REVIEW: Linear Algebra & Probability.

- Matrix
- Eigen Vectors / Eigenvalues
- SVD (Singular Value Decomposition).

Section 11.4. pp 330 (Hopcroft & Kannan)

Random Variable : $X : \Omega \to \mathbb{R}$

A real-valued function on a set of possible outcomes ← Sample Space, Ω

Random Outcome ⟶ Property or measurements on the random outcome.

Probability Space $(\Omega, \mathcal{F}, P)$

Ω = Set (of possible outcomes) {No Additional Structure}

$\mathcal{F}$ = σ-field of subsets of Ω {Nonempty, closed under complement & countable union.

P = Measure on the measurable space. $(\Omega, \mathcal{F})$

$P(\Omega) = 1 \Rightarrow P(\phi) = 0$

$A \in \mathcal{F}, \; P(A) \in [0,1]$

$\{\omega \in \Omega : X(\omega) < a\} \in \mathcal{F}$ = Event

$P_X(B) = P(X^{-1}(B))$ B = Borel subset of the real line.

a) Kolmogorov's 0-1-Law

b) Law of Large Numbers / [Strong Law of Large Numbers.]

Events in the asymptotic σ-field has probability 0 or 1.

$\{X_n, n \geq 1\}$: Sequence of centered independent r.v.s.

$$S_n = X_1 + \cdots + X_n$$

$$\sum_{n \geq 1} \frac{E(X_n^2)}{n^2} < \infty \quad \Rightarrow \quad \frac{S_n}{n} \to 0 \quad a.s.$$

$$\left[\begin{array}{l} E(|X_1|) < \infty \Rightarrow \frac{S_n}{n} \to E(X_1) \text{ a.s.} \\ E(|X_1|) = \infty \Rightarrow \limsup_{n \to \infty} \frac{|S_n|}{n} = +\infty \quad \text{a.s.} \end{array} \right.$$

History:
a) Hilbert's Sixth Problem (1900)
b) Borel's Paradox:
Related Baye's Theorem.
c) Banach-Tarski Paradox (also, Axiom of Choice)
d) Savage's Utility Problem.

PDF : Probability distribution function:

$P_X(B)$

$\cancel{P[a<X<}$ $\quad P[a < X < b] = P_X(a,b) = P(\{\omega \in \Omega : a < X(\omega) < b\})$

<u>Sample Space Ω</u>, Set of outcomes of a probability experiment.

E.g. n cointosses $\to (H+T)^n = 2^n$-long strings over alphabet $\Sigma = \{H, T\}$

<u>Real Random Variable</u> $X : \Omega \to \mathbb{R}$, Real-valued function on Ω.

E.g. Number of heads among the n-cointosses.

$$\left. \begin{array}{ll} X = k & \binom{n}{k} \text{ possibilities.} \\ & = \binom{n-1}{k-1} + \binom{n-1}{k} = \frac{n!}{(n-k)!\,k!} \end{array} \right\} 0 \leq k \leq n.$$

<u>Event</u>: Subset of Ω (in the σ-field, if Ω is not finite).

E.g. A specific subset of 3-tosses; with even # of heads.

Event = {HHT, HTH, THH, TTT}

<u>Conditional Probability</u> (of event F conditioned on event G)

$$\Pr(F|G)\Pr(G) = \Pr(F \wedge G) \qquad \left\{ \Pr(F|G) = \frac{\Pr(F\wedge G)}{\Pr(G)} \right.$$

$$\left\{ \begin{array}{l} = \dfrac{\Pr(F\wedge G)}{\Pr(F\wedge G) + \Pr(\neg F \wedge G)} \\[2ex] = \dfrac{\Pr(G|F)\Pr(F)}{\Pr(G|F)\Pr(F) + \Pr(G|\neg F)\Pr(\neg F)} \end{array} \right.$$

<u>Independence of events F and G</u>

$$\Pr(F|G) = \Pr(F)$$

$$\Pr(F\wedge G) = \Pr(F|G)\cdot\Pr(G) = \Pr(F)\cdot\Pr(G)$$

<u>Mutual Independence of events $F_1, F_2, \cdots, F_n$</u>

$$\Pr(F_1 \wedge F_2 \wedge \cdots \wedge F_n) = \Pr(F_1)\Pr(F_2)\cdots\Pr(F_n)$$

$$\left\{ = \Pr(F_1 | F_2 \wedge\cdots\wedge F_n)\Pr(F_2 | F_3\wedge\cdots\wedge F_n)\cdots\Pr(F_{n-1}|F_n)\Pr(F_n) \right\}$$

<u>Indicator Variable</u> $\quad \mathbf{1}_{\text{Event}} = \begin{cases} 1 & \text{if event happens} \\ 0 & \text{otherwise} \end{cases}$

