STOR 435

Distributions

Probability Relationships

$A \subset B$	If A occurs, so does B
$A = B$	iff $A \subset B$ and $B \subset A$
$A \cup B$	A or B occurs, or both occur
$A \cap B$	both A and B occur
$A \cup B = \emptyset$	A and B are `mutually exclusive`

Addition Rule

P(A ∪ B) = P(A) + P(B) - P(A ∩ B)
P(A ∪ B) = P(A) + P(B) iff $A \cap B = \emptyset$

Conditional Probability and Independence

Division Rule, or Conditional Probability

$$P(A|B) = \begin{cases} \frac{P(A \cap B)}{P(B)} & \text{if } P(B) > 0 \\ 0, & \text{if } P(B) = 0 \end{cases}$$

Multiplication Rule

P(A ∩ B) = P(A) P(B|A)

Independence

Two events A and B are independent if any are satisfied:
P(A ∩ B) = P(A) P(B)
P(A|B) = P(A)
P(B|A) = P(B)

Discrete RV Independence

Discrete RV's X_1, ..., X_n are said to be independent iff

$$P(X_1 = x_1, ..., X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i), \forall \{x_1, ..., x_n\}$$

Continuous RV Independence

Continous RV's X_1, ..., X_n are said to be independent iff

$$P(X_1 \leq x_1, ..., X_n \leq x_n) = \prod_{i=1}^{n} P(X_i \leq x_i), \forall (x_1, ..., x_n) \in \mathbb{R}^n$$

Binomial RV and Distribution

Bernoulli trials

X ∼ Bin(n, p)
Let X = total number of Heads among n tosses, then
P(X = k) $= \binom{n}{k} p^k q^{n-k}$

Normal RV and Distribution

Density

- Continuous RV X
- Density curve $f(x)$ for X that satisfies:
  - $f(x) \geq 0, \forall x \in \mathbb{R}$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$
  - P(a ≤ X ≤ b) = $\int_a^b$ f(x) dx
- cdf (cumulative distribution function)
  - $F(x) = \int_{-\infty}^{x} f(u)du, x \in \mathbb{R}$
- $\frac{dF(x)}{dx} = f(x)$, ∀ x at which $\frac{dF(x)}{dx}$ exists

Normal Distribution

- X ∼ N(μ, σ^2)
  - μ ::= mean(location)
  - σ^2 ::= variance (scale)
  - σ ::= std dev
- density $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{(2\sigma^2)}}$

Standard Normal Distribution

- Z ∼ N(0, 1)
- density ϕ(z) $= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$
- cdf Φ(z) = $\int_{-\infty}^{z} \phi(u)du$
- Φ(a, b) = Φ(b) - Φ(a)
- Φ(-1, 1) ≈ 0.6826
- Φ(-2, 2) ≈ 0.9544
- Φ(-3, 3) ≈ 0.9975

Standardization

- X ∼ N(μ, σ^2) $\iff$ Z = $\frac{X-\mu}{\sigma}$ ∼ N(0, 1)
- If X ∼ N(μ, σ^2), then P(a ≤ X ≤ b) = $\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$

Binomial Approximated by Normal or Poisson

Normal Approx to Binomial

- Let X ∼ Bin(n, p), Z ∼ N(0, 1). For a ≤ b and large n,
- Applies to cumulative binomial probabilities

$$\begin{aligned} P(a \leq X \leq b) &\approx P(\frac{a-\mu-0.5}{\sigma} \leq Z \leq \frac{b-\mu+0.5}{\sigma}) \\ &= \Phi(\frac{b-\mu+0.5}{\sigma}) - \Phi(\frac{a-\mu-0.5}{\sigma}) \end{aligned}$$

where μ = np, $\sigma = \sqrt{npq}$

Poisson approximation

- Conditions: Let X ∼ Bin(n,p) where n is large, and either p or q is small, so $\sqrt{npq} < 3$
- Applies to single-point binomial probabilities

P(X = k) $\approx e^{-\mu}\frac{\mu^k}{k!}$, k=0, 1, 2...

