Lecture 6:

Recap $\mu(f) = \int f \, d\mu$ defined in steps:

1) indicator functions $\int I_A \, d\mu = \mu(A)$
2) step functions $\sum_k a_k I_{A_k} : \sum_k a_k \mu(A_k)$

3) for $f \in m\Sigma^+$ we set

$$\int f \, d\mu = \sup_{\substack{g \le f \\ g \text{ step function}}} \int g \, d\mu.$$

3*) for $f \in m\Sigma$ we set $\int f \, d\mu = \int f^+ d\mu - \int f^- d\mu$

Important Question: When can limits & integrals be exchanged?

$$\left.\begin{array}{l} \text{1) Monotone Convergence Theorem} \\ \text{2) Dominated Convergence Theorem} \end{array}\right\} \Rightarrow \int \lim_n f_n \, d\mu = \lim_n \int f_n \, d\mu$$

Let $f \in m\Sigma^+$. We consider the restricted integral:

Define $\lambda(A) = \lambda_{f,\mu}(A) = \int_A f \, d\mu = \int f I_A \, d\mu.$

We also write $\mu(f; A)$ for this.

This defines a measure: • $\lambda(\emptyset) = 0$

• $\lambda(A) \geq 0$ for all $A \in \Sigma$.

• σ-additive as for disjoint (A_i)

$$\lambda\left(\bigcup_{i=1}^{\infty} A_i\right) = \int f I_{\bigcup_{i=1}^{\infty} A_i} \, d\mu$$

$$= \int f \sum_{i=1}^{\infty} I_{A_i} \, d\mu \quad \text{by disjointness}$$

$$= \int f \lim_{n\to\infty} \sum_{i=1}^{n} I_{A_i} \, d\mu = \int \lim_{n\to\infty} \sum_{i=1}^{n} f I_{A_i} \, d\mu$$

$$= \lim_{n\to\infty} \int \sum_{i=1}^{n} f I_{A_i} \, d\mu \quad \text{by MCT}$$

$$= \sum_{i=1}^{\infty} \int f I_{A_i} \, d\mu = \sum_{i=1}^{\infty} \lambda(A_i).$$

The function f is called the density of λ.
We write $f = \frac{d\lambda}{d\mu}$.

Note: If $\mu(A) = 0$ then $\lambda(A) = \int_A f \, d\mu = 0$, all null sets of μ are null sets of λ.

Def$^{\underline{n}}$ We say a measure μ is σ-finite if there exist $A_i \in \Sigma$, $i \in \mathbb{N}$ s.t. $\bigcup A_i = S$ and $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$.

Radon-Nikodym Theorem:

If λ, μ are two σ-finite measures on (S, Σ) that satisfy $\mu(A) = 0 \Rightarrow \lambda(A) = 0$ then there exists a density $f = \frac{d\lambda}{d\mu}$ such that $\lambda(A) = \int_A f \, d\mu$ for all $A \in \Sigma$.

Example: If we take μ to be the Lebesgue measure on $\mathbb{R}$ and $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, then $\lambda(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ is the probability that a normal random variable lies in A, i.e. is the probability measure associated with a standard random variable.

Expectations

Expectations (expected values / mean / arithmetic mean) are integrals relative to probability measures.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X a random variable. We define

$$\mathbb{E}(X) = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \int X \, d\mathbb{P}.$$

If it exists, we say X is integrable wrt $\mathbb{P}$:
$$\left(\int_\Omega |X| \, d\mathbb{P} < \infty \right)$$

Example: Consider the finite probability space $\Omega = \{1,2,3,4,5,6\}$ with $\mathbb{P}(\{i\}) = 1/6$ for all i:
Every function $f: \Omega \to \mathbb{R}$ is a finite sum of indicator functions:

$$f(x) = f(1)\, I_{\{1\}}(x) + f(2)\, I_{\{2\}}(x) + \ldots + f(6)\, I_{\{6\}}(x)$$

$$\Rightarrow \int f \, d\mathbb{P} = \frac{1}{6} \cdot f(1) + \frac{1}{6} \cdot f(2) + \ldots + \frac{1}{6} f(6) = \frac{f(1) + \ldots + f(6)}{6}$$

All theorems on integrals become theorems on expected values. For example, assume $X_n \to X$ a.s.

- If $0 \le X_n \uparrow X$, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$ (monotone convergence)
- If $|X_n| \le Y$ with $\mathbb{E}(Y) < \infty$, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$ (dominated convergence)
- $\mathbb{E}(X) = \mathbb{E}(\lim_{n\to\infty} X_n) \le \liminf_{n\to\infty} \mathbb{E}(X_n)$ (Fatou's lemma)

We also define $\mathbb{E}(X; E)$ as for $\mu(f; E)$:

$$\mathbb{E}(X; E) = \mathbb{E}(X \cdot I_E) = \int_F X \, d\mathbb{P}$$

Markov's inequality: Let Z be a random variable with values in G (a.s.), and let $g: G \to [0, \infty]$ be a nondecreasing measurable function. Then,

$$\mathbb{E}(g(Z)) \ge \mathbb{E}(g(Z) I_{\{Z \ge c\}}) = \mathbb{E}(g(Z); Z \ge c)$$
$$\ge \mathbb{E}(g(c); Z \ge c) = g(c)\, \mathbb{E}(I_{\{Z \ge c\}}) = g(c)\, \mathbb{P}(Z \ge c)$$

We can write as $\mathbb{P}(Z \ge c) \le \frac{1}{g(c)} \mathbb{E}(g(Z))$.

