18.175 Theory of Probability
Fall 2008

Section 12

Levy's Continuity Theorem. Poisson Approximation. Conditional Expectation.

Let us start with the following bound.

Lemma 27 *Let X be a real-valued r.v. with distribution $\mathbb{P}$ and let*

$$f(t) = \mathbb{E}e^{itX} = \int e^{itx} d\mathbb{P}(x).$$

Then,

$$\mathbb{P}\Big(|X| > \frac{1}{u}\Big) \leq \frac{7}{u}\int_0^u (1 - \mathbf{Re} f(t))dt.$$

Proof. Since

$$\mathbf{Re} f(t) = \int \cos tx d\mathbb{P}(x)$$

we have

$$\begin{aligned}
\frac{1}{u}\int_0^u \int_{\mathbb{R}} (1-\cos tx) d\mathbb{P}(x) dt &= \frac{1}{u}\int_{\mathbb{R}}\int_0^u (1-\cos tx) dt d\mathbb{P}(x) \\
&= \int_{\mathbb{R}} \Big(1 - \frac{\sin xu}{xu}\Big) d\mathbb{P}(x) \\
&\geq \int_{|xu|\geq 1} \Big(1 - \frac{\sin xu}{xu}\Big) d\mathbb{P}(x) \\
\Big\{\text{since } \frac{\sin y}{y} < \frac{\sin 1}{1} \text{ if } y > 1\Big\} \quad &\geq (1-\sin 1)\int_{|xu|\geq 1} 1 d\mathbb{P}(x) \geq \frac{1}{7}\mathbb{P}\Big(|X| \geq \frac{1}{u}\Big).
\end{aligned}$$

□

Theorem 28 *(Levy continuity) Let (X_n) be a sequence of r.v. on $\mathbb{R}^k$. Suppose that*

$$f_n(t) = \mathbb{E}e^{i(t,X_n)} \to f(t)$$

and $f(t)$ is continuous at 0 along each axis. Then there exists a probability distribution $\mathbb{P}$ such that

$$f(t) = \int e^{i(t,x)} d\mathbb{P}(x)$$

and $\mathcal{L}(X_n) \to \mathbb{P}$.

Proof. By Lemma 19 we only need to show that $\{\mathcal{L}(X_n)\}$ is uniformly tight. If we denote

$$X_n = (X_{n,1}, \ldots, X_{n,k})$$

then the c.f.s along the i^{th} coordinate:

$$f_n^i(t_i) := f_n(0, \ldots, t_i, 0, \ldots 0) = \mathbb{E}e^{it_i X_{n,i}} \to f(0, \ldots, t_i, \ldots 0) =: f^i(t_i).$$

Since $f_n(0) = 1$ and, therefore, $f(0) = 1$, for any $\varepsilon > 0$ we can find $\delta > 0$ such that for all $i \le k$

$$|f^i(t_i) - 1| \le \varepsilon \quad \text{if} \quad |t_i| \le \delta.$$

This implies that for large enough n

$$|f_n^i(t_i) - 1| \le 2\varepsilon \quad \text{if} \quad |t_i| \le \delta.$$

Using previous Lemma,

$$\mathbb{P}\Big(|X_{n,i}| > \frac{1}{\delta}\Big) \le \frac{7}{\delta} \int_0^\delta \Big(1 - \mathbf{Re} f_n^i(t_i)\Big) dt_i \le \frac{7}{\delta} \int_0^\delta \big|1 - f_n^i(t_i)\big| \, dt_i \le 7 \cdot 2\varepsilon.$$

The union bound implies that

$$\mathbb{P}\Big(|X_n| > \frac{\sqrt{k}}{\delta}\Big) \le 14k\varepsilon$$

and $\{\mathcal{L}(X_n)\}_{n \ge 1}$ is uniformly tight.