<u>Independence of r.v.s X and Y</u>

$$\forall_{A,B \subseteq \mathbb{R}} \quad \Pr(X\in A \wedge Y\in B) = \Pr(X\in A)\Pr(X\in B)$$

$$\forall a,b \in \mathbb{R} \quad \Pr(X=a \wedge Y=b) = \Pr(X=a)\Pr(X=b)$$

<u>Mutual Independence of r.v.s $X_1, X_2, \cdots, X_n$</u>

$$\forall A_1, A_2 \cdots A_n \subseteq \mathbb{R} \quad \Pr(X_1\in A_1 \wedge X_2 \in A_2 \cdots \wedge X_n\in A_n) = \Pr(X_1\in A_1)\Pr(X_2\in A_2)\cdots\Pr(X_n\in A_n)$$

$$\forall a_1, a_2, \cdots A_n \in \mathbb{R} \quad \Pr(x_1=a_1 \wedge \cdots \wedge x_n=a_n) = \Pr(x_1=a_1)\cdots\Pr(x_n=a_n)$$

General Random Variable

Function from Ω to any set

$$X : \Omega \to S$$

E.g. Random Graph

$$G(n,p) : \binom{n}{2} \text{ coin tosses} \to G = (V,E)$$

$$|V| = n.$$

$$A_G = \begin{cases} a_{ij} = 1 = a_{ji} & i<j \wedge (i,j)^{th} \text{ cointoss} = H \\ 0 & \text{o.w.} \end{cases}$$

Probability Density Function pdf $p(x)$, such that

$$Pr(a < x < b) = \int_a^b p(x)\,dx.$$

Cumulative Distribution Function cdf $f(a)$

$$f(a) = \int_{-\infty}^{a} p(x)\,dx. = Pr(x \leq a).$$

Mean or Expectation $E(X)$ of a r.v. X.

$$E(X) = \int_{-\infty}^{\infty} x\, p(x)\, dx = \mu(x) \left\{ \text{or } \sum_x p(x) = \mu(x) \right\}$$

$$E\left(\mathbb{1}_{a<x<b}\right) = \int_{-\infty}^{\infty} \mathbb{1}_{a<x<b}\, p(x)\, dx$$

$$= \int_a^b p(x)\,dx = Pr(a < x < b)$$

Let x be a non-negative r.v. Then

$$\mathbb{1}_{x \geq t} \leq \frac{x}{t}$$

$$E\left(\mathbb{1}_{x \geq t}\right) \leq E\left(\frac{x}{t}\right) = \frac{E(x)}{t}$$

$$Pr(x \geq t) \leq E(x)/t \leftarrow \boxed{\text{Markov Inequality}}$$

<u>Variance</u> $\quad Var(x) = E(X - E(X))^2 = \sigma^2(x)$

<u>k^{th} Moments</u> $\quad \mu_k(x) = E(x^k)$

$$Var(x) = \mu_2(x) - \mu_1(x)^2 = E(X^2) - E(X)^2 = \sigma^2$$

$$Pr[(x-\mu)^2 \geq a^2\sigma^2] \leq \frac{E[(x-\mu)^2]}{a^2\sigma^2} \quad \text{by Markov Ineq.}$$

$$= \frac{1}{a^2}$$

$$\therefore \quad Pr[|x-\mu| \geq a\sigma] \leq \frac{1}{a^2} \quad \leftarrow \text{Chebyshev's Inequality}$$

<u>Linearity of Expectations</u> [No independence assumed.]

$$E\left(\sum x_i\right) = \sum E(x_i)$$

Expectation of sum of r.v.s is sum of expectations.

Variance of sum of ~~r.v.s.~~ independent r.v.s is the sum of variances. [Independence must be assumed.]

$$Var\left(\sum x_i\right) = \sum var(x_i), \quad \text{if } x_i\text{'s are pairwise independent r.v.s.}$$

<u>i.i.d</u> (Mutually) Independent Identically Distributed r.v.s.

Probability Distributions

◇ Bernoulli Trials

$$x \in \{0,1\}$$
$$Pr[x=1] = p \quad \{Pr[x=0] = q = 1-p\}$$

◇ Binomial Distribution $\quad X \sim B(n,p) \qquad x \in \{0,1,\dots,n\}$

$$Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

In n independent Bernoulli trials, exactly k successes are observed.

$$\mu = np \qquad \sigma^2 = npq$$

◇ Poisson Distribution. $\quad X \sim \text{Poisson}(\lambda)$

$$Pr(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Average rate of an event per unit time $= \lambda$
Exactly k events are observed in a unit time

$$X \sim Bin(n, \tfrac{\lambda}{n})$$

$$Pr(X=k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n^{(k)}}{k!} \frac{\lambda^k}{n^k} \left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}\cdot\lambda} \left(1-\frac{\lambda}{n}\right)^{-k}$$

$$\approx e^{-\lambda} \frac{\lambda^k}{k!} \frac{n^{(k)}}{n^k} \frac{1}{\left(1-\frac{\lambda}{n}\right)^k}$$

$$\lim_{n\to\infty} \left(\quad\right) \to e^{-\lambda} \frac{\lambda^k}{k!}$$

$$\mu = \lambda \qquad \sigma^2 = \lambda$$

Gaussian Distribution.

$$X \sim \mathcal{N}(\mu, \sigma^2) \qquad p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \mu = \mu, \quad \sigma^2 = \sigma^2$$

$$\left\{ \begin{array}{l} X \sim \mathcal{N}(0,1) \\ p(x) = \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} \end{array} \right.$$

Central Limit Theorem.

