Lecture 26

(18 November 2024)

Course Review

Basics of Probability (Module 1)

Different Approaches to Probability:

Classical Approach

$$P(E) = \frac{\text{no. of outcomes favourable to event } E}{\text{total no. of possible outcomes}}$$

Relative Frequency Approach

$$P(E) = \lim_{n \to \infty} \frac{n_E}{n}, \quad \text{where}$$

n_E = no. of times E occurs

n = total no. of trials

Axiomatic Approach

Ω - Sample space

$\mathcal{F}$ - Event space

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

(iii) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Probability law: $P : \mathcal{F} \to [0,1]$

(i) (Non-negativity). $P(E) \geq 0$

(ii) (Normalization). $P(\Omega) = 1$

(iii) (Additivity). $A_1, A_2, \ldots$ disjoint

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

Infinite union and Intersection:

Given a sequence $A_1, A_2, \ldots$, i.e., $(A_n)_{n \in \mathbb{N}}$,

$$\bigcup_{n=1}^{\infty} A_n = \{x \in \Omega : x \in A_n \text{ for some } n \in \mathbb{N}\},$$

$$\bigcap_{n=1}^{\infty} A_n = \{x \in \Omega : x \in A_n \ \forall n \in \mathbb{N}\}.$$

Properties of Probability Law

(1) If $A \subseteq B$ then $P(A) \leq P(B)$.

(2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(3) Continuity of Probability

For a sequence of events $A_1, A_2, \ldots$,

$$\lim_{n \to \infty} P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right).$$

(i) $A_1 \subseteq A_2 \subseteq \ldots$

$$\lim_{n \to \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right).$$

(ii) $B_1 \supseteq B_2 \supseteq \ldots$

$$\lim_{n \to \infty} P(B_n) = P\left(\bigcap_{n=1}^{\infty} B_n\right).$$

Conditional Probability

$$P(A|B) = \frac{P(A\cap B)}{P(B)}, \quad P(B) > 0.$$

Independence

$$P(A\cap B) = P(A)P(B)$$

Events $A_1, A_2, \ldots, A_n$ are (mutually) independent if

$$P\left(\bigcap_{i\in \mathcal{I}} A_i\right) = \prod_{i\in \mathcal{I}} P(A_i), \quad \mathcal{I} \subseteq [1:n].$$

A, B are conditionally independent given C with $P(C) > 0$ if

$$P(A\cap B | C) = P(A|C)P(B|C).$$

Total Probability Theorem:

$A_1, A_2, \ldots, A_n$ be a partition of Ω such that $P(A_i) > 0, \forall i \in [1:n]$.

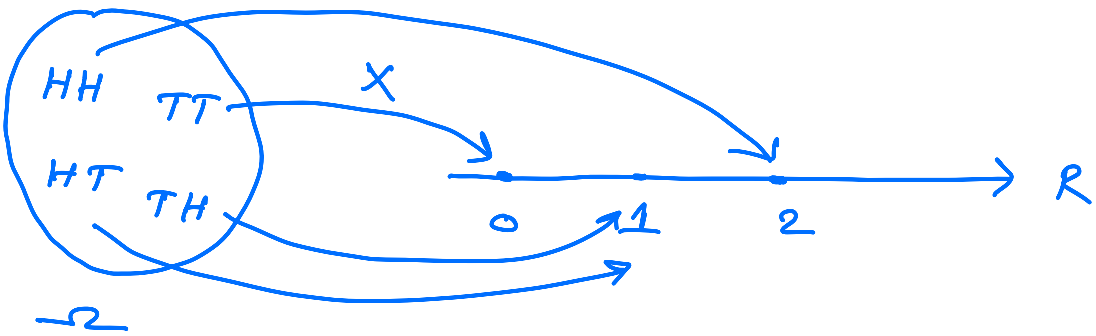
$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

Bayes' Theorem

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}.$$

Discrete Random Variables (Module 2)

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



Cumulative Distribution Function (CDF)

$$F_X(x) = P(X \leq x) \quad \forall x \in \mathbb{R}.$$

Defining Properties of a CDF

(a) If $x < y$ then $F_X(x) \leq F_X(y)$.

(b) $\lim_{x \to -\infty} F_X(x) = 0$, $\quad \lim_{x \to \infty} F_X(x) = 1$.

(c) F_X is right-continuous i.e., $\lim_{\varepsilon \to 0^+} F_X(x+\varepsilon) = F_X(x)$.

Types of Random Variables

Discrete RV: X takes countable values in $\mathbb{R}$.

