

1. THE PROHOROV METRIC

Throughout Sections 1–4, (S, d) is a metric space (d denoting the metric), $\mathcal{B}(S)$ is the σ -algebra of Borel subsets of S , and $\mathcal{P}(S)$ is the family of Borel probability measures on S . We topologize $\mathcal{P}(S)$ with the Prohorov metric

$$(1.1) \quad \rho(P, Q) = \inf \{ \varepsilon > 0: P(F) \leq Q(F^\varepsilon) + \varepsilon \text{ for all } F \in \mathcal{C} \},$$

where $\mathcal{C}$ is the collection of closed subsets of S and

$$(1.2) \quad F^\varepsilon = \left\{ x \in S: \inf_{y \in F} d(x, y) < \varepsilon \right\}.$$

To see that ρ is a metric, we need the following lemma.

1.1 Lemma Let $P, Q \in \mathcal{P}(S)$ and $\alpha, \beta > 0$. If

$$(1.3) \quad P(F) \leq Q(F^\alpha) + \beta$$

for all $F \in \mathcal{C}$, then

$$(1.4) \quad Q(F) \leq P(F^\alpha) + \beta$$

for all $F \in \mathcal{C}$.

Proof. Given $F_1 \in \mathcal{C}$, let $F_2 = S - F_1^\alpha$, and note that $F_2 \in \mathcal{C}$ and $F_1 \subset S - F_2^\alpha$. Consequently, by (1.3) with $F = F_2$,

$$(1.5) \quad P(F_1^\alpha) = 1 - P(F_2) \geq 1 - Q(F_2^\alpha) - \beta \geq Q(F_1) - \beta,$$

implying (1.4) with $F = F_1$. □

It follows immediately from Lemma 1.1 that $\rho(P, Q) = \rho(Q, P)$ for all $P, Q \in \mathcal{P}(S)$. Also, if $\rho(P, Q) = 0$, then $P(F) = Q(F)$ for all $F \in \mathcal{C}$ and hence for all $F \in \mathcal{B}(S)$; therefore, $\rho(P, Q) = 0$ if and only if $P = Q$. Finally, if $P, Q, R \in \mathcal{P}(S)$, $\rho(P, Q) < \delta$, and $\rho(Q, R) < \varepsilon$, then

$$(1.6) \quad \begin{aligned} P(F) &\leq Q(F^\delta) + \delta \leq Q(\overline{F^\delta}) + \delta \\ &\leq R((\overline{F^\delta})^\varepsilon) + \delta + \varepsilon \leq R(F^{\delta+\varepsilon}) + \delta + \varepsilon \end{aligned}$$

for all $F \in \mathcal{C}$, so $\rho(P, R) \leq \delta + \varepsilon$, proving the triangle inequality.

The following theorem provides a probabilistic interpretation of the Prohorov metric when S is separable.

1.2 Theorem Let (S, d) be separable, and let $P, Q \in \mathcal{P}(S)$. Define $\mathcal{M}(P, Q)$ to be the set of all $\mu \in \mathcal{P}(S \times S)$ with marginals P and Q (i.e., $\mu(A \times S) = P(A)$ and $\mu(S \times A) = Q(A)$ for all $A \in \mathcal{B}(S)$). Then

$$(1.7) \quad \rho(P, Q) = \inf_{\mu \in \mathcal{M}(P, Q)} \inf \{ \varepsilon > 0: \mu\{(x, y): d(x, y) \geq \varepsilon\} \leq \varepsilon \}.$$

Proof. If for some $\varepsilon > 0$ and $\mu \in \mathcal{M}(P, Q)$ we have

$$(1.8) \quad \mu\{(x, y): d(x, y) \geq \varepsilon\} \leq \varepsilon,$$

then

$$(1.9) \quad \begin{aligned} P(F) &= \mu(F \times S) \\ &\leq \mu((F \times S) \cap \{(x, y): d(x, y) < \varepsilon\}) + \varepsilon \\ &\leq \mu(S \times F^c) + \varepsilon = Q(F^c) + \varepsilon \end{aligned}$$

for all $F \in \mathcal{C}$, so $\rho(P, Q)$ is less than or equal to the right side of (1.7).

