ON EXTREME ORDER STATISTICS

BY JOHN LAMPERTI

Dartmouth College

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of independent random variables with the common distribution function $F(x) = \Pr(X_i \leqq x)$. Define the order statistics Y_n^k for $n \geqq k$ by

$$Y_n^k = k\text{th largest among } (X_1, X_2, \cdots, X_n). \tag{1.1}$$

(We will usually write M_n for Y_n^1, the maximum.) The random variables Y_n^k have, of course, been the subject of many papers, and in particular their limiting behavior as $n \to \infty$ has been thoroughly investigated. A survey of results in this area (with some new ones) was recently published by Barndorff-Nielsen [1]. However, it does not seem that $\{Y_n^k\}$ has previously been studied *as a stochastic process* (with n for the parameter), and to make such a study, with emphasis on limit theorems, is the object of the present paper.

The problem of limiting distributions for M_n has been treated very completely by Gnedenko in [3]. He determined all the non-degenerate distribution functions $G(x)$ which can appear in

$$\lim_{n\to\infty} \Pr[(M_n - a_n)/b_n \leqq x] = G(x) \tag{1.2}$$

for some choice of the constants $b_n > 0$, a_n; such a function must be of the same type as one of the laws Φ_α, Ψ_α or Λ defined in [3]. (These distributions can be conveniently found in [2], Equation (1.1).) Conditions on F insuring that (1.2) holds are also given in [3]. In the present work, the form of the law $G(x)$ will usually not be important and it is convenient to simply take (1.2) as the basic hypothesis. (But it is useful to note that all limit distribution functions are continuous.)

Following a procedure which is common enough in other contexts we define the stochastic processes

$$m_n(t) = (M_{[nt]} - a_n)/b_n, \qquad t \geqq 1/n, \tag{1.3}$$

where [u] means the greatest integer not exceeding u; it is technically convenient to define $m_n(t) = m_n(1/n)$ for $0 \leqq t \leqq 1/n$. We shall show in the next section that whenever (1.2) holds,

$$\lim_{n\to\infty} \{m_n(t)\} = \{m(t)\} \tag{1.4}$$

in the sense of convergence of finite-dimensional (f.d.) joint distributions; $\{m(t)\}$ is a Markov process with increasing path functions. In a sense, there is only one process $\{m(t)\}$ despite the variety of possible limit laws G. In Section 3 the result (1.4) is strengthened by proving two versions of an "invariance principle",

Received 20 November 1963; revised 4 April 1964.

showing that the laws of various functionals of $\{m_n(t)\}$ (and so of $\{M_n\}$) converge to the laws of the same functionals applied to $\{m(t)\}$. Results of Yu. V. Prokhorov [5] and A. V. Skorohod [6] play a prominent role here.

The immediately preceding paper [2] by M. Dwass is also concerned with this situation, and obtains results complementary to ours. Roughly the points of view are that Dwass analyzes the structure of the limiting processes $\{m(t)\}$ (which he calls "extremal processes"), while here the passage to the limit is of primary concern, as well as the study of $\{Y_n^k\}$ with $k > 1$. I appreciate the courtesy of Professor Dwass and the editors of the *Annals* in arranging for the two papers to be published together.

The next subject taken up in this paper (Section 4) is the joint limiting behavior of M_n and Y_n^2, the two largest members of the sample, considered as a two-dimensional process in a manner analogous to (1.3). The fifth and final section concerns $\{Y_n^k\}$; we point out the momentarily surprising fact that $\{Y_n^k\}$, for each k, is a Markov process with stationary transition probabilities. It is then shown that when (1.2) holds $\{Y_n^k\}$ converges to a limiting process in a manner similar to the behavior of $\{M_n\}$; once again the transition function for the limit process can be rather simply written in terms of $G(x)$. The invariance principles of Section 3 apply to these limit theorems as well.

2. The limiting processes for $\{M_n\}$.

THEOREM 2.1. *Suppose that* (1.2) *holds for a non-degenerate distribution* $G(x)$. *Then the* f.d. *laws of the processes* $\{m_n(t)\}$ *defined by* (1.3) *converge on the parameter interval* $(0, \infty)$ *to those of a Markov process* $\{m(t)\}$ *such that*

$$\Pr(m(t) \leqq x) = G(x)^t \tag{2.1}$$

and

$$\begin{aligned} \Pr(m(t+s) \leqq y \mid m(s) = x) &= 0 && \text{if } y < x, \\ &= G(y)^t && \text{if } y \geqq x. \end{aligned} \tag{2.2}$$

