MTH 664
Lectures 8 - 11

Yevgeniy Kovchegov
Oregon State University
Topics:

- Expectation.
- Jensen’s inequality.
- Convergence theorems.
- Size biasing.
- Conditional probability. Independence.
- Conditional expectation.
Expectation.

If \( \int_{\Omega} |X(\omega)| \, dP(\omega) \) exists and is finite, then

\[
E[X] = \int_{\Omega} X(\omega) \, dP(\omega) = \int_{\mathbb{R}} x \, d\mu(x)
\]

Properties:

- \( E[X + Y] = E[X] + E[Y] \)
- \( E[aX + b] = aE[X] + b \) for any \( a, b \in \mathbb{R} \)
- If \( X(\omega) \leq Y(\omega), \ \forall \omega \in \Omega \), then \( E[X] \leq E[Y] \)
- If \( P\left(\{\omega \in \Omega : X(\omega) \leq Y(\omega)\}\right) = 1 \), then \( E[X] \leq E[Y] \)
- If \( g(x) \in \mathcal{B} \), then

\[
E[g(X)] = \int_{\Omega} g(X(\omega)) \, dP(\omega) = \int_{\mathbb{R}} g(x) \, d\mu(x)
\]
Jensen’s inequality.

A function $\varphi(x)$ is said to be **convex** over an interval $I$, the domain of the function, if

$$\varphi(\lambda a + (1 - \lambda)b) \leq \lambda \varphi(a) + (1 - \lambda)\varphi(b)$$

for all $\lambda \in [0, 1]$ and all real $a$ and $b$ in $I$.

**Jensen’s inequality:** Suppose $\varphi$ is convex. Then

$$\varphi(E[X]) \leq E[\varphi(X)]$$

**Proof.** Let $\rho = E[X]$. There is a line $\ell(x) = ax + b$ such that

$$\ell(x) \leq \varphi(x) \quad \text{and} \quad \ell(\rho) = \varphi(\rho)$$

Then

$$\varphi(\rho) = \ell(\rho) = E[\ell(X)] \leq E[\varphi(X)]$$
**Jensen’s inequality.**

**Jensen’s inequality:** Suppose ϕ is convex. Then
\[ \varphi(E[X]) \leq E[\varphi(X)] \]

**Examples:**

- \( E[X^2] \geq (E[X])^2 \)
- \( E[e^{aX}] \geq e^{aE[X]} \)
- If \( X > 0 \) then \( E[X^3] \geq (E[X])^3 \)
- If \( X > 0 \) then \( E[X \cdot \ln(X)] \geq E[X] \cdot \ln(E[X]) \)
  as \( \varphi(x) = x \ln(x) \) is convex for \( x > 0 \).
- If \( X > 0 \) then \( E[\ln(X)] \leq \ln(E[X]) \) as \( \varphi(x) = \ln(x) \) is concave for \( x > 0 \).
Markov and Chebyshev inequalities.

**Markov inequality:** If $X \geq 0$ and $E[X] < \infty$, then for any $a > 0$,

$$P(X \geq a) \leq \frac{E[X]}{a}$$

**Chebyshev inequality:** If $E[X] < \infty$ and $Var(X) < \infty$, then for any $a > 0$,

$$P(|X - \mu| \geq a) \leq \frac{Var(X)}{a^2}$$
Convergence theorems.

Fatou's lemma: If $X_n \geq 0$ then $\liminf_{n \to \infty} E[X_n] \geq E\left[\liminf_{n \to \infty} X_n\right]$.

Monotone convergence theorem: If $0 \leq X_n \uparrow X$ then
$$E[X_n] \uparrow E[X].$$

Dominated convergence theorem: If $X_n \to X$ a.s. (with probability one), $|X_n| \leq Y$ for all $n$, and $E[Y] < \infty$, then
$$E[X_n] \to E[X].$$
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 \( A \in \mathcal{F} \),
\[
\mu(A) = 0 \text{ implies } \nu(A) = 0
\]
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 \( \int g \, d\nu = \int g \cdot \frac{d\nu}{d\mu} \, d\mu \)
Size biasing.

Recall the following examples of using Jensen’s inequality:

Let \( \rho = E[X] \) and \( \sigma^2 = Var(X) \), then

- If \( X > 0 \) then \( E[X \cdot \ln(X)] \geq \rho \cdot \ln \rho \) as \( \varphi(x) = x \ln(x) \) is convex for \( x > 0 \).

- If \( X > 0 \) then \( E[\ln(X)] \leq \ln(\rho) \) as \( \varphi(x) = \ln(x) \) is concave for \( x > 0 \).

Question:

- Can we bound \( E[X \cdot \ln(X)] \) from above using \( \rho \) and \( \sigma^2 \)?

The bound needs to be small even as we take \( \rho = E[X] \) to \( +\infty \), provided \( \frac{\sigma}{\rho} \) stays bounded. Of course, we can use \( X > \ln X \) to obtain \( E[X \cdot \ln(X)] \leq \rho^2 \left(1 + \frac{\sigma^2}{\rho^2}\right)\), but it is too coarse.
Size biasing.

Question: Can we bound $E[X \cdot \ln(X)]$ from above using $\rho$ and $\sigma^2$?

