Section 17

Metrics for convergence of laws.
Empirical measures.

**Levy-Prohorov metric.** Consider a metric space \((S,d)\). For a set \(A \subseteq S\) let us denote by
\[
A^\varepsilon = \{ y \in S : d(x,y) < \varepsilon \text{ for some } x \in A \}
\]
its \(\varepsilon\)-neighborhood. Let \(\mathcal{B}\) be a Borel \(\sigma\)-algebra on \(S\).

**Definition.** If \(P, Q\) are probability distributions on \(\mathcal{B}\) then
\[
\rho(P, Q) = \inf \{ \varepsilon > 0 : P(A) \leq Q(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{B} \}
\]
is called the Levy-Prohorov distance between \(P\) and \(Q\).

**Lemma 34** \(\rho\) is a metric on the set of probability laws on \(\mathcal{B}\).

**Proof.** 1. First, let us show that \(\rho(Q,P) = \rho(P,Q)\). Suppose that \(\rho(P,Q) > \varepsilon\). Then there exists a set \(A\) such that \(P(A) > Q(A^\varepsilon) + \varepsilon\). Taking complements gives
\[
Q(A^{\varepsilon \text{cc}}) > P(A^\varepsilon) + \varepsilon \geq P(A^{\varepsilon \text{cc}}) + \varepsilon,
\]
where the last inequality follows from the fact that \(A^\varepsilon \supseteq A^{\varepsilon \text{cc}}\):
\[
a \in A^{\varepsilon \text{cc}} \implies d(a, A^{\varepsilon \text{cc}}) < \varepsilon \implies d(a, b) < \varepsilon \text{ for some } b \in A^{\varepsilon \text{cc}}
\]
\[
\{ \text{since } b \notin A^{\varepsilon}, d(b, A) \geq \varepsilon \}
\]
\[
\implies d(a, A) > 0 \implies a \notin A \implies a \in A^\varepsilon.
\]
Therefore, for a set \(B = A^{\varepsilon \text{cc}}, Q(B) > P(B^\varepsilon) + \varepsilon\). This means that \(\rho(Q,P) > \varepsilon\) and, therefore, \(\rho(Q,P) \geq \rho(P,Q)\). By symmetry, \(\rho(Q,P) \leq \rho(P,Q)\) and \(\rho(Q,P) = \rho(P,Q)\).

2. Next, let us show that if \(\rho(P,Q) = 0\) then \(P = Q\). For any set \(F\) and any \(n \geq 1\),
\[
P(F) \leq Q(F^{\frac{1}{n}}) + \frac{1}{n}.
\]
If \(F\) is closed then \(F^{\frac{1}{n}} \downarrow F\) as \(n \to \infty\) and by continuity of measure
\[
P(F) \leq Q\left( \bigcap F^{\frac{1}{n}} \right) = Q(F).
\]
Similarly, \(P(F) \geq Q(F)\) and, therefore, \(P(F) = Q(F)\).
3. Finally, let us prove the triangle inequality

$$\rho(P, \mathbb{R}) \leq \rho(P, Q) + \rho(Q, \mathbb{R}).$$

If $\rho(P, Q) < x$ and $\rho(Q, \mathbb{R}) < y$ then for any set $A$,

$$P(A) \leq Q(A^x) + x \leq \mathbb{R}((A^x)^y) + y \leq \mathbb{R}(A^{x+y}) + x + y,$$

which means that $\rho(P, \mathbb{R}) \leq x + y$.

**Bounded Lipschitz metric.** Given probability distributions $P, Q$ on the metric space $(S, d)$ we define a bounded Lipschitz distance between them by

$$\beta(P, Q) = \sup\left\{ \left| \int f dP - \int f dQ \right| : \|f\|_{BL} \leq 1 \right\}.$$

**Lemma 35** $\beta$ is a metric on the set of probability laws on $B$.

**Proof.** $\beta(P, Q) = \beta(Q, P)$ and the triangle inequality are obvious. It remains to prove that $\beta(P, Q) = 0$ implies $P = Q$. Given a closed set $F$, the sequence of functions $f_m(x) = md(x, F) \wedge 1$ converges $f_m \to I_U$, where $U = F^c$. Obviously, $\|f_m\|_{BL} \leq m + 1$ and, therefore, $\int f_m dP = \int f_m dQ$. Letting $m \to \infty$ proves that $P(U) = Q(U)$.

