References

[1] R. Bartoszyński and E. Pleszczyńska, *Reducibility of statistical structures and decision problems*, Preprint No 93, Inst. Math. of Polish. Acad. Sci., Warszawa 1976; also this volume pp. 29–38.

[2] R.L. Dobrushyn, *Systems of random variables defined by conditional distributions*, Theory Prob. Appl. 15 (1970), pp. 469–497 (in Russian).

[3] R.M. Dudley, *Distances of probability measures and random variables*, Ann. Math. Statist. 39 (1968), pp. 1563–1572.

[4] R. Engelking, *Topologia ogólna*, PWN, Warszawa 1976 (in Polish); *Outline of general topology*, Wiley, New York 1968.

[5] F.R. Hampel, *A qualitative definition of robustness*, Ann. Math. Statist. 42 (1971), pp. 1887–1896.

[6] —, *Robust estimation: A condensed partial survey*, Z. Warscheinlichkeitstheorie verw. Geb. 27 (1973), pp. 87–104.

[7] K. Kuratowski, *Topologie I* (second ed.), Warszawa-Wrocław 1948.

[8] L.D. Meshalkin, *On the robustness of some characterizations of the normal distribution*, Ann. Math. Statist. 39 (1968), pp. 1747–1750.

[9] K. Musiał, *Existence of proper conditional probabilities*, Z. Wahrscheinlichkeitstheorie verw. Geb. 22 (1972), pp. 8–12.

[10] J. Pfanzagl, *On the existence of regular conditional probabilities*, ibid. 11 (1969), pp. 244–256.

[11] Yu. V. Prokhorov, *Convergence of random processes and limit theorems in probability theory*, Theory Prob. Appl. 1 (1956), pp. 157–214.

[12] D.A. Sibley, *A metric for weak convergence of distribution functions*, Rocky Mountain J. of Math. 1 (1971), pp. 427–430.

[13] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. 36 (1965), pp. 423–439.

[14] S.M. Ulam, *A collection of mathematical problems*, Interscience Publishers, No 8, Inc., New York 1960.

[15] V.M. Zolotarev, *On the continuity of stochastic sequences generated by recurrent procedures*, Theory Prob. Appl. 20.4 (1975), pp. 834–847 (in Russian).

[16] —, *Stability effect of characterizations of probability distributions*, Zapisky nautchnyh seminarov LOMI. 61 (1976), pp. 38–55 (in Russian).

[17] —, *Metric distances in spaces of random variables and their distributions*, Math. Sbornik 101 (143) (1976), pp. 416–454 (in Russian).

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

ON WEAK CONVERGENCE OF SEQUENCES OF MEASURES

HARALD BERGSTRÖM

Chalmers University of Technology and University of Göteborg,
Department of Mathematics, Göteborg, Sweden

1. Introduction

In [1] A. D. Alexandroff has presented an exhaustive study of weak convergence of finite measures and finite-signed measures on normal spaces and completely normal spaces. In [2] I gave a survey of Alexandroff's theory, considering measures and I there also gave some applications to weak convergence of stochastic processes into the C-space and the D-space. Here, using these theorems, I shall give a more complete presentation of the theory of weak convergence of probability measures on the C-space and D-space.

Alexandroff's main tools are linear functionals. Since weak convergence means convergence of linear functionals, it seems natural to rely heavily upon linear functionals throughout in the theory.

Alexandroff has two main theorems which I here state for measures. Here a *measure* means a finitely additive non-negative set function on an algebra in a σ-topological space. A measure is called *σ-smooth* if it is σ-additive.

THEOREM 1.1 (Alexandroff's first theorem). *Let ψ be the Stone vector lattice of bounded continuous functions from a normal σ-topological space S into the real number field $\boldsymbol{R}$ and L a non-negative bounded linear functional from ψ into $\boldsymbol{R}$. Then L determines uniquely a regular measure μ on the algebra generated by the closed sets and μ satisfies the relation*

$$L(f) = \int_S f(x)\mu(dx), \quad f \in \psi. \tag{i}$$

COROLLARY. *For a metric σ-topological space Theorem* 1.1 *remains true if $\bar{\psi}$ is changed into the Stone vector lattice ψ of uniformly continuous functions from S into $\boldsymbol{R}$ and* (i) *still holds for $f \in \psi$.*

THEOREM 1.2 (Alexandroff's second theorem). *Let S be a completely normal space and $\mathscr{S}$ the algebra generated by the closed sets. Let $\{\mu_n\}_{n=1}^{+\infty}$ be a sequence of σ-smooth finite measures on S. If $\{\mu_n\}_{n=1}^{+\infty}$ converges weakly to a measure μ on $\mathscr{S}$, then μ is σ-smooth.*

The first theorem is not difficult to prove. The second one is much deeper and requires results from the set theory and the functional analysis which, though elementary, are rather complicated.

