$$P\left(\bigcap_{i=1}^{n} X^{-1}(A_i)\right) = \prod_{i=1}^{n} P(X_i^{-1}(A_i)) \quad \text{(To show)}$$

$A_i = ((-\infty, x] : x \in \mathbb{R}\}$

From def. of indep of random vectors
$(A_i : i \in [n])$ are indep.
Remark implies that $(\sigma(A_i) : i \in [n])$ are indep $\blacksquare$

Exercise: Define when two r.v $X : \Omega \to \mathbb{R}^n$, $Y : \Omega \to \mathbb{R}^n$ are independent.

$P(A(x) \cap A(y)) = P(A(x))P(A(y))$

1. Discrete Random Vectors.

If a random vector $X : \Omega \to \underbrace{x_1 \times \dots \times x_n}_{\text{countable}} \subseteq \mathbb{R}^n$ then it is called a "discrete random vector".

Prob. mass function $P_X(x) = P\left(\bigcap_{i=1}^{n} X_i^{-1}\{x_i\}\right)$

For an indep. random vector $P_X(x) = \prod_{i=1}^{n} P_{X_i}(x_i)$ $\forall x \in \mathbb{R}^n$

Ex: (Multiple coin tosses) $\Omega = (H, T)^n$

$E_i = \{\omega_i \in \Omega : \omega_i = H\}$ $\mathcal{F} = \sigma(E_1, E_2, \dots)$

$P(E_1 \cap E_2 \cdots \cap E_n) = p^n$

$A_n = \{\omega \in \Omega : \omega_i = H \text{ for some } i \in [n]\}$

$= \bigcup_{i \in [n]} E_i \in \sigma(E_1, \dots)$

$B_n = \{\omega \in \Omega : \omega_1, \omega_2 \cdots \omega_{n-1} = T, \omega_n = H\}$

$= E_1^c \cap E_2^c \cap \cdots E_{n-1}^c \cap E_n \in \sigma(E_1 \cdots)$

$P(B_n) = (1-p)^{n-1} p$

$P(A_n) = 1 - p^n$ $\quad \mathcal{F}_n = \sigma(E_1, \dots E_n)$

If $A \in \mathcal{F}_n$, then $P(A) = \sum_{\omega \in A} p^{k_n(\omega)} (1-p)^{n-k_n(\omega)}$

Ex: [Finite coin tosses] $\Omega = \{H, T\}^n$ $\mathcal{F} = 2^\Omega$

$P(\omega) = \frac{1}{2^n}$, $\forall \omega \in \Omega$

$X : \Omega \to \mathbb{R}^n$ s.t $X_i(\omega) = 1_{\{\omega_i = H\}}$

There is a bijection b/w Ω and $\{0,1\}^n$

For any $x \in [0,1]^n$

$$F_X(x) = P\left(\bigcap_{i=1}^n \{X_i \le x_i\}\right)$$

$$= \begin{cases} 1, & \text{if all } x_i \ge 1 \\ \frac{1}{2}, & \text{if } x_i \in [0,1] \text{ for some } i \in [n] \\ 0, & \text{if } x_i < 0 \text{ for some } i \in [n] \end{cases}$$

$$N(x) = \sum_{i=1}^n 1\{0 \le x_i < 1\}$$

$$F_{X_i}(x_i) = \begin{cases} 1 & \text{if } x_i \ge 1 \\ \frac{1}{2} & \text{if } 0 \le x_i < 1 \\ 0 & \text{if } x_i < 0 \end{cases}$$

Check: $F_X(x) = \prod_{i=1}^n F_{X_i}(x_i)$ is an "i.i.d" vector

Continuous Random Vector

1.3 Properties of distribution function

i. $x, y \in \mathbb{R}^n$, $x_i \le y_i \ \forall\ i \in [n]$

ii. $F_X(x)$ is right continuous at all points $x \in \mathbb{R}^n$

iii. $\lim_{x_i \to -\infty} F_X(x) = 0$ $\quad \lim_{x_j \to \infty} F_X(x) = 1$ $\ \forall\ j \in [n]$

Proof: (i) $A(x) \subseteq A(y)$

$$\bigcap_{i=1}^n X_i^{-1}(-\infty, x_i] \subseteq \bigcap_{i=1}^n X_i^{-1}(-\infty, y_i]$$

(ii) $(A(x_m)) \downarrow \quad A(x) = \bigcap_{m \in \mathbb{N}} A(x_m)$

(iii) $A(x) = \phi$ if $x_i = -\infty$ for some $i \in [n]$
$A(x) = \Omega$ if $x_i = \infty$ $\forall$ $i \in [n]$

Lecture 6

Recap: Prob. space Ω: sample space / space of outcomes

$\mathcal{F}_\Omega$: Event space $\subseteq 2^\Omega$ is a σ-alg if [1. $\Omega \in \mathcal{F}_\Omega$

2. If $A \in \mathcal{F}$, $A^c \in \mathcal{F}$ 3. If $A_i \in \mathcal{F}$ $\forall n \in N$ then $\bigcup_{n \in N} A \in \mathcal{F}$

P: Probability function $\mathcal{F}_\Omega \to [0,1]$

[1 - Non-negativity 2 - σ-Additivity 3 - Certainity]

Properties of probability
1. $P(\phi) = 0$
2. Finite additivity
3. Monotonicity
4. Inclusion - exclusion
5. Continuity * (Lim of sets)

- $A = B$ if $A \subseteq B$ & $B \subseteq A$
- $A \subseteq B$) Take an arbitrary element $\omega \in A$, and show that $\omega \in B$.
- Order is a relation $\omega_1, \omega_2 \in A$ $\omega_1 \leq \omega_2$ order

$(\omega_1, \omega_2) \in R \subseteq A \times A$

$A \subseteq B$ [set inclusion is a partial order]

