§ Chapter 1. Crashcourse on Probability Theory

Def 1.1.1. (σ-field) Let Ω be a set. A σ-algebra on Ω is a non-empty set $\mathcal{F}$ of subsets of Ω such that

① if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$

② if $A_1, A_2, \cdots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Examples.

(1) Coin-toss space (2 tosses) $\Omega = \{HT, HH, TT, TH\}$, Ω is the set of all outcomes for a random expiriment.

$\omega \in \Omega \triangleq$ one specific outcome, e.g. $HH \in \Omega$

define $A = \{$the first toss is $H\} = \{HT, HH\}$

define $\mathcal{F} = \{\{HT, HH\}, \{TT, TH\}, \Omega, \phi\}$

Remark: σ- algebra contains all events I want to assign probabilities to.

(2) $F = \{\Omega, \phi\}$

(3) Bowl σ-algebra is the smallest σ-algebra which contains all open subsets of $\mathbb{R}$ (eg. (a,b) is an open subset of $\mathbb{R}$, which equals $\{x \in \mathbb{R}; a < x < b\}$

Def 1.12 (probability measures) Let F be a σ-algebra on Ω. A measure μ on (Ω, F) is a function $\mu: F \to [0,\infty)$, such that

a measure

<1> $\mu(\phi)=0$

<2> if $A_1, A_2, \cdots \in F$ are disjoint (i.e. $A_i \cap A_j = \phi$), then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

A probability measure $\mathbb{P}$ is such a function (measure) on (Ω, F) and satisfies $\mathbb{P}(\Omega)=1$ (implies that $\mathbb{P}: F \to [0,1]$)

Examples:

① $\mu(A)=0 \quad \forall A \quad \Rightarrow$ a measure, but not a prob. measure

② The lebesgue measure on $\mathbb{R}$ $leb(a,b] = b-a \quad a<b$ is a measure, but not a prob. measure.

Notations:

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

$\Omega \triangleq$ sample space (space of all outcomes)

$\omega \in \Omega \triangleq$ outcome

$A \subseteq \Omega \triangleq$ event

Let $A \in F$ if $\mathbb{P}(A)=1$, then A is called an almost true event, and if $\mathbb{P}(A)=0$, the A is called a null event.

Def 1.2.1 (random variable). A random variable (r.v.) $X: \Omega \to \mathbb{R}^{(-\infty,+\infty)}$ is a function s.t. $\{\omega \in \Omega : X(\omega) \le t\} \in F$ for all $t \in \mathbb{R}$. (*)

We say that X is F-measurable if (*) holds.

The cumulative distribution function (cdf) $F_X: \mathbb{R} \to [0,1]$ is defined as

$F_X(t) = \mathbb{P}(X \le t) = P(\{w \in \Omega : X(w) \le t\})$ → obviously F_X is a prob. measure

Remarks:

<1> we often write $\mathbb{P}(X \in (s,t)) = \mathbb{P}(\{X \in \Omega : X(w) \in (s,t)\})$

<2> if f is a reasonable function and X is a r.v., then

used to test r.v. $f(X)$ is a r.v.

(eg. X, Y are r.v. $\Rightarrow X+Y,\ X/Y,\ X^2,\ \ln(x), \cdots$ are r.v.)

Def 1.2.2 $\sigma(X) \triangleq$ the smallest σ-algebra which contains $\{X \le t\}\ \forall t \in \mathbb{R}$

$\triangleq$ the σ-algebra generated by X

→ X is $\sigma(X)$-measurable

Def/prop. 1.6.1 Let $X_1, \cdots, X_N$ be r.v. following statements are equivalent

