue measure of this set is at most δ/ϵ.

We come now to the proof of our theorem. We have

$$P(x_N \leq a) \geq P(x_N \leq a; y_N \leq a + \delta) \geq P(x_N \leq a) - P(|x_N - y_N| > \delta),$$

$$P(y_N \leq a + \delta) \geq P(x_N \leq a; y_N \leq a + \delta) \geq P(y_N \leq a + \delta) - P(|x_N - y_N| > \delta) - P(a \leq x_N \leq a + 2\delta).$$

Therefore

$$P(x_N \leq a; y_N \leq a + \delta) = P(x_N \leq a) - \bar{\theta}_N P(|x_N - y_N| > \delta)$$
$$= P(y_N \leq a + \delta) - \bar{\theta}'_N P(|x_N - y_N| > \delta) - \bar{\theta}'_N P(a \leq x_N \leq a + 2\delta),$$

where $0 \leq \bar{\theta}_N \leq 1$, $0 \leq \bar{\theta}'_N \leq 1$. Hence

$$P(y_N \leq a + \delta) = P(x_N \leq a) + \theta_N P(|x_N - y_N| > \delta) + \theta'_N[F_N(a + 2\delta) - F_N(a)]$$

where $|\theta_N|, |\theta'_N| \leq 1$.

By hypothesis we have $P(|x_N - y_N| \geq 1/m) \leq 1/m$ for almost all N and every integer m. Hence we can choose a sequence $\{\delta_N\}$ with $\delta_N > 0$ in such a way that $\lim_{N\to\infty} \delta_N = 0$, $\lim_{N\to\infty} P(|x_N - y_N| \geq \delta_N) = 0$. We can then choose N_ϵ so that $P(|x_N - y_N| \geq \delta_N) \leq \epsilon/3$ for $N \geq N_\epsilon$. Applying Lemma 2 we see that except for a set of measure at most $6\,\delta_N/\epsilon$ we have $F_N(a + 2\delta_N) - F_N(a) \leq \epsilon/3$. Similarly the set of points for which $g_N(a + \delta_N) - g_N(a) \geq \epsilon/3$ has at most the Lebesgue measure $3\,\delta_N/\epsilon$. Hence, except in a set of points whose measure is at most $9\,\delta_N/\epsilon$, we have

$$|G_N(a) - F_N(a)| \leq \epsilon,$$

and this completes the proof of Theorem 4.

THEOREM 4a. *Let* $\operatorname{plim}_{N\to\infty}(x_N - y_N) = 0$. *Let* F_N, G_N *be the distribution functions of* x_N, y_N *respectively. Furthermore, let* $R_N(\epsilon)$ *be the set of points inside an r-dimensional cube where* $|F_N - G_N| \geq \epsilon$ *and let* $M_N(\epsilon)$ *be the Lebesgue measure of* $R_N(\epsilon)$, *then* $\lim_{N\to\infty} M_N(\epsilon) = 0$.

We prove first

LEMMA 2a. *Let* $\delta = (\delta^1, \delta^2, \cdots, \delta^r) > 0$ *and max.* $\delta^i = d$. *Let* I *be the cube defined by* $(-A \leq x^i \leq A, i = 1, 2, \cdots r)$. *Let furthermore* f *be a d.f. Then the Lebesgue measure of the points* a *in* I *for which* $f(a + \delta) - f(a) \geq \epsilon$ *is at most* $dr^2 A^{r-1}/\epsilon$.