$A_n \uparrow$ $\qquad \lim_n A_n = \bigcup_n A_n$

$A_n \downarrow$ $\qquad \lim_n A_n = \bigcap_n A_n$

- Conditional Analysis & independence

$A, B \in \mathcal{F}_\Omega$ & $P(A) > 0$ $\quad P(B|A) = \dfrac{P(A \cap B)}{P(A)}$

- $P(\cdot \,|A) : \mathcal{F}_\Omega \to [0,1]$ is a prob. measure $(\Omega', \mathcal{F}', P')$ defined on $\Omega' = A$
$\mathcal{F}' = \{B \cap A : B \in \mathcal{F}_\Omega\}$

- Independence
  Two collection of events $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{F}$ are indep.
  if $P(A \cap B) = P(A) P(B)$ for any $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$

  Arbitrary collection $\{\mathcal{A}_i : i \in I\}$ are indep if
  $$P(\bigcap_{i \in F} A_i) = \prod_{i \in F} P(A_i)$$ for any finite set $F \subseteq I$ and $A_i \in \mathcal{A}_i$ for all $i \in F$.

- Law of total Probabilities: For a partition $(A_n \in \mathcal{F} : n \in \mathbb{N})$ of Ω, then
  $$P(B) = \sum_{n \in \mathbb{N}} P(B \cap A_n) \quad \text{for any event } B \in \mathcal{F}$$
  $$= \sum_{n \in \mathbb{N}} P(B | A_n) P(A_n) \quad \text{only if } P(A_n) > 0 \ \forall n \in \mathbb{N}$$

- Conditional independence $C \in \mathcal{F}$ $P(C) > 0$
  A collection $\mathcal{A} \subseteq \mathcal{F}$ is conditionally indep. given C if
  $$P(\bigcap_{i \in F} A_i | C) = \prod_{i \in F} P(A_i | C) \quad \text{for } A_i \in \mathcal{F} \ \forall i \in F \text{ finite}$$

Ex: Let A, B not be indep $[P(A \cap B) \neq P(A)P(B)]$ $P(A) > 0$
will show that A and B are cond. indep. given A.
That is, $P(A \cap B | A) = P(A|A) P(B|A)$
$$\text{L.H.S } P(A \cap B | A) = \frac{P(A \cap B)}{P(A)} = P(B|A) \frac{P(A \cap A)}{P(A)}$$
$$= P(B|A) P(A|A)$$

Ex: Let $A, B \in \mathcal{F}$ be indep events.
$P(A \cap B) > 0$, $P(A \cap B) < 1$.
The events A & B are not conditionally indep given $A \cup B$
$\therefore$ $P(A \cap B | A \cup B) \neq P(A | A \cup B) P(B | A \cup B)$

$$L.H.S = \frac{P(A\cap B)}{P(A\cup B)} = \frac{P(A)\,P(B)}{P(A\cup B)} = P(A \mid A\cup B)\,P(B)$$

If $P(B) \neq P(B \mid A\cup B)$ then we are done.

- Random variable $(\Omega, \mathcal{F}, P)$

$X : \Omega \to \mathbb{R}$ s.t. $X^{-1}(-\infty, x] \in \mathcal{F}$

[This implies that $X^{-1}(B) \in \mathcal{F}$ for any $B \in \mathcal{B}(\mathbb{R})$]

$\sigma((-\infty, x] : x \in \mathbb{R}) = \mathcal{B}(\mathbb{R})$

$\mathcal{B}$ is generated by $(-\infty, x]$; $X^{-1}(\mathcal{B})$ is generated by $X^{-1}(-\infty, x]$

$\sigma(X) = \sigma(X^{-1}(-\infty, x] : x \in \mathbb{R})$

$F_X(x) := P \circ X^{-1}(-\infty, x]$

- $F_X(x) \leq F_X(y)$ $\quad x \leq y$
- right cont. at all points $x \in \mathbb{R}$
- $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$

Random Vectors

$\mathcal{B}(\mathbb{R}^n)$

Defn: $\pi_i : \mathbb{R}^n \to \mathbb{R}$

$A \subseteq \mathbb{R}$, then $\pi_i^{-1}(A) = \mathbb{R} \times \dots A \times \mathbb{R}$

$$\mathcal{B}(\mathbb{R}^n) = \sigma(\pi_i^{-1}(-\infty, x] : x \in \mathbb{R}, i \in [n])$$

Random Vector : $X : \Omega \to \mathbb{R}^n$ s.t. $B \in \mathcal{B}(\mathbb{R}^n)$ $X^{-1}(B) \in \mathcal{F}$

[Measurability: $f : \mathbb{R}^n \to \mathbb{R}^m$ is Borel measurable

iff $f^{-1}(B_m) \in \mathcal{B}(\mathbb{R}_n)$ where $B_m \in \mathcal{B}(\mathbb{R}_m)$]

$$A(x) = \bigcap_{i=1}^{n} \{X_i \leq x_i\} \in \mathcal{B}(\mathbb{R}^n) = \bigcap_{i=1}^{n} A_i(x_i)$$

$\{X_i^{-1}(-\infty, x_i)\}$

$F_X(x) = P(A(x))$

$F_{X_i}(x_i) = P(A_i(x_i))$

$$= \lim_{\substack{x_j \to \infty \\ j \neq i}} F_X(x)$$

Independence: $F_X(x) = F_{X_1}(x_1)\,F_{X_2}(x_2) \cdots F_{X_n}(x_n)$

This implies that $(\sigma(X_1), \sigma(X_2) \dots \sigma(X_n))$ are indep. collections.

