MTH 664 - Lectures 12 & 13

Yevgeniy Kovchegov
Oregon State University
Topics:

- Conditional probability. Independence.
- Conditional expectation.
- Modes of convergence.
Conditional probability. Independence.

Consider a probability space \((\Omega, \mathcal{F}, P)\).

- For two events \(A\) and \(B\) in \((\Omega, \mathcal{F})\) such that \(P(B) > 0\),
  
  \[ P(A|B) = \frac{P(A \cap B)}{P(B)} \]

  is the \textbf{conditional probability} of \(A\) given \(B\).

- Events \(A\) and \(B\) in \((\Omega, \mathcal{F})\) are \textbf{independent} if
  
  \[ P(A \cap B) = P(A)P(B) \]

  Thus, if \(P(B) > 0\), \(P(A|B) = P(A)\).

- Two \(\sigma\)-algebras \(\mathcal{G}_1 \subseteq \mathcal{F}\) and \(\mathcal{G}_2 \subseteq \mathcal{F}\) are said to be \textbf{independent} if all pairs of events \(A \in \mathcal{G}_1\) and \(B \in \mathcal{G}_2\) are independent.
Conditional expectation.

Consider a probability space \((\Omega, \mathcal{F}, P)\) and a random variable \(X \in \mathcal{F}\).

Let \(\mathcal{G} \subseteq \mathcal{F}\) be a smaller \(\sigma\)-algebra.

**Definition.** Conditional expectation \(E[X|\mathcal{G}]\) is a unique function from \(\Omega\) to \(\mathbb{R}\) satisfying:

1. \(E[X|\mathcal{G}]\) is \(\mathcal{G}\)-measurable

2. \(\int_A E[X|\mathcal{G}] \, dP(\omega) = \int_A X \, dP(\omega)\) for all \(A \in \mathcal{G}\)

The existence and uniqueness of \(E[X|\mathcal{G}]\) comes from the Radon-Nikodym theorem.
Change of variables: Radon-Nikodym derivative.

Let $\mu$ and $\nu$ be probability measures on $(\Omega, \mathcal{F})$. We say that $\nu$ is absolutely continuous with respect to $\mu$ if for $A \in \mathcal{F}$,

$$\mu(A) = 0 \implies \nu(A) = 0$$

Abbreviate: $\nu \ll \mu$

Radon-Nikodym Theorem: If $\nu \ll \mu$, there is a function $f \in \mathcal{F}$ such that

$$\int_A f \, d\mu = \nu(A)$$

Such function $f$, called the Radon-Nikodym derivative is usually denoted by $\frac{d\nu}{d\mu}$.

Thus $\int g \, d\nu = \int g \cdot \frac{d\nu}{d\mu} \, d\mu$
Conditional expectation.

The existence and uniqueness of $E[X|\mathcal{G}]$ comes from the Radon-Nikodym theorem: let for all $A \in \mathcal{G}$,

$$\nu(A) = \int_A X(\omega) \, dP(\omega)$$

Then $\nu \ll P$ on $(\Omega, \mathcal{G})$. Thus, by the Radon-Nikodym theorem, there is a function $Y(\omega) = \frac{d\nu}{dP}(\omega) \in \mathcal{G}$ such that

$$\int_A Y(\omega) \, dP(\omega) = \nu(A) = \int_A X(\omega) \, dP(\omega)$$

We let $E[X|\mathcal{G}] = Y$. Then

1. $E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable

2. $\int_A E[X|\mathcal{G}] \, dP(\omega) = \int_A X \, dP(\omega)$ for all $A \in \mathcal{G}$
Conditional expectation.

Example. Consider a simple random walk of a particle on $\mathbb{Z}$ starting at the origin, where the particle moves right with probability $p$ and left with probability $1 - p$, independently at each step.

Each path can be written as an infinite sequence of Bernoulli trials representing left (L) and right (R) moves:

$$LLLRRRLRLRLLLLRLRLRLLRLLLRLR\ldots$$

which can be mapped on a $[0, 1)$ interval as a binary point

$$0.0001110101000010101001000101\ldots$$

Without loss of generality we let $\Omega = [0, 1)$ and the random walk generates a probability measure $\mu$. For example, the event that the path starts with $LR$ corresponds to $\left[\frac{1}{4}, \frac{1}{2}\right) \subseteq \Omega$ and has probability

$$P(LR) = \mu[1/4, 1/2) = p(1 - p)$$
Conditional expectation.

