1. THE PROHOROV METRIC

Throughout Sections 1–4, \((S, d)\) is a metric space (\(d\) denoting the metric), \(\mathcal{B}(S)\) is the \(\sigma\)-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

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

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

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

To see that \(\rho\) is a metric, we need the following lemma.

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

\[
(1.3) \quad P(F) \leq Q(F^c) + \beta
\]

for all \(F \in \mathcal{C}\), then

\[
(1.4) \quad Q(F) \leq P(F^c) + \beta
\]

for all \(F \in \mathcal{C}\).

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

\[
(1.5) \quad P(F_1^c) = 1 - P(F_2) \geq 1 - Q(F_2^c) - \beta = Q(F_1) - \beta,
\]

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

\[
(1.6) \quad P(F) \leq Q(F^c) + \delta \leq Q(F^c) + \delta
\]

\[
\leq R((F^c)^c) + \delta + \varepsilon \leq R(F^c + \epsilon) + \delta + \varepsilon
\]

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
\begin{equation}
\mu\{(x, y): d(x, y) \geq \varepsilon\} \leq \varepsilon,
\end{equation}
then
\begin{equation}
P(F) = \mu(F \times S)
\leq \mu((F \times S) \cap \{(x, y): d(x, y) < \varepsilon\}) + \varepsilon
\leq \mu(S \times F^*) + \varepsilon = Q(F^*) + \varepsilon
\end{equation}
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. \hfill \Box

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

\begin{equation}
\{d(X, Y) \geq \delta + \varepsilon\} \subseteq \{X \in E_0\} \cup \left\{\zeta < \max \left[\frac{\varepsilon}{P(E_i)}: P(E_i) > 0\right]\right\},
\end{equation}
and
\begin{equation}
\nu\{d(X, Y) \geq \delta + \varepsilon\} \leq \delta + \varepsilon.
\end{equation}

The proof of this lemma depends on another lemma.

1.4 Lemma Let $\mu$ be a finite positive Borel measure on $S$, and let $p_i \geq 0$ and $A_i \in \mathcal{B}(S)$ for $i = 1, \ldots, n$. Suppose that
\begin{equation}
\sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A_i\right) \quad \text{for all } I \subseteq \{1, \ldots, n\}.
\end{equation}

Then there exist positive Borel measures $\lambda_1, \ldots, \lambda_n$ on $S$ such that $\lambda_i(A) = \lambda_i(S) = p_i$ for $i = 1, \ldots, n$ 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 $\lambda_1$ on $\mathcal{B}(S)$ by $\lambda_1(A) = \ldots
p_1 \mu(A \cap A_i)/\mu(A_i). Then \lambda_1(A_i) = \lambda_1(S) = p_1, and since \mu(A) \leq \mu(A \cap A_i) \leq \mu(A) \text{ for all } A \in \mathcal{A}(S). Suppose now that the lemma holds with n replaced by m for m = 1, \ldots, n - 1 and that \mu, p_i, and A_i (1 \leq i \leq n) satisfy (1.13). Define \eta on \mathcal{A}(S) by \eta(A) = \mu(A \cap A_n)/\mu(A_n), and let \epsilon_0 be the largest \epsilon such that

\begin{equation}
\sum_{i \in I} p_i \leq (\mu - \epsilon \eta) \left( \bigcup_{i \in I} A_i \right) \quad \text{for all } I \subset \{1, \ldots, n - 1\}.
\end{equation}

**Case 1.** \(\epsilon_0 \geq p_n\). Let \lambda_n = p_n \eta and put \mu' = \mu - \lambda_n. Since \mu_n \leq \mu(A_n) by (1.13), \mu' is a positive Borel measure on S, so by (1.14) (with \epsilon = p_n) and the induction hypothesis, there exist positive Borel measures \lambda_1, \ldots, \lambda_{n-1} on S such that \lambda_i(A_i) = \lambda_i(S) = p_i for i = 1, \ldots, n - 1 and \sum_{i=1}^{n-1} \lambda_i(A) \leq \mu'(A) \text{ for all } A \in \mathcal{A}(S). Also, \lambda_n(A_n) = \lambda_n(S) = p_n, so \lambda_1, \ldots, \lambda_n have the required properties.