$X_1, X_2, \dots$ iid with $E(X_i) = \mu$, $Var(X_i) = \sigma^2$

Then $$\frac{\sum_{i=1}^{n} x_i - n\mu}{\sqrt{n}\,\sigma} \sim \mathcal{N}(0,1)$$

Let $x_1, \dots, x_n$ be independent random variables (Bernoulli)

$$x_i \sim \text{Bernoulli}(p)$$

$$S = \sum_{i=1}^{n} x_i \qquad m = \mu(S) = np$$

Chernoff Bounds.

$$\forall_{\delta > 0} \quad \text{Prob}\left[S > (1+\delta)m\right] \le \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^m$$

$$\forall_{\delta > 2e-1} \quad \text{Prob}\left[S > (1+\delta)m\right] \le \left(\frac{e}{1+\delta}\right)^{(1+\delta)m}$$

$$\forall_{0 < r \le 1} \quad \text{Prob}\left[S < (1-r)m\right] < \left[\frac{e^{-r}}{(1+r)^{(1+r)}}\right]^m < e^{-\frac{r^2 m}{2}}$$

Hoeffding Bounds

$$\text{Prob}\left[S - m \ge \epsilon\right] \le e^{-2\epsilon^2/n}$$

Martingale: A sequence of r.v.s $z_1, z_2 \cdots z_n$ s.t $E(z_{i+1} | z_1, \cdots z_i) = z_i$ wit prob $= 1$.

Azuma-Hoeffding Bounds.

Parametric Family of Distributions.

$$\mathcal{P} \quad \leftarrow \quad \theta \in \Theta \subset \mathbb{R}^d$$

IDENTIFIABILITY Mapping from $\Theta \to \mathcal{P}$ is one-to-one.

Likelihood Function:

$$L(\theta) = p_{X|\Theta}(x|\theta)$$

θ = unobservable.

$T(x)$ = Statistics (Eg $\bar{x}, s^2$, etc.) = Observable.

$\phi(X) = \hat{\theta}$ = Estimator of θ

Maximum Likelihood Estimator (MLE)

$$\hat{\theta} = \sup_{\theta \in \Theta} L(\theta)$$

Maximum A Posteriori Estimator (MAP)

$$\hat{\theta} = \sup_{\theta \in \Theta} L(\theta)\, p(\theta)$$

$p(x|\theta)\, p(\theta)$
$= p(\theta|x)\, p(x)$

Limit of a Bayes Estimator under 0-1 Loss function

Minimum Mean Square Error Estimator (MMSE)

$$\hat{\theta} = \int \theta\, L(\theta)\, p(\theta)\, d\theta$$

Minimizes
$E_{\theta,x}\left[(\hat{\theta} - \theta)^2\right]$

Bayesian

Under Laplace Rule of Indifference
(Uninformative Prior)

$$\hat{\theta} = \int \theta\, L(\theta)\, d\theta$$

Empirical Bayes: $\hat{\theta} = \int \theta\, L(\theta)\, p_{\eta_x^*}(\theta)\, d\theta$

$p_\eta(\theta)$ = family of priors

Best (Empirical) prior.

$$\underbrace{x_1, x_2, \ldots, x_n}_{\text{observed.}} \qquad \underbrace{\text{mean} = \mu_0, \quad \text{variance} = \sigma_0^2}_{\text{unobserved}}$$

$$L(\mu, \sigma^2; x_1, \ldots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum_{j=1}^{n} (x_j - \mu)^2}{2\sigma^2}\right)$$

$$l = -\log_e L(\mu, \sigma^2; x_1, \cdots x_n) = \frac{n}{2} \ln(2\pi) + \frac{n}{2} \ln(\sigma^2) + \frac{1}{2\sigma^2} \sum_{j=1}^{n} (x_j - \mu)^2$$

Negative LogLikelihood.

~~$\arg\inf_{\mu} l = \arg\inf_{\mu}$~~

$$\arg\min_{\mu} l = \arg\min_{\mu} \sum_{j=1}^{n} (x_j - \mu)^2$$

$$\frac{\partial}{\partial\mu} \sum_{j=1}^{n} (x_j - \mu)^2 = \sum_{j=1}^{n} x_j - n\mu = 0$$

$$\hat{\mu}_0 = \sum_{j=1}^{n} x_j / n$$

$$\frac{\partial}{\partial\sigma^2} l = \frac{n}{2\sigma^2} - \left[\frac{1}{2} \sum_{j=1}^{n} (x_j - \mu)^2\right] \frac{1}{\sigma^4}$$

$$\Rightarrow \quad n = \sum_{j=1}^{n} (x_j - \mu)^2 / \sigma^2$$

$$\hat{\sigma}_0^2 = \sum_{j=1}^{n} (x_j - \hat{\mu}_0)^2 / n$$