2. The C-space and the D-space

We shall apply Alexandroff's theorems to the space C of real-valued bounded continuous functions on $[0,1]$ and the space D of all real-valued bounded functions x on $[0,1]$ such that $x(t-)$ exists (finite) for $t\in(0,1]$, $x(t+)$ exists finite and $x(t) = x(t+)$ for $t\in[0,1)$.

In the C-space we use the uniform metric

$$||x-y|| = \sup_{t\in[0,1]} |x(t)-y(t)|$$

and in the D-space the simple Skorohod metric defined by

$$\varrho(x,y) = \inf_{\lambda\in\Lambda} \max\{||x-y\circ\lambda||,\ ||\lambda-\bar{\varepsilon}||\} \tag{2.1}$$

where Λ is the set of continuous strictly increasing functions λ (on $[0,1]$) with $\lambda(0)=0$, $\lambda(1)=1$, and $\bar{\varepsilon}$ is the identity function $\bar{\varepsilon}(t)=t$ on $[0,1]$.

We shall represent the functions in the C-space by infinite series of Schauder functions and the functions in the D-space by sequences of modified Schauder functions and related sequences. The system of Haar functions on $[0,1]$ is given by

$$f_0(t) = 1,\qquad f_{k2^{-n}}(t) = \begin{cases} 2^{(n-1)/2} & \text{for}\quad t\in[(k-1)2^{-n}, k2^{-n}),\\ -2^{(n-1)/2} & \text{for}\quad t\in[k2^{-n}, (k+1)2^{-n}),\\ 0 & \text{otherwise},\end{cases}$$

and for k odd $<2^n$ $(k>0)$. Denote by $\hat{N}$ the set of all points $k2^{-n}$, k odd $<2^n$, $k>0$, $n=1,2,\ldots$, together with 0 and 1, and by $\hat{N}_r$ the set of the points $k2^{-n}$, k odd $<2^n$, $k>0$, $n=1,2,\ldots,r$, together with 0 and 1. Then the Haar functions can be denoted by f_ν, $\nu\in\hat{N}$ with $f_1\equiv 0$. The Schauder functions g_ν are the integrated Haar functions

$$g_\nu(t) = \int_0^t f_\nu(\tau)\,d\tau.$$

The formal Schauder series of a function x on $[0,1]$ is then given by (2.1)

$$x(t) = x(0)+\sum_{\nu\in\hat{N}} c_\nu g_\nu(t)$$

with

$$c_0 = x(1),\qquad c_{k2^{-n}} = 2^{(n-1)/2}[2x(k2^{-n})-x((k+1)2^{-n})-x((k-1)2^{-n})].$$

Besides (2.1) we consider the partial sums

$$x^{(r)}(t) = x(0)+\sum_{\nu\in\hat{N}_r} c_\nu g_\nu(t). \tag{2.2}$$

Generally, the expansion (2.1) is only formal but we have the following classical result ([4], p. 50).

THEOREM 2.1 *For any real-valued function x the relation $x(t)=x^{(r)}(t)$ holds for $t\in\hat{N}_r$. Series* (2.1) *is convergent and equal to $x(t)$ at any continuity point t of x, and uniformly convergent if and only if x is continuous on* $[0,1]$, *i.e. if $x\in C$.*

Let $T=(t_1,t_2,\ldots,t_j)$, $t_1<t_2<\ldots<t_j$ be a set of points on $[0,1]$. The vector $\tilde{x}^T=[x(t_1),\ldots,x(t_j)]$ is called the *projection* of the real-valued function x. Introducing the projection operator $\pi(T)$, we write

$$\tilde{x}^{(T)} = \pi(T)x.$$

For given T these projections, $x\in C$, generate a vector space C^T and clearly $\pi(T)$ maps C onto C^T. For the particular projections belonging to $T=\hat{N}_r$ we use the notation $\pi^{(r)}$ and write $C^{(r)}$ instead of C^T. We also consider the mapping $V^{(r)}$ of $C^{(r)}$ into C defined by

$$V^{(r)}\tilde{x}^{(r)} = x^{(r)}$$

where $x^{(r)}$ is given by (2.2). Then $V^{(r)}\pi^{(r)}$ is a mapping of C into C. The following lemma is easy to prove.