PROOF. We use the obvious fact that $\Pr(M_n \leqq y) = F(y)^n$. Now for any $t > 0$, and sufficiently large n,

$$\Pr(m_n(t) \leqq x) = \Pr(M_{[nt]} \leqq b_n x + a_n) = [F(b_n x + a_n)^n]^{[nt]/n}.$$

The quantity inside the outer brackets converges to $G(x)$ as $n \to \infty$, for that quantity is $\Pr[(M_n - a_n)/b_n \leqq x]$ and we assume (1.2). Thus we have

$$\lim_{n\to\infty} \Pr(m_n(t) \leqq x) = G(x)^t, \tag{2.3}$$

at least at continuity points of the right-hand side. But $G(x)$ must be of the same type as one of the laws Φ_α, Ψ_α or Λ; hence G is continuous and (2.3) holds for all values of x.

Next we shall show that the f.d. laws have limits, and evaluate them. Consider the event $\{m_n(t_1) \leqq x_1, m_n(t_2) \leqq x_2, \cdots, m_n(t_k) \leqq x_k\}$, $0 < t_1 < t_2 \cdots < t_k$. If for some $i < j$ we have $x_i \geqq x_j$, the condition $m_n(t_i) \leqq x_i$ can be omitted without changing the event. This is clear since, due to the fact that $m_n(t)$ is increasing in t, that condition is implied by $m_n(t_j) \leqq x_j$. Thus the only k-dimensional events which are not really of lower dimension are those with $x_1 < x_2 \cdots < x_k$. In such a case we will see that

$$\lim \Pr(m_n(t_1) \leqq x_1, \cdots, m_n(t_k) \leqq x_k) = G(x_1)^{t_1} G(x_2)^{t_2 - t_1} \cdots G(x_k)^{t_k - t_{k-}} \tag{2.4}$$

The reason for this is amply illustrated by the case $k = 2$. Take $0 < t_1 < t_2$ and $x_1 < x_2$; then

$$\Pr(m_n(t_1) \leqq x_1, m_n(t_2) \leqq x_2) = \Pr(m_n(t_1) \leqq x_1) \Pr(m_n(t_2) \leqq x_2 \mid m_n(t_1) \leqq x_1)$$
$$= \Pr(m_n(t_1) \leqq x_1) \Pr(X_l \leqq b_n x_2 + a_n,\ 1 \leqq l \leqq [nt_2] \mid X_l \leqq b_n x_1 + a_n,$$
$$1 \leqq l \leqq [nt_1]) = \Pr(m_n(t_1) \leqq x_1) \Pr(X_l \leqq b_n x_2 + a_n,\ [nt_1] + 1 \leqq l \leqq [nt_2])$$
$$= \Pr(m_n(t_1) \leqq x_1) \Pr(X_l \leqq b_n x_2 + a_n,\ 1 \leqq l \leqq [nt_2] - [nt_1]).$$

But by (2.3), as $n \to \infty$ the first factor tends to $G(x_1)^{t_1}$ and the second to $G(x_2)^{t_2 - t_1}$ which yields (2.4) with $k = 2$.

To complete the proof of the theorem, it is only necessary to verify that (2.2) does define a Markov transition function when $G(y)$ is of one of the limiting types, that this transition function is consistent with (2.1) and that together they generate the family of joint f.d. laws obtained in (2.4). This verification is trivial and will be omitted; we only comment that the form of G is actually irrelevant here.

Remarks. The convergence of $\{m_n(t)\}$ to $\{m(t)\}$ is almost the same as "attraction to a semi-stable process" defined in [4]; the only things which may be lacking are the requirements of [4] that $b_n \to \infty$ and that $t = 0$ be included in the convergence. In case $G = \Phi_\alpha$ we do have necessarily $b_n \to \infty$ and $m(0) = 0$ and do obtain a semi-stable Markov process for $\{m(t)\}$, but when $G = \Psi_\alpha$ or $G = \Lambda$ this is no longer true. In fact, the paths of $\{m(t)\}$ come from $-\infty$ as t increases from 0 in these cases, so that $\{m(t)\}$ cannot be a semi-stable process in the sense of [4]. Of course, these examples may be regarded as indicating that the definitions in [4] should be broadened.

It is clear in what sense there is only one limit process $\{m(t)\}$: for any two cases there is a continuous one-to-one change of variable on the state space (real line) which carries all the f.d. laws of one process into those of the other. The following theorem "explains" this and the form of the limit laws by showing how they follow from certain qualitative properties obviously possessed by $\{m_n(t)\}$ and which should be inherited in the limit.