PROOF: Let $f_1(x^1), f_2(x^2), \cdots f_r(x^r)$ be the marginal distributions of $x^1, x^2, \cdots x^r$ respectively. It follows from Lemma 2 that the linear Lebesgue measure of those numbers a^i for which $f_i(a^i + \delta^i) - f_i(a^i) \geq \epsilon/r$ is smaller than rd/ϵ. We form the set $(x^i = a^i \;\&\; x \subset I)$ for every such a^i and for $i = 1, 2, \cdots r$. The

Lebesgue measure of the sum $R(\epsilon)$ of all these sets is at most r^2dA^{r-1}/ϵ. We shall show that $R(\epsilon)$ contains all points a inside I for which $f(a + \delta) - f(a) \geq \epsilon$. We have

$$f(a^1 + \delta^1, a^2 + \delta^2, \cdots, a^r + \delta^r) - f(a^1, a^2, \cdots a^r) = \Delta_1 + \Delta_2 + \cdots + \Delta_r,$$

where $\Delta_i = f(a^1, a^2, \cdots a^{i-1}, a^i + \delta^i, \cdots a^r + \delta^r) - f(a^1, \cdots a^i, a^{i+1} + \delta^{i+1}, \cdots a^r + \delta^r)$. If $f(a + \delta) - f(a) \geq \epsilon$ then we must have for at least one i

$$\Delta_i \geq \epsilon/r.$$

But Δ_i is the probability of a subset of the set $T = (a^i \leq x^i \leq a^i + \delta^i)$ and $f_i(a^i + \delta^i) - f_i(a^i)$ is the probability of T itself. Hence

$$\epsilon/r \leq \Delta_i \leq f_i(a^i + \delta^i) - f_i(a^i),$$

and if $(a^1, a^2, \cdots a^r)$ is in I then it is contained in $R(\epsilon)$. Hence Lemma 2a is proved.

The proof of Theorem 4a using Lemma 2a is similar to that of Theorem 4 and therefore it is omitted.

The Jordan measure of a set R with respect to the distribution function F is defined as follows. We consider only intervals whose boundary points are continuity points of F. We cover R with the sum I of a finite number of intervals. (The intervals themselves may also be infinite. For instance the sets $a \leq x < \infty$, $a < x < \infty$ are also considered intervals.) We consider $M(I) = \int_I dF$ for every I covering R. The g.l.b. of all such $M(I)$ is called the exterior Jordan measure $M(R)$ of R. Similarly we consider all sums $\bar{I}$ of a finite number of intervals which are contained in R. The l.u.b. of $\int_{\bar{I}} dF$ is called the interior Jordan measure $\bar{M}(R)$ of R. If $M(R) = \bar{M}(R)$ then $M(R)$ is called the Jordan measure of R.

LEMMA 3. *Let $F_N(x)$ be a sequence of d.f.'s such that $\lim_{N\to\infty} F_N(x) = F(x)$ in every continuity point of $F(x)$. Let $h(x)$ be a bounded function such that the discontinuity points of $h(x)$ have the Jordan measure 0 with respect to F and such that $\int_{-\infty}^{+\infty} h(x)\, dF_N(x)$ and $\int_{-\infty}^{+\infty} h(x)\, dF(x)$ exist. Then $\lim_{N\to\infty} \int_{-\infty}^{+\infty} h(x)\, dF_N(x) = \int_{-\infty}^{+\infty} h(x)\, dF(x)$.*

PROOF: There is only an enumerable set of hyperplanes parallel to the plane $x^i = 0$ which have positive probability with respect to F. Hence we can find for every δ an interval net whose cells have a diameter at most δ and such that the boundary points of every cell are continuity points of F.

We first determine a closed finite interval I such that $\int_I dF(x) \geq 1 - \frac{\epsilon}{2}$ and such that the boundary points of I are continuity points of F. We further determine a sum I' of a finite number of open intervals such that I' contains all discontinuity points of h, $\int_{I'} dF(x) \leq \frac{\epsilon}{2}$, and such that the boundary of I' does not contain any discontinuity points of F. All this is possible by hypothesis and because the set of hyperplanes with positive probability is enumerable. Let R be the subset of I consisting of all points of I which are not contained in I'. R is a closed set and can be decomposed into a finite number of intervals. The function h is continuous in R and therefore uniformly continuous. We can therefore cover R by a finite set of intervals such that the variation of h in every interval is less than ϵ and such that the boundary points of each interval are continuity points of F. Let $I_1, I_2, \cdots I_k$ be such a finite set of intervals. Let x_j be any point in I_j. We have