Transformation of random vectors

$f: \mathcal{X} \to \mathcal{Y}$ $\quad \mathcal{F} \subseteq 2^{\mathcal{X}}$ $\quad \mathcal{G} \subseteq 2^{\mathcal{Y}}$

If for any $G \in \mathcal{G}$, $f^{-1}(G) \in \mathcal{F}$, then f is $(\mathcal{G}, \mathcal{F})$ measurable.

$\mathcal{X} = \mathbb{R}^m$ Associated σ-algebra is $\mathcal{B}(\mathbb{R}^m)$

$f: \mathbb{R}^m \to \mathbb{R}^n$ is measurable (Borel measurable)

$f^{-1}(B_m) \in \mathcal{B}(\mathbb{R}^n)$ where $B_m \in \mathcal{B}(\mathbb{R}^m)$

Prop: If $X: \Omega \to \mathbb{R}$ is an r.v on $(\Omega, \mathcal{F}, P)$ and $g: \mathbb{R} \to \mathbb{R}$ is Borel measurable fn. then $g \circ X: \Omega \to \mathbb{R}$ is a r.v.

Proof $Y = g \circ X$ is a r.v

$(-\infty, y] \in \mathcal{B}(\mathbb{R})$

Then,
$$
\begin{aligned}
Y^{-1}(-\infty, y] &= X^{-1} \circ g^{-1}(-\infty, y] \\
&= \{\omega \in \Omega : g(x) \leq y\} \\
&= \{\omega \in \Omega : x \in g^{-1}(-\infty, y]\} \\
&= X^{-1} \circ g^{-1}(-\infty, y] \\
&= X^{-1}(\underbrace{\mathcal{B}(\mathbb{R})}_{\in \mathcal{B}(\mathbb{R})}) \in \mathcal{F}
\end{aligned}
$$

Ex.: Let $g: \mathbb{R} \to \mathbb{R}$ be a monotone increasing function. Then $g \circ X$ is a r.v for X a r.v

$$\underbrace{g^{-1}(-\infty, x]}_{\text{set inverse mapping}} = (-\infty, \underbrace{g^{-1}}_{\substack{\text{operator overloading} \\ \text{inverse of a fun.}}}(x)]$$

$$
\begin{aligned}
F_Y(y) &= P \circ X^{-1} \circ g^{-1}(-\infty, y] \\
&= P \circ X^{-1}(-\infty, g^{-1}(y)] = F_X(g^{-1}(y))
\end{aligned}
$$

Eg.: $g: \mathbb{R}_+ \to \mathbb{R}_+$ $\quad g(x) = e^{-\theta x}$ $\quad g^{-1}(y) = -\frac{1}{\theta} \ln y$

$$F_Y(y) = F_X\left(-\frac{1}{\theta} \ln y\right)$$

Prop: [Indep of functions of r.v] If $X: \Omega \to \mathbb{R}^n$ is an indep. r.v on $(\Omega, \mathcal{F}, P)$, and $g_i: \mathbb{R} \to \mathbb{R}$ (Borel measurable) $\forall i \in [n]$

Then $(g_i(x_i) : i \in [n])$ are indep random vector

Probability is always of events not of outcomes or r.v.

Proof: It suffices to show that

$$P\left(\bigcap_{i=1}^{n}\{g_i(x_i)\le x_i\}\right) = \prod_{i=1}^{n} P(\{g_i(x_i)\le x_i\})$$

L.H.S $\quad P\left(\bigcap_{i=1}^{n} X_i^{-1}\circ \underbrace{g_i^{-1}(-\infty,x_i]}_{\in \mathcal{B}(\mathbb{R})}\right) = \prod_{i=1}^{n} P(X_i^{-1}\circ g_i^{-1}(-\infty,x_i])$

$\equiv$ RHS ∎

$\therefore$ any function/transformation on collection of random var. can not make them independent.

2. Function of random Vectors

Prop: $X: \Omega \to \mathbb{R}^n$ defined on $(\Omega, \mathcal{F}, P)$

$g: \mathbb{R}^n \to \mathbb{R}^m$ Borel measurable (hidden are $\mathcal{B}(\mathbb{R}^n), \mathcal{B}(\mathbb{R}^m)$)

This has been designed to make the transformations as random vectors.

$\sigma(g(x)) \subseteq \sigma(x) \quad g: \mathbb{R}\to\mathbb{R} \quad g(x)=C$

For a bijection it will be equal.

$$B(y) := \bigcap_{j=1}^{m}\{x\in\mathbb{R}^m : g_j(x)\le y_j\}\in \mathcal{B}(\mathbb{R}^m) \ \forall\, y\in\mathbb{R}^m$$

Then, $Y \equiv g.X: \Omega \to \mathbb{R}^m$ is a random vector.

Further, $F_Y(y) = P(X^{-1}(B(y)))$

Ex: Sum of Random Variables

Lecture: 7 Expectation of R.V

$(\Omega, \mathcal{F}, P)$ probability space

- N trials of a random experiment.

$X_1, X_2, X_3 \ldots X_n$ real values associated with N trials.

$X: \Omega \to \mathcal{X}^N$ where $\underset{\sim}{\mathcal{X}} \subseteq \mathbb{R}$

discrete $\to$ discrete random vector

$$\hat{m} = \frac{1}{N} \sum_{i=1}^{N} X_i(N)$$

$$\hat{P}(x) = \frac{1}{N} \sum_{i=1}^{N} 1\{X_i(\omega) = x\}$$

$$\hat{m} = \frac{1}{N} \sum_{i=1}^{N} \Big[\underbrace{\sum_{x \in \mathcal{X}} x\, 1_{\{X_i(\omega) = x\}}}_{X_i(\omega)} \Big]$$

$$= \sum_{x \in \mathcal{X}} x \Big[\frac{1}{N} \sum_{i=1}^{N} 1\{X_i(\omega) = x\} \Big]$$

$$= \sum_{x \in \mathcal{X}} x\, \hat{P}_X(x)$$

Defn: For a discrete random variable $X: \Omega \to \mathcal{X} \subseteq \mathbb{R}$ taking only finitely many values x having PMF $P_X: \mathcal{X} \to [0,1]$ is called a "simple random variable".