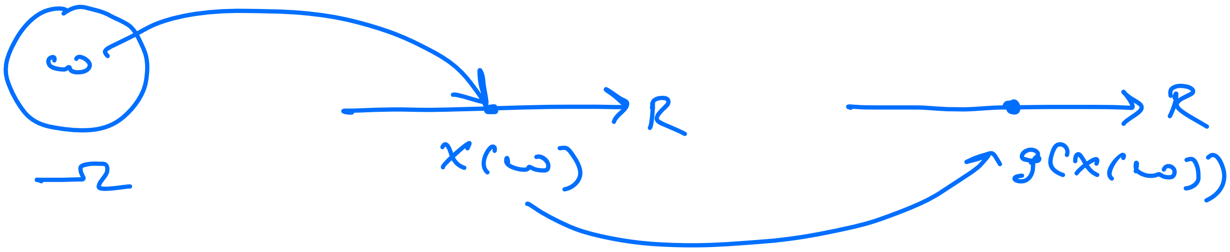
Continuous RV: $F_X(x) = \int_{-\infty}^{x} f_X(u)\, du \quad \forall x \in \mathbb{R}$.

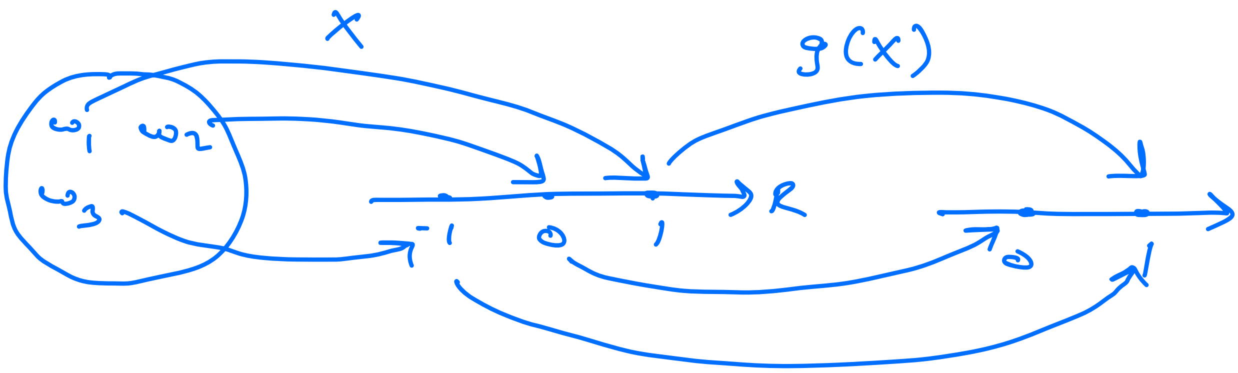
A discrete RV is associated with a pmf p_X.

$$p_X(x) = P(\{\omega : X(\omega) = x\}) = P(X = x), \quad \sum_x p_X(x) = 1.$$

Functions of RVs

$$X : \Omega \to \underline{R} \qquad g : R \to R$$





$$Y = g(X) \implies p_Y(y) = \sum_{x : g(x) = y} p_X(x).$$

Expectation

$$E[X] = \sum_x x \, p_X(x)$$

$$E[g(X)] = \sum_x g(x) \, p_X(x) \quad [\text{LOTUS}]$$

$$\text{Var}(X) = E[(X - E[X])^2].$$

Examples of Discrete RVs

Bernoulli RV: $P_X(1) = p = 1 - P_X(0)$

$E[X] = p$, $\text{var}(X) = p(1-p)$

Binomial RV: $P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$.

$E[X] = np$, $\text{var}(X) = np(1-p)$

Geometric RV: $P_X(k) = (1-p)^{k-1} p$, $k = 1, 2, \ldots$

$E[X] = 1/p$, $\text{var}(X) = \dfrac{1-p}{p^2}$.

Poisson RV: $P_X(k) = \dfrac{e^{-\lambda} \cdot \lambda^k}{k!}$, $k = 0, 1, 2, \ldots$

$E[X] = \text{var}(X) = \lambda$.