$$\begin{aligned}
| H_N | = \left| \int_{-\infty}^{+\infty} h(x)\, dF_N(x) - \int_{-\infty}^{+\infty} h(x)\, dF(x) \right| &= \left| \sum_{j=1}^{k} \int_{I_j} [h(x) - h(x_j)]\, dF_N(x) \right. \\
&\quad - \sum_{j=1}^{k} \int_{I_j} [h(x) - h(x_j)]\, dF(x) + \sum_{j=1}^{k} h(x_j) \left[\int_{I_j} dF_N(x) - \int_{I_j} dF(x) \right] \\
&\quad \left. + \int_{x \notin R} h(x)\, dF_N(x) - \int_{x \notin R} h(x)\, dF(x) \right| \\
&\leq \epsilon + \epsilon + \sum_{j=1}^{k} h(x_j) \left[\int_{I_j} dF_N(x) - \int_{I_j} dF(x) \right] \\
&\quad + \max. h(x) \left[\int_{x \notin R} dF_N(x) + \epsilon \right].
\end{aligned}$$

But $\lim_{N \to \infty} \int_R dF_N(x) \geq 1 - \epsilon$. Hence

$$\lim. \sup. H_N \leq 2\epsilon + 2\epsilon \max. h(x) .$$

Since ϵ was arbitrary, we must have $\lim_{N \to \infty} H_N = 0$.

We are now prepared to prove

THEOREM 5. Let $d\infty(x_N) = d(x)$. *Let* $g(x)$ *be a Borel measurable function such that the set* R *of discontinuity points of* $g(x)$ *is closed and* $P(x \subset R) = 0$. *Then* $d\infty[g(x_N)] = d[g(x)]$.

PROOF: Let F_N be the d.f. of x_N, F the d.f. of x, F_{Ng}, F_g the d.f.'s of $g(x_N)$, $g(x)$ resp. Then $\lim_{N \to \infty} F_N = F$ in every cont. point of F. Let $h(x)$ be defined as follows:

$$h(x) = 1 \text{ if } g(x) \leq a ,$$

$$h(x) = 0 \text{ if } g(x) > a .$$

The discontinuities of h are contained in the set M of all points where $g(x) = a$ and is continuous or where $g(x)$ is discontinuous. The set R of discontinuity points of $g(x)$ is closed and of measure 0 with respect to F. We can therefore subtract from M a sum R^* of a finite number of open intervals of arbitrarily small measure with respect to F which contains all discontinuity points of $g(x)$. This difference set M' is closed and contains only points where $g(x) = a$ and

$x \not\subset R$. If a is a continuity point of F_g then the Borel measure of M' with respect to F is 0. Since M' is closed, its Jordan measure is also 0. Hence the Jordan measure of the discontinuity points of $h(x)$ is 0 if a is a continuity point of F_g. Since $g(x)$ is Borel measurable, $\int_{-\infty}^{+\infty} h(x)\, dF_N(x) = F_{Ng}(a)$ and $\int_{-\infty}^{+\infty} h(x)\, dF(x) = F_g(a)$ exist for every a. Hence by Lemma 3 $\lim_{N\to\infty} F_{Ng}(a) = F_g(a)$ in every continuity point of F_g and this proves our theorem.

3. Corollaries and applications. COROLLARY 2. *If* $\operatorname{plim}_{N\to\infty} (x_N - y_N) = 0$, $d\infty(y_N) = d(y)$ *and if f is continuous except in a set R for which* $\lim_{N\to\infty} P(y_N \subset R) = 0$ *then* $\operatorname{plim}_{N\to\infty} f(x_N) - f(y_N) = 0$.