Theorem 2.2. *Suppose $\{x_t\}$ is, for $0 < t < \infty$, a Markov process such that $\Pr(-\infty \leqq a < x_t < b \leqq +\infty) = 1$. Suppose $\{x_t\}$ has a stationary transition probability function $p_t(x, E)$, that $x_t \to_p a$ as $t \to 0$, that $p_t(x, (a, y]) = 0$ for $y < x$, and that $p_t(x, (c, b))$ is independent of x so long as $x \leqq c$. Then there exists a distribution function $G(y)$ with $G(a) = 0$ and $G(b-) = 1$ such that*

(2.5) $$\Pr(x_t \leqq y) = G(y)^t$$

and

(2.6) $$\begin{aligned} p_t(x, (a, y]) &= 0 \qquad \text{if } y < x. \\ &= G(y)^t \qquad \text{if } y \geqq x. \end{aligned}$$

PROOF. The condition $p_t(x, (a, y]) = 0$ when $y < x$ obviously is an expression of the requirement that x_t increases a.s.; using it, when $x \leqq y$ we can write the Chapman-Kolmogorov equation as

$$p_{t+s}(x, (a, y]) = \int_a^y p_t(x, du) p_s(u, (a, y]).$$

But $p_s(u, (a, y]) = 1 - p_s(u, (y, b))$ can be replaced by $p_s(x, (a, y])$ in the integral because of our last assumption. We then obtain

$$p_{t+s}(x, (a, y]) = p_t(x, (a, y]) p_s(x, (a, y]),$$

so that for some function $G(x, y)$,

$$p_t(x, (a, y]) = G(x, y)^t.$$

But again by the last assumption, for $x \leqq y$ the left side does not depend on x; we thus replace $G(x, y)$ by $G(y)$ and have established (2.6). It is evident that $G(y)$ must be increasing, right continuous, and that $G(b-) = 1$ since x_t must a.s. be inside the interval (a, b).

Consider now the equation

$$\Pr(x_t \leqq y) = \int_a^y p_{t-\epsilon}(u, (a, y])\, d \Pr(x_\epsilon \leqq u).$$

By the above, the right side equals $G(y)^{t-\epsilon} \Pr(x_\epsilon \leqq y)$. But since $x_t \to a$ in law as $t \to 0$, if we let $\epsilon \to 0$ we obtain exactly relation (2.5). From this it is clear that $G(a+) = 0$ and the proof is complete.

REMARKS. These processes generalize the class $\{m(t)\}$; when G is continuous they are "equivalent", via a transformation of the state space, to any of the extremal processes. Any process with transition function (2.6) is a pure jump process; the jump rate slows down as x_t increases, and an infinite number of jumps can occur only in the neighborhood of $(t = 0, x = a)$ and $(t = \infty, x = b)$. Under our hypotheses there will always be infinitely many jumps in both of these cases.

3. Invariance principles. In each instance of the convergence of the finite-dimensional laws of $\{m_n(t)\}$ to those of $\{m(t)\}$ given in Theorem 2.1, the processes $\{m_n(t)\}$ have non-decreasing path functions. This automatically implies the convergence of the laws of certain functionals $f(m_n(t))$ to the laws of $f(m(t))$. For instance, if f is the supremum over a finite interval $[r, s]$, $f(m_n(t))$ is the same as $m_n(s)$ so convergence of the laws is already contained in Theorem 2.1. More generally, we have

THEOREM 3.1. *Suppose the* f.d. *laws of a sequence of real stochastic processes*

$\{x_n(t)\}$ *converge to the* f.d. *laws of a process* $\{x(t)\}$, *and that all these processes have path functions which are almost surely finite and non-decreasing for* $r \leqq t \leqq s$, *where* r *and* s *are not fixed points of discontinuity for* $\{x(t)\}$. *Then at all continuity points of the right-hand side,*

$$\lim_{n\to\infty} \Pr\,(f(x_n(t)) \leqq x) = \Pr\,(f(x(t)) \leqq x) \tag{3.1}$$

for any functional f *which is defined for all increasing functions on* $[r, s]$ *and continuous with respect to weak convergence at almost all paths of the process* $\{x(t)\}$.

(A sequence of non-decreasing functions $y_n(t)$ "converges weakly" to $y(t)$ if $\int_r^s h(t)\,dy_n(t) \to \int_r^s h(t)\,dy(t)$ for every continuous, bounded h; equivalently if there is ordinary convergence at each t_0 where $y(t)$ is continuous.)