LEMMA 2.1. *$\pi^{(r)}$ is a continuous mapping of C onto $C^{(r)}$ and $V^{(r)}$ a continuous mapping of $C^{(r)}$ into C_j, consequently $V^{(r)}\pi^{(r)}$ is a continuous mapping of C into C.*

Put

$$d_r(x) = ||x-x^{(r)}||.$$

LEMMA 2.2. *The function d_r is a uniformly continuous mapping of C into $\boldsymbol{R}$.*

Also this lemma is easily verified. Clearly, $x\in D$ cannot, like $x\in C$, always be well approximated by the continuous functions $x^{(r)}$ since $x\in D$ may have jumps. However, defining

$$\hat{x}^{(r)}(t) = x^{(r)}(t^{(r)}) = x(t^{(r)})$$

where $t^{(r)}$ is the number in $\hat{N}_r$ closest to the left of t ($t^{(s)}=t$ if $t\in\hat{N}_r$), we get a pure step function which agrees with x at the point $t\in N_r$. We may use more general step functions, formed in the same way as $\hat{x}^{(r)}$, as approximations of x. Let $T^{(r)}$ be a set of finitely many different points $t_j^{(r)}$ on $[0,1]$,

$$0 = t_0^{(r)} < t_1^{(r)} < \ldots < t_{j_r}^{(r)} = 1,\qquad r = 1,2,\ldots,$$

$$\max_{j=1,\ldots,j_r} t_j^{(r)}-t_{j-1}^{(r)} = h(r).$$

We call $T^{(r)}$ a *net* of points on $[0,1]$ and say that $T^{(r+1)}$ is a *refinement of* $T^{(r)}$ if all points in $T^{(r)}$ belong to $T^{(r+1)}$. We shall only consider nets $T^{(r)}$ such that $T^{(r+1)}$ is a refinement of $T^{(r)}$ for any $r=1,2,\ldots$ and also assume that $h(r)\to 0$ $(r\to+\infty)$.

Note that then $\bigcup_{r=1}^{+\infty} T^{(r)}$ is a dense set on [0, 1]. We call $\{T^{(r)}\}$ a *dense increasing sequence of nets*.

Clearly, $T^{(r)}$ determines a projection $\hat{\pi}(T^{(r)})$ which projects D onto a space $D^{(r)}$ of vectors $\{x(t_j^{(r)})\}$. We write shorter $\hat{\pi}^{(r)}$ instead of $\hat{\pi}(T^{(r)})$ also in the general case where $T^{(r)}$ is not necessarily $\hat{N}_r$ and put $\hat{\pi}^{(r)}x = \tilde{x}^{(r)}$, $\hat{\pi}^{(r)}D = D^{(r)}$. Further we define the mapping $\hat{V}^{(r)}$ from $D^{(r)}$ into D by $\hat{V}_r\tilde{x}^{(r)} = \hat{x}^{(r)}$ where $\hat{x}^{(r)}$ is the step function equal to $x(t)$ for $t \in T^{(r)}$ and equal to $x(t')$ for $t' \leqslant t < t''$ where t' and t'' are neighbour points in $T^{(r)}$.

LEMMA 2.3. *The projection $\hat{\pi}^{(r)}$ of D onto $D^{(r)}$ is measurable and the mapping $\hat{V}^{(r)}$ of $D^{(r)}$ into D is continuous. It is possible to choose a sequence $\{\hat{\pi}^{(r)}\}$ to a given probability measure μ on D such that $\hat{\pi}^{(r)}$ is almost surely continuous (μ) for $r = 1, 2, \ldots$*

Proof. The fact that $\hat{V}^{(r)}$ is continuous easily follows. Indeed, if for a sequence $\{x_n^{(r)}\}$ in $D^{(r)}$, $\tilde{x}_n^{(r)} \to \tilde{x}^{(r)}$ $(n \to +\infty)$, $\tilde{x}^{(r)} \in D^{(r)}$, then $\varrho(\hat{x}_n^{(r)}, \hat{x}^{(r)}) \to 0$ $(n \to +\infty)$ since $\varrho(\hat{x}_n^{(r)}, \hat{x}^{(r)}) \leqslant \|\hat{x}_n^{(r)} - \hat{x}^{(r)}\|$. The statements about $\hat{\pi}^{(r)}$ can be proved as in [3], p. 121.