Special case: $\mathbb{P}(Z \ge c) \le \mathbb{E}(Z)/c$ for non-negative Z.

If $Z : \Omega \to \mathbb{N} \cup \{0\}$ we get

$$\mathbb{P}(Z \neq 0) = \mathbb{P}(Z \geq 1) \leq \mathbb{E}(Z).$$

Another special case is $g(x) = e^{\theta x}$:

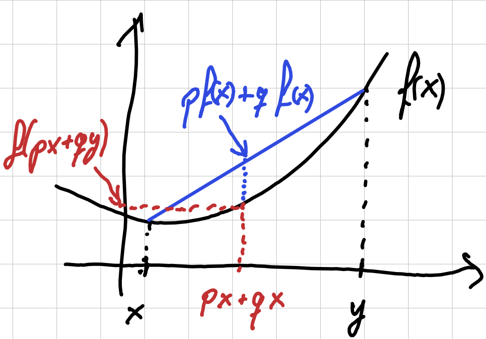
$$\mathbb{P}(Z \geq c) \leq e^{-\theta c} \mathbb{E}(e^{\theta Z}) \quad \text{for all } \theta > 0.$$

Jensen's inequality

A function $f : I \to \mathbb{R}$ is said to be convex (interval)

if $f(px + qy) \leq p f(x) + q f(y) \quad \forall p, q \geq 0$ s.t. $p + q = 1$

$\forall x, y \in I$.



Examples:

$x \mapsto e^x$

$x \mapsto x^r$ for $r \geq 1$

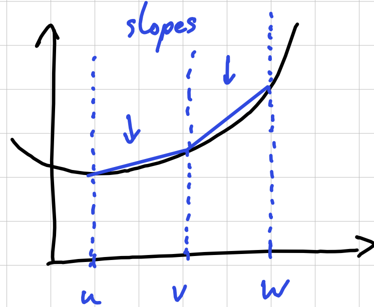
$x \mapsto 1$

Jensen's inequality: Let $f : I \to \mathbb{R}$ be convex and $X : \Omega \to I$ be a random variable. Then

$$\mathbb{E}(f(X)) \geq f(\mathbb{E}X).$$

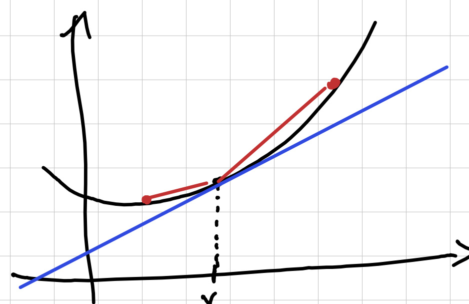
Proof: We can rewrite the convexity condition as $\dfrac{f(v)-f(u)}{v-u} \leq \dfrac{f(w)-f(v)}{w-v}$

for $u<v<w$.



This implies that left and right derivative exist. We get

$f(x) \geq f(v) + m(x-v)$ for any m between left & right derivative.



Substituting gives

$$f(X) \geq f(\mathbb{E}X) + m(X-\mathbb{E}X)$$

$$\Downarrow$$

$$\mathbb{E}(f(X)) \geq \mathbb{E}\Big(\underbrace{f(\mathbb{E}(X))}_{\text{const}} + m\big(X - \underbrace{\mathbb{E}(X)}_{\text{const}}\big)\Big)$$

$$= f(\mathbb{E}(X)) + m\big(\mathbb{E}(X) - \underbrace{\mathbb{E}(\mathbb{E}(X))}_{=\mathbb{E}(X)}\big)$$

$$= f(\mathbb{E}(X)).$$

$\square$

L^p Norm : For $p \geq 1$ we define

$$\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$$

(norm for $p \geq 1$, def. makes "sense" for $p > 0$)

$L^p(\Omega, F, P)$ is the space of all random variables X for which $\|X_p\| < \infty$.

Let $f(x) = x^{r/p}$ for $r \geq p \geq 1$.

f is convex on $[0, \infty)$. Jensen's ineq. gives

$\mathbb{E}(Y^{r/p}) \geq \mathbb{E}(Y)^{r/p}$ for non-neg r.v. Y.

Letting $Y = |X|^p$ we get

$$\mathbb{E}(|X|^r) \geq \mathbb{E}(|X|^p)^{r/p} \Rightarrow \mathbb{E}(|X|^r)^{1/r} \geq \mathbb{E}(|X|^p)^{1/p}.$$

So if $\|X\|_r < \infty$ then $\|X\|_p < \infty$.

i.e. $L^r(\Omega, \mathcal{F}, P) \subseteq L^p(\Omega, \mathcal{F}, P)$!

The family $L^2(\Omega, F, P)$ is important!