SOCIAL NETWORK

March 24 2015

LECTURE # 6

Review: Probability Theory.

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

A real valued function on a sample space, Ω

↪ A set of possible outcomes.

Random Outcome

⇒ Property or measurements on the random outcomes.

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

Ω = Set of possible outcomes [No Additional Structure.

$\mathcal{F}$: σ-field of subsets of Ω [Nonempty, closed under complement" & Countable union.

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

$\mathcal{P}(\Omega) = 1 \Rightarrow \mathcal{P}(\phi) = 0$

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

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

B = Borel subset of the real line.

$P_X(B) \triangleq P(X^{-1}(B))$

HISTORY. a) Hilbert's SIXTH PROBLEM (1900)

b) Borel's Paradox

(Related to Baye's Theorem).

c) Banach-Tarski Paradox.

(Also Axiom of Choice).

d) Savage's Utility Problem.

Solution → Kolmogorov

1) Kolmogorov's 0-1 Law

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

2) Law of Large Numbers //

Strong Law of Large Numbers.

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

$$S_n = x_1 + x_2 + \cdots + x_n$$

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

$$\begin{cases} E(|X_1|) < \infty \Rightarrow \dfrac{S_n}{n} \to E(X_1) \text{ a.s.} \\ E(|X_1|) = \infty \Rightarrow \limsup_{n \to \infty} \dfrac{|S_n|}{n} = +\infty \text{ a.s.} \end{cases}$$

Probability Distribution Function (PDF)

$\mathcal{P}_X(B) = P(X^{-1}(B))$ ← PDF.

$$P[a<x<b] = P_x(a,b) = P\{\omega \in \Omega \mid a < X(\omega) < b\}$$

Open Interval = Borel.

Sample Space, Ω = Set of outcomes of a Probability Experiment

Example

n: cointosses $\rightarrow (H+T)^n$

$= 2^n$ n-long strings over the alphabet $\Sigma = \{H,T\}$

random genomes of length $n \rightarrow (a+t+c+g)^n$

$= 4^n$ n-long strings over the alphabet $\Sigma = \{a,t,c,g\}$

Real Random Variable $X : \Omega \rightarrow \mathbb{R}$

Real-valued functions on Ω.

Example : Number of heads among the n-coin tosses

$X = k$ $\quad \binom{n}{k}$ possiblities

$$\frac{n!}{k!\,(n-k)!} \quad \left\{ \binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \right.$$

k = Head $\quad$ k = Tail

$X' = k-1$ $\quad$ $X' = k$

Event: Subset of Ω. $\left\{\begin{array}{l}\text{In the } \sigma\text{-field,} \\ \text{if } \Omega \text{ is not finite.}\end{array}\right.$

Example: A specific subset of 3-tosses with even number of heads.

$$\{TTT, HHT, HTH, THH\}$$

$$\# \binom{3}{0} + \binom{3}{2} = 1 + 3 = 4.$$

Conditional Probability. (of events F conditioned on event G).

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

$\boxed{\text{Baye's Rule:}}$

$$Pr(F|G) = \frac{Pr(F \wedge G)}{Pr(G)} = \frac{Pr(F \wedge G)}{Pr(G \wedge F) + Pr(G \wedge \neg F)}$$

$$= \frac{Pr(G|F)\;Pr(F)}{Pr(G|F)\,Pr(F) + Pr(G|\neg F)\;Pr(\neg F)}$$

Independence (of events F and G)

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

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

$$= Pr(F) \cdot Pr(G).$$

Mutual Independence (of events $F_1, F_2 \cdots F_n$)

$$\Pr(F_1 \wedge F_2 \wedge \cdots \wedge F_n)$$
$$= \Pr(F_1 \mid F_2 \wedge \cdots \wedge F_n) \Pr(F_2 \mid F_3 \wedge \cdots \wedge F_n) \cdots \Pr(F_n)$$
$$= \Pr(F_1) \Pr(F_2) \cdots \Pr(F_n).$$

Indicator Variable

$$\mathbb{1}_{Event} = \begin{cases} 1 & \text{if event happens} \\ 0 & \text{otherwise.} \end{cases}$$

Independence of Random Variables X & Y.