Now put

$$\hat{d}^{(r)}(x) = \varrho(x, \hat{x}^{(r)}).$$

Then we have

LEMMA 2.4. *$\hat{d}^{(r)}$ is a measurable function from D into $\mathbf{R}$ and $\hat{d}^{(r)}(x) \to 0$ $(r \to +\infty)$. For a probability measure μ on D the sequence of projections $\hat{\pi}^{(r)}$ may be chosen such that $\hat{d}_r$ is almost surely continuous (μ) for $r = 1, 2, \ldots$*

Proof. The fact that $\hat{d}^{(r)}(x) \to 0$ $(r \to +\infty)$ easily follows.

Indeed, to $x \in D$ there exist points τ_j, $0 = \tau_0 < \tau_1 < \ldots < \tau_m = 1$ for any given number $\varepsilon > 0$, such that the oscillation of x on any open interval (τ_i, τ_{i+1}) is smaller than ε. Take r sufficiently large and let $\tau_i^{(r)}$ be the point in $T^{(r)}$ closest to the left of τ_i. For sufficiently large r there are points $t_{j-3}^{(j)}, t_{j-2}^{(j)}, t_0^{(j)}, t_{j+1}^{(j)}, t_{j+2}^{(j)}$ in $T^{(r)}$ such that $\tau_{j-1} < t_{j-3}^{(j)} < t_{j-2}^{(j)} < t_{j-1}^{(j)} < t_j^{(j)} \leqslant \tau_j < t_{j+1}^{(j)} < t_{j+2} < t_{j+3} < \tau_{j+1}$ for $0 < j < j+1 < 1$, where $t_{j-3}^{(j)}, \ldots, t_{j+3}^{(j)}$ are the closest points to τ_j satisfying these inequalities. Define $\lambda(\tau_j) = \tau_j^{(r)} = t_j^{(r)}$, $\lambda(0) = 0$, $\lambda(1) = 1$, $\lambda(t_{j-2}) = t_{j-2}$, $\lambda(t_{j+2}) = t_{j+2}$, λ continuous and linear on $[t_{j-2}, \tau_j]$ and $[\tau_j, t_{j+2}]$. In this way λ will be defined on certain intervals. On the remaining intervals we define $\lambda(t) = t$. It then follows that

$$|x(\lambda^{(r)}(t)) - \hat{x}^{(r)}(t)| < \varepsilon, \qquad |\lambda^{(r)}(t) - t| \leqslant h(r)$$

where $h(r) \to 0$, $r \to +\infty$ for any $\varepsilon > 0$. Thus $\varrho(x, \hat{x}^{(r)}) \to 0$, $r \to +\infty$, according to definition (2.1) of ϱ.

By Lemma 2.3, $\hat{\pi}^{(r)}$ is measurable for any sequence of projections considered there, and it is possible to choose the sequence for a given μ such that $\hat{\pi}^{(r)}$ is almost continuous μ for all r. Now consider the mapping

$$\tilde{x}^{(r_1)} \to \varrho(\hat{x}^{(r_1)}, \hat{x}^{(r)}), \qquad r_1 > r, \qquad \tilde{x}^{(r)} = \hat{\pi}^{(r)}x, \qquad \hat{x}^{(r)} = \hat{V}^{(r)}\tilde{x}^{(r)}. \tag{2.2}$$

We claim that this mapping is continuous. Indeed, if $\tilde{x}_n^{(r_1)} \to \tilde{x}^{(r_1)}$ $(n \to +\infty)$, then $x_n(t) \to x(t)$ for those t which determine $\hat{\pi}^{(r_1)}$ and thus since $r_1 > r$,

$$|\varrho(\hat{x}^{(r_1)}, \hat{x}^{(r)}) - \varrho(\hat{x}_n^{(r_1)}, \hat{x}_n^{(r)})| \leqslant \varrho(\hat{x}^{(r_1)}, \hat{x}_n^{(r_1)}) + \varrho(\hat{x}^{(r)}, \hat{x}_n^{(r)})$$
$$\leqslant \|\hat{x}^{(r_1)} - \hat{x}_n^{(r_1)}\| + \|\hat{x}^{(r)} - \hat{x}_n^{(r)}\| \to 0 \qquad (n \to +\infty).$$