The mean or expectation of a simple r.v X is defined as

$$\mathbb{E}[X] = \mathbb{E}X = \sum_{x \in \mathcal{X}} x\, P_X(x)$$

Remark:

$A_x = X^{-1}\{x\} \in \mathcal{F}$ for each $x \in \mathcal{X}$

$A := (A_x : x \in \mathcal{X})$ is a finite partition of Ω, for simple r.v

$$1 \quad P(\Omega) = P\Big(\bigcup_x A_x\Big) = \sum_{x \in \mathcal{X}} P(A_x)$$

$$P_X(x) := P(A_x)$$

$$X(\omega) = \sum_{x \in \mathcal{X}} x\, 1_{A_x}(\omega)$$

$$\mathbb{E}[X] = \sum_x x \underbrace{P_X(x)}_{P(A_x)} = \sum_x x\, \mathbb{E}\, 1_{A_x}(\omega) = \sum_x x \int_\Omega 1_{A_x}(\omega) P(d\omega)$$

$$= \int_\Omega \sum x\, 1_{A_x}(\omega)\, P(d\omega)$$

$$= \int_\Omega X(\omega)\, P(d\omega)$$

Theorem: Given $X:\Omega\to\mathbb{R}_+$ a r.v on $(\Omega,\mathcal{F},P)$
There exists a seq. of non-decreasing, non-negative sample r.v

$Y:\Omega\to\mathbb{R}_+^{\mathbb{N}}$ s.t $\forall\omega\in\Omega$

$0\le Y_n(\omega)\le Y_{n+1}(\omega)\quad\forall n\in\mathbb{N}$

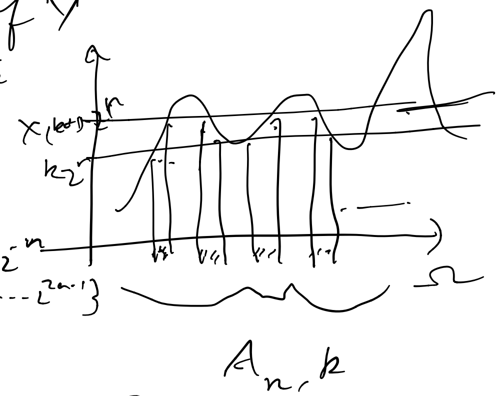
$\lim_n Y_n(\omega)=X(\omega)$

$(\mathbb{E}[Y_n]\in\mathbb{R}_+ : n\in\mathbb{N})$ is non-decreasing
$\lim_n \mathbb{E}Y_n\in\mathbb{R}_+\cup\{\infty\}$ exists, indep. of Y

Proof: $A_{n,k}=X^{-1}(k2^{-n},(k+1)2^{-n})\in\mathcal{F}$

$$Y_n(\omega)=\sum_{k=0}^{2^{2n}-1}k2^{-n}\,\mathbb{1}_{A_{n,k}}(\omega)$$

$$=\begin{cases}k2^{-n} & k2^{-n}<X(\omega)\le(k+1)2^{-n}\\ & k\in\{0,1,\dots,2^{2n-1}\}\\ 0 & X(\omega)>2^{n}\end{cases}$$



$$\bigcup_{k=0}^{2^{2n}-1}A_{n,k}=X^{-1}\left(\bigcup_{k=0}^{2^{2n}-1}(k2^{-n},(k+1)2^{-n}]\right)$$

$$=X^{-1}(0,2^{n}]$$

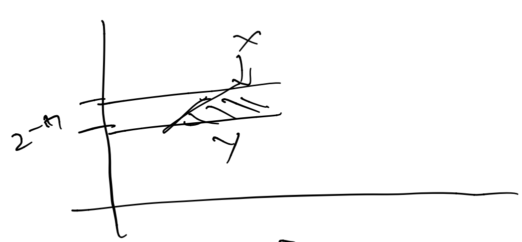
$$\lim_{n\to\infty}\bigcup_{k=0}^{2^{2n}-1}A_{n,k}=\Omega\qquad[\text{If } A_{0,n}=[0,2^{-n}]]$$

$(k+1)2^{-n}=2(k+1)2^{-(n+1)}$ $\quad Y_{n+1}>Y_n$

$2(k+1)2^{-(n+1)}$ $\quad Y_{n+1}=2(k+1)2^{-(n+1)}$

$Y_n=k2^{-n}$ $\quad k2^{-n}=2k\cdot2^{-(n+1)}$ $\quad Y_{n+1}=Y_n$

$$X(\omega)=\lim_n Y_n(\omega)$$



Defn: (Expectation): $E(X):=\lim_n\mathbb{E}[Y_n]$

Defn: [Expectation] $X = X_+ - X_-$

$$X_+ = \max\{X, 0\} = X \vee 0$$

$$X_- = \max\{-X, 0\} = -X \vee 0$$

If $\mathbb{E}[X_+]$ & $\mathbb{E}[X_-]$ are finite

$$\mathbb{E}[X] := \mathbb{E}[X_+] - \mathbb{E}[X_-]$$

Thm: $X : \Omega \to \mathbb{R}$ defined of $(\Omega, \mathcal{F}, P)$

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \, dF_X(x) \underset{\text{If } x \text{ ct.}}{=} \int_{x \in \mathbb{R}} x f_X(x) \, dx$$

Proof: We will prove it for discrete rv and take lim & prove it for +ve rv & from positive we can go to real r.v.

