18.175 Theory of Probability
Fall 2008

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 ε-neighborhood. Let $\mathcal{B}$ be a Borel σ-algebra on S.

Definition. If $\mathbb{P}, \mathbb{Q}$ are probability distributions on $\mathcal{B}$ then

$$\rho(\mathbb{P}, \mathbb{Q}) = \inf\{\varepsilon > 0 : \mathbb{P}(A) \leq \mathbb{Q}(A^{\varepsilon}) + \varepsilon \text{ for all } A \in \mathcal{B}\}$$

is called the *Levy-Prohorov distance* between $\mathbb{P}$ and $\mathbb{Q}$.

Lemma 34 *ρ is a metric on the set of probability laws on $\mathcal{B}$.*

Proof. 1. First, let us show that $\rho(\mathbb{Q}, \mathbb{P}) = \rho(\mathbb{P}, \mathbb{Q})$. Suppose that $\rho(\mathbb{P}, \mathbb{Q}) > \varepsilon$. Then there exists a set A such that $\mathbb{P}(A) > \mathbb{Q}(A^{\varepsilon}) + \varepsilon$. Taking complements gives

$$\mathbb{Q}(A^{\varepsilon c}) > \mathbb{P}(A^c) + \varepsilon \geq \mathbb{P}(A^{\varepsilon c \varepsilon}) + \varepsilon,$$

where the last inequality follows from the fact that $A^c \supseteq A^{\varepsilon c \varepsilon}$:

$$\begin{aligned} a \in A^{\varepsilon c \varepsilon} \Longrightarrow d(a, A^{\varepsilon c}) < \varepsilon \quad &\Longrightarrow \quad d(a, b) < \varepsilon \text{ for some } b \in A^{\varepsilon c} \\ &\qquad \Big\{\text{since } b \notin A^{\varepsilon}, d(b, A) \geq \varepsilon\Big\} \\ &\Longrightarrow \quad d(a, A) > 0 \Longrightarrow a \notin A \Longrightarrow a \in A^c. \end{aligned}$$

Therefore, for a set $B = A^{\varepsilon c}$, $\mathbb{Q}(B) > \mathbb{P}(B^{\varepsilon}) + \varepsilon$. This means that $\rho(\mathbb{Q}, \mathbb{P}) > \varepsilon$ and, therefore, $\rho(\mathbb{Q}, \mathbb{P}) \geq \rho(\mathbb{P}, \mathbb{Q})$. By symmetry, $\rho(\mathbb{Q}, \mathbb{P}) \leq \rho(\mathbb{P}, \mathbb{Q})$ and $\rho(\mathbb{Q}, \mathbb{P}) = \rho(\mathbb{P}, \mathbb{Q})$.

2. Next, let us show that if $\rho(\mathbb{P}, \mathbb{Q}) = 0$ then $\mathbb{P} = \mathbb{Q}$. For any set F and any $n \geq 1$,

$$\mathbb{P}(F) \leq \mathbb{Q}(F^{\frac{1}{n}}) + \frac{1}{n}.$$

If F is closed then $F^{\frac{1}{n}} \downarrow F$ as $n \to \infty$ and by continuity of measure

$$\mathbb{P}(F) \leq \mathbb{Q}\Big(\bigcap F^{\frac{1}{n}}\Big) = \mathbb{Q}(F).$$

Similarly, $\mathbb{P}(F) \geq \mathbb{Q}(F)$ and, therefore, $\mathbb{P}(F) = \mathbb{Q}(F)$.

3. Finally, let us prove the triangle inequality

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

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

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

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

$\square$

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

$$\beta(\mathbb{P},\mathbb{Q}) = \sup\Big\{\Big|\int f d\mathbb{P} - \int f d\mathbb{Q}\Big| : ||f||_{\mathrm{BL}} \le 1\Big\}.$$

Lemma 35 *β is a metric on the set of probability laws on $\mathcal{B}$.*

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

$\square$

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

Theorem 40 *(Ulam) If (S,d) is separable then for any law $\mathbb{P}$ on $\mathcal{B}$ there exists a closed totally bounded set $K \subseteq S$ such that $\mathbb{P}(S \setminus K) \le \varepsilon$. 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 $m \ge 1$, $S = \bigcup_{i=1}^{\infty} \bar{B}\Big(x_i, \frac{1}{m}\Big)$, where $\bar{B}$ denotes a closed ball, and by continuity of measure, for large enough $n(m)$,

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

If we take

$$K = \bigcap_{m\ge1} \bigcup_{i=1}^{n(m)} \bar{B}\Big(x_i, \frac{1}{m}\Big)$$

then

$$\mathbb{P}(S \setminus K) \le \sum_{m\ge1} \frac{\varepsilon}{2^m} = \varepsilon.$$

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

$\square$

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

1. *$\mathbb{P}_n \to \mathbb{P}$.*
2. *For all $f \in BL(S,d)$, $\int f d\mathbb{P}_n \to \int f d\mathbb{P}$.*
3. *$\beta(\mathbb{P}_n, \mathbb{P}) \to 0$.*
4. *$\rho(\mathbb{P}_n, \mathbb{P}) \to 0$.*

Proof. 1$\Longrightarrow$2. Obvious.