Poisson approximation for a Binomial:

$Y \sim \text{Binomial}(n, p)$. As $n \to \infty$, while $np = \lambda$ (const), we have

$$\lim_{n\to\infty} P_Y(k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}.$$

Jointly Discrete RVs

(X, Y) takes countable no. of values in $\mathbb{R}^2$.

$$p_{XY}(x,y) = P(X=x, Y=y)$$

$$p_X(x) = \sum_y p_{XY}(x,y), \quad p_Y(y) = \sum_x p_{XY}(x,y).$$

$$Z = g(X,Y) \Rightarrow p_Z(z) = \sum_{(x,y):\, g(x,y)=z} p_{XY}(x,y).$$

Independence $\quad p_{XY}(x,y) = p_X(x)\, p_Y(y)$

X and Y are independent $\Rightarrow g(X)$ and $h(Y)$ are independent.

n RVs $X_1, X_2, \ldots, X_n$ are independent if

$$p_{X_1 X_2 \ldots X_n}(x_1, \ldots, x_n) = \prod_i p_X(x_i), \quad \forall x_1, x_2, \ldots, x_n.$$

Some Properties

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i]$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$$

If X & Y are independent

(i) $Var(X+Y) = Var(X) + Var(Y)$,

(ii) $Z = X+Y$, $p_Z(z) = \sum_x p_X(x)\, p_Y(z-x)$

$$= p_X * p_Y$$

(convolution)

Conditioning

$p_{X|A}(x) = P(X=x|A)$, $p_{X|Y}(x|y) = P(X=x|Y=y)$,

$p_Y(y) > 0$.

$$E[X|A] = \sum_x x\, p_{X|A}(x)$$

$$E[X|Y=y] = \sum_x x\, p_{X|Y}(x|y)$$

$$E[g(X)|Y=y] = \sum_x g(x)\, p_{X|Y}(x|y)$$

Total Expectation Theorem: $A_1, A_2, \ldots, A_n$ - Partition

$$E[X] = \sum_{i=1}^{n} E[X|A_i]\, P(A_i) = \sum_y E[X|Y=y]\, p_Y(y).$$

$\phi(y) = E[X|Y=y] \Longrightarrow \phi(Y) \equiv E[X|Y]$ is a RV

$$E[E[X|Y]] = E[X].$$

Continuous RVs (Module 3)

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du, \quad P(X \in B) = \int_B f_X(u)\,du$$

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1, \quad f_X(x) = \frac{dF_X(x)}{dx}.$$

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

$$\mathrm{Var}(X) = E[X^2] - E[X]^2$$

If $Y \geq 0$, $E[Y] = \int_0^{\infty} P(Y > y)\,dy.$

Examples of Continuous RVs

Uniform RV: $f_X(x) = \begin{cases} 1/(b-a), & a \leq x \leq b, \\ 0 & \text{o.w.} \end{cases}$

$$E[X] = \frac{a+b}{2}, \quad \mathrm{Var}(X) = (b-a)^2/12.$$

Exponential RV: $f_X(x) = \lambda e^{-\lambda x}, \; x \geq 0.$

$$E[X] = 1/\lambda, \quad \mathrm{Var}(X) = 1/\lambda^2.$$

Gaussian RV: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$

$$E[X] = \mu, \quad \mathrm{Var}(X) = \sigma^2.$$

<u>Joint CDF</u>. $F_{XY}(x,y) = P(X \le x, Y \le y)$

(X, Y can be either discrete or continuous)

X, Y are jointly continuous if

$$F_{XY}(x,y) = \int_{v=-\infty}^{y} \int_{u=-\infty}^{x} f_{XY}(u,v)\, du\, dv.$$

$$f_{XY}(x,y) = \frac{d^2 F_{XY}(x,y)}{dx\, dy}.$$

$$P((X,Y) \in B) = \int_{(x,y)\in B} f_{XY}(x,y)\, dx\, dy$$

$$F_{X|A}(x) = P(X \le x | A) = \int_{-\infty}^{x} f_{X|A}(u)\, du$$

conditional CDF

$$f_{X|Y}(x|y) \triangleq \lim_{\Delta y \to 0} f_{X|\{y \le Y \le y+\Delta y\}}(x)$$

$$= f_{XY}(x,y) / f_Y(y).$$

$$E[X|Y=y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

$$E[X|A] = \int_{-\infty}^{\infty} x\, f_{X|A}(x)\, dx$$

Total Expectation Theorems

(1) $A_1, A_2, \ldots$ is a partition.

$$E[X] = \sum_{i=1}^{n} E[X \mid A_i]\, P(A_i).$$

(2) $$E[X] = \int_{-\infty}^{\infty} E[X \mid Y=y]\, f_Y(y) = E[E[X \mid Y]].$$

Independence $f_{XY}(x,y) = f_X(x) f_Y(y) \quad \forall x, y.$

More generally two RVs X and Y (either continuous or discrete) are independent if

$$F_{XY}(x,y) = F_X(x)\, F_Y(y) \quad \forall x, y.$$

Bayes' Rule

X is discrete Y is continuous

$$P_X(x)\, f_{Y|X}(y|x) = f_Y(y)\, P_{X|Y}(x|y)$$

$$P_{X|Y}(x|y) = \lim_{\Delta y \to 0} P(X=x \mid y \le Y \le y+\Delta y)$$

— Application in Binary MAP detection.