Since the mapping $x \to \hat{x}^{(r_1)}$ is measurable and the mapping (2.2) is continuous and thus measurable, the mapping $x \to \varrho(x^{(r_1)}, x^{(r)})$ is measurable for any $r_1 > r$. Letting $r_1 \to +\infty$ and observing that

$$|\varrho(x, \hat{x}^{(r)}) - \varrho(\hat{x}^{(r_1)}, \hat{x}^{(r)})| \leqslant \varrho(x, x^{(r_1)}) = \hat{d}^{(r_1)}(x) \to 0 \ (r_1 \to +\infty),$$

we conclude that the mapping $x \to \varrho(x, \hat{x}^{(r)})$ is measurable.

Let now μ be a probability measure and let $\hat{\pi}^{(r)}$ be chosen such that it is almost surely continuous for $r = 1, 2, \ldots$ This means that there exists a subset $D' \subset D$ with $\mu(D') = 1$ such that for any $x \in D'$ and any sequence $\{x_n\}_{n=1}^{+\infty}$ with $x_n \in D'$ and $\varrho(x_n, x) \to 0$ $(n \to +\infty)$ we have $\hat{\pi}^{(r)}x_n \to \hat{\pi}^{(r)}x (n \to +\infty)$. Since $\hat{V}^{(r)}$ is a continuous mapping of $D^{(r)}$ into D, this implies

$$\hat{x}_n^{(r)} = \hat{V}^{(r)}\hat{\pi}^{(r)}x_n \to \hat{V}^{(r)}\hat{\pi}^{(r)}x = \hat{x}^{(r)} \qquad (n \to +\infty).$$

Hence

$$|\varrho(x_n\hat{x}_n^{(r)}) - \varrho(x, \hat{x}^{(r)})| \leqslant \varrho(x_n, x) + \varrho(\hat{x}_n^{(r)}, \hat{x}^{(r)}) \to 0 \qquad (n \to +\infty),$$

which proves that $\hat{d}^{(r)}(x) = \varrho(x, \hat{x}^{(r)})$ is almost surely continuous (μ).

3. Weak convergence in the *C*-space and the *D*-space

Consider a sequence $\{\mu_m\}_{m=1}^{+\infty}$ of probability measures on the C-space or the D-space. For these spaces we use the same metrics as in Section 2. We now give our main theorem for the D-space and the corresponding theorem for the C-space as a corollary.

THEOREM 3.1. *Let $\{\mu_m\}_{m=1}^{+\infty}$ be a sequence of probability measures on D and $\{\hat{\pi}^{(r)}\}_{r=1}^{+\infty}$ a sequence of projections belonging to an increasing dense sequence $\{T^{(r)}\}$ of nets, defined in Section 2. Suppose*

(i) *that the sequence of probability measures $\mu_m^{(r)} = \mu_m(\hat{\pi}^{(r)-1}\cdot)$ for any $r = 1, 2, \ldots$, converges weakly to a measure $\mu^{(r)}$ on $D^{(r)} = \hat{\pi}^{(r)}D$ as $m \to +\infty$,*

(ii) $\lim\limits_{r \to +\infty} \limsup\limits_{m \to +\infty} \mu_m\{x: \hat{d}_r(x) \geqslant \varepsilon\} = 0;$

then $\{\mu_m\}_{m=1}^{+\infty}$ converges to a probability measure μ on D. Conversely, if $\{\mu_m\}_{m=1}^{+\infty}$ converges weakly to a probability measure μ on D, then sequences $\hat{\pi}^{(r)}$ for suitable nets $T^{(r)}$ for μ satisfy conditions (i) *and* (ii) *and then $\{\mu_m^{(r)}\}_{m=1}^{+\infty}$ converges weakly to $\mu(\hat{\pi}^{(r)-1}(\cdot))$ (which necessarily are probability measures).*

Remark. Nets $T^{(r)}$ are suitable for μ if the corresponding mappings $\hat{\pi}^{(r)}$ are almost surely continuous.

Corollary. *The theorem holds true for the C-space and for the projections $\pi^{(r)}$ which are continuous.*

Proof. Let $\overline{\psi}$ be the Stone vector lattice of bounded uniformly continuous functions from D into $\boldsymbol{R}$ and let $\hat{\pi}^{(r)}$ and $\hat{V}^{(r)}$ be the mappings defined in Section 2.