Size biasing: $\rho = E_{\mu}[X] = \int_{\mathbb{R}} x \, d\mu(x) > 0$. Let

$$\nu(A) = \int_A \frac{x}{\rho} d\mu(x) \quad \forall A \in \mathcal{B}$$

Then $\nu$ is a probability measure over $(\mathbb{R}, \mathcal{B})$, and $\nu \ll \mu$

Notations: $E_{\mu}[g(X)] = \int_{\mathbb{R}} g(x) \, d\mu(x)$ and

$$E_{\nu}[g(X)] = \int_{\mathbb{R}} g(x) \, d\nu(x) = \int_{\mathbb{R}} g(x) \cdot \frac{x}{\rho} \, d\mu(x) = \frac{1}{\rho} \cdot E_{\mu}[g(X) \cdot X]$$

So,

$$E_{\mu}[X \cdot \ln(X)] = \rho \cdot E_{\nu}[\ln(X)] \leq \rho \cdot \ln \left( E_{\nu}[X] \right) = \rho \cdot \ln \left( \rho + \frac{\sigma^2}{\rho} \right)$$
**Size biasing.**

\[
\rho \cdot \ln \rho \leq E[X \cdot \ln(X)] \leq \rho \cdot \ln \left( \rho + \frac{\sigma^2}{\rho} \right) = \rho \cdot \left( \ln \rho + \ln \left( 1 + \frac{\sigma^2}{\rho^2} \right) \right)
\]

Hence \( \lim_{\rho \to +\infty} \frac{E[X \cdot \ln(X)]}{\rho \cdot \ln \rho} = 1 \).

**Example:** Let \( X \) be an **exponential** random variable with parameter \( \lambda = 1/\rho \). Then

\[
E[X] = \rho \quad \text{and} \quad Var(X) = \sigma^2 = \rho^2
\]

So, \( \frac{\sigma}{\rho} = 1 \) and

\[
\rho \cdot \ln \rho \leq E[X \cdot \ln(X)] \leq \rho \cdot \ln(2\rho) = \rho \cdot \left( \ln \rho + \ln 2 \right)
\]

Observe here,

\[
d\nu(x) = \frac{x}{\rho} \cdot d\mu(x) = \frac{x}{\rho^2} e^{-x/\rho} 1_{x \geq 0} \ dx
\]
Size biasing.

Radon-Nikodym derivative: \( \int g \, d\nu = \int g \cdot \frac{d\nu}{d\mu} \, d\mu \)

**Example:** Let \( X \) be a Poisson random variable with parameter \( \rho > 0 \). Then \( \mu(x) = \sum_{k=0}^{\infty} e^{-\rho} \cdot \frac{\rho^k}{k!} \delta_k(x) \) and \( E[X] = \rho \). Let

\[
\nu(A) = \int_A \frac{x}{\rho} \, d\mu(x) \quad \forall A \in \mathcal{B} \quad \text{i.e.} \quad \frac{d\nu}{d\mu} = \frac{x}{\rho}
\]

then

\[
\nu(x) = \sum_{k=0}^{\infty} \frac{k}{\rho} \cdot e^{-\rho} \frac{\rho^k}{k!} \delta_k(x) = \sum_{k=1}^{\infty} e^{-\rho} \frac{\rho^{k-1}}{(k-1)!} \delta_k(x)
\]

Thus \( X \big|_{\nu} \overset{\text{dist}}{=} X \big|_{\mu} + 1 \), and

\[
E_{\mu}[g(X + 1)] = E_{\nu}[g(X)] = \frac{1}{\rho} \cdot E_{\mu}[g(X) \cdot X]
\]

**Application:** \( E_{\mu}[X^2] = \rho \cdot E_{\mu}[X+1] = \rho^2 + \rho \), implying \( Var(X) = \rho \).
Next find \( E_{\mu}[X^3] \), and so on
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 conditional probability of \(A\) given \(B\).

- Events \(A\) and \(B\) in \((\Omega, \mathcal{F})\) are 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 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 \( A \in \mathcal{F} \),

\[
\mu(A) = 0 \quad \text{implies} \quad \nu(A) = 0
\]

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:

$$LLLRRRLRLRLL...$$

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

$$0.00011101010000101001000101...$$

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 $[1/4, 1/2) \subset \Omega$ 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 $\mathcal{F}_k$ be a $\sigma$-algebra of the events that depend on the outcome of the first $k$ steps (Bernoulli trials). Then

$$\mathcal{F}_1 = \left\{ \emptyset, [0, 1/2), [1/2, 1), \Omega = [0, 1) \right\}$$

$$\mathcal{F}_2 = \left\{ \text{generated by } [0, 1/4), [1/4, 1/2), [1/2, 3/4), \text{ and } [3/4, 1) \right\}$$

and so on. Here $X(\omega) = \begin{cases} 1 & \text{on } [0, 1/4) \\ 2 & \text{on } [1/4, 1/2) \\ 3 & \text{on } [1/2, 3/4) \\ 4 & \text{on } [3/4, 1) \end{cases} \notin \mathcal{F}_2$
Conditional expectation.

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

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

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