The reverse inequality is an immediate consequence of the following lemma. $\square$

1.3 Lemma Let S be separable. Let $P, Q \in \mathcal{P}(S)$, $\rho(P, Q) < \varepsilon$, and $\delta > 0$. Suppose that $E_1, \dots, E_N \in \mathcal{B}(S)$ are disjoint with diameters less than δ and that $P(E_0) \leq \delta$, where $E_0 = S - \bigcup_{i=1}^N E_i$. Then there exist constants $c_1, \dots, c_N \in [0, 1]$ and independent random variables $X, Y_0, \dots, Y_N$ (S -valued) and ξ ($[0, 1]$ -valued) on some probability space $(\Omega, \mathcal{F}, \nu)$ such that X has distribution P , ξ is uniformly distributed on $[0, 1]$,

$$(1.10) \quad Y = \begin{cases} Y_i & \text{on } \{X \in E_i, \xi \geq c_i\}, \quad i = 1, \dots, N, \\ Y_0 & \text{on } \{X \in E_0\} \cup \bigcup_{i=1}^N \{X \in E_i, \xi < c_i\} \end{cases}$$

has distribution Q ,

$$(1.11) \quad \{d(X, Y) \geq \delta + \varepsilon\} \subset \{X \in E_0\} \cup \left\{ \xi < \max \left[\frac{\varepsilon}{P(E_i)} : P(E_i) > 0 \right] \right\},$$

and

$$(1.12) \quad \nu\{d(X, Y) \geq \delta + \varepsilon\} \leq \delta + \varepsilon.$$

The proof of this lemma depends on another lemma.

1.4 Lemma Let μ be a finite positive Borel measure on S , and let $p_i \geq 0$ and $A_i \in \mathcal{B}(S)$ for $i = 1, \dots, n$. Suppose that

$$(1.13) \quad \sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A_i\right) \quad \text{for all } I \subset \{1, \dots, n\}.$$

Then there exist positive Borel measures $\lambda_1, \dots, \lambda_n$ on S such that $\lambda_i(A_i) = \lambda_i(S) = p_i$ for $i = 1, \dots, n$ and $\sum_{i=1}^n \lambda_i(A) \leq \mu(A)$ for all $A \in \mathcal{B}(S)$.

Proof. Note first that it involves no loss of generality to assume that each $p_i > 0$.

We proceed by induction on n . For $n = 1$, define λ_1 on $\mathcal{B}(S)$ by $\lambda_1(A) =$

$p_1\mu(A \cap A_1)/\mu(A_1)$. Then $\lambda_1(A_1) = \lambda_1(S) = p_1$, and since $p_1 \leq \mu(A_1)$ by (1.13), we have $\lambda_1(A) \leq \mu(A \cap A_1) \leq \mu(A)$ for all $A \in \mathcal{B}(S)$. Suppose now that the lemma holds with n replaced by m for $m = 1, \dots, n-1$ and that μ , p_i , and A_i ($1 \leq i \leq n$) satisfy (1.13). Define η on $\mathcal{B}(S)$ by $\eta(A) = \mu(A \cap A_n)/\mu(A_n)$, and let ε_0 be the largest ε such that

$$(1.14) \quad \sum_{i \in I} p_i \leq (\mu - \varepsilon\eta)\left(\bigcup_{i \in I} A_i\right) \quad \text{for all } I \subset \{1, \dots, n-1\}.$$

CASE 1. $\varepsilon_0 \geq p_n$. Let $\lambda_n = p_n\eta$ and put $\mu' = \mu - \lambda_n$. Since $p_n \leq \mu(A_n)$ by (1.13), μ' is a positive Borel measure on S , so by (1.14) (with $\varepsilon = p_n$) and the induction hypothesis, there exist positive Borel measures $\lambda_1, \dots, \lambda_{n-1}$ on S such that $\lambda_i(A_i) = \lambda_i(S) = p_i$ for $i = 1, \dots, n-1$ and $\sum_{i=1}^{n-1} \lambda_i(A) \leq \mu'(A)$ for all $A \in \mathcal{B}(S)$. Also, $\lambda_n(A_n) = \lambda_n(S) = p_n$, so $\lambda_1, \dots, \lambda_n$ have the required properties.