□

CLT describes how sums of independent r.v.s are approximated by normal distribution. We will now give a simple example of a different approximation. Consider independent Bernoulli random variables $X_i^n \sim B(p_i^n)$ for $i \le n$, i.e. $\mathbb{P}(X_i^n = 1) = p_i^n$ and $\mathbb{P}(X_i^n = 0) = 1 - p_i^n$. If $p_i^n = p > 0$ then by CLT

$$\frac{S_n - np}{\sqrt{np(1-p)}} \to \mathcal{N}(0, 1).$$

However, if $p = p_i^n \to 0$ fast enough then, for example, the Lindeberg conditions will be violated. It is well-known that if $p_i^n = p_n$ and $np_n \to \lambda$ then S_n has approximately Poisson distribution Π_λ with p.f.

$$f(k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ for } k = 0, 1, 2, \ldots$$

Here is a version of this result.

Theorem 29 *Consider independent $X_i \sim B(p_i)$ for $i \le n$ and let*

$$S_n = X_1 + \ldots + X_n \text{ and } \lambda = p_1 + \ldots + p_n.$$

Then for any subset of integers $B \subseteq \mathbb{Z}$,

$$|\mathbb{P}(S_n \in B) - \Pi_\lambda(B)| \le \sum_{i \le n} p_i^2.$$

Proof. The proof is based on the construction on "one probability space". Let us construct Bernoulli r.v. $X_i \sim B(p_i)$ and Poisson r.v. $X_i^* \sim \Pi_{p_i}$ on the same probability space as follows. Let us consider a probability space $([0,1], \mathcal{B}, \lambda)$ with Lebesque measure λ. Define

$$X_i = X_i(x) = \begin{cases} 0, & 0 \le x \le 1 - p_i, \\ 1, & 1 - p_i < x \le 1. \end{cases}$$

Clearly, $X_i \sim B(p_i)$. Let us construct X_i^* as follows. If for $k \ge 0$ we define

$$c_k = \sum_{0 \le l \le k} \frac{(p_i)^l}{l!} e^{-p_i}$$

then

$$X_i = X_i(x) = \begin{cases} 0, & 0 \le x \le c_0, \\ 1, & c_0 < x \le c_1, \\ 2, & c_1 < x \le c_2, \\ \dots \end{cases}$$

Clearly, $X_i^* \sim \Pi_{p_i}$. When $X_i \neq X_i^*$? Since $1 - p_j \le e^{-p_j} = c_0$, this can only happen for

$$1 - p_i < x \le c_0 \text{ and } c_1 < x \le 1,$$

i.e.

$$\mathbb{P}(X_j \neq X_j^*) = e^{p_j} - (1 - p_j) + (1 - e^{-p_j} - p_j e^{-p_j}) = p_j(1 - e^{-p_j}) \le p_j^2$$

We construct pairs (X_i, X_i^*) on separate coordinates of a product space, thus, making them independent fo $i \le n$. It is well-known that $\sum_{i \le n} X_i^* \sim \Pi_\lambda$ and, finally, we get

$$\mathbb{P}(S_n \neq S_n^*) \le \sum_{j \le n} \mathbb{P}(X_j \neq X_j^*) \le \sum_{j \le n} p_j^2.$$

□

Conditional expectation. Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a random variable such that $\mathbb{E}|X| < \infty$. Let $\mathcal{A}$ be a σ-subalgebra of $\mathcal{B}$, $\mathcal{A} \subseteq \mathcal{B}$.

Definition. $Y = \mathbb{E}(X|\mathcal{A})$ is called *conditional expectation* of X given $\mathcal{A}$ if

1. $Y : \Omega \to \mathbb{R}$ is measurable on $\mathcal{A}$, i.e. if B is a Borel set on $\mathbb{R}$ then $Y^{-1}(B) \in \mathcal{A}$.

2. For any set $A \in \mathcal{A}$ we have $\mathbb{E}X\mathrm{I}_A = \mathbb{E}Y\mathrm{I}_A$, where $I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A. \end{cases}$

Definition. If X, Z are random variables then conditional expectation of X given Z is defined by

$$Y = \mathbb{E}(X|Z) = \mathbb{E}(X|\sigma(Z)).$$

Since Y is measurable on $\sigma(Z)$, $Y = f(Z)$ for some measurable function f.