3$\Longrightarrow$4. In fact, we will prove that

$$\rho(\mathbb{P}_n,\mathbb{P}) \le 2\sqrt{\beta(\mathbb{P}_n,\mathbb{P})}. \tag{17.0.1}$$

Given a Borel set $A\subseteq S$, consider a function

$$f(x) = 0 \vee \Bigl(1-\frac{1}{\varepsilon}d(x,A)\Bigr) \;\text{ such that }\; \mathrm{I}_A \le f \le \mathrm{I}_{A^\varepsilon}.$$

Obviously, $||f||_{\mathrm{BL}} \le 1+\varepsilon^{-1}$ and we can write

$$\begin{aligned}
\mathbb{P}_n(A) &\le \int f d\mathbb{P}_n = \int f d\mathbb{P} + \Bigl(\int f d\mathbb{P}_n - \int f d\mathbb{P}\Bigr)\\
&\le \mathbb{P}(A^\varepsilon) + (1+\varepsilon^{-1})\sup\Bigl\{\Bigl|\int f d\mathbb{P}_n - \int f d\mathbb{P}\Bigr| : ||f||_{\mathrm{BL}}\le 1\Bigr\}\\
&= \mathbb{P}(A^\varepsilon) + (1+\varepsilon^{-1})\beta(\mathbb{P}_n,\mathbb{P}) \le \mathbb{P}(A^\delta)+\delta,
\end{aligned}$$

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

$$\beta\le 1 \Longrightarrow \rho\le 2\sqrt{\beta};\quad \beta\ge 1\Longrightarrow \rho\le 1\le 2\sqrt{\beta}.$$

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

$$\mathbb{P}_n(A)\le \mathbb{P}(A^{\varepsilon_n})+\varepsilon_n \quad\text{for all measurable } A\subseteq S.$$

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

$$\limsup_{n\to\infty}\mathbb{P}_n(A)\le \limsup_{n\to\infty}\Bigl(\mathbb{P}(A^{\varepsilon_n})+\varepsilon_n\Bigr)=\mathbb{P}(A).$$

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

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

$$f(x)=0\vee\Bigl(1-\frac{1}{\varepsilon}d(x,K)\Bigr) \text{ with } ||f||_{\mathrm{BL}}\le 1+\frac{1}{\varepsilon}$$

then

$$\mathbb{P}_n(K^\varepsilon)\ge\int f d\mathbb{P}_n\to\int f d\mathbb{P}\ge\mathbb{P}(K)\ge 1-\varepsilon,$$

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

$$B=\Bigl\{f: ||f||_{\mathrm{BL}(S,d)}\le 1\Bigr\},\;\; B_K=\Bigl\{f\big|_K : f\in B\Bigr\}\subseteq C(K),$$

where $f\big|_K$ 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_∞. 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_∞. In any case, given $\varepsilon>0$, we can find $f_1,\ldots,f_k\in B$ such that for all $f\in B$

$$\sup_{x\in K}|f(x)-f_j(x)|\le\varepsilon \quad\text{for some } j\le k.$$

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

$$\begin{aligned}
|f(x)-f_j(x)| &\le |f(x)-f(y)|+|f(y)-f_j(y)|+|f_j(y)-f_j(x)|\\
&\le ||f||_{\mathrm{L}}d(x,y)+\varepsilon+||f_j||_{\mathrm{L}}d(x,y)\le 3\varepsilon.
\end{aligned}$$

Therefore, for any $f \in B$,

$$\begin{aligned} \Big|\int f d\mathbb{P}_n - \int f d\mathbb{P}\Big| &\leq \Big|\int_{K^\varepsilon} f d\mathbb{P}_n - \int_{K^\varepsilon} f d\mathbb{P}\Big| + ||f||_\infty \big(\mathbb{P}_n(K^{\varepsilon c}) + \mathbb{P}(K^{\varepsilon c})\big) \\ &\leq \Big|\int_{K^\varepsilon} f d\mathbb{P}_n - \int_{K^\varepsilon} f d\mathbb{P}\Big| + 2\varepsilon + \varepsilon \\ &\leq \Big|\int_{K^\varepsilon} f_j d\mathbb{P}_n - \int_{K^\varepsilon} f_j d\mathbb{P}\Big| + 3\varepsilon + 3\varepsilon + 2\varepsilon + \varepsilon \\ &\leq \Big|\int f_j d\mathbb{P}_n - \int f_j d\mathbb{P}\Big| + 3\varepsilon + 3\varepsilon + 3\varepsilon + 2\varepsilon + \varepsilon \\ &\leq \max_{1\leq j\leq k}\Big|\int f_j d\mathbb{P}_n - \int f_j d\mathbb{P}\Big| + 12\varepsilon. \end{aligned}$$

Finally,

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

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 μ be the law of X_i on S. Let us define the random *empirical measures* μ_n on the Borel σ-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 μ_n converges to μ weakly almost surely,*

$$\mathbb{P}\big(\omega : \mu_n(\cdot)(\omega) \to \mu \text{ weakly}\big) = 1.$$

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 $C_b(S, d) = C_b(S, e)$ 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_∞ 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, $\int f_m d\mu_n \to \int f_m d\mu$ 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.

□