CASE 2. $\varepsilon_0 < p_n$. Put $\mu' = \mu - \varepsilon_0\eta$, and note that μ' is a positive Borel measure on S . By the definition of ε_0 , there exists $I_0 \subset \{1, \dots, n-1\}$ (nonempty) such that

$$(1.15) \quad \sum_{i \in I} p_i \leq \mu'\left(\bigcup_{i \in I} A_i\right) \quad \text{for all } I \subset I_0$$

with equality holding for $I = I_0$. By the induction hypothesis, there exist positive Borel measures λ_i on S , $i \in I_0$, such that $\lambda_i(A_i) = \lambda_i(S) = p_i$ for each $i \in I_0$ and $\sum_{i \in I_0} \lambda_i(A) \leq \mu'(A)$ for all $A \in \mathcal{B}(S)$. Let $p'_i = p_i$ for $i = 1, \dots, n-1$ and $p'_n = p_n - \varepsilon_0$. Put $B_0 = \bigcup_{i \in I_0} A_i$, define μ'' on $\mathcal{B}(S)$ by $\mu''(A) = \mu'(A) - \mu'(A \cap B_0)$, and let $I_1 = \{1, \dots, n\} - I_0$. Then, for all $I \subset I_1$,

$$(1.16) \quad \begin{aligned} \sum_{i \in I} p'_i + \mu'(B_0) &= \sum_{i \in I \cup I_0} p'_i \\ &\leq \mu'\left(\bigcup_{i \in I \cup I_0} A_i\right) \\ &= \mu'\left(\bigcup_{i \in I} A_i\right) + \mu'(B_0) - \mu'\left(\bigcup_{i \in I} A_i \cap B_0\right) \\ &= \mu''\left(\bigcup_{i \in I} A_i\right) + \mu'(B_0). \end{aligned}$$

Here, equality in the first line holds because equality in (1.15) holds for $I = I_0$, while the inequality in the second line follows from (1.14) if $n \notin I$ and from (1.13) if $n \in I$; more specifically, if $n \in I$, then

$$(1.17) \quad \sum_{i \in I \cup I_0} p'_i \leq \mu\left(\bigcup_{i \in I \cup I_0} A_i\right) - \varepsilon_0 = \mu'\left(\bigcup_{i \in I \cup I_0} A_i\right).$$

By (1.16),

$$(1.18) \quad \sum_{i \in I} p_i \leq \mu''\left(\bigcup_{i \in I} A_i\right) \quad \text{for all } I \subset I_1,$$

so by the induction hypothesis, there exist positive Borel measures λ'_i on S , $i \in I_1$, such that $\lambda'_i(A_i) = \lambda'_i(S) = p_i$ for each $i \in I_1$ and $\sum_{i \in I_1} \lambda'_i(A) \leq \mu''(A)$ for all $A \in \mathcal{B}(S)$. Finally, let $\lambda_i = \lambda'_i$ for $i \in I_1 - \{n\}$ and $\lambda_n = \lambda'_n + \varepsilon_0 \eta$. Then $\lambda_i(A_i) = \lambda_i(S) = p_i$ for each $i \in I_1$, hence for $i = 1, \dots, n$, and

$$\begin{aligned} (1.19) \quad \sum_{i=1}^n \lambda_i(A) &= \sum_{i \in I_0} \lambda_i(A) + \sum_{i \in I_1} \lambda_i(A) \\ &= \sum_{i \in I_0} \lambda_i(A \cap B_0) + \sum_{i \in I_1} \lambda'_i(A) + \varepsilon_0 \eta(A) \\ &\leq \mu'(A \cap B_0) + \mu''(A) + \varepsilon_0 \eta(A) \\ &= \mu'(A) + \varepsilon_0 \eta(A) \\ &= \mu(A) \end{aligned}$$

for all $A \in \mathcal{B}(S)$, so $\lambda_1, \dots, \lambda_n$ again have the required properties. $\square$

1.5 Corollary Let μ be a finite positive Borel measure on S , and let $p_i \geq 0$ and $A_i \in \mathcal{B}(S)$ for $i = 1, \dots, n$. Let $\varepsilon > 0$, and suppose that

$$(1.20) \quad \sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A_i\right) + \varepsilon \quad \text{for all } I \subset \{1, \dots, n\}.$$

Then there exist positive Borel measures $\lambda_1, \dots, \lambda_n$ on S such that $\lambda_i(A_i) = \lambda_i(S) \leq p_i$ for $i = 1, \dots, n$, $\sum_{i=1}^n \lambda_i(S) \geq \sum_{i=1}^n p_i - \varepsilon$, and $\sum_{i=1}^n \lambda_i(A) \leq \mu(A)$ for all $A \in \mathcal{B}(S)$.