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

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

Mutual Independence of r.v.s. $X_1, \cdots, X_n$.

$$\forall_{A_1, A_2, \cdots A_n}$$
$$\Pr(X_1 \in A_1 \wedge X_2 \in A_2 \wedge \cdots X_n \in A_n)$$
$$= \Pr(X_1 \in A_1) \Pr(X_2 \in A_2) \cdots \Pr(X_n \in A_n)$$

$$\forall a_1, a_2, \cdots, a_n$$
$$\Pr(X_1 = a_1 \wedge X_2 = a_2 \wedge \cdots \wedge X_n = a_n)$$
$$= \Pr(X_1 = a_1) \Pr(X_2 = a_2) \cdots \Pr(X_n = a_n).$$

General Random Variable

Function from Ω to any arbitrary Set.

$$X : \Omega \to S$$

Example:
Random Graph

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

$$G = (V, E) ; \quad |V| = n$$

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

Probability Density Function. pdf $p(x)$

such that

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

Cumulative Density Function cdf $c(a)$

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

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

$$E(X) = \int_{-\infty}^{\infty} x \, p(x) \, dx = \mu(x)$$

$$\left\{ \text{or } \sum x \, p(x) \text{ if } \Omega = \text{finite} \right\}$$

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

$$= \int_{a}^{b} p(x) \, dx = \Pr[a<x<b].$$

Let x be a nonnegative r.v. Then.

$$t \leq x \iff 1 \leq \frac{x}{t}$$

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

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

$$\Pr[x \geq t] \leq \frac{E(x)}{t}$$ } Markov Inequality.

Variance $Var(X) = E(X - E(X))^2 = \sigma^2(X)$

k^{th} Moment $\mu_k(x) = E(X^k)$

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

$$Pr[(x-\mu)^2 \geq a^2\sigma^2] \leq \frac{E[(x-\mu)^2]}{a^2\sigma^2} = \frac{1}{a^2}$$

(by Markov Ineq)

$$Pr[\,|x-\mu| \geq a\sigma\,] \leq \frac{1}{a^2}$$ Chebyshev Inequality.

Linearity of Expectations

(No independence assumed.)

$$E(\Sigma x_i) = \Sigma E(x_i)$$

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

But.

$$Var(\Sigma x_i) = \Sigma Var(x_i)$$ if x_i's are pairwise independent r.v.s.

Probability Distributions.

- **Bernoulli**

$$X \sim \text{Bernoulli}(p)$$
$$X \in \{0,1\}$$
$$Pr[X=1] = p$$
$$Pr[X=0] = q = 1-p$$

- **Binomial** $X \sim B(n,p)$ $\quad X \in \{0,1,2,\dots,n\}$

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

In n independent trials, exactly k successes are observed:

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

- **Poisson** $X \sim \text{Poisson}(\lambda)$

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

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

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

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

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

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

$$\left.\lim_{n\to\infty}(\;) \to e^{-\lambda}\frac{\lambda^k}{k!}\right\} \quad \mu = \lambda, \quad \sigma^2 = \lambda$$

Let $x_1, x_2, \dots, x_n$ be independent Bernoulli random variables

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

Random Walk.

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

Chernoff Bound:

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

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

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

Hoeffding Bounds

$$\text{Prob}\left[S_n - m_n \ge \epsilon \right] < e^{-2\epsilon^2/n}$$

Gaussian $X \sim \mathcal{N}(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mu : \mu, \quad \sigma^2 = \sigma^2$$

CLT (Central Limit Theorem)

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

Then

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

Random Graph $G(n, p)$ Model
{ Erdös - Renyi }

A graph $G = (V, E)$ is constructed by connecting every pair of nodes uniformly randomly.

$$\forall_{u, v \in V} \quad [(u, v) \in E] \sim \text{Bernoulli}(p) \text{ iid.}$$

For every pair of nodes $u, v \in V$, an edge $(u, v) \in E$ is included in the graph with probability p independent from every other edge.

$|V| = n \qquad \langle |E| \rangle = \binom{n}{2} p$

Expected number of edges.

Expected degree of a $G \in G(n,p)$

$$\langle d \rangle = \frac{2 \langle |E| \rangle}{\langle |V| \rangle} = \frac{2 \binom{n}{2} p}{n} = (n-1)p$$

$d(v) \sim \text{Binomial}(n-1, p)$

$$\Pr[d(v) = k] = \binom{n-1}{k} p^k (1-p)^{n-k-1}$$

$$\approx \frac{(np)^k}{k!} e^{-np}$$

if expected degree $\bar{d}$ is held constant (independent of n): n = large, np = const. $= \lambda$

$d(v) \sim \text{Poisson}(\lambda) \qquad \lambda = \bar{d} = np.$

Poisson Approximation.

Phase Transition: 0.1 Laws.

Small $p \qquad p < \frac{(1-\epsilon)\ln n}{n} \qquad G(n,p)$ = Disconnected a.s.

Large $p \qquad p > \frac{(1+\epsilon)\ln n}{n} \qquad G(n,p)$ = Connected a.s.