PROOF. This follows quite easily from a theorem of Prokhorov ([5], Theorem 1.12). Under the Lévy metric for monotonic functions, which generates the topology of weak convergence, the space to which the paths of the processes $\{x_n(t)\}$ and $\{x(t)\}$ belong is a complete, separable metric space. The set S of functions ξ in this space such that $-\infty < A \leqq \xi(r) \leqq \xi(s) \leqq B < +\infty$ is a compact set by Helly's theorem for any A and B; clearly for suitable A and B, $\Pr\,(x(t) \,\varepsilon\, S) \geqq 1 - \epsilon$. By the assumed convergence of f.d. laws, we also have

$$\lim \Pr\,(x_n(t) \,\varepsilon\, S) \geqq 1 - \epsilon. \tag{3.2}$$

But (3.2), with the convergence of all f.d. laws, is just what is needed for Prokhorov's theorem; the convergence (3.1) for the indicated class of functionals f is the result.

REMARKS. This result applies to any case of Theorem 2.1 provided $r > 0$ and $s < \infty$; if $r = 0$ it applies as it stands only when the limit law G is of the same type as one of the Φ_α so that $m(0)$ is finite. Of course, in the other cases we can apply the theorem to the convergence of $\{\exp m_n(t)\}$ to $\{\exp m(t)\}$ with $r = 0$ and the proper interpretation of continuity for f. The convention following (1.3) is necessary when $r = 0$ in order that $m_n(t)$ be defined on the entire interval.

Theorem 3.1 makes it seem quite natural to use the weak topology when studying the convergence of processes with increasing paths. However, the "natural" topology may not be the most advantageous one in a particular case. Consider, for instance, the situation of our Theorem 2.1 and let the functional L be the greatest jump; i.e.,

$$L(\xi) = \sup_{r\leqq t\leqq s}\,[\xi(t+0) - \xi(t-0)], \qquad 0 < r < s < \infty.$$

This functional is not continuous at *any* discontinuous increasing function $\xi(t)$ with respect to weak convergence, since one can always approximate discontinuous functions by continuous ones in that topology. However, it is still true that in our case

$$\lim_{n\to\infty} \Pr\,(L(m_n) \leqq x) = \Pr\,(L(m) \leqq x).$$

To prove this we note that L *is* continuous with respect to the stronger "J_1

topology" of Skorokhod [6], in which $\xi_n(t) \to \xi(t)$ provided there exists a sequence of continuous, strictly increasing functions $\lambda_n(t)$ mapping $[r, s]$ onto itself such that

$$\lim_{n\to\infty} |\lambda_n(t) - t| = 0, \qquad \lim_{n\to\infty} |\xi_n(\lambda_n(t)) - \xi(t)| = 0$$

uniformly for $t \in [r, s]$. The fact stated above about the size of the greatest jump is included in the following

THEOREM 3.2. *Let $\{m_n(t)\}$ converge to $\{m(t)\}$ in the manner of Theorem 2.1, where $m(r)$ and $m(s)$ are finite and the limit process $\{m(t)\}$ is assumed separable. Then the conclusion (3.1) of Theorem 3.1 holds for any functional f which is J_1 continuous* a.e. (*with respect to $\{m(t)\}$, $r \leqq t \leqq s$*).

PROOF. According to Theorem 3.2.1 of [7], the conclusion will follow provided we can show that for each $\epsilon > 0$,

$$\lim_{c\to 0} \limsup_{n\to\infty} \Pr(\Delta[c, m_n(t)] > \epsilon) = 0, \tag{3.3}$$

where, for monotonic functions $x(t)$,

$$\Delta(c, x(t)) = \sup_{r\leqq t\leqq s} \min\{|x(t) - x((t-c)')|, |x((t+c)'') - x(t)|\}. \tag{3.4}$$

(Here $(t-c)'$ means $\max(r, t-c)$; $(t+c)''$ means $\min(t+c, s)$.) In demonstrating (3.3) two cases arise; the first when $r = 0$, so that G must be of the type of some Φ_α, and the second when $r > 0$. We shall give the proof in the first case as it is a trifle harder, and briefly sketch the necessary changes for the second.