Example. Without loss of generality we let \( \Omega = [0, 1) \) and the random walk generates a probability measure \( \mu \). For example, the event that the path starts with \( LR \) corresponds to \( \left[ \frac{1}{4}, \frac{1}{2} \right) \subseteq \Omega \) and has probability

\[
P(LR) = \mu[1/4, 1/2) = p(1 - p)
\]

Let \( F_k \) be a \( \sigma \)-algebra of the events that depend on the outcome of the first \( k \) steps (Bernoulli trials). Then

\[
F_1 = \left\{ \emptyset, [0, 1/2), [1/2, 1), \Omega = [0, 1) \right\}
\]

\[
F_2 = \left\{ \text{generated by } [0, 1/4), [1/4, 1/2), [1/2, 3/4), \text{ and } [3/4, 1) \right\}
\]

and so on. Here \( X(\omega) = \left\{ \begin{array}{ll} 1 & \text{on } [0, 1/4) \\ 2 & \text{on } [1/4, 1/2) \\ 3 & \text{on } [1/2, 3/4) \\ 4 & \text{on } [3/4, 1) \end{array} \right\} \in F_2 \)
Conditional expectation.

$$X(\omega) = \begin{cases} 
1 & \text{on } [0, 1/4) \\
2 & \text{on } [1/4, 1/2) \\
3 & \text{on } [1/2, 3/4) \\
4 & \text{on } [3/4, 1) 
\end{cases}$$

$$E[X|\mathcal{F}_1](\omega) = \begin{cases} 
1 \cdot (1-p) + 2 \cdot p = 1 + p & \text{on } [0, 1/2) \\
3 \cdot (1-p) + 4 \cdot p = 3 + p & \text{on } [1/2, 1) 
\end{cases}$$
Conditional expectation.

\[ X(\omega) = \begin{cases} 
1 & \text{on } [0, \ 1/4) \\
2 & \text{on } [1/4, \ 1/2) \\
3 & \text{on } [1/2, \ 3/4) \\
4 & \text{on } [3/4, \ 1) 
\end{cases} \]

\[ E[X|\mathcal{F}_1](\omega) = \begin{cases} 
1 \cdot (1 - p) + 2 \cdot p = 1 + p & \text{on } [0, \ 1/2) \\
3 \cdot (1 - p) + 4 \cdot p = 3 + p & \text{on } [1/2, \ 1) 
\end{cases} \]

Check:

\[
\int_{[0, \ 1/2)} X(\omega) \, d\mu(\omega) = \int_{[0, \ 1/4)} X(\omega) \, d\mu(\omega) + \int_{[1/4, \ 1/2)} X(\omega) \, d\mu(\omega)
\]

\[ = \int_{[0, \ 1/4)} 1 \, d\mu(\omega) + \int_{[1/4, \ 1/2)} 2 \, d\mu(\omega) = 1 \cdot \mu[0, \ 1/4] + 2 \cdot \mu[1/4, \ 1/2) \]

\[ = P(LL) + 2 \cdot P(LR) = (1 - p)^2 + 2(1 - p)p = (1 - p)(1 + p) = 1 - p^2 \]

and

\[
\int_{[0, \ 1/2)} E[X|\mathcal{F}_1](\omega) \, d\mu(\omega) = (1 + p) \cdot \mu[0, \ 1/2) = (1 + p) \cdot P(L) = 1 - p^2
\]
Conditional expectation.

\[ X(\omega) = \begin{cases} 
1 & \text{on } [0, 1/4) \\
2 & \text{on } [1/4, 1/2) \\
3 & \text{on } [1/2, 3/4) \\
4 & \text{on } [3/4, 1) 
\end{cases} \]

\[ E[X|\mathcal{F}_1](\omega) = \begin{cases} 
1 \cdot (1 - p) + 2 \cdot p = 1 + p & \text{on } [0, 1/2) \\
3 \cdot (1 - p) + 4 \cdot p = 3 + p & \text{on } [1/2, 1) 
\end{cases} \]

Check:

\[
\int_{[0, 1/2)} X(\omega) \, d\mu(\omega) = 1 - p^2 = \int_{[0, 1/2)} E[X|\mathcal{F}_1](\omega) \, d\mu(\omega)
\]

Similarly,

\[
\int_{[1/2, 1)} X(\omega) \, d\mu(\omega) = (3 + p)p = \int_{[1/2, 1)} E[X|\mathcal{F}_1](\omega) \, d\mu(\omega)
\]

Finally, observe the nesting of (temporal) \(\sigma\)-algebras, called filtration:

\[ \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots \subset \mathcal{F}_\infty = \mathcal{B} \]
Modes of convergence.

Let \((\Omega, \mathcal{F}, P)\) be a probability space, and \(X_1, X_2, \ldots, X\) are random variables over \((\Omega, \mathcal{F})\).

• We say that \(X_n\) converges to \(X\) \textbf{P-almost everywhere} (\textit{P-a.e.}) if

\[
P \left\{ \omega \in \Omega : \limsup_{n \to \infty} |X_n(\omega) - X(\omega)| > 0 \right\} = 0
\]

Since \(P\) is a probability measure, we can also say that \(X_n\) converges to \(X\) \textbf{P-almost surely} (\textit{P-a.s.}).

• Given \(p > 0\). We say that \(X_n\) converges to \(X\) \textbf{in } \(L^p(\Omega, \mathcal{F}, P)\) if

\[
\lim_{n \to \infty} ||X_n - X||_{L^p} = \lim_{n \to \infty} \left( E[|X_n - X|^p] \right)^{1/p} = 0
\]

• We say that \(X_n\) converges to \(X\) \textbf{in probability} (or in \textit{P-measure}) if for all \(\epsilon > 0\),

\[
\lim_{n \to \infty} P(|X_n - X| \geq \epsilon) = 0
\]
Modes of convergence.