Discrete RV's

Univariate models

- range(or support) of RV X : set of all possible values for X
- transformations : a new RV Y is defined via composition Y = g(X)

Multivariate Models

- independence : X and Y are independent iff P(X = x, Y = y) = P(X = x) P(Y = y), ∀ x, y

Expectations

- E(X) = $\sum_x$ x P(X = x)
- E(Y) = E[g(X)] = $\sum_x$ g(x) P(X = x)
- For constant c, E(c) = c
- For constant c and RV X, E(cX) = c E(X)
- For RV's X_1, ..., X_n, E(X_1 + ... + X_n) = E(X_1) + ... + E(X_n)
- If RV's X and Y are independent, then E(XY) = E(X) E(Y)

X ∼ Uniform{x_1, ..., x_K}	E(X) = $\frac{x_1+...+x_K}{K}$
X ∼ Poisson(μ)	E(X) = μ
X ∼ Bin(n,p)	E(X) = np
X ∼ Geom(p)	E(X) = 1/p

Summaries of Distributions

- Mean: E(X) = μ
- Covariance: Cov(X, Y) = E(XY) - E(X) E(Y)
- Variance: Var(X) = Cov(X, X) = E(X^2) - μ^2 = σ^2
- SD(X) = $\sqrt{Var(X)} = \sigma$
- Correlation: Corr(X,Y) = $\frac{Cov(X,Y)}{SD(X)SD(Y)} = \rho$
- X and Y are said to be uncorrelated if Corr(X,Y) = 0
- If X and Y are independent, then Corr(X,Y) = 0 and Corr(X,Y) = 0

Tail-sum formulas

$$E(X) = \begin{cases} \sum_{k=1}^{n} P(X \geq k), & \text{if range of X is \{0, 1, ..., n\}} \\ \sum_{k=1}^{\infty} P(X \geq k), & \text{if range of X is \{0, 1, ..., \}} \end{cases}$$

SD and Normal Approx

Variance and SD Properties

- Var(X) ≥ 0 for any RV X
- For constant c, Var(C) = 0
- For constant c and RV X, Var(cX) = c^2 Var(X)
- Var(X + Y) = Var(X) + Var(Y) + 2 Cov(X, Y), in general
- Var(X + Y) = Var(X) + Var(Y) iff X and Y are uncorrelated
- Let $X_1, ..., X_n$ be iid RV's with a common mean μ and variance σ^2, denote $\bar{X} = \frac{1}{n}(X_1 + ... + X_n)$
  - E($\bar{X}$) = μ
  - Var($\bar{X}$) = $\frac{\sigma^2}{n}$

X $\sim$ Uniform$\{x_1, ..., x_K\}$ Var(X) $= \frac{1}{K}\sum_{k=1}^{K} x_k^2 - (\bar{x})^2$
X $\sim$ Poisson(μ) Var(X) $= \mu$
X $\sim$ Bin(n,p) Var(X) = npq

Central Limit Theorem

Let $X_1, X_2, ...$ be a sequence of iid RV's with (the same) mean μ and variance σ^2. Denote $S_n = X_1 + ... + X_n$. For any a $\leq$ b,

$$\lim_{n\to\infty} P(a \leq \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq b) = \Phi(b) - \Phi(a)$$

Under assumptions for CLT, with large n and Z $\sim$ N(0, 1),

$$P(a \leq S_n \leq b) \approx P(\frac{a - n\mu - \frac{\delta}{2}}{\sqrt{n\sigma^2}} \leq Z \leq \frac{b - n\mu + \frac{\delta}{2}}{\sqrt{n\sigma^2}})$$

, where terms $\pm\frac{\delta}{2}$ are referred to as correction for continuity, where $\delta > 0$ is the largest number such that all possible values of X_1 are multiples of δ e.g. $\delta = 1$ when X $\sim$ Bin(n,p) or X $\sim$ Poisson(μ)

Geometric Distribution

Basic example

Toss a coin(p) repeatedly and let X be total number of tosses needed to observe the first H, denoted by X $\sim$ Geom(p)

P(X = k) $= q^{k-1}p$
P(X $\geq$ k) $= q^{k-1}$
E(X) $= \frac{1}{p}$
Var(X) $= \frac{q}{p^2}$

Infinite Sum Rule

If a countable collection of events $\{A_1, A_2, ...\}$ partition A, then

$$P(A) = \sum_{k=1}^{\infty} P(A_k)$$

Negative Binomial Distribution

Basic example

Toss a coin(p) repeatedly and let T_r be the total number of tosses needed to observe the rth H, where r is a positive integer. Note the range of T_r is {r, r+1, ...}
Let $X = T_r - r$ be the total number of T's before seeing the rth H. X $\sim$ NegBin(r,p), with

$$P(X = k) = \binom{k + r - 1}{r - 1} q^k p^{r-1} p$$

$E(T_r)$ $= \frac{r}{p}$
E(X) $= \frac{rq}{p}$
Var(T_r) $= \frac{rq}{p^2}$
Var(X) $= \frac{rq}{p^2}$

Density and Expectation

Plug-in formula for Expectation

Let X have a density f and g be a known function.