We can assume without loss of generality that $m(0) = 0$; hence $G(x) > 0$ for $x > 0$. Since $m_n(t)$ increases only by jumps, it is easy to see that the event $\Delta(c, m_n(t)) > \epsilon$ is contained in the union of the two events

$$A_n = \{m_n(0) < -\epsilon/2\}$$

and

$$B_n = \{m_n(0) \geqq -\epsilon/2 \text{ and } \exists t \leqq s: m_n(t) \geqq \epsilon/2, m_n((t+c)'') > m_n(t) > m_n((t-c)')\}.$$

Because $\lim \Pr(A_n) = 0$, we need only consider B_n. Let T_n denote the first t such that $m_n(t) \geqq \epsilon/2$; then B_n in turn is contained in the union of the events

$$C_n = \{m_n(T_n + c) > m_n(T_n)\}$$

and

$$D_n = \{\text{for some } t \in [T_n, s], m_n(\cdot) \text{ has at least two jumps in the interval } [t, (t+2c)'']\}.$$

Let us consider C_n. In order for this event to occur, we must have

$$b_n^{-1}\{\max(X_{[T_n n]+1}, \cdots, X_{[(T_n+c)n]}) - a_n\} > m_n(T_n) \geqq \epsilon/2.$$

Since the random variables in question are independent of T_n, from (2.3) and

(1.3) we obtain lim sup Pr $(C_n) \leqq 1 - G(\epsilon/2)^c$ which tends to 0 with c since $G(\epsilon/2) > 0$.

In order to bound Pr (D_n), we proceed as follows:

$$\begin{aligned}\Pr(D_n) &\leqq E(\text{number of pairs of jumps of } m_n(t) \text{ of type considered}) \\ &= \sum_{0 \leqq i < j \leqq ns} \Pr(X_i, X_j \text{ constitute such a pair of jumps}) \\ &\leqq \sum_{\substack{0 \leqq i < j \leqq ns \\ j-i \leqq [2cn]}} \Pr\left(\frac{X_i - a_n}{b_n} > \epsilon/2, \frac{X_j - a_n}{b_n} > \epsilon/2\right) \\ &= [1 - F(b_n(\epsilon/2) + a_n)]^2 \sum_{\substack{0 \leqq i < j \leqq ns \\ j-i \leqq [2cn]}} 1.\end{aligned}$$

The second factor is of course only the number of terms in the sum; it is clearly less than $2cn^2s$. As for the first factor, recall that (1.2) states that

$$\lim_{n\to\infty} \{1 - [1 - F(b_n x + a_n)]\}^n = G(x),$$

and that in our case $G(\epsilon/2) > 0$. These things mean that

$$\lim_{n\to\infty} n[1 - F(b_n(\epsilon/2) + a_n)] = -\log G(\epsilon/2) < \infty,$$

so that

$$[1 - F(b_n(\epsilon/2) + a_n)]^2 = O(n^{-2}).$$

It follows that $\limsup_{n\to\infty} \Pr(D_n) = O(c)$ and the proof of the theorem in case one is complete.

In the second case, $r > 0$, the initial covering of the event $\Delta(c, m_n(t) > \epsilon$ is by the union of $A_n' = \{G(m_n(r)) < \delta\}$ and $B_n' = \{G(m_n(r)) \geqq \delta$ and for some $t \,\varepsilon\, [r, s]$ $m_n(\cdot)$ has at least two jumps in $[t, (t + 2c)'']\}$. For small enough δ the limit of Pr (A_n') can be made small; with δ fixed B_n' is treated just as was D_n above. Thus (3.3) can be proved even more simply than before. Incidentally, if $r = 0$ but $G = \Psi_\alpha$ or Λ, we can similarly show convergence of $\{\exp m_n(t)\}$ to $\{\exp m(t)\}$ in the present sense of the J_1 topology.

4. A joint limit theorem for M_n and Y_n^2. In this section we shall always assume that (1.2) holds for suitable normalizing constants $b_n > 0$, a_n and a non-degenerate limit law G. We extend our earlier notation as follows:

$$y_{kn}(t) = (Y^k_{[nt]} - a_n)/b_n, \qquad t \geqq k/n, \tag{4.1}$$

with the same convention as before when $0 \leqq t < k/n$. The main result obtained here will show that $\{m_n(t), y_{2n}(t)\}$ converges as $n \to \infty$ to a two-dimensional Markov process whose transition law can be simply stated in terms of the function G.