Proof. Let $S' = S \cup \{\Delta\}$, where Δ is an isolated point not belonging to S . Extend μ to a Borel measure on S' by defining $\mu(\{\Delta\}) = \varepsilon$. Letting $A'_i = A_i \cup \{\Delta\}$ for $i = 1, \dots, n$, we have

$$(1.21) \quad \sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A'_i\right) \quad \text{for all } I \subset \{1, \dots, n\}.$$

By Lemma 1.4, there exist positive Borel measures $\lambda'_1, \dots, \lambda'_n$ on S' such that $\lambda'_i(A'_i) = \lambda'_i(S') = p_i$ for $i = 1, \dots, n$ and $\sum_{i=1}^n \lambda'_i(A) \leq \mu(A)$ for all $A \in \mathcal{B}(S')$. Let λ_i be the restriction of λ'_i to $\mathcal{B}(S)$ for $i = 1, \dots, n$. Then $\lambda_i(A_i) = \lambda'_i(A_i) \leq \lambda'_i(A'_i) = p_i$ and $\lambda_i(S - A_i) = \lambda'_i(S' - A'_i) = 0$ for $i = 1, \dots, n$. Also,

$$(1.22) \quad \sum_{i=1}^n \lambda_i(S) = \sum_{i=1}^n [p_i - \lambda'_i(\{\Delta\})] \geq \sum_{i=1}^n p_i - \mu(\{\Delta\}) = \sum_{i=1}^n p_i - \varepsilon$$

and $\sum_{i=1}^n \lambda_i(A) = \sum_{i=1}^n \lambda'_i(A) \leq \mu(A)$ for all $A \in \mathcal{B}(S)$. $\square$

Proof of Lemma 1.3 Let $P, Q, \varepsilon, \delta$, and $E_0, \dots, E_N$ be as in the statement of the lemma. Let $p_i = P(E_i)$ and $A_i = E_i^c$ for $i = 1, \dots, N$. Then

$$(1.23) \quad \sum_{i \in I} p_i \leq P\left(\overline{\bigcup_{i \in I} E_i}\right) \leq Q\left(\bigcup_{i \in I} A_i\right) + \varepsilon \quad \text{for all } I \subset \{1, \dots, N\},$$

so by Corollary 1.5, there exist positive Borel measures $\lambda_1, \dots, \lambda_N$ on S such that $\lambda_i(A_i) = \lambda_i(S) \leq p_i$ for $i = 1, \dots, N$,

$$(1.24) \quad \sum_{i=1}^N \lambda_i(S) \geq \sum_{i=1}^N p_i - \varepsilon,$$

and $\sum_{i=1}^N \lambda_i(A) \leq Q(A)$ for all $A \in \mathcal{B}(S)$. Define $c_1, \dots, c_N \in [0, 1]$ by $c_i = (p_i - \lambda_i(S))/p_i$, where $0/0 = 0$, and note that $(1 - c_i)P(E_i) = \lambda_i(S)$ for $i = 1, \dots, N$ and $P(E_0) + \sum_{i=1}^N c_i P(E_i) = 1 - \sum_{i=1}^N \lambda_i(S)$. Consequently, there exist $Q_0, \dots, Q_N \in \mathcal{P}(S)$ such that

$$(1.25) \quad Q_i(B)(1 - c_i)P(E_i) = \lambda_i(B), \quad i = 1, \dots, N,$$

and

$$(1.26) \quad Q_0(B)\left(P(E_0) + \sum_{i=1}^N c_i P(E_i)\right) = Q(B) - \sum_{i=1}^N \lambda_i(B)$$

for all $B \in \mathcal{B}(S)$.