**Case 2.** \(\epsilon_0 < p_n\). Put \mu' = \mu - \epsilon_0 \eta, and note that \mu' is a positive Borel measure on S. By the definition of \epsilon_0, there exists \(I_0 \subset \{1, \ldots, n-1\}\) (nonempty) such that

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

with equality holding for \(I = I_0\). By the induction hypothesis, there exist positive Borel measures \lambda_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) \text{ for all } A \in \mathcal{A}(S). Let \(p'_i = p_i\) for \(i = 1, \ldots, n - 1\) and \(p'_n = p_n - \epsilon_0\). Put \(B_0 = \bigcup_{i \in I_0} A_i\), define \mu'' on \mathcal{A}(S) by \mu''(A) = \mu'(A) - \mu'(A \cap B_0), and let \(I_1 = \{1, \ldots, n\} - I_0\). Then, for all \(I \subset I_1\),

\begin{equation}
\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 \cap I_0} A_i \cap B_0 \right) = \mu'' \left( \bigcup_{i \in I} A_i \right) + \mu'(B_0).
\end{equation}

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

\begin{equation}
\sum_{i \in I \cup I_0} p_i \leq \mu \left( \bigcup_{i \in I \cup I_0} A_i \right) - \epsilon_0 = \mu' \left( \bigcup_{i \in I \cup I_0} A_i \right).
\end{equation}
By (1.16),
\begin{equation}
\sum_{i \in I} p_i \leq \mu''(\bigcup_{i \in I} A_i) \quad \text{for all } I \subseteq I_1,
\end{equation}
so by the induction hypothesis, there exist positive Borel measures \( \lambda'_i \) on \( S \), \( i \in I_1 \), such that \( \lambda'_i(A_i) = \lambda'_i(S) = p_i \) for each \( i \in I_1 \) and \( \sum_{i \in I_1} \lambda'_i(A_i) \leq \mu''(A) \) for all \( A \in \mathcal{B}(S) \). Finally, let \( \lambda_i = \lambda'_i \) for \( i \in I_1 - \{n\} \) and \( \lambda_n = \lambda'_n + \epsilon_0 n \). Then \( \lambda_i(A_i) = \lambda_i(S) = p_i \) for each \( i \in I_1 \), hence for \( i = 1, \ldots, n \), and

\begin{equation}
\sum_{i = 1}^n \lambda_i(A) = \sum_{i = 0}^n \lambda'_i(A) + \sum_{i \in I_1} \lambda'_i(A) = \mu(A) + \epsilon_0 \eta(A)
\end{equation}

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

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

\begin{equation}
\sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A_i\right) + \epsilon \quad \text{for all } I \subseteq \{1, \ldots, n\}.
\end{equation}

Then there exist positive Borel measures \( \lambda_1, \ldots, \lambda_n \) on \( S \) such that \( \lambda_i(A_i) = \lambda_i(S) \leq p_i \) for \( i = 1, \ldots, n \), \( \sum_{i = 1}^n \lambda_i(S) \geq \sum_{i = 1}^n p_i - \epsilon \), 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 \( \Delta \) is an isolated point not belonging to \( S \). Extend \( \mu \) to a Borel measure on \( S' \) by defining \( \mu(\{\Delta\}) = \epsilon \). Letting \( A_i' = A_i \cup \{\Delta\} \) for \( i = 1, \ldots, n \), we have

\begin{equation}
\sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A_i\right) \quad \text{for all } I \subseteq \{1, \ldots, n\}.
\end{equation}

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

\begin{equation}
\sum_{i = 1}^n \lambda_i(S) = \sum_{i = 1}^n \left[ p_i - \lambda'_i(\Delta) \right] \geq \sum_{i = 1}^n p_i - \mu(\{\Delta\}) = \sum_{i = 1}^n p_i - \epsilon
\end{equation}

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 $E_0, \ldots, E_N$ be as in the statement of the lemma. Let $p_i = P(E_i)$ and $A_i = E_i$ for $i = 1, \ldots, N$. Then

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

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

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

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

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

and

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

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

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

(1.27) $\nu\{Y \in B\} = \sum_{i=1}^{N} Q_i(B)(1 - c_i)P(E_i)$

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

$\quad = Q(B)$

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 $i = 1, \ldots, N$, we have

(1.28) $\{d(X, Y) \geq \delta + \varepsilon\} \subset \{X \in E_0\} \cup \bigcup_{i=1}^{N} \{X \in E_i, \xi < c_i\}$

$\quad \subset \{X \in E_0\} \cup \left\{ \xi < \max\left\{ \frac{\varepsilon}{P(E_i)} : P(E_i) > 0 \right\} \right\}$,
I. THE MOTTOS

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

\[
(1.29) \quad v\{d(X, Y) \geq \delta + \varepsilon\} \leq P(E_0) + \sum_{i=1}^{N} c_i P(E_i) \\
= P(E_0) + \sum_{i=1}^{N} \left( p_i - \lambda_i(S) \right) \\
\leq \delta + \varepsilon. \tag*{\Box}
\]

1.6 Corollary Let \((S, d)\) be separable. Suppose that \(X_n, n = 1, 2, \ldots, \) and \(X\) are \(S\)-valued random variables defined on the same probability space with distributions \(P_n, n = 1, 2, \ldots, \) 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 \(n = 1, 2, \ldots\), let \(\mu_n\) 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. \(\Box\)