(a) Suppose that (i) and (ii) are satisfied and put

$$L_m^{(r)}(f) = \int_D f(\hat{V}^{(r)}\hat{\pi}^{(r)}x)\mu_m(dx) \qquad (f \in \overline{\psi}), \tag{3.1}$$

$$L_m(f) = \int_D f(x)\mu_m(dx) \qquad (f \in \overline{\psi}), \tag{3.2}$$

$$L_m^{(r)}(f) = \int_{D^{(r)}} f(\hat{V}^{(r)}\tilde{x}^{(r)})\mu_m^{(r)}(d\tilde{x}^{(r)}), \tag{3.3}$$

and, by assumption (i), let $\mu_m^{(r)}$ converge weakly to $\mu^{(r)}$ as $m \to +\infty$. Since $v^{(r)}$ is a continuous mapping of $D^{(r)}$ into D, we thus get

$$\lim_{m\to+\infty} L_m^{(r)}(f) = L^{(r)}(f) = \int_{D^{(r)}} f(\hat{V}^{(r)}\tilde{x}^{(r)})\mu^{(r)}(d\tilde{x}^{(r)}). \tag{3.4}$$

We now use the obvious inequality

$$|L_{m_1}(f)-L_{m_2}(f)| \leqslant |L_{m_1}(f)-L_{m_1}^{(r)}(f)|+|L_{m_1}^{(r)}(f)-L_{m_2}^{(r)}(f)|+|L_{m_2}^{(r)}(f)-L_{m_2}(f)|. \tag{3.5}$$

We have

$$|L_m(f)-L_m^{(r)}(f)| = \left|\int_D [f(x)-f(x^{(r)})]\mu_m(dx)\right| \tag{3.6}$$

$$\leqslant \sup_{\varrho(x,\hat{x}^{(r)})<\varepsilon} |f(x)-f(\hat{x}^{(r)})|+2\|f\| \int_{\varrho(x,\hat{x}^{(r)})\geqslant\varepsilon} \mu_m(dx).$$

By (ii)

$$\int_{\varrho(x,\hat{x}^{(r)})\geqslant\varepsilon} \mu_m(dx) = \mu_m\{x\colon d_r(x) \geqslant \varepsilon\}$$

tends to 0 as first $m \to +\infty$ and then $r \to +\infty$. This is true for any $\varepsilon > 0$. The first term on the right-hand side of (3.6) is arbitrarily small for sufficiently small $\varepsilon > 0$, since f is uniformly continuous. Hence

$$\lim_{r\to+\infty} \limsup_{m\to=\infty} |L_m(f)-L_m^{(r)}(f)| = 0. \tag{3.7}$$

Since $\mu_m^{(r)}$ converges weakly to $\mu^{(r)}$, we find by (3.3) and the definition of weak convergence

$$\lim_{\substack{m_1\to+\infty\\ m_2\to+\infty}} |L_{m_1}^{(r)}(f)-L_{m_2}^{(r)}(f)| = 0 \tag{3.8}$$

for $r = 1, 2, \ldots$ Combining (3.5), (3.7) and (3.8), we get

$$\lim_{\substack{m_1\to+\infty\\ m_2\to+\infty}} |L_{m_1}(f)-L_{m_2}(f)| = 0 \quad \text{for} \quad f \in \overline{\psi}.$$

Hence

$$L(f) = \lim_{m\to+\infty} L_m(f), \quad f \in \overline{\psi}, \tag{3.9}$$

exists and clearly L is a bounded non-negative linear functional. By Alexandroff's first theorem it defines a measure μ such that

$$L(f) = \int f(x)\mu(dx), \quad f \in \overline{\psi},$$

and by Alexandroff's second theorem μ is a σ-smooth measure since the μ_m are σ-smooth. Choosing $f(x) \equiv 1$ in (3.9), we find that μ is a probability measure since the μ_m are probability measures.

(b) Suppose that $\{\mu_m\}_{m=1}^{+\infty}$ converges weakly to a probability measure μ on D. Since $\hat{V}^{(r)}$ is a continuous mapping of D onto $D^{(r)}$, $\hat{V}^{(r)-1}F' = F$ is a closed set for any closed set F' in $D^{(r)}$. Then the relation

$$\limsup_{m\to+\infty} \mu_m(\hat{V}^{(r)-1}F') \leqslant \mu(\hat{V}^{(r)-1}F')$$

is a consequence of the weak convergence of μ_m to μ. On the other hand, this relation for any closed set F' in $D^{(r)}$ implies the weak convergence of $\mu_m(\hat{V}^{(r)-1}(\cdot))$ to $\hat{\mu}(\hat{V}_m^{(r)-1}(\cdot))$. [This holds true for normal spaces. For metric spaces see [3], p. 12.]