The law $P$ on $(S, d)$ is tight if for any $\varepsilon > 0$ there exists a compact $K \subseteq S$ such that $P(S \setminus K) \leq \varepsilon$.

**Theorem 40** (Ulam) If $(S, d)$ is separable then for any law $P$ on $B$ there exists a closed totally bounded set $K \subseteq S$ such that $P(S \setminus K) \leq \varepsilon$. If $(S, d)$ is complete and separable then $K$ is compact and, therefore, every law is tight.

**Proof.** Consider a sequence $\{x_1, x_2, \ldots\}$ that is dense in $S$. For any $m \geq 1$, $S = \bigcup_{i=1}^{\infty} \bar{B}(x_i, \frac{1}{m})$, where $\bar{B}$ denotes a closed ball, and by continuity of measure, for large enough $n(m)$,

$$P(S \setminus \bigcup_{i=1}^{n(m)} \bar{B}(x_i, \frac{1}{m})) \leq \frac{\varepsilon}{2^m}.$$

If we take

$$K = \bigcap_{m \geq 1} \bigcup_{i=1}^{n(m)} \bar{B}(x_i, \frac{1}{m})$$

then

$$P(S \setminus K) \leq \sum_{m \geq 1} \frac{\varepsilon}{2^m} = \varepsilon.$$

$K$ is closed and totally bounded by construction. If $S$ is complete, $K$ is compact.

**Theorem 41** Suppose that either $(S, d)$ is separable or $P$ is tight. Then the following are equivalent.

1. $P_n \to P$.
2. For all $f \in BL(S, d)$, $\int f dP_n \to \int f dP$.
3. $\beta(P_n, P) \to 0$.
4. $\rho(P_n, P) \to 0$.
**Proof.** 1⇒2. Obvious.

3⇒4. In fact, we will prove that

\[ \rho(\mathbb{P}_n, \mathbb{P}) \leq 2\sqrt{\beta(\mathbb{P}_n, \mathbb{P})}. \]  \hspace{1cm} (17.0.1)

Given a Borel set \( A \subseteq S \), consider a function

\[ f(x) = 0 \vee \left( 1 - \frac{1}{\varepsilon} d(x, A) \right) \] such that \( I_A \leq f \leq I_{A^c} \).

Obviously, \( \|f\|_{BL} \leq 1 + \varepsilon^{-1} \) and we can write

\[
\begin{align*}
\mathbb{P}_n(A) &\leq \int f \, d\mathbb{P}_n = \int f \, d\mathbb{P} + \left( \int f \, d\mathbb{P}_n - \int f \, d\mathbb{P} \right) \\
&\leq \mathbb{P}(A^c) + (1 + \varepsilon^{-1}) \sup \left\{ \left| \int f \, d\mathbb{P}_n - \int f \, d\mathbb{P} \right| : \|f\|_{BL} \leq 1 \right\} \\
&= \mathbb{P}(A^c) + (1 + \varepsilon^{-1}) \beta(\mathbb{P}_n, \mathbb{P}) \leq \mathbb{P}(A^c) + \delta,
\end{align*}
\]

where \( \delta = \max(\varepsilon, (1 + \varepsilon^{-1}) \beta(\mathbb{P}_n, \mathbb{P})) \). This implies that \( \rho(\mathbb{P}_n, \mathbb{P}) \leq \delta \). Since \( \varepsilon \) is arbitrary we can minimize \( \delta = \delta(\varepsilon) \) over \( \varepsilon \). If we take \( \varepsilon = \sqrt{3} \) then \( \delta = \max(\sqrt{3}, \beta + \sqrt{3}) = \beta + \sqrt{3} \) and

\[ \beta \leq 1 \implies \rho \leq 2\sqrt{3}; \quad \beta \geq 1 \implies \rho \leq 1 \leq 2\sqrt{3}. \]