Non-negative sequence Y_n.

$$Y_n = \sum_{k=0}^{2^{2n}-1} k \cdot 2^{-n} \mathbb{1}_{A_{n,k}}(\omega) \quad \left[A_{n,k} = X^{-1}(k2^{-n}, (k+1)2^{-n})\right]$$

$$\mathbb{E}[Y_n] = \sum_{k=0}^{2^{2n}-1} k 2^{-n} P(A_{n,k})$$

$$\left[F_X((k+1)2^{-n}) - F_X(k2^{-n})\right]$$

$$\mathbb{E}[Y_n] = \sum_{k=0}^{2^{2n}-1} k 2^{-n} \left[F_X((k+1)2^{-n}) - F_X(k2^{-n})\right]$$

$$\mathbb{E}[X] = \lim_n \mathbb{E}[Y_n] = \int_{x \in \mathbb{R}} x \, dF_X(x) \quad \blacksquare$$

- simple rv can only take finitely many values.
- discrete rv can take countable infinite values.

2. Properties of Expectation

1. Linearity: $X : \Omega \to \mathbb{R}^n$, $\mathbb{E}\left[\sum_{i=1}^n \alpha_i x_i\right] = \sum_{i=1}^n \alpha_i \mathbb{E} X_i$

2. Monotonicity: $P\{X \geq Y\} = 1$ then $\mathbb{E}X \geq \mathbb{E}Y$

3. Functions of r.v: $g\ \mathbb{R} \to \mathbb{R}$ then $\mathbb{E}g(x) = \int_{\mathbb{R}} g(x)\, dF_X(x)$

4. Continuous r.v: $\mathbb{E}X = \int_{\mathbb{R}} x f(x)\, dx$

5. discrete r.v: $\mathbb{E}X = \sum_{x \in X} x\, P_X(x)$

6. Integration by parts

$$\mathbb{E}X = \int_{x \geq 0} (1 - F_X(x))\, dx - \int_{x < 0} F_X(x)\, dx$$

Proof: i $X: \Omega \to \mathbb{R}^n$, $g: \mathbb{R}^n \to \mathbb{R}$ is Borel measurable.

$$x \mapsto \sum_{i=1}^{n} \alpha_i x_i \quad \text{for some } \alpha \in \mathbb{R}^n.$$

Proof of 1 from Parimal's Notes.

(ii) (Monotonicity) $X(\omega) \geq Y(\omega)$ $\forall$ outcomes ω

or $\underset{\text{event}}{E} = \{ \underbrace{X(\omega) - Y(\omega)}_{\text{random variable}} \geq 0 \}$ then $P(E) = 1$

If something happens with a probability 1, it is called an almost sure event. (a.s)

$E[X - Y] = E[X] - E[Y]$

It suffices to show that if $Z \geq 0$ a.s; then $\mathbb{E}Z \geq 0$

If Z is simple $Z := \sum_{z \in Z} z\, 1_{\{Z = z\}}$

$$\mathbb{E}Z = \sum_z z\, P_Z(z) \geq 0$$

$\therefore$ expectation is +ve.

So proved ☺.

3. $g: \mathbb{R} \to \mathbb{R}$ is Borel measurable. $g(x)$ is a r.v

$$\mathbb{E}\, g(x) = \int_x g(x)$$

Look at proofs from the notes

Lecture 8: Moments

Ex:- [Absolute value function] $|\cdot| : \mathbb{R} \to \mathbb{R}_+$
$x \to |x|$

$$|\cdot|^{-1}(-\infty, x] = \{ y \in \mathbb{R} : |y| \le x \}$$
$$= \begin{cases} \phi & \text{if } x \le 0 \\ [-x, x] & \text{if } x \ge 0 \end{cases}$$

This implies that absolute value is a Borel measurable function.
Inverse of half open sets is a Borel set.

Lemma: If $\mathbb{E}|X|$ is finite, then $\mathbb{E}X$ exists & is finite.
Proof: $|X|$ is a random variable, $|X| \ge 0$
$\mathbb{E}|X| \ge 0$ exists
$$\mathbb{E}|X| = \underbrace{\mathbb{E}X_+ + \mathbb{E}X_-}_{\text{both finite}}$$
$\mathbb{E}X = \mathbb{E}X_+ - \mathbb{E}X_-$ exists & is finite.

Cor: If $g: \mathbb{R} \to \mathbb{R}$ Borel measurable, $\mathbb{E}|g(X)|$ is finite
$X: \Omega \to \mathbb{R}$ is a r.v then $\mathbb{E}g(X)$ is exists & is finite.

Ex: [Polynomial fn.] for any $k \in \mathbb{N}$ $g_k: \mathbb{R} \to \mathbb{R}$; $x \mapsto x^k$
show that g_k is Borel measurable

This will imply $g_k(\omega) = x^k$ is a r.v.

Def. Let $X: \Omega \to \mathbb{R}$ r.v defined on $(\Omega, \mathcal{F}, P)$
The k^{th} moment of X is defined as
$$m_k := \mathbb{E}g_k(X) = \mathbb{E}X^k$$

Remark: m_1 is called mean of the event
If $\mathbb{E}|X|^k$ is finite, then m_k exists (?)
Ex:- If $|X| \le 1$, then $|X|^k \le 1$, a.s.
$\mathbb{E}|X|^k \le 1$ from monotonicity of $\mathbb{E}$.

This implies m_k exists and is finite $\forall\, k \in \mathbb{N}$.