We begin with the easily derived formula

$$\begin{aligned}\Pr(M_n \leqq x, Y_n^2 \leqq y) &= F(x)^n && \text{if } y \geqq x, \\ &= F(y)^n + n[\mathrm{F}(x) - F(y)]F(y)^{n-1} && \text{if } y < x.\end{aligned} \tag{4.2}$$

Now since we are assuming (1.2), we have

$$\lim_{n\to\infty} F(b_n y + a_n)^n = G(y) \tag{4.3}$$

for all y, which in turn means that

$$\lim_{n\to\infty} n[1 - F(b_n y + a_n)] = -\log G(y) \tag{4.4}$$

exists; the limit is $+\infty$ when $G(y) = 0$. Combining these facts with (4.2) we obtain

$$\begin{aligned}\lim_{n\to\infty} \Pr\left(\frac{M_n - a_n}{b_n} \leqq x, \frac{Y_n^2 - a_n}{b_n} \leqq y\right) \\ = \begin{cases} G(x) & \text{if } y \geqq x; \\ G(y)\{1 + \log G(x)/G(y)\} & \text{if } y < x. \end{cases}\end{aligned} \tag{4.5}$$

In particular, under condition (1.2) we have

$$\lim_{n\to\infty} \Pr\,[(Y_n^2 - a_n)/b_n \leqq y] = G(y)\{1 - \log G(y)\}. \tag{4.6}$$

(In both (4.5) and (4.6) the limit is 0 for values of y such that $G(y) = 0$.) Note that these new limit laws are determined from that for the maximum in a simple way which is the same regardless of the extreme type to which $G(y)$ belongs. We shall see that the same thing is true of the limiting processes.

It is clear that $\{(M_n, Y_n^2)\}$ is a Markov process with stationary transition probabilities. It is by finding limits for these that we will prove

THEOREM 4.1. *Assume that* (1.2) *holds. Then there is a Markov process* $\{(m(t), y_2(t)\}$ *in the plane to which the processes* $\{(m_n(t), y_{2n}(t)\}$ *defined above tend in the sense of convergence of the finite dimensional joint distributions. The limit process is characterized by*

$$\begin{aligned}\Pr\,(m(t) \leqq x, y_2(t) \leqq y) &= G(x)^t & \text{if } y \geqq x, \\ &= G(y)^t\{1 + t \log G(x)/G(y)\} & \text{if } y < x\end{aligned} \tag{4.7}$$

for each $t \geqq 0$, *together with the following*: for $x \geqq y$, $x' \geqq y'$, $x' \geqq x$ *and* $y' \geqq y$ *we have*

$$\begin{aligned}\Pr\,(m(t+s) \leqq x', y_2(t+s) \leqq y' \mid m(t) = x, y_2(t) = y) \\ = \begin{cases} G(y')^s & \text{if } y \leqq y' \leqq x, \\ G(y')^s\{1 + s \log G(x')/G(y')\} & \text{if } y' > x. \end{cases}\end{aligned} \tag{4.8}$$

In addition if $y' \geqq x'$ *the right-hand side is* $G(x')^s$; *if either* $x' < x$ *or* $y' < y$ *the right-hand side is* 0.

REMARK. A verbal description may help in visualizing the transitions described by (4.8). The process never enters the region $y_2(t) > m(t)$. From $m(t) = x \geqq y_2(t) = y$, one possible transition which can occur with positive probability in time s is to stand still. Another possibility is that $m(t+s) = m(t)$ but $y_2(t+s)$ takes a new value y' on the segment (y, x); finally $m(t+s)$ and $y_2(t+s)$ may take on values (x', y') which lie in the region $x' \geqq y' \geqq x$. The total probability of these alternatives is one. It is clear that a separable version of this process has path functions which for $t > 0$ move by jumps either straight upwards or up and to the right, but always below $y = x$ (on the x-y plane, with x horizontal representing $m(t)$). Depending on the type of G, $m(0)$ and $y_2(0)$ may be improper random variables $(-\infty)$.

PROOF OF THF THEOREM. Equation (4.5) is equivalent to stating that the joint law of $(m_n(t), y_{2n}(t))$ converges to that given in (4.7) when $t = 1$; it is very simple to check that the proof holds also for other values of t. It is also straightforward to verify that

$$\lim_{n\to\infty} \Pr\,(m_n(t+s) \leqq x', y_{2n}(t+s) \leqq y' \mid m_n(t) = x, y_{2n}(t) = y) = f_s(x, y; x', y') \tag{4.9}$$