PROOF: Let I be a closed interval such that $P(y_N \subset I) \geq 1 - \epsilon/2$. Let I' be a sum of open intervals containing all discontinuity points of $f(x)$ in I and such that $P(y_N \subset I') \leq \epsilon/2$ for sufficiently large N. The set J of points of I which are not points of I' is a closed set. Hence f is uniformly continuous in J and $P(y_N \subset J) \geq 1 - \epsilon$ for sufficiently large N. In Theorem 2 we put $R_N(\epsilon) = J$, $f_N = f$. Then all conditions of Theorem 2 are satisfied and it follows that $\operatorname{plim}_{N\to\infty} [f(x_N) - f(y_N)] = 0$.

If, moreover, the set of discontinuity points of f is closed then by Theorems 3 and 5 $d\infty[f(x_N)] = d\infty[f(y_N)] = d[f(y)]$.

Special cases of Corollary 2 have been proved by J. L. Doob and W. G. Madow (2).

Theorem 5 is very useful in deriving limit distributions.

It follows for instance from Theorem 5 that if $d\infty(x_N) = d(x)$, $d\infty(y_N) = d(y)$, where x, y are independently and normally distributed with mean 0 and equal variances, then $d\infty(x_N/y_N) = d(x/y)$. That is to say the distribution of x_N/y_N converges to a Cauchy distribution.

It also follows from Theorem 5 that under very general conditions the limit distribution of $t = \sqrt{N}(\bar{x} - \mu)/s$ is normal. ($\bar{x}$ = sample mean, μ = population mean, s^2 = sample variance.) For we have under very general conditions $d\infty\sqrt{N}(\bar{x} - \mu) = d(\xi)$, plim $s = \sigma$, where ξ is normally distributed with variance σ^2.

Applying Theorem 5 it can also easily be shown that under very general conditions the limit distribution of T^2 is a chi-square distribution if the means of all variates are 0. Hotelling's T^2 (the generalized Student ratio) for a p-variate distribution is defined as follows:

$$T^2 = N \sum_{i=1}^{p} \sum_{j=1}^{p} A_{ij} \xi_i \xi_j \quad \text{where} \quad \| A_{ij} \| = \| s_{ij} \|^{-1}, \qquad \xi_i = \bar{x}^i,$$

where s_{ij} is the sample covariance between x^i and x^j.

We have $d\infty(A_{ij}) = d(\sigma^{ij})$, where $\| \sigma_{ij} \|^{-1} = \| \sigma^{ij} \|$. If $E(x^i) = 0$ for $i = 1, 2, \cdots p$ then $d\infty(\sqrt{N}\,\xi_i) = d(\eta_i)$ where the η_i have a joint normal distribution with covariance matrix $\|\sigma_{ij}\|$. Hence

$$d\infty(T^2) = d\left[\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma^{ij}\eta_i\eta_j\right] = d\left(\sum_{i=1}^{p}\eta_i'^2\right),$$

where the η_i' are normally and independently distributed with variance 1. Hence the distribution of T^2 converges to a chi-square distribution with p degrees of freedom.

If the samples are drawn from a sequence of populations $\{\pi_N\}$ all with the same covariance matrix and such that $\lim_{N\to\infty}\sqrt{N}\mu_{iN} = \mu_i$ where μ_{iN} is the mean value of the ith variate in the Nth population, then one sees in exactly the same way that the limit distribution of T^2 is a non-central square distribution with p degrees of freedom.

The limit distribution of T^2 has been derived by W. G. Madow (2).

COROLLARY 3. *Let x_N, y_N be r-dimensional vectors $d\infty(y_N) = d(y)$ and $x_N - y_N = O_p[f(N)]$ with $\lim_{N\to\infty} f(N) = 0$. Let $g(x)$ be a function admitting continuous jth derivatives except in a set R with $\lim_{N\to\infty} P(y_N \subset R) = 0$. Let*

$$T_j(x, a) = \sum_{i=1}^{r}\left(\frac{\partial g}{\partial x^i}\right)_{x=a}(x^i - a^i) + \cdots + \left[\sum_{i=1}^{r}(x^i - a^i)\left(\frac{\partial}{\partial x^i}\right)_{x=a}\right]^j g,$$