4⇒1. Suppose that \( \rho(\mathbb{P}_n, \mathbb{P}) \to 0 \) which means that there exists a sequence \( \varepsilon_n \downarrow 0 \) such that

\[ \mathbb{P}_n(A) \leq \mathbb{P}(A^{\varepsilon_n}) + \varepsilon_n \] for all measurable \( A \subseteq S \).

If \( A \) is closed, then \( \bigcap_{n \geq 1} A^{\varepsilon_n} = A \) and, by continuity of measure,

\[ \lim_{n \to \infty} \sup \mathbb{P}_n(A) \leq \lim_{n \to \infty} \sup \left( \mathbb{P}(A^{\varepsilon_n}) + \varepsilon_n \right) = \mathbb{P}(A). \]

By the portmanteau theorem, \( \mathbb{P}_n \to \mathbb{P} \).

2⇒3. If \( \mathbb{P} \) is tight, let \( K \) be a compact such that \( \mathbb{P}(S \setminus K) \leq \varepsilon \). If \( (S, d) \) is separable, by Ulam’s theorem, let \( K \) be a closed totally bounded set such that \( \mathbb{P}(S \setminus K) \leq \varepsilon \). If we consider a function

\[ f(x) = 0 \vee \left( 1 - \frac{1}{\varepsilon} d(x, K) \right) \] with \( \|f\|_{BL} \leq 1 + \frac{1}{\varepsilon} \)

then

\[ \mathbb{P}_n(K^c) \geq \int f \, d\mathbb{P}_n - \int f \, d\mathbb{P} \geq \mathbb{P}(K) \geq 1 - \varepsilon, \]

which implies that for \( n \) large enough, \( \mathbb{P}_n(K^c) \geq 1 - 2\varepsilon. \) This means that all \( \mathbb{P}_n \) are essentially concentrated on \( K^c \). Let

\[ B = \left\{ f : \|f\|_{BL(S, d)} \leq 1 \right\}, \quad B_K = \left\{ f|_K : f \in B \right\} \subseteq C(K), \]

where \( f|_K \) denotes the restriction of \( f \) to \( K \). If \( K \) is compact then, by the Arzela-Ascoli theorem, \( B_K \) is totally bounded with respect to \( d_\infty \). If \( K \) is totally bounded then we can isometrically identify functions in \( B_K \) with their unique extensions to the completion \( K' \) of \( K \) and, by the Arzela-Ascoli theorem for the compact \( K' \), \( B_K \) is again totally bounded with respect to \( d_\infty \). In any case, given \( \varepsilon > 0 \), we can find \( f_1, \ldots, f_k \in B \) such that for all \( f \in B \)

\[ \sup_{x \in K} |f(x) - f_j(x)| \leq \varepsilon \] for some \( j \leq k \).

This uniform approximation can also be extended to \( K^c \). Namely, for any \( x \in K^c \) take \( y \in K \) such that \( d(x, y) \leq \varepsilon \). Then

\[
\begin{align*}
|f(x) - f_j(x)| &\leq |f(x) - f(y)| + |f(y) - f_j(y)| + |f_j(y) - f_j(x)| \\
&\leq ||f||_1 d(x, y) + \varepsilon + ||f_j||_1 d(x, y) \leq 3\varepsilon.
\end{align*}
\]