① $X_1, \cdots, X_N$ are independent

② $\mathbb{P}(X_1 \le t_1, \cdots X_N \le t_N) = \mathbb{P}(X_1 \le t_1) \cdot \mathbb{P}(X_2 \le t_2) \cdots \mathbb{P}(X_N \le t_N)$

for all $t_1, \cdots t_N \in \mathbb{R}$

③ if $X_1, \cdots X_N$ have joint pdf/pmf $f_{X_1, \cdots X_N}(x_1, \cdots x_N)$

then $f_{X_1, X_2, \cdots X_N}(x_1, \cdots x_N) = f_{X_1}(x_1) \cdots f_{X_N}(x_N)$

④ For any bounded Continuous function $g_1, \cdots g_N : \mathbb{R} \to \mathbb{R}$

$\mathbb{E}[g_1(X_1) \cdots g_N(X_N)] = \mathbb{E}[g_1(X_1)] \cdots \mathbb{E}[g_N(X_N)]$

e.g. if X, Y are independent $E[XY] = E[X]E[Y]$

Additional Contents from the Lecture

1.3. Expectations and Variances.

Let X be a random variable on $(\Omega, \mathcal{F}, P)$

$$E[X] = \sum_{i=1}^{n} x_i \underbrace{P(X = x_i)}_{\text{pmf}} = \int_{-\infty}^{\infty} x \underbrace{\mu(x)}_{\text{pdf}}\, dx .$$

A random variable X is integrable iff $E[|X|] < \infty$, and is square-integrable iff $E[X^2] < \infty$.

The variance of an integrable random variable X, is

$$Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

The covariance of square-integrable random variable X and Y is.

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X] \cdot E[Y]$$

if neither X nor Y is almost surely constant, their correlation is then

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}}$$

Theorem 1.3.4. Let the function $g: \mathbb{R} \to \mathbb{R}$ be such that $g(x)$ is integrable.

① If X is a discrete r.v. with pmf p_X, then

$$E[g(X)] = \sum_{t \in S} g(t)\, p_X(t)$$

② If X is an absolutely continuous r.v. with pdf f_X, then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$$

More generally, if X is a r.v. in $\mathbb{R}^n$ with density function f_X and $g: \mathbb{R}^n \mapsto \mathbb{R}$, then

$$E[g(X)] = \int_{\mathbb{R}^n} g(x) f_X(x)\,dx$$

Theorem 1.3.2 (Jensen's inequality). Let X be a random variable and $g: \mathbb{R} \mapsto \mathbb{R}$ be convex function. Then

$$E[g(X)] \geq g(E[X])$$

Theorem 1.3.3 (Hölder's inequality). Let X, Y be r.v.s and let $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$. If $X \in L^p$ and $Y \in L^q$, then

$$E[XY] \leq E[|X|^p]^{\frac{1}{p}} \cdot E[|Y|^q]^{\frac{1}{q}}$$

(For $p \geq 1$, the space L^p is the collection of r.v.s such that $E[|X|^p] < \infty$)

The case when $p=2$ is called the Cauchy-Schwarz inequality

1.7. Probability Inequalities

Theorem 1.7.1 (Markov's inequality). Let X be a positive r.v. then

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

for all $\varepsilon > 0$.

Theorem 1.7.2 (Tschebycheff's inequality). Let X be a r.v. with $E[X]=\mu$ and $Var[X]=\sigma^2$. Then

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

for all $\varepsilon > 0$

1.8. Fundamental probability results

Theorem 1.8.7 (A strong law of large numbers). Let $X_1, X_2, \cdots$ be independent and i.i.d. integrable r.v.s. with common mean $E[X_i]=\mu$. Then.

$$\frac{X_1+X_2+\cdots+X_n}{n} \to \mu \quad \text{almost surely}$$

Theorem 1.8.8. (Central limit theorem). Let $X_1, X_2, \cdots$ be i.i.d. with $E[X_i]=\mu$ and $Var[X_i]=\sigma^2$ for each $i=1,2,\cdots n$, let

$$Z_n = \frac{X_1+X_2+\cdots+X_n - n\mu}{\sigma \cdot \sqrt{n}}$$

Then $Z_n \to Z$ in distribution, where $Z \sim N(0,1)$