Now, define g_ε as that continuous function which is equal to 0 on $(-\infty, 0]$, equal to 1 on $t \geqslant \varepsilon$, linear on $[0, \varepsilon]$. Choose $T^{(r)}$ such that $\hat{\pi}^{(r)}$ is almost surely continuous. Since $\hat{d}_r(x) = \varrho(x, \hat{x}^{(r)})$, then by Lemma 2.4 $\hat{d}_r$ is an almost surely continuous function from D into $\boldsymbol{R}$, the function $f_r(x) = g_\varepsilon(\hat{d}_r(x))$ is a bounded almost surely continuous function from D into $\boldsymbol{R}$ and $0 \leqslant f_\varepsilon(x) \leqslant 1$ for all x, $f_r(x) = 1$ for $\hat{d}_r(x) \geqslant \varepsilon$. Hence

$$\mu_m[\hat{d}_r(x) \geqslant \varepsilon] \leqslant \int_D f_r(x)\mu_m(dx) \to \int_D f_r(x)\mu(dx) \qquad (m \to +\infty).$$

But $f_r(x) \to 0$ $(r \to +\infty)$ since $d_r(x) \to 0$ $(r \to +\infty)$. Thus

$$\lim_{r\to+\infty} \limsup_{m\to+\infty} \mu_m[x\colon \hat{d}_r(x) \geqslant \varepsilon] = 0$$

for any $\varepsilon > 0$.

Thus we have proved the theorem and the corollary is obtained by this proof.

References

[1] A. D. Alexandroff, *Additive set-functions in abstract spaces*, (a) Mat. sb. 8 (1940), pp. 307–348; (b) Mat. sb. 9 (1941), pp. 563–628; (c) Mat. sb. 13 (1943), pp. 169–238.

[2] H. Bergström, *On weak convergence in normal spaces*, To appear in Symposia Mathematica Vol. XXI (Roma).

[3] P. Billingsley, *Convergence of probability measures*, John Wiley & Sons Inc., New York, London, Sydney, Toronto 1968.
[4] S. Kaczmarz, H. Steinhaus, *Theorie der Ortogonalreihen*, Chelsea Publishing Company 1951.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

SERIATION WITH APPLICATIONS IN PHILOLOGY

L. BONEVA

Institute of Mathematics and Mechanics, Bulgarian Academy of Sciences, Sofia, Bulgaria

1. Introduction

Seriation has turned out to be a common problem not only in archaeology but in philology and other fields as well. Generally speaking, the basic idea of the recently developed new mathematical, statistical and computing methods connected with seriation was the reconstruction of the "true" chronological order of a set of objects using only the available nonmetric information about the similarities (or dissimilarities) between pairs of objects. These methods have done a good service to all problems dealing with a great amount of data for numerous objects about which only a chronological ordering is needed. We are going to discuss here the SKK-*method*, which we call so in honour of the names of the three most famous men (Shepard–Kruskal–Kendall) who took part in creating the "main body" of this useful technique.

In fact, the seriation problem was formulated for the first time by the English archaeologist Flinders Petrie [19] at the very end of the last century. He was confronted with a very difficult problem — to find an approximate dating for 4000 prehistoric Egyptian graves, each containing pottery, jewellery and other objects permitting a final classification into types of varieties. Evidently, a chronological trend of these types is to be expected according to which the approximate dating of the graves might be done. Actually, Petrie managed to arrange 900 graves containing a total amount of 800 varieties. The weakest point of his laborious work is the "reverse connection" between graves and varieties, i.e. the varieties were classified according to the graves in which they were found, while the graves were ordered according to the varieties they contained. However, he is to be thanked for the so-called "Petrie's Concentration Principle", which shortly states that the more close together in temporal order two graves are the more likely they are to contain varieties of the same or similar types.

A second merit of Petrie's work should not be omitted. It is he who gave the initial impulse (though it resounded about 50 years later) to many mathematicians, such as Robinson [20], Shepard [23], Kruskal [17], [18], Kendall [10]–[16], Sibson