Let $X, Y_0, \dots, Y_N$, and ξ be independent random variables on some probability space $(\Omega, \mathcal{F}, \nu)$ with $X, Y_0, \dots, Y_N$ having distributions $P, Q_0, \dots, Q_N$ and ξ uniformly distributed on $[0, 1]$. We can assume that $Y_1, \dots, Y_N$ take values in $A_1, \dots, A_N$, respectively. Defining Y by (1.10), we have by (1.25) and (1.26),

$$(1.27) \quad \begin{aligned} \nu\{Y \in B\} &= \sum_{i=1}^N Q_i(B)(1 - c_i)P(E_i) \\ &\quad + Q_0(B)\left(P(E_0) + \sum_{i=1}^N c_i P(E_i)\right) \\ &= Q(B) \end{aligned}$$

for all $B \in \mathcal{B}(S)$. Noting that $\{X \in E_i, \xi \geq c_i\} \subset \{X \in E_i, Y \in A_i\} \subset \{d(X, Y) < \delta + \varepsilon\}$ for $i = 1, \dots, N$, we have

$$(1.28) \quad \begin{aligned} \{d(X, Y) \geq \delta + \varepsilon\} &\subset \{X \in E_0\} \cup \bigcup_{i=1}^N \{X \in E_i, \xi < c_i\} \\ &\subset \{X \in E_0\} \cup \left\{\xi < \bigvee_{i=1}^N c_i\right\} \\ &\subset \{X \in E_0\} \cup \left\{\xi < \max \left[\frac{\varepsilon}{P(E_i)} : P(E_i) > 0 \right]\right\}, \end{aligned}$$

where the third containment follows from $p_i - \lambda_i(S) \leq \varepsilon$ for $i = 1, \dots, N$ (see (1.24)). Finally, by the first containment in (1.28) and by (1.24),

$$\begin{aligned}
 (1.29) \quad v\{d(X, Y) \geq \delta + \varepsilon\} &\leq P(E_0) + \sum_{i=1}^N c_i P(E_i) \\
 &= P(E_0) + \sum_{i=1}^N (p_i - \lambda_i(S)) \\
 &\leq \delta + \varepsilon. \quad \square
 \end{aligned}$$

1.6 Corollary Let (S, d) be separable. Suppose that X_n , $n = 1, 2, \dots$, and X are S -valued random variables defined on the same probability space with distributions P_n , $n = 1, 2, \dots$, and P , respectively. If $d(X_n, X) \rightarrow 0$ in probability as $n \rightarrow \infty$, then $\lim_{n \rightarrow \infty} \rho(P_n, P) = 0$.

Proof. For $n = 1, 2, \dots$, let μ_n be the joint distribution of X_n and X . Then $\lim_{n \rightarrow \infty} \mu_n\{(x, y): d(x, y) \geq \varepsilon\} = 0$ for every $\varepsilon > 0$, so the result follows from Theorem 1.2. $\square$

The next result shows that the metric space $(\mathcal{P}(S), \rho)$ is complete and separable whenever (S, d) is. We note that while separability is a topological property, completeness is a property of the metric.

1.7 Theorem If S is separable, then $\mathcal{P}(S)$ is separable. If in addition (S, d) is complete, then $(\mathcal{P}(S), \rho)$ is complete.

Proof. Let $\{x_n\}$ be a countable dense subset of S , and let δ_x denote the element of $\mathcal{P}(S)$ with unit mass at $x \in S$. We leave it to the reader to show that the probability measures of the form $\sum_{i=1}^N a_i \delta_{x_i}$ with N finite, a_i rational, and $\sum_{i=1}^N a_i = 1$, comprise a dense subset of $\mathcal{P}(S)$ (Problem 3).