72
Therefore, for any \( f \in B \),

\[
\left| \int f d\mathbb{P}_n - \int f d\mathbb{P} \right| \leq \left| \int_{K^c} f d\mathbb{P}_n - \int_{K^c} f d\mathbb{P} \right| + \|f\|_\infty (\mathbb{P}_n(K^{cc}) + \mathbb{P}(K^{cc})) \\
\leq \left| \int_{K^c} f d\mathbb{P}_n - \int_{K^c} f d\mathbb{P} \right| + 2\varepsilon + \varepsilon \\
\leq \left| \int_{K^c} f_j d\mathbb{P}_n - \int_{K^c} f_j d\mathbb{P} \right| + 3\varepsilon + 3\varepsilon + 2\varepsilon + \varepsilon \\
\leq \left| \int f_j d\mathbb{P}_n - \int f_j d\mathbb{P} \right| + 3\varepsilon + 3\varepsilon + 2\varepsilon + \varepsilon \\
\leq \max_{1 \leq j \leq k} \left| \int f_j d\mathbb{P}_n - \int f_j d\mathbb{P} \right| + 12\varepsilon.
\]

Finally,

\[
\beta(\mathbb{P}_n, \mathbb{P}) = \sup_{f \in B} \left| \int f d\mathbb{P}_n - \int f d\mathbb{P} \right| \leq \max_{1 \leq j \leq k} \left| \int f_j d\mathbb{P}_n - \int f_j d\mathbb{P} \right| + 12\varepsilon
\]

and, using assumption 2, \( \limsup_{n \to \infty} \beta(\mathbb{P}_n, \mathbb{P}) \leq 12\varepsilon \). Letting \( \varepsilon \to 0 \) finishes the proof. □

**Convergence of empirical measures.** Let \((\Omega, \mathbb{P})\) be a probability space and \(X_1, X_2, \ldots : \Omega \to S\) be an i.i.d. sequence of random variables with values in a metric space \((S, d)\). Let \(\mu\) be the law of \(X_i\) on \(S\). Let us define the random *empirical measures* \(\mu_n\) on the Borel \(\sigma\)-algebra \(\mathcal{B}\) on \(S\) by

\[
\mu_n(A)(\omega) = \frac{1}{n} \sum_{i=1}^{n} I(X_i(\omega) \in A), \ A \in \mathcal{B}.
\]

By the strong law of large numbers, for any \(f \in C_b(S)\),

\[
\int f d\mu_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i) \to \mathbb{E} f(X_1) = \int f d\mu \text{ a.s.}
\]

However, the set of measure zero where this convergence is violated depends on \(f\) and it is not obvious that the convergence holds for all \(f \in C_b(S)\) with probability one.

**Theorem 42** (Varadarajan) Let \((S, d)\) be a separable metric space. Then \(\mu_n\) converges to \(\mu\) weakly almost surely,

\[
\mathbb{P}\left( \omega : \mu_n(\cdot)(\omega) \to \mu \text{ weakly} \right) = 1.
\]

**Proof.** Since \((S, d)\) is separable, by Theorem 2.8.2 in R.A.P., there exists a metric \(e\) on \(S\) such that \((S, e)\) is totally bounded and \(e\) and \(d\) define the same topology, i.e. \(e(s_n, s) \to 0\) if and only if \(d(s_n, s) \to 0\). This, of course, means that \(C_0(S, d) = C_0(S, e)\) and weak convergence of measures does not change. If \((T, e)\) is the completion of \((S, e)\) then \((T, e)\) is compact. By the Arzela-Ascoli theorem, \(BL(T, e)\) is separable with respect to the \(d_{\infty}\) norm and, therefore, \(BL(S, e)\) is also separable. Let \((f_m)\) be a dense subset of \(BL(S, e)\). Then, by the strong law of large number,

\[
\int f_m d\mu_n = \frac{1}{n} \sum_{i=1}^{n} f_m(X_i) \to \mathbb{E} f_m(X_1) = \int f_m d\mu \text{ a.s.}
\]

Therefore, on the set of probability one, \( \int f_m d\mu_n \to \int f d\mu \) for all \(m \geq 1\). Since \((f_m)\) is dense in \(BL(S, e)\), on the same set of probability one, \( \int f d\mu_n \to \int f d\mu \) for all \(f \in BL(S, e)\). Since \((S, e)\) is separable, the previous theorem implies that \(\mu_n \to \mu\) weakly. □