then

$$g(x_N) - g(y_N) - T_j(x_N, y_N) = o_p\{[f(N)]^j\}.$$

Since the jth derivatives are continuous except in a set of limit measure 0 we can determine a closed set $R(\epsilon)$ on which they are uniformly continuous and so that $P(y_N \subset R(\epsilon)) \geq 1 - \epsilon$ for sufficiently large N. Then for every sequence with $a_N - b_N = O(f(N))$, $b_N \subset R(\epsilon)$ we have

$$g(a_N) - g(b_N) - T_j(a_N, b_N) = o[f(N)^j].$$

Hence Corollary 3 follows from Theorem 1.

Corollary 3 was first proved by W. G. Madow [2] and J. L. Doob [1] for the important case that y_N is a constant.

The following example will illustrate Corollary 3. Let x, y be normally and independently distributed random variables with mean 0 and variance 1; $\{z_N\}$, $\{z_N'\}$ sequences of random variables with $\operatorname{plim}_{N\to\infty}\sqrt{N}\,z_N = \operatorname{plim}_{N\to\infty}\sqrt{N}\,z_N' = 1$. Let $x_N = x + z_N$, $y_N = y + z_N'$. We consider the function $g(x, y) = x^3/3 + y^3/3 + 2x - 2y + 5$. Applying Corollary 1 it is easy to verify that $g(x_N, y_N) - g(x, y) = \Omega_p[1/\sqrt{N}]$, $z_N = O_p(1/\sqrt{N})$, $z_N' = O_p(1/\sqrt{N})$. Hence applying Corollary 3 for $j = 1$ we have

$$g(x_N, y_N) - g(x, y) - (x^2 + 2)z_N - (y^2 - 2)z_N' = o_p(1/\sqrt{N}).$$

Multiplying by $\sqrt{N}$ we have

$$[g(x_N, y_N) - g(x, y)]\sqrt{N} - [(x^2 + 2)z_N + (y^2 - 2)z_N']\sqrt{N} = o_p(1).$$

This is equivalent to

$$\operatorname*{plim}_{N\to\infty}[\sqrt{N}(g(x_N, y_N) - g(x, y))] = x^2 + y^2 .$$

Hence the distribution of $\sqrt{N}(g(x_N, y_N) - g(x, y))$ converges to the chi-square distribution with 2 degrees of freedom.

If $\operatorname*{plim}_{N\to\infty} x_N = a$ and $\{\sigma_N\}$ is a sequence of numbers with $\lim_{N\to\infty} \sigma_N = 0$ such that $d\infty[(x_N^i - a^i)/\sigma_N] = d(\xi_i)$ where the ξ_i are constants or random variables and if g admits continuous first derivatives at $x = a$ at least one of which is different from 0, then putting $\left(\frac{\partial g}{\partial x^i}\right)_{x=a} = g_i$, we have

$$g(x_N) - g(a) = g_1(x_N^1 - a^1) + \cdots + g_r(x_N^r - a^r) + o_p(\sigma_N) .$$

Hence applying Theorems 3 and 5 we have

$$\text{(i)} \qquad d\infty\left[\frac{g(x_N) - g(a)}{\sigma_N}\right] = d(g_1\xi_1 + \cdots + g_r\xi_r).$$

That is to say the distribution of $[g(x_N) - g(a)]/\sigma_N$ converges to the distribution of $\sum_{i=1}^{p} g_i\xi_i$ in all continuity points of the latter. A corresponding result can be obtained from Corollary 3 if all first derivatives are 0 at $x = a$ and at least one second derivative is different from 0 and so forth.

A method of deriving limiting distributions and limit standard deviations based on (i) is known as the δ-method and has been extensively applied in statistical literature.

REFERENCES

[1] J. L. Doob, "The limiting distribution of certain statistics." *Annals of Math. Stat.*, Vol. 6 (1935).

[2] W. G. Madow, "Limiting distribution of quadratic and bilinear forms." *Annals of Math. Stat.*, Vol. 11 (1940).