To prove completeness it is enough to consider sequences $\{P_n\} \subset \mathcal{P}(S)$ with $\rho(P_{n-1}, P_n) < 2^{-n}$ for each $n \geq 2$. For $n = 2, 3, \dots$, choose $E_1^{(n)}, \dots, E_{N_n}^{(n)} \in \mathcal{P}(S)$ disjoint with diameters less than 2^{-n} and with $P_{n-1}(E_0^{(n)}) \leq 2^{-n}$, where $E_0^{(n)} = S - \bigcup_{i=1}^{N_n} E_i^{(n)}$. By Lemma 1.3, there exists a probability space $(\Omega, \mathcal{F}, v)$ on which are defined S -valued random variables $Y_0^{(n)}, \dots, Y_{N_n}^{(n)}$, $n = 2, 3, \dots$, $[0, 1]$ -valued random variables $\xi^{(n)}$, $n = 2, 3, \dots$, and an S -valued random variable X_1 with distribution P_1 , all of which are independent, such that if the constants $c_1^{(n)}, \dots, c_{N_n}^{(n)} \in [0, 1]$, $n = 2, 3, \dots$, are appropriately chosen, then the random variable

$$(1.30) \quad X_n = \begin{cases} Y_i^{(n)} & \text{on } \{X_{n-1} \in E_i^{(n)}, \xi^{(n)} \geq c_i^{(n)}\}, \quad i = 1, \dots, N_n, \\ Y_0^{(n)} & \text{on } \{X_{n-1} \in E_0^{(n)}\} \cup \bigcup_{i=1}^{N_n} \{X_{n-1} \in E_i^{(n)}, \xi^{(n)} < c_i^{(n)}\} \end{cases}$$

has distribution P_n and

$$(1.31) \quad \nu\{d(X_{n-1}, X_n) \geq 2^{-n+1}\} \leq 2^{-n+1},$$

successively for $n = 2, 3, \dots$. By the Borel–Cantelli lemma,

$$(1.32) \quad \nu\left\{\sum_{n=2}^{\infty} d(X_{n-1}, X_n) < \infty\right\} = 1,$$

so by the completeness of (S, d) , $\lim_{n \rightarrow \infty} X_n \equiv X$ exists a.s. Letting P be the distribution of X , Corollary 1.6 implies that $\lim_{n \rightarrow \infty} \rho(P_n, P) = 0$. $\square$

As a further application of Lemma 1.3, we derive the so-called Skorohod representation.

1.8 Theorem Let (S, d) be separable. Suppose P_n , $n = 1, 2, \dots$, and P in $\mathcal{P}(S)$ satisfy $\lim_{n \rightarrow \infty} \rho(P_n, P) = 0$. Then there exists a probability space $(\Omega, \mathcal{F}, \nu)$ on which are defined S -valued random variables X_n , $n = 1, 2, \dots$, and X with distributions P_n , $n = 1, 2, \dots$, and P , respectively, such that $\lim_{n \rightarrow \infty} X_n = X$ a.s.

Proof. For $k = 1, 2, \dots$, choose $E_1^{(k)}, \dots, E_{N_k}^{(k)} \in \mathcal{B}(S)$ disjoint with diameters less than 2^{-k} and with $P(E_0^{(k)}) \leq 2^{-k}$, where $E_0^{(k)} = S - \bigcup_{i=1}^{N_k} E_i^{(k)}$, and assume (without loss of generality) that $\varepsilon_k \equiv \min_{1 \leq i \leq N_k} P(E_i^{(k)}) > 0$. Define the sequence $\{k_n\}$ by

$$(1.33) \quad k_n = \max \{1\} \cup \left\{k \geq 1 : \rho(P_n, P) < \frac{\varepsilon_k}{k}\right\},$$

and apply Lemma 1.3 with $Q = P_n$, $\varepsilon = \varepsilon_{k_n}/k_n$ if $k_n > 1$ and $\varepsilon = \rho(P_n, P) + 1/n$ if $k_n = 1$, $\delta = 2^{-k_n}$, $E_i = E_i^{(k_n)}$, and $N = N_{k_n}$ for $n = 1, 2, \dots$. We conclude that there exists a probability space $(\Omega, \mathcal{F}, \nu)$ on which are defined S -valued random variables $Y_0^{(n)}, \dots, Y_{N_n}^{(n)}$, $n = 1, 2, \dots$, a random variable ξ uniformly distributed on $[0, 1]$, and an S -valued random variable X with distribution P , all of which are independent, such that if the constants $c_1^{(n)}, \dots, c_{N_n}^{(n)} \in [0, 1]$, $n = 1, 2, \dots$, are appropriately chosen, then the random variable