exists and is given by (4.8) and its addenda. To see this, first note that obviously $m_n(t)$ and $y_{2n}(t)$ are increasing so that the limit is 0 if $x' < x$ or $y' < y$. If $y' \geqq x'$ the condition $y_{2n}(t+s) \leqq y'$ makes no restriction on the set $m_n(t+s) \leqq x'$; the information $y_{2n}(t) = y$ is useless in this case and the situation in (4.9) reduces to that of Section 2: the random variables $X_{[nt]+1}, \cdots, X_{[n(t+s)]}$ must all be at most $b_n x' + a_n$ in order for the event in question to occur. The limiting probability is thus $G(x')^s$ as asserted. If $y \leqq y' \leqq x \leqq x'$, the conditional event in (4.9) occurs if and only if all the random variables mentioned just above are at most $b_n y' + a_n$, so that the limiting (conditional) probability is $G(y')^s$ as in the first part of (4.8). Finally if $y \leqq x \leqq y' \leqq x'$, at most one of those random variables may exceed $b_n y' + a_n$ and that one must not exceed $b_n x' + a_n$. The probability of this is evaluated in the limit by the argument leading to (4.6) and the result is the second expression in (4.8). Thus (4.9) is proved and f_s is determined.

We wish to show now that the joint distribution of $\{m_n(t), y_{2n}(t); m_n(t+s), y_{2n}(t+s)\}$ tends to a limit given by combining (4.7) and (4.8) appropriately. That is, using the fact that $\{m_n(t), y_{2n}(t)\}$ is a Markov process, we must verify that

$$\lim_{n\to\infty} \int_A \Pr\,(z_n(t+s) \in B \mid z_n(t) = z)\, d\Pr\,(z_n(t) = z) = \int_A f_s(z, B)\, d\mu_t(z), \tag{4.10}$$

where $z_n(t) = (m_n(t), y_{2n}(t))$, $z = (x, y)$, f_s is the right-hand side of (4.9) (equal, as we have seen, to the right side of (4.8)), μ_t is the measure generated by the right side of (4.7), and both A and B are sets of the form $\{x \leqq \alpha, y \leqq \beta\}$. We have already shown pointwise convergence of the integrands to $f_s(z, B)$, and convergence of the integrators which implies weak convergence of their induced measures. By using the following simple lemma, we can easily complete the proof of (4.10).

LEMMA. *Suppose μ_n is a sequence of Lebesgue-Stieltjes probability measures in R^k which converges to μ, also with total mass 1, in the usual weak sense. Let $f_n(x)$ be a uniformly bounded sequence of Borel functions such that $f_n(x) \to f(x)$ uniformly on compact sets. Suppose $f(x)$ is continuous* a.e. (μ). *Then*

$$\lim_{n\to\infty} \int f_n(x)\, d\mu_n(x) = \int f(x)\, d\mu(x). \tag{4.11}$$

To apply the lemma to our case we must know that $f_s(\cdot, B)$ is continuous a.e. (μ_t) and that the limit (4.9) is uniform in (x, y) on compact sets. The first is immediate; discontinuities of $f_s(x, y; \alpha, \beta)$ occur only on the lines $x = \alpha$, $y = \beta$ on which μ_t, given by the right side of (4.7), puts measure 0. (Recall that G is continuous.) The uniformity of convergence is just as easy, for the values of the conditioning variables x, y only determine which of two limiting processes is to be used—both of which are then independent of (x, y). The convergence is thus in fact uniform over the whole plane. The proof of (4.10) is complete; higher dimensional joint laws can be handled similarly by an induction argument.

REMARKS. The theorem is now demonstrated; it *follows* that the transition function in (4.8) satisfies the Chapman-Kolmogorov equation and that it is consistent with (4.7). It is clear that similar limiting results can be developed for the joint behavior of $Y^l_{[nt]}$, $l = 1, 2, \cdots, k$, but the results seem to become increasingly cumbersome.

5. The limiting processes for $\{Y^k_n\}$. The purpose of this section is to generalize Theorem 2.1 to include the limiting behavior of the processes $\{y_{kn}(t)\}$ defined in (4.1) using the k the largest members of the samples. The main result is

THEOREM 5.1. *Assume $F(x)$ continuous and that (1.2) holds for a non-degenerate distribution $G(x)$. Then the finite-dimensional laws of $\{y_{kn}(t)\}$ converge as $n \to \infty$ to those of a Markov process $\{y_k(t)\}$ whose state space is the interval I where $0 < G(x) < \infty$. For $x, y, z \in I$ the limit process satisfies*

$$\Pr(y_k(t) \leqq x) = G(x)^t \sum_{l=0}^{k-1} [-t \log G(x)]^l / l!, \tag{5.1}$$

and has transition probabilities

$$\begin{aligned} &\Pr(y_k(t+s) \leqq z \mid y_k(t) = y) \\ &\quad = 0 && \text{if } z < y, \\ &\quad = \sum_{l=0}^{k-1} b[l; k-1, \log G(z)/\log G(y)] H_s^{k-l}(z), && \text{if } z \geqq y, \end{aligned} \tag{5.2}$$

where $H^k_t(x)$ denotes the right-hand side of (5.1).