Lemma: If m_N is finite for some $N \in \mathbb{N}$, then m_k is finite $\forall\, k \in [N]$

Proof: $|x^k| = |x^k| \mathbb{1}_{\{|x^k| \le 1\}} + |x^k| \mathbb{1}_{\{|x^k| \ge 1\}}$

$$\le \mathbb{1}_{\{|x^k| \le 1\}} + |x|^N \mathbb{1}_{\{|x^k| > 1\}}$$

$$\le 1 + |x|^N \qquad (k \le N)$$

$$\mathbb{E}|x^k| \le 1 + \mathbb{E}|x|^N$$

$$m_k \le 1 + m_N$$

QED

Cor: If m_p is finite for some $p \ge 1$, then $m_{p'}$ is finite $\forall\, p'$, $1 \le p' \le p$.

2. L^p spaces:

For a prob. space $(\Omega, \mathcal{F}, P)$ and $p \ge 1$, define

$$L^p \triangleq \{ X : \mathbb{E}|X|^p < \infty \}$$

set of r.v — r.v

Remark $1 \le p' \le p$ then $L^p \subseteq L^{p'}$ [If $x \in L^p$, then $m_p < \infty$ for cor. $m_{p'} < \infty$ i.e. $X \in L^{p'}$]

Defn :- $\|\cdot\|_p : L^p \to \mathbb{R}_+$ $\quad x \mapsto \|x\|_p \equiv (\mathbb{E}|x^p|)^{1/p}$

[We will show that this is a norm]

3. Moment generating functions

[All of these defined have certain properties]

Prob. space $(\Omega, \mathcal{F}, P)$

Random variables $X : \Omega \to \mathbb{R}$ $\quad F_X : \mathbb{R} \to [0,1]$

Ex: $g_\theta : \mathbb{R} \to \mathbb{R}_+$ is Borel measurable $x \mapsto e^{\theta x}$

$g_\theta(X) = e^{\theta X}$ is a pos. r.v.

$h_\theta : \mathbb{R} \to \mathbb{C}$

$x \mapsto e^{j\theta x} = \cos\theta x + j \sin\theta x \quad j = \sqrt{-1}$

$h_\theta(x) = e^{j\theta x}$ is a complex valued random variable

{∃ diff. mgf}

Def: [MGF] $M_X : \mathbb{R} \to \mathbb{R}_+$

$$\theta \longmapsto M_X(\theta) = \mathbb{E}[e^{\theta X}] \quad \forall\, \theta \in \mathbb{R}$$

for which this is finite

$$= \int_{x \in \mathbb{R}} e^{\theta x} d_x F(x)$$

Defn. [Characteristic function] $\phi_X : \mathbb{R} \to \mathbb{C}$

$$\theta \to \phi_X(\theta) = \mathbb{E}\big[\underbrace{e^{j\theta x}}_{h_\theta(x)}\big] \quad \forall\, \theta \in \mathbb{R}$$

Defn. [Z-trans. of PMP] $X : \Omega \to \mathcal{X}$ discrete

$$\Psi_X : \mathbb{C} \to \mathbb{C}$$

$$z \longmapsto \mathbb{E}[z^X] = \sum_{x \in \mathcal{X}} z^x P_X(x)$$

Thm: Two rv have the same dist. fun. iff they have the same characteristic function (also holds true for other MGFs)

Proofs: Necessity is easy, sufficiency is difficult,

X, Y with F_X $\quad \mathbb{E}\, e^{j\theta X} = \mathbb{E}\, e^{j\theta Y}$

If $M_X(\theta) = M_Y(\theta)$, can we prove $F_X = F_Y$?

Lemma: If $\mathbb{E} X^k$ exists & is finite for $k \in \mathbb{N}$, then derivatives of ϕ_X upto order k exist and are continuous.

$$\phi_X^k(0) = j^k\, \mathbb{E}[X^k]$$

$$\Big[\ \phi_X(\theta) = \mathbb{E}[e^{j\theta x}] \qquad = 1 + j\theta x + \frac{j^2\theta^2}{\underline{2}} x^2 + \dots$$

$$\phi_X^{(k)} \underset{?}{=} \mathbb{E}\left[\frac{d^k}{d\theta^k} e^{j\theta x}\right] = \mathbb{E}\left[j^k x^k + \frac{j^{k+1}\theta x^{k+1}}{\underline{k+1}} \dots\right]$$

under some regulatory condition

$$\phi_X^k(0) = j^k\, \mathbb{E}[X^k]$$

$$\psi_X^{(k)}(1) = \frac{d^k}{dz^k}\mathbb{E}[z^X]\Big|_{z=1}$$

under regu. condition
$$= \mathbb{E}\left[\frac{d^k z^X}{dz^k}\right]$$

$$= \mathbb{E}\left[X(X-1)(X-2)\cdots(X-k+1)z^{(X-k)}\Big|_{z=1}\right]$$

$$= \mathbb{E}\left[X(X-1)\cdots\cdots(X-k+1)\right]$$

4 - Central Moments

Ex:- $h_k : \mathbb{R} \to \mathbb{R}$ is Borel Measurable
$x \mapsto (x-m_1)^k$

Defn: k^{th} central moment of r.v X
$$\sigma_k := \mathbb{E}\, h_k(X) = \mathbb{E}(X-m_1)^k$$

$\sigma_2 = \mathbb{E}(X-m_1)^2$ is called the variance & denoted by σ^2.