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 \(\delta_x\) 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 \(n = 2, 3, \ldots\), choose \(E_1^{(n)}, \ldots, E_N^{(n)} \in \mathcal{B}(S)\) disjoint with diameters less than \(2^{-n}\) and with \(P_{n-1}(E_0^{(n)}) \leq 2^{-n}\), where \(E_0^{(n)} = S - \bigcup_{i=1}^{N} E_i^{(n)}\). By Lemma 1.3, there exists a probability space \((\Omega, \mathcal{F}, \nu)\) on which are defined \(S\)-valued random variables \(Y_1^{(n)}, \ldots, Y_n^{(n)}\), \(n = 2, 3, \ldots, [0, 1]\)-valued random variables \(\xi^{(n)}, n = 2, 3, \ldots, \) and an \(S\)-valued random variable \(X_1\) with distribution \(P_1\), all of which are independent, such that if the constants \(c_1^{(n)}, \ldots, c_N^{(n)} \in [0, 1], n = 2, 3, \ldots, \) are appropriately chosen, then the random variable

\[
X_n = \begin{cases} 
Y_i^{(n)} & \text{on } \{X_{n-1} \in E_i^{(n)}, \xi^{(n)} \geq c_i^{(n)}\}, \\
Y_0^{(n)} & \text{on } \{X_{n-1} \in E_0^{(n)}\} \cup \bigcup_{i=1}^{N} \{X_{n-1} \in E_i^{(n)}, \xi^{(n)} < c_i^{(n)}\}
\end{cases}
\tag{1.30}
\]
102 CONVERGENCE OF PROBABILITY MEASURES

has distribution $P_n$ and

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

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

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

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

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

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

Proof. For $k = 1, 2, \ldots$, choose $E_1^{(k)}, \ldots, E_N^{(k)} \in \mathcal{B}(S)$ 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

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

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 $k_n = 1$, $\delta = 2^{-k_n} / (k_n - 1)$, $E = E_{k_n}^{(k_n)}$, and $N = N_{k_n}$ for $n = 1, 2, \ldots$. We conclude that there exists a probability space $(\Omega, \mathcal{F}, \nu)$ on which are defined $S$-valued random variables $Y_0^{(n)}, \ldots, Y_{N_k}^{(n)}$, $n = 1, 2, \ldots$, a random variable $\xi$ 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 $c_1^{(n)}, \ldots, c_{N_k}^{(n)} \in [0, 1]$, $n = 1, 2, \ldots$, are appropriately chosen, then the random variable

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

has distribution $P_n$ and

$$(1.35) \quad \left\{d(X_n, X) \geq 2^{-k_n} + \frac{\varepsilon_{k_n}}{k_n}\right\} \subset \{X \in E_0^{(k_n)}\} \cup \left\{\xi < \frac{1}{k_n}\right\} \text{ if } k_n > 1$$
for \( n = 1, 2, \ldots \). If \( K_n \equiv \min_{m \geq n} k_m > 1 \), then

\[
\nu \left( \bigcup_{m=n}^{\infty} \left\{ d(X_m, X) \geq 2^{-k_m} + \frac{\epsilon_{k_m}}{k_m} \right\} \right) 
\leq \sum_{k=K_n}^{\infty} \nu \{ X \in E^{(k)}_0 \} + \nu \left\{ \zeta < \frac{1}{K_n} \right\}
\leq 2^{-k_n+1} + \frac{1}{K_n},
\]

and since \( \lim_{n \to \infty} K_n = \infty \), we have \( \lim_{n \to \infty} X_n = X \) 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 \( h: S \to S' \) be Borel measurable. Suppose that \( P_n, n = 1, 2, \ldots \), and \( P \) in \( \mathcal{P}(S) \) satisfy \( \lim_{n \to \infty} \rho(P_n, P) = 0 \), and define \( Q_n, n = 1, 2, \ldots \), and \( Q \) in \( \mathcal{P}(S') \) by

\[
Q_n = P_n h^{-1}, \quad Q = Ph^{-1}. \tag{1.37}
\]

(By definition, \( Ph^{-1}(B) = P\{ s \in S : h(s) \in B \} \).) Let \( C_n \) be the set of points of \( S \) at which \( h \) is continuous. If \( P(C_h) = 1 \), then \( \lim_{n \to \infty} \rho'(Q_n, Q) = 0 \), where \( \rho' \) 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 \( X_n, n = 1, 2, \ldots \), and \( X \) with distributions \( P_n, n = 1, 2, \ldots \), and \( P \), respectively, such that \( \lim_{n \to \infty} X_n = X \) a.s. Since \( \nu \{ X \in C_h \} = 1 \), we have \( \lim_{n \to \infty} h(X_n) = h(X) \) a.s., and by Corollary 1.6, this implies that \( \lim_{n \to \infty} \rho'(Q_n, Q) = 0 \). \( \square \)

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 \( \lim_{n \to \infty} x_n = x \). 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 \( K \subset S \)