(The notation $b(l; n, p)$ means the probability of l successes in n Bernoulli trials with probability p for success on each trial.)

This theorem is easily obtained from some facts about order statistics which do not involve (1.2).

LEMMA. *For each k the sequence $\{Y^k_n, n = k, k+1, \cdots\}$ is a Markov process with stationary transition probabilities. These are given by*

$$\begin{aligned} &\Pr(Y^k_{n+m} \leqq z \mid Y^k_n = y) \\ &= 0, && \text{for } z < y, \\ &= \sum_{l=0}^{k} b[l; k-1, (1-F(z))/(1-F(y))] \Pr(Y^{k-l}_n \leqq z), && \text{for } z \geqq y. \end{aligned} \tag{5.3}$$

The lemma follows from the known fact that given $Y_n^k = y$, the $n - k$ observations less than y and the $k - 1$ observations greater than y form conditionally independent sets of random variables, the variables in each set being themselves independently distributed according to the law F but conditioned to be less (respectively greater) than y. Under these conditions, when m new random variables are added to the sample the probability that $Y_{n+m}^k \leqq z$ can be easily written in each of the (disjoint) cases in which exactly l of the $k - 1$ original observations exceeding y are in addition less than z; the right-hand side of (5.3) is the result.

Proof of the Theorem. The limit distributions of Y_n^k as $n \to \infty$ were obtained by Smirnov in [7], but we shall briefly discuss them here. Assume that (1.2) holds; no distinction into separate cases depending on the type of G is required. It is very easy to see that

$$\Pr(Y_n^k \leqq x) = F(x)^n + nF(x)^{n-1}[1 - F(x)] + \cdots + \binom{n}{k-1} F(x)^{n-k+1}[1 - F(x)]^{k-1}. \tag{5.4}$$

But assumption (1.2) has the equivalent forms (4.3) and (4.4); combining these with (5.4) we obtain

$$\lim_{n\to\infty} \Pr[(Y_n^k - a_n)/b_n \leqq x] = G(x)\sum_{l=0}^{k-1}[-\log G(x)]^l/l!. \tag{5.5}$$

It is also easily seen that if Y_n^k is replaced by $Y_{[nt]}^k$ in (5.5), the limit is given by the same formula with $G(x)$ replaced throughout by $G(x)^t$. In view of the definition (4.1), this statement is the same as (5.1).

It is now very easy, using (5.3) and the results above, to prove that the transition probabilities of $\{Y_k^n\}$ converge to those given in (5.2). The convergence of the finite-dimensional distributions can then be deduced using the lemma of Section 4, for it is not hard to show that the transition probabilities converge to their limits uniformly in the initial state y, and the limits are continuous everywhere except at $y = z$ while $H_t^k(x)$ is continuous everywhere. We shall not go into more detail. Incidentally, when $k = 2$ the limiting process can also be obtained from Theorem 4.1, and the result is readily seen to agree with that above.

Final Remarks. When G is of the limiting type Φ_α, the limiting processes of Theorem 5.1 provide new examples of semi-stable Markov processes; like $\{m(t)\}$ they are jump processes with non-decreasing paths. (More precisely, a separable version has these properties.) It is worth noting that the invariance principles developed above in Section 3 apply also to the convergence of $\{y_{kn}(t)\}$ to $\{y_k(t)\}$ for $k > 1$. In fact Theorem 3.1 applies (subject to the same restrictions as in the maximum case) simply because $\{y_{kn}(t)\}$ has increasing paths, while for Theorem 3.2 an examination of the proof shows that the same approach can be carried through when $k > 1$. Thus in principle we obtain many new limit theorems for functionals of order statistics; which ones (if any) will prove useful remains to be seen.

REFERENCES

[1] Barndorff-Nielsen, O. (1963). On the limit behavior of extreme order statistics. *Ann. Math. Statist.* **34** 992–1002.

[2] Dwass, M. (1964). Extremal stochastic processes. *Ann. Math. Statist.* **35** 1718–1725.

[3] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Ann. of Math.* **44** 423–453.

[4] Lamperti, J. W. (1962). Semi-stable stochastic processes. *Trans. Amer. Math. Soc.* **104** 62–78.

[5] Prokhorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appl.* **1** 157–214. (English translation).

[6] Skorohod, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 261–289. (English translation).

[7] Smirnov, N. V. (1952). Limit distributions for the terms of a variational series. *Amer. Math. Soc. Transl.* No. 67.