Lemma: $\sigma_1 = \mathbb{E}(X-m_1) = 0$

$$\sigma^2 = \sigma_2 = \mathbb{E}(X-m_1)^2 \geq 0$$
$$= \mathbb{E}X^2 + m_1^2 - 2(\mathbb{E}X)m_1$$
$$= m_2 - m_1^2 \geq 0$$

Rem: $m_2 \geq m_1^2$
$L^2 \subseteq L^1$

5. Inequalities

$X : \Omega \to \mathbb{R}$ r.v. on $(\Omega, \mathcal{F}, P)$

[Markov's inequality] $f : \mathbb{R} \to \mathbb{R}_+$ monotonically non-decr.

$$P\{X \geq \epsilon\} \leq \frac{\mathbb{E} f(X)}{f(\epsilon)}$$

Proof: f is Borel measurable & hence $f(X)$ is a r.v

$$f(x) = f(x)\left(\underbrace{1_{\{x \geq \epsilon\}} + 1_{\{x < \epsilon\}}}\right)$$
$$\geq f(x)\, 1_{\{x \geq \epsilon\}}$$
$$\geq f(\epsilon)\, 1_{\{x \geq \epsilon\}}$$

$$\mathbb{E}(f(X)) \geq f(\epsilon)\, P\{X \geq \epsilon\}$$ ∎

Corollary: [Markov] $X : \Omega \to \mathbb{R}_+$ $f : \mathbb{R}_+ \to \mathbb{R}_+$, $x \mapsto x$

$$P(X \geq \epsilon) \leq \frac{\mathbb{E}X}{\epsilon} \quad \forall \epsilon > 0$$

Corollary [Chebyshev] $X : \Omega \to \mathbb{R}$ with $\overbrace{m_1 = \mu \quad m_2 = \sigma^2}^{\text{finite}}$

Then $P\{ |\underbrace{X - \mu}_{Y}| \geq \epsilon \} \leq \dfrac{\mathbb{E} f(Y)}{f(\epsilon)} = \dfrac{\mathbb{E}(X-\mu)^2}{\epsilon^2} = \dfrac{\sigma^2}{\epsilon^2}$

$f : y \mapsto y^2$

Corollary [Chernoff] $X : \Omega \to \mathbb{R}$ with $M_X(\theta) = \mathbb{E}\, e^{\theta X}$ finite for some $\theta > 0$

Then $P\{X \geq \epsilon\} \leq \dfrac{\mathbb{E} f(X)}{f(\epsilon)} = \dfrac{\mathbb{E}\, e^{\theta X}}{e^{\theta \epsilon}}$

$f : x \mapsto e^{\theta x}$

$$= e^{-\theta \epsilon} M_X(\theta)$$

$$P\{X \geq \epsilon\} \leq \inf_{\theta} e^{-\theta \epsilon} M_X(\theta)$$

6. Correlation & Covariance

Ex:- $g : \mathbb{R}^2 \to \mathbb{R}$ is Borel measurable
$(x, y) \mapsto xy$

Defn: X, Y r.v. then XY is a r.v. Correlation b/w 2 random variables X & Y is defined by $\mathbb{E}XY$.

If $\mathbb{E}XY = \mathbb{E}X\, \mathbb{E}Y$, then X & Y are uncorrelated.

Lemma: If X & Y are Ind then they are uncorrelated.

Proof
$$X = \sum x\, 1_{A_x} \qquad A_x := X^{-1}\{x\}$$
$$Y = \sum y\, 1_{A_y} \qquad A_y := Y^{-1}\{y\}$$
$$XY = \sum_{(x,y) \in X \times Y} xy\, 1_{A_x \cap B_y}$$
$$\mathbb{E}XY = \sum_{(x,y)} xy\, P(A_x \cap B_y)$$

Ex [Uncorrelated but Dependent]

$X : \Omega \to \mathbb{R}$ zero mean
$g : \mathbb{R} \to \mathbb{R}$ even, incr. in $y \in \mathbb{R}_+$

This implies that g is Borel Measurable

$Y = g(x)$ is a r.v.

Show that, X, Y are uncorrelated by dep r.v.

$x, y \in \mathbb{R}_+$ $\qquad x > g^{-1}(y)$ and $F_X(x) \leq 1$

$B_y = \{ Y \leq y \} = \{ g(x) \leq y \} = \{ -g^{-1}(y) \leq x \leq g^{-1}(y) \}$

$= \{ X \in [-x_y, x_y] \}$

$F_{X,Y}(x, y) = P\{ X \leq x, Y \leq y \}$

$= P(A_x \cap B_y) = P(B_y)$

$\neq P(A_x) P(B_y)$

Lecture - 9 Correlation

Theorem (AM $\geq$ GM) $\mathbb{E}[XY] \leq \frac{1}{2}(\mathbb{E}X^2 + \mathbb{E}Y^2)$ with equality if $X = Y$

[Recall : $\sqrt{xy} \leq \frac{x+y}{2} \Rightarrow \sqrt{x^2y^2} \leq \frac{x^2+y^2}{2}$]

Proof: $(X-Y)^2 \geq 0$

This implies $\mathbb{E}(X-Y)^2 \geq 0$.

$$\mathbb{E}(X^2 + Y^2 - 2XY) = \mathbb{E}(X^2 + \mathbb{E}Y^2) - 2\mathbb{E}(XY) \geq 0$$

Theorem: [Cauchy-Schwarz inq.] $\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2 \mathbb{E}Y^2}$ with equality if $x = \alpha y$, for some $\alpha \in \mathbb{R}$

Proof: $W := \frac{|X|}{\sqrt{\mathbb{E}X^2}}$ $\quad V := \frac{|Y|}{\sqrt{\mathbb{E}Y^2}}$

$\mathbb{E}W^2 = 1 \quad \mathbb{E}V^2 = 1$

$\mathbb{E}WV \leq \frac{1}{2}(\mathbb{E}(W^2) + \mathbb{E}V^2)$

$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2 \mathbb{E}Y^2}$

2. Covariance

$$\mathrm{Cov}(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y]$$

Lemma: If X & Y are correlated, then $\mathrm{Cov}(X,Y) = 0$

Proof: $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

$\mathrm{Cov}(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - EY]$

$= \mathbb{E}(XY - XEY - YEX + \mathbb{E}X\mathbb{E}Y)$

$= \mathbb{E}(XY) - 2EYEX + EX\,\mathbb{E}Y = 0$ ∎

Lemma: If $X: \Omega \to \mathbb{R}^n$ a random vector (uncorrelated) $\mathbb{E}X_iX_j = \mathbb{E}X_i\mathbb{E}X_j$