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx = \int_{\mathbb{R}} g(x)f(x)dx$$

Special Cases:
Mean: E(X) $= \mu$
Variance: Var(X) $= \sigma^2$
kth moments: E(X^k) $= \int_{\mathbb{R}} x^k f(x)dx$, where k is a pos int

Uniform Distributions

Suppose X $\sim$ Uniform[a,b] with density $f(x) = \frac{1}{b-a}, x \in [a, b]$.
E(X) $= \frac{a+b}{2}$
Var(X) $= \frac{(b-a)^2}{12}$

Cauchy Distribution

Suppose X $\sim$ Cauchy with density $f(x) = \frac{1}{\pi(1+x^2)}, x \in \mathbb{R}$ Then E(X) does not exist.

Exponential and Gamma Distributions

Exponential Distribution

Let "lifetime" X $\sim$ Exp(λ) with density $f(x) = \lambda e^{-\lambda x}, x \geq 0$, where $\lambda > 0$ is a constant parameter.

- Survival Function: P(X > x) = 1 - F(X) = $e^{-\lambda x}, x \geq 0$ with cdf:
- $$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$
- $E(X) = \frac{1}{\lambda}$
- $Var(X) = \frac{1}{\lambda^2}$
- Memoryless Property: P(X > x + t | X > x) = P(X > t), $\forall$ x $\geq$ 0, t $\geq$ 0.

Gamma Distribution

Let X_1, ..., X_n be iid Exp(λ), and $S_n = X_1 + ... + X_n$. Then $S_n \sim$ Gamma(n, λ) has density:

$$f_{S_n}(x) = \frac{e^{-\lambda x}\lambda^n x^{n-1}}{(n-1)!}, x \geq 0$$

- E(S_n) $= \frac{n}{\lambda}$
- Var(S_n) $= \frac{n}{\lambda^2}$

CDF

Definition

$$F(x) = P(X \leq x), x \in \mathbb{R}$$

If X has a discrete distribution,

$$F(x) = \sum_{u:u\leq x} P(X = u) \iff P(X = x) = F(x) - F(x-)$$

where

$$F(x-) = \lim_{u \to x^-} F(u)$$

If X has a density, then F is a continuous curve with,

$$F(x) = \int_{-\infty}^{x} f(u)du \iff f(x) = \frac{dF(x)}{dx}$$

Also, there are RV's whose distributions are neither continous nor continuous.

cdf Properties

- Every cdf F is a non-decreasing and right-continous function, and satisfies:

  $$0 \leq F(x) \leq 1; \lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$$

- Interval Probabilities via cdf:
  - $P(a < X < b) = F(b-) - F(a)$
  - $P(a \leq X < b) = F(b-) - F(a-)$
  - $P(a < X \leq b) = F(b) - F(a)$
  - $P(a \leq X <\leq) = F(b) - F(a-)$

Multivariate Continuous Distributions

Basics

- For random vector (X_1, ... , X_n):
  - joint cdf: F(x_1, ... , x_n) = P($X_1 \leq x_1$, ..., $X_n \leq x_n$)
  - joint density: f(x_1, ..., x_n) that satisfies:
    * f(x_1, ... , x_n) $\geq 0, \forall$ (x_1, ..., x_n) $\in \mathbb{R}^n$;
    * $\int_{\mathbb{R}^n} f(x_1, ..., x_n)dx_1...dx_n = 1$
- Connections: for (x_1, ... , x_n) $\in \mathbb{R}^n$,
  - F(x_1, ... , x_n) = $\int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} f(u_1, ..., u_n)du_1...du_n$
  - f(x_1, ... , x_n) $= \frac{\partial^n F(x_1,...,x_n)}{\partial x_1...\partial x_n}$
- infinitesimal probability
  - P($X_1 \in dx_1, ..., X_n \in dx_n$) $\approx f(x_1, ..., x_n)dx_1...dx_n$
- domain probability: for a domain D $\subset \mathbb{R}^n$,
  - P((X_1, ..., X_n) $\in D$) $= \int_D f(x_1, ..., x_n)dx_1...dx_n$
- 1d-marginals: for i = 1, ..., n,
  - $f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} f(u_1, ..., u_{i-1}, x_i, u_{i+1}, ..., u_n)du_1...du_{i-1}du_{i+1}...du_n$
  - $F_{X_i}(x_i) = \lim_{u_j\to\infty, j\neq i} F(u_1, ..., u_{i-1}, x_i, u_{i+1}, ..., u_n)$
- independence
  $X_1, ..., X_n$ are independent

  $$\iff f(x_1, ..., x_n) = f_{X_1}(x_1)...f_{X_n}(x_n), \forall(x_1, ..., x_n) \in \mathbb{R}^n$$

  $$\iff F(x_1, ..., x_n) = F_{X_1}(x_1)...F_{X_n}(x_n), \forall(x_1, ..., x_n) \in \mathbb{R}^n$$