Properties of conditional expectation.

1. (Existence of conditional expectation.) Let us define

$$\mu(A) = \int_A X d\mathbb{P} \quad \text{for} \quad A \in \mathcal{A}.$$

$\mu(A)$ is a σ-additive signed measure on $\mathcal{A}$. Since X is integrable, if $\mathbb{P}(A) = 0$ then $\mu(A) = 0$ which means that μ is absolutely continuous w.r.t. $\mathbb{P}$. By Radon-Nikodym theorem, there exists $Y = \frac{d\mu}{d\mathbb{P}}$ measurable on $\mathcal{A}$ such that for $A \in \mathcal{A}$

$$\mu(A) = \int_A X d\mathbb{P} = \int_A Y d\mathbb{P}.$$

By definition $Y = \mathbb{E}(X|\mathcal{A})$.

2. (Uniqueness) Suppose there exists $Y' = \mathbb{E}(X|\mathcal{A})$ such that $\mathbb{P}(Y \neq Y') > 0$, i.e.

$$\mathbb{P}(Y > Y') > 0 \text{ or } \mathbb{P}(Y < Y') > 0.$$

Since both Y, Y' are measurable on $\mathcal{A}$ the set $A = \{Y > Y'\} \in \mathcal{A}$. One one hand, $\mathbb{E}(Y - Y')\mathrm{I}_A > 0$. On the other hand,

$$\mathbb{E}(Y - Y')\mathrm{I}_A = \mathbb{E}X\mathrm{I}_A - \mathbb{E}X\mathrm{I}_A = 0$$

- a contradiction.

3. $\mathbb{E}(cX + Y|\mathcal{A}) = c\mathbb{E}(X|\mathcal{A}) + \mathbb{E}(Y|\mathcal{A})$.

4. If σ-algebras $\mathcal{C} \subseteq \mathcal{A} \subseteq \mathcal{B}$ then

$$\mathbb{E}(\mathbb{E}(X|\mathcal{A})|\mathcal{C}) = \mathbb{E}(X|\mathcal{C}).$$

Consider a set $C \in \mathcal{C} \subseteq \mathcal{A}$. Then

$$\mathbb{E}\mathrm{I}_C(\mathbb{E}(\mathbb{E}(X|\mathcal{A})|\mathcal{C})) = \mathbb{E}\mathrm{I}_C\mathbb{E}(X|\mathcal{A}) = \mathbb{E}\mathrm{I}_C X \quad \text{and} \quad \mathbb{E}\mathrm{I}_C(\mathbb{E}(X|\mathcal{C})) = \mathbb{E}X\mathrm{I}_C.$$

We conclude by uniqueness.

5. $\mathbb{E}(X|\mathcal{B}) = X$, $\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}X$, $\mathbb{E}(X|\mathcal{A}) = \mathbb{E}X$ if X is independent of $\mathcal{A}$.

6. If $X \le Z$ then $\mathbb{E}(X|\mathcal{A}) \le \mathbb{E}(Z|\mathcal{A})$ a.s.; proof is similar to proof of uniqueness.

7. (Monotone convergence) If $\mathbb{E}|X_n| < \infty, \mathbb{E}|X| < \infty$ and $X_n \uparrow X$ then $\mathbb{E}(X_n|\mathcal{A}) \uparrow \mathbb{E}(X|\mathcal{A})$. Since

$$\mathbb{E}(X_n|\mathcal{A}) \le \mathbb{E}(X_{n+1}|\mathcal{A}) \le \mathbb{E}(X|\mathcal{A})$$

there exists a limit

$$g = \lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{A}) \le \mathbb{E}(X|\mathcal{A}).$$

Since $\mathbb{E}(X_n|\mathcal{A})$ are measurable on $\mathcal{A}$, so is $g = \lim \mathbb{E}(X_n|\mathcal{A})$. It remains to check that

$$\text{for any set } A \in \mathcal{A}, \ \ \mathbb{E}g\mathrm{I}_A = \mathbb{E}X\mathrm{I}_A.$$