$a \in \mathbb{R}^n$

then $\mathrm{Var}(\sum_{i=1}^n a_i x_i) = \sum_{i=1}^n a_i^2 \mathrm{Var}(x_i)$

Proof: $\mathbb{E}[(\sum_{i=1}^n a_i x_i - \sum a_i Ex_i)^2]$

$$\mathbb{E}\left[\sum_{i=1}^{n} a_i(x_i - \mathbb{E}x_i)\right]^2 = \mathbb{E}\left(\sum_{i,j} a_i a_j (x_i - \mathbb{E}x_i)(x_j - \mathbb{E}x_j)\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} a_i^2 (x_i - \mathbb{E}x_i)^2 + \sum_{i \neq j} a_i a_j (x_i - \mathbb{E}x_i)(x_j - \mathbb{E}x_j)\right]$$

$$= \sum_{i=1}^{n} a_i \operatorname{Var}(x_i) + \sum_{i \neq j} a_i a_j \underbrace{\operatorname{Covar}(x_i, x_j)}_{=0} \quad \blacksquare$$

Defn: (Correlation coefficient) $\rho_{XY} := \dfrac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$

Thm: $\rho_{X,Y} \leq 1$ with equality if $X = \alpha Y + \beta$

$\alpha = \sqrt{\dfrac{\operatorname{Var}(X)}{\operatorname{Var}(Y)}}$, $\beta = \mathbb{E}X - \alpha\mathbb{E}Y$

Proof: $W := \dfrac{X - \mathbb{E}X}{\sqrt{\operatorname{Var}(X)}}$ $\qquad V := \dfrac{Y - \mathbb{E}Y}{\sqrt{\operatorname{Var}(Y)}}$

$\mathbb{E}W^2 = 1$ $\qquad \mathbb{E}V^2 = 1$

$\mathbb{E}W = 0$ $\qquad \mathbb{E}V = 0$

From AM $\geq$ GM, $\underbrace{\mathbb{E}WV}_{= \rho_{XY}} \leq 1$

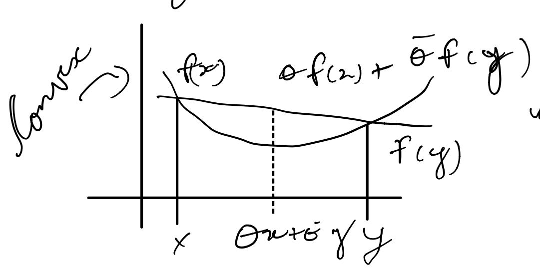
This implies that $\mathbb{E}\dfrac{(X - \mathbb{E}X)(Y - \mathbb{E}Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} \leq 1$ $\blacksquare$

Generalisations:

$f: \mathbb{R} \to \mathbb{R}$ is convex if

$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$

$\forall x, y \in \mathbb{R}$, $\theta \in [0,1]$.



Theorem: [Jensen] If $f: \mathbb{R} \to \mathbb{R}$ convex and r.v X defined on $(\Omega, \mathcal{F}, P)$ then,

$$f(\mathbb{E}X) \leq \mathbb{E}(f(x))$$

Proof: f convex then Borel measurable.

X r.v then $f(x)$ is a r.v.

If $X: \Omega \to \mathcal{X} \subset \mathbb{R}$ be a simple r.v.

We will show this by induction on $|\mathcal{X}|$.

For $|\mathcal{X}| = 1$, this is trivially true. (Check Proof from notes)

Theorem 3.3: Holder's inequality: Consider two random variables X, Y, such that $\mathbb{E}|X|^p$ & $\mathbb{E}|Y|^p$ are finite for $p, q \geq 1$, such that $\frac{1}{p} + \frac{1}{q} = 1$. Then,

$$\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{\frac{1}{p}} \times (\mathbb{E}|Y|^q)^{\frac{1}{q}}$$

Proof: Recall that $f(z) = e^z$ is a convex function. Therefore for random variable $Z \in \{p \ln V, q \ln W\}$ with PMF $(\frac{1}{p}, \frac{1}{q})$, it follows from Jensen's inequality that

$$VW = f(\mathbb{E}Z) \leq \mathbb{E}(f(Z)) = \frac{V^p}{p} + \frac{W^q}{q}$$

Taking expectation on both sides, we get from monotonicity of $\mathbb{E}$ that

$$\mathbb{E}VW = \frac{\mathbb{E}V^p}{p} + \frac{\mathbb{E}W^q}{q}$$ Taking $V = \frac{|X|}{(\mathbb{E}X)^{\frac{1}{p}}}$, $W = \frac{|Y|}{(\mathbb{E}Y)^{\frac{1}{q}}}$

$$\mathbb{E}|XY| = (\mathbb{E}X^p)^{\frac{1}{p}} (\mathbb{E}Y^q)^{\frac{1}{q}}$$

L^p spaces

Definition 4.1 We define a function $\|\ \|_p : L^p \to \mathbb{R}_+$ defined by $\|\ \|_p(X) = \|X\|_p \triangleq (\mathbb{E}X^p)^{\frac{1}{p}}$ for any $X \in L^p$ and real $p \geq 1$.

Definition 4.2 Given a Vector space of random varia- a norm on the vector space is a map $f: V \to \mathbb{R}_+$ such that

homogeneity: $f(ax) = |a| f(x)\ \forall a \in \mathbb{R}$ and $x \in V$

sub-additivity: $f(x+y) \leq f(x) + f(y)$

point-separating: $f(x) \geq 0\ \forall\ x \in V$.

See rest from notes.

L^2 is Hilbert Space.