$$(1.34) \quad X_n = \begin{cases} Y_i^{(n)} & \text{on } \{X \in E_i^{(k_n)}, \xi \geq c_i^{(n)}\}, \quad i = 1, \dots, N_{k_n}, \\ Y_0^{(n)} & \text{on } \{X \in E_0^{(k_n)}\} \cup \bigcup_{i=1}^{N_{k_n}} \{X \in E_i^{(k_n)}, \xi < c_i^{(n)}\} \end{cases}$$

has distribution P_n and

$$(1.35) \quad \left\{d(X_n, X) \geq 2^{-k_n} + \frac{\varepsilon_{k_n}}{k_n}\right\} \\ \subset \{X \in E_0^{(k_n)}\} \cup \left\{\xi < \frac{1}{k_n}\right\} \quad \text{if } k_n > 1$$

for $n = 1, 2, \dots$. If $K_n \equiv \min_{m \geq n} k_m > 1$, then

$$\begin{aligned}
 (1.36) \quad & \nu \left(\bigcup_{m=n}^{\infty} \left\{ d(X_m, X) \geq 2^{-k_m} + \frac{\varepsilon_{k_m}}{k_m} \right\} \right) \\
 & \leq \sum_{k=K_n}^{\infty} \nu \{ X \in E_0^{(k)} \} + \nu \left\{ \xi < \frac{1}{K_n} \right\} \\
 & \leq 2^{-K_n+1} + \frac{1}{K_n},
 \end{aligned}$$

and since $\lim_{n \rightarrow \infty} K_n = \infty$, we have $\lim_{n \rightarrow \infty} X_n = X$ a.s. $\square$

We conclude this section by proving the continuous mapping theorem.

1.9 Corollary Let (S, d) and (S', d') be separable metric spaces, and let $h: S \rightarrow S'$ be Borel measurable. Suppose that $P_n, n = 1, 2, \dots$, and P in $\mathcal{P}(S)$ satisfy $\lim_{n \rightarrow \infty} \rho(P_n, P) = 0$, and define $Q_n, n = 1, 2, \dots$, and Q in $\mathcal{P}(S')$ by

$$(1.37) \quad Q_n = P_n h^{-1}, \quad Q = P h^{-1}.$$

(By definition, $P h^{-1}(B) = P\{s \in S: h(s) \in B\}$.) Let C_h be the set of points of S at which h is continuous. If $P(C_h) = 1$, then $\lim_{n \rightarrow \infty} \rho'(Q_n, Q) = 0$, where ρ' is the Prohorov metric on $\mathcal{P}(S')$.

Proof. By Theorem 1.8, there exists a probability space $(\Omega, \mathcal{F}, \nu)$ on which are defined S -valued random variables $X_n, n = 1, 2, \dots$, and X with distributions $P_n, n = 1, 2, \dots$, and P , respectively, such that $\lim_{n \rightarrow \infty} X_n = X$ a.s. Since $\nu\{X \in C_h\} = 1$, we have $\lim_{n \rightarrow \infty} h(X_n) = h(X)$ a.s., and by Corollary 1.6, this implies that $\lim_{n \rightarrow \infty} \rho'(Q_n, Q) = 0$. $\square$

2. PROHOROV'S THEOREM

We are primarily interested in the convergence of sequences of Borel probability measures on the metric space (S, d) . A common approach for verifying the convergence of a sequence $\{x_n\}$ in a metric space is to first show that $\{x_n\}$ is contained in some compact set and then to show that every convergent subsequence of $\{x_n\}$ must converge to the same element x . This then implies that $\lim_{n \rightarrow \infty} x_n = x$. We use this argument repeatedly in what follows, and, consequently, a characterization of the compact subsets of $\mathcal{P}(S)$ is crucial. This characterization is given by the theorem of Prohorov that relates compactness to the notion of tightness.

A probability measure $P \in \mathcal{P}(S)$ is said to be *tight* if for each $\varepsilon > 0$ there exists a compact set $K \subset S$ such that $P(K) \geq 1 - \varepsilon$. A family of probability measures $\mathcal{M} \subset \mathcal{P}(S)$ is *tight* if for each $\varepsilon > 0$ there exists a compact set $K \subset S$