Functions of RVs

$Y = g(X), \quad X \sim f_X$

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(x_i)|},$$

$x_1, x_2, \ldots, x_n$ are solutions of $g(x) = y$.

$Z = g_1(X, Y) \quad W = g_2(X, Y), \quad (X, Y) \sim f_{XY}$

$$f_{ZW}(z, w) = \sum_{i=1}^{n} \frac{f_{XY}(x_i, y_i)}{|J(x_i, y_i)|},$$

where $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ are solutions of

$g_1(x, y) = z \quad g_2(x, y) = w.$

Moment Generating Function

$$M_X(s) = E[e^{sX}].$$

Characteristic Function

$$\phi_X(t) = E[e^{itX}].$$

Tail Bounds and Limit Theorems (Module 3)

Markov's Inequality

If $X \geq 0$, then $P(X \geq a) \leq \dfrac{E[X]}{a}$, for all $a > 0$

Chebyshev's Inequality

$$P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}, \quad \text{for all } c > 0.$$

Chernoff Bounds

$$P(X \geq a) \leq \inf_{s>0} \frac{E[e^{sX}]}{e^{as}}.$$

Convergence

$X_n \xrightarrow{P} X$ if $\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$, for every $\varepsilon > 0$.

WLLN: $X_i \sim$ i.i.d. mean μ $\Rightarrow \sum_{i=1}^{n} X_i / n \xrightarrow{P} \mu$.

$X_n \xrightarrow{D} X$ if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for all points x at which F_X is continuous.

CLT: $\left(\sum_{i=1}^{n} X_i - n\mu\right) / \sqrt{n}\sigma \xrightarrow{D} N(0, 1)$.

$X_n \xrightarrow{a.s.} X$ if $P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$.

SLLN: $X_i \sim$ i.i.d. with mean $\mu \Rightarrow \sum_{i=1}^{n} X_i / n \xrightarrow{a.s.} \mu$.

$X_n \xrightarrow{m.s.} X$ if $\lim_{n\to\infty} E((X_n - X)^2) = 0$.

Hierarchy of convergence

$$
\begin{array}{ccc}
X_n \xrightarrow{a.s.} X & & \\
 & \searrow\!\!\Rightarrow & \\
 & & X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X. \\
 & \nearrow\!\!\Rightarrow & \\
X_n \xrightarrow{m.s.} X & &
\end{array}
$$

No other implications hold in general.

Random Processes (Module 5)

A collection of RVs indexed by time is a called a random process

X_t , $t \in \mathbb{R}$, X_n , $n \in \mathbb{Z}$.

Mean $\mu_X(t) = E[X_t]$

(Auto) correlation $R_X(t_1, t_2) = E[X_{t_1} X_{t_2}]$

(Auto) covariance $C_X(t_1, t_2) = R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$

$= \mathrm{cov}(X_{t_1}, X_{t_2})$.

Bernoulli process

$X_i \sim$ i.i.d. with $P_X(1) = p = 1 - P_X(0)$.

Poisson Process N_t

(1) $N_0 = 0$

(2) $0 \le t_1 < t_2 < \dots < t_n$, $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$ are independent

(3) $N_{t+\tau} - N_t \sim \mathrm{Poisson}(\lambda\tau)$.

Strict-Sense Stationary Process (SSS):

$$F_{X_{t_1}, X_{t_2} \dots X_{t_r}}(x_1, x_2, \dots, x_r) = F_{X_{t_1+\tau}, \dots, X_{t_r+\tau}}(x_1, x_2, \dots, x_r),$$

$\forall\ r \in \mathbb{N},\ t_1, t_2, \dots, t_r \in \mathbb{R},\ \tau \in \mathbb{R}.$

Wide-Sense Stationary Process (WSS):

$$\mu_X(t_1) = \mu_X(t_2),\ \forall\ t_1, t_2$$

$$R_X(t_1, t_2) = R_X(t_1 - t_2)$$

So, Autocorrelation function $R_X(\tau)$.

Power Spectral Density

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau)\, e^{-i2\pi f \tau}\, d\tau.$$