Since $X_n\mathrm{I}_A \uparrow X\mathrm{I}_A$ and $\mathbb{E}(X_n|\mathcal{A})\mathrm{I}_A \uparrow g\mathrm{I}_A$, by monotone convergence theorem,

$$\mathbb{E}X_n\mathrm{I}_A \uparrow \mathbb{E}X\mathrm{I}_A \quad \text{and} \quad \mathbb{E}\mathrm{I}_A\mathbb{E}(X_n|\mathcal{A}) \uparrow \mathbb{E}g\mathrm{I}_A.$$

But since $\mathbb{E}\mathrm{I}_A\mathbb{E}(X_n|\mathcal{A}) = \mathbb{E}X_n\mathrm{I}_A$ this implies that $\mathbb{E}g\mathrm{I}_A = \mathbb{E}X\mathrm{I}_A$ and, therefore, $g = \mathbb{E}(X|\mathcal{A})$ a.s.

8. (Dominated convergence) If $|X_n| \le Y, \mathbb{E}Y < \infty$, and $X_n \to X$ then

$$\lim \mathbb{E}(X_n|A) = \mathbb{E}(X|A).$$

We can write,

$$-Y \le g_n = \inf_{m\ge n} X_m \le X_n \le h_n = \sup_{m \ge n} X_m \le Y.$$

Since

$$g_n \uparrow X, \ \ h_n \downarrow X, \ \ |g_n| \le Y, |h_n| \le Y$$

by monotone convergence

$$\mathbb{E}(g_n|\mathcal{A}) \uparrow \mathbb{E}(X|\mathcal{A}), \ \ \mathbb{E}(h_n|\mathcal{A}) \downarrow \mathbb{E}(X|\mathcal{A}) \Longrightarrow \mathbb{E}X_n|A) \to \mathbb{E}(X|A).$$

9. If $\mathbb{E}|X| < \infty, \mathbb{E}|XY| < \infty$ and Y is measurable on $\mathcal{A}$ then

$$\mathbb{E}(XY|\mathcal{A}) = Y\mathbb{E}(X|\mathcal{A}).$$

We can assume that $X, Y \geq 0$ by decomposing $X = X^+ - X^-, Y = Y^+ - Y^-$. Consider a sequence of simple functions

$$Y_n = \sum w_k \mathrm{I}_{C_k}, \;\; C_k \in \mathcal{A}$$

measurable on $\mathcal{A}$ such that $0 \leq Y_n \uparrow Y$. By monotone convergence theorem, it is enough to prove that

$$\mathbb{E}(X\mathrm{I}_{C_k}|\mathcal{A}) = \mathrm{I}_{C_k}\mathbb{E}(X|\mathcal{A}).$$

Take $B \in \mathcal{A}$. Since $BC_k \in \mathcal{A}$,

$$\mathbb{E}\mathrm{I}_B\mathrm{I}_{C_k}\mathbb{E}(X|\mathcal{A}) = \mathbb{E}\mathrm{I}_{BC_k}\mathbb{E}(X|\mathcal{A}) = \mathbb{E}X\mathrm{I}_{BC_k} = \mathbb{E}(X\mathrm{I}_{C_k})\mathrm{I}_B.$$

10. (Jensen's inequality) If $f : \mathbb{R} \to \mathbb{R}$ is convex then

$$f(\mathbb{E}(X|\mathcal{A})) \leq \mathbb{E}(f(X|\mathcal{A})).$$

By convexity,

$$f(X) - f(\mathbb{E}(X|\mathcal{A})) \geq \partial f(\mathbb{E}(X|\mathcal{A}))(X - \mathbb{E}(X|\mathcal{A})).$$

Taking condition expectation of both sides,

$$\mathbb{E}(f(X)|\mathcal{A}) - f(\mathbb{E}(X|\mathcal{A})) \geq \partial f(\mathbb{E}(X|\mathcal{A}))(\mathbb{E}(X|\mathcal{A}) - \mathbb{E}(X|\mathcal{A})) = 0.$$

$\square$