**Lemma.** \( X_n \to X \) \( P \)-almost surely if and only if
\[
P( \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon \}) = 0
\]
for any \( \epsilon > 0 \).

**Proof.** Observe that
\[
\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon \} = \{\omega \in \Omega : \exists \{n_k\} \text{ s.t. } |X_{n_k}(\omega) - X(\omega)| \geq \epsilon \}
\]

Next, notice that
\[
\{\omega \in \Omega : \exists \{n_k\} \text{ s.t. } |X_{n_k}(\omega) - X(\omega)| \geq \epsilon \} \subseteq \{\omega \in \Omega : \limsup_{n \to \infty} |X_n(\omega) - X(\omega)| \geq \epsilon \}
\]
and
\[
\{\omega \in \Omega : \limsup_{n \to \infty} |X_n(\omega) - X(\omega)| > \epsilon \} \subseteq \{\omega \in \Omega : \exists \{n_k\} \text{ s.t. } |X_{n_k}(\omega) - X(\omega)| \geq \epsilon \}
\]

W.l.o.g. we let \( \epsilon \in \mathbb{Q}^+ \).
Modes of convergence.

Theorem. (a). Either almost sure convergence or $L^p$-convergence implies convergence in probability.

(b). Conversely, if $Y_n := \sup_{j: j \geq n} |X_j| \to 0$ in probability, then $X_n \to 0$ $P$-almost surely.

Proof.

• The proof that $L^p$-convergence implies convergence in probability is one of the homework problems.

• Note that $X_n \to X$ $P$-almost surely if and only if

$$P( \cap_{N=1}^{\infty} \cup_{n=N}^{\infty} \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\} ) = 0$$

for any $\epsilon > 0$ (see Lemma). Thus, since $P$ is continuous from above,

$$X_n \to X \ P-a.s. \iff \lim_{N \to \infty} P( \cup_{n=N}^{\infty} \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\} ) = 0$$

Now, since

$$P( \{\omega \in \Omega : |X_N(\omega) - X(\omega)| \geq \epsilon\} ) \leq P( \cup_{n=N}^{\infty} \{\omega \in \Omega : |X_n(\omega) - X(\omega) \geq \epsilon\} ),$$

$X_n \to X \ P-a.s.$ implies $X_n \to X$ in probability. Thus we proved part (a) of the theorem.
Modes of convergence.

Theorem. (a). Either almost sure convergence or $L^p$-convergence implies convergence in probability.

(b). Conversely, if $Y_n := \sup_{j: j \geq n} |X_j| \to 0$ in probability, then $X_n \to 0$ $P$-almost surely.

Proof. (continued)

• Suppose $Y_n := \sup_{j: j \geq n} |X_j| \to 0$ in probability, then

$$P( \cap_{N=1}^{\infty} \cup_{n=N}^{\infty} \{ \omega \in \Omega : \sup_{j: j \geq n} |X_j| \geq \epsilon \} ) = \lim_{N \to \infty} P( \{ \omega \in \Omega : \sup_{j: j \geq N} |X_j| \geq \epsilon \} ) = 0$$

While we need to show that $X_n \to 0$ $P$-almost surely, i.e.

$$P( \cap_{N=1}^{\infty} \cup_{n=N}^{\infty} \{ \omega \in \Omega : |X_n(\omega)| \geq \epsilon \} ) = 0$$

However,

$$\{ \omega \in \Omega : |X_n(\omega)| \geq \epsilon \} \subseteq \{ \omega \in \Omega : \sup_{j: j \geq n} |X_j(\omega)| \geq \epsilon \}$$

implying the latter probability is indeed equal to 0. Thus we proved part (b) of the theorem. □
Modes of convergence.

There is actually part (c) of the theorem that states

Theorem. (c). If $X_n \to 0$ in probability and $|X_n| \leq Y$ (P-a.s.) for some $Y \in L^P(\Omega, \mathcal{F}, P)$, then $X_n \to 0$ in $L^p$.

Proof. For $\epsilon > 0$,

$$E[|X_n|^p] = \int_{|X_n| \geq \epsilon} |X_n(\omega)|^p dP(\omega) + \int_{|X_n| < \epsilon} |X_n(\omega)|^p dP(\omega) \leq \int |Y^p(\omega)| dP(\omega) + \epsilon^p$$

Let $A_M = \{\omega \in \Omega : Y^p(\omega) > M\}$. Then

$$\int_{|X_n| \geq \epsilon} Y^p(\omega) dP(\omega) \leq \int M dP(\omega) + \int_{Y^p \leq M} Y^p(\omega) dP(\omega) \leq M \cdot P(|X_n| \geq \epsilon) + \int_{Y^p > M} Y^p(\omega) dP(\omega),$$

where $\int_{A_M} Y^p(\omega) dP(\omega) \to 0$ by the Dominated Convergence Theorem as $0 \leq \phi_M(\omega) = 1_{A_M} \cdot Y^p(\omega) \to 0$ P – a.s. and is dominated by $L^1$-integrable constant function $Y^p$. 

$\square$