MTH 664 - Lectures 6 and 7

Yevgeniy Kovchegov
Oregon State University
Topics:

• Measure and Integral.

• Random variables.

• Expectation.

• Jensen’s inequality.

• Convergence theorems.
Measure and Integral.

Definition 1. A collection $\mathcal{F}$ of subsets of $\Omega$ is a $\sigma$-algebra if

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$
3. If $A_1, A_2, A_3, \ldots \in \mathcal{F}$, then $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$

The elements of $\mathcal{F}$ are said to be $\mathcal{F}$-measurable.

Note: $\mathcal{F}$ is an algebra if the latter condition is substituted with $\bigcup_{j=1}^{N} A_j \in \mathcal{F}$ for all $N < \infty$.

If $\Omega$ is a topological space (e.g. $\mathbb{R}^n$), the $\sigma$-algebra generated by all open sets is called a Borel $\sigma$-algebra, and the elements are called the Borel sets.
Measure and Integral.

Definition 2. A set function $\mu : \mathcal{F} \rightarrow [0, \infty]$ is called a measure on a measurable space $(\Omega, \mathcal{F})$ if

1. $\mu(\emptyset) = 0$

2. For any sequence of disjoint (mutually non-intersecting) subsets $A_1, A_2, \ldots,$

$$\mu \left( \bigcup_{j=1}^{\infty} A_j \right) = \sum_{j=1}^{\infty} \mu(A_j)$$

If $\mu(\Omega) = 1$, than $\mu$ is a probability measure. In this case, $\mu$ will satisfy all the axioms of probability, and a measurable subset $A \in \mathcal{F}$ is an event.
Measure and Integral.

**Definition 3.** Suppose $\mathcal{G}$ is a collection of subsets of $\Omega$. A set function $\mu : \mathcal{G} \to \mathbb{R}$ is **countably additive** on $\mathcal{G}$ if for any sequence of disjoint (mutually non-intersecting) subsets $A_1, A_2, \ldots$ in $\mathcal{G}$,

$$
\mu \left( \bigcup_{j=1}^{\infty} A_j \right) = \sum_{j=1}^{\infty} \mu(A_j)
$$

Famous examples.

• **Dirac point-mass.** $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if $x \notin A$

• **Lebesgue measure** $\mu$ on $\Omega = \mathbb{R}^d$.

$$
m \left( [a_1, b_1) \times \cdots \times [a_d, b_d) \right) = \prod_{j=1}^{d} (b_j - a_j)
$$
Probability measure.

Let $P$ be a probability measure on $(\Omega, \mathcal{F})$. Then the following properties can be shown.

(1.) **Monotonicity:** If $A \subseteq B$ then $P(B) \geq P(A)$

(2.) **Subadditivity:** $P(\bigcup_{j=1}^{\infty} A_j) \leq \sum_{j=1}^{\infty} P(A_j)$

(3.) **Continuity from below:** If $A_1 \subseteq A_2 \subseteq \ldots$ then
$$\lim_{j \to \infty} P(A_j) = P(\bigcup_{j=1}^{\infty} A_j)$$

(4.) **Continuity from above:** If $A_1 \supseteq A_2 \supseteq \ldots$ then
$$\lim_{j \to \infty} P(A_j) = P(\bigcap_{j=1}^{\infty} A_j)$$
Product spaces.

If \((\Omega_j, \mathcal{F}_j, P_j)\) \((j = 1, 2, \ldots, d)\) are probability spaces with corresponding probability measures. Let

\[
\Omega = \Omega_1 \times \cdots \times \Omega_d = \{ (\omega_1, \ldots, \omega_d) : \omega_j \in \Omega_j \}
\]

be the product (sample) space. Then \(\{ A_1 \times \cdots \times A_d : A_i \in \mathcal{F}_i \}\) generates a \(\sigma\)-algebra over \(\Omega\),

\[
\mathcal{F} = \mathcal{F}_1 \times \cdots \times \mathcal{F}_d
\]

together with the probability (product) measure

\[
P(A_1 \times \cdots \times A_d) = P_1(A_1) \cdot P_2(A_2) \cdots P_d(A_d)
\]

Examples.

- Roll two dice.
- Toss a coin \(d\) times.
Random variables.

A real valued function $X : \Omega \to \mathbb{R}$ is said to be a random variable defined on $(\Omega, \mathcal{F})$ if it is $\mathcal{F}$-measurable function, i.e.

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

for any Borel set $B \subset \mathbb{R}$.

Denote: $X(\omega) \in \mathcal{F}$

Example.

- Indicator variable. Take $E \in \mathcal{F}$. Let $1_E(\omega) = \begin{cases} 1 & \omega \in E \\ 0 & \omega \notin E \end{cases}$
Random variables.

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

If \(X\) is a random variable, then \(X\) induces a probability measure on \((\mathbb{R}, \mathcal{B})\),

\[
\mu(B) = P\left(X^{-1}(B)\right) = P\left(\{\omega \in \Omega : X(\omega) \in B\}\right)
\]

for any \(B \in \mathcal{B}\) (Borel \(\sigma\)-algebra).

This probability measure \(\mu\) is called the (probability) distribution of r.v. \(X\).

While the (cumulative) distribution function is defined as

\[
F(a) = P(X \leq a) = P\left(\{\omega \in \Omega : X(\omega) \in (-\infty, a]\}\right) = \mu\left((-\infty, a]\right)
\]
**Distribution function.**

The (cumulative) **distribution function** is defined as

\[ F(a) = P(X \leq a) = \mu((-\infty, a]) \]

Properties:

I. \( F \) is nondecreasing

II. \( \lim_{a \to \infty} F(a) = 1, \quad \lim_{a \to -\infty} F(a) = 0 \)

III. \( F \) is right continuous: \( F(a+) = \lim_{x \downarrow a} F(x) = F(a) \)

IV. \( F(a-) = \lim_{x \uparrow a} F(x) = P(X < a) \)

V. \( P(X = a) = F(a) - F(a-) \)

Note: properties I-III suffice for a function \( F \) to be a distribution function for some r.v.
Continuous random variables.

$X$ is a continuous random variable if

$$F(a) = \int_{-\infty}^{a} f(x) \, dx,$$

for all $a \in \mathbb{R}$, where $f$ is said to be the probability density function.

There

$$P(X = a) = \lim_{\epsilon \downarrow 0} \int_{a-\epsilon}^{a+\epsilon} f(x) \, dx = 0$$

In other words, distribution $\mu$ contains no point-mass components.
Discrete random variables.

$X$ is a **discrete random variable** if there are countably many values

$$a_1, a_2, \ldots \in \mathbb{R}$$

such that

$$\mu(E) = \sum_j p_j \cdot \delta_{a_j}(E),$$

where $p_j > 0$ and $\sum_j p_j = 1$.

There

$$F(a_j) - F(a_j-) = p_j$$

and $F(x) - F(x-) \equiv 0$ if $x \notin \{a_1, a_2, \ldots\}$. 
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 \( P\left( \left\{ \omega \in \Omega : X(\omega) \leq Y(\omega) \right\} \right) = 1 \), 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 \( \mathcal{I} \), 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 \( \mathcal{I} \).

**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 \( \varphi \) 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 \( x > 0 \).