- expectation : for a known function g

  $$E[g(X_1, ..., X_n)] = \int_{\mathbb{R}^n} g(x_1, ..., x_n)f(x_1, ..., x_n)dx_1...dx_n$$

MultiVar Uniform Dist

A random vector $(X_1, ..., X_n)$ is said to be uniformly distributed over a bounded domain $D \subset \mathbb{R}^n$ if:

$$P((X_1, ..., X_n) \in A) = \frac{volume(A)}{volume(b)}, \forall A \subset D$$

Rectangular Uniform Domains

For a rectangular domain D = [a,b] × [c,d], (X,Y) is uniformly distributed over D $\iff$ X ~ Uniform[a,b] and Y ~ Uniform[c,d], and X and Y are independent.

Independent Normal RV's

- If X = $\sigma Z + \mu$ with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, then $Z \sim N(0,1) \iff X \sim N(\mu, \sigma^2)$
- Let $X_1, ..., X_n$ be independent with $X_i \sim N(\mu_i, \sigma_i^2), i = 1, ..., n.$ Then $S_n \triangleq X_1 + \cdots + X_n \sim N(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2)$

Covariance and Correlation

Def's and Properties

- $Cov(X,Y) = E[(X - \mu_X)(Y - \mu_y)] = E(XY) - E(X)E(Y)$
- $Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)} = \rho$
- Properties
  - $Cov(X,X) = Var(X)$ and $Cov(X,Y) = Cov(Y,X)$
  - $Cov(I_A, I_B) = P(A \cap B) - P(A)P(B)$
  - $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
  - X and Y are independent ⇒ Corr(X, Y) = 0; but not vice versa
  - (bilinearity) for constants $\{a_i\}, \{b_j\}$ and RV's $\{X_i\}, \{Y_j\}$,
    - $Cov(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j Cov(X_i, Y_j)$

Multinomial Distribution

Let $X_1, ..., X_n$ be iid with $P(X_1 = m) = p_m, m = 1, ..., M.$ Denote the frequency of category m in n trials

$$N_m = I_{\{X_1 = m\}} + ... + I_{\{X_n = m\}}$$

Then

$$P(N_1 = n_1, ..., N_M = n_M) = \frac{n!}{n_1!...n_M!} p_1^{n_1} ... p_M^{n_M}$$

where $n = n_1 + ... + n_M$

- $N_1 + \cdots + N_M = n$ and $p_1 + \cdots + p_M = 1$ Hence, free params are n and p_m, $m = 1, \cdots, M-1$
- Bin(n, p) is a special case of multinomial distribution with $M = 2, p_1 = p, and p_2 = q = 1 - p$
- $N_m \sim Bin(n, p_m); N_k + N_m \sim Bin(n, p_k + p_m) and (N_k, N_m, n - N_k - N_m)$ follows a multinomial distribution with parameters $\{n; p_k, p_m, 1 - p_k - p_m\}$

Hypergeometric Distribution

A box contains B black balls and R red balls. Sample n balls without replacement. Write $N + B + R$ Let X denote the number of black balls in the sample, which follows a hypergeometric distribution denoted by $X \sim HG(n; N, B)$ with

$$P(X = b) = \frac{\binom{B}{b}\binom{R}{r}}{\binom{N}{n}}, wherer = n - b$$

In calculating E(X) and Var(X), we write $X = I_{A_i} + \cdots + I_{A_n}$ where A_i = {a black ball appears in the ith draw}
$E(X) = \sum_{i=1}^n P(A_i) = np$ where $p = \frac{B}{N}, q = \frac{R}{N}$

$$Var(X) = \frac{N-n}{N-1} npq$$

Conditional Dist and Expectation: Discrete Case

Basics

Denote the sets of atoms for discrete RV's X and Y by D_X and D_Y respectively. Given $x \in D_X$, the collection $\{P(Y = y | X = x), y \in D_Y\}$ is called the conditional distribution of Y given $X = x$

Definition

Assume (X, Y) follows a discrete joint distribution. Then for $x \in D_X$,

$$E(Y|X = x) = \sum_y y P(Y = y | X = x)$$

For a known function g, we have the plug-in formula:

$$E[g(X,Y)|X = x] = \sum_y g(x,y) P(Y = y | X = x).$$

Averaging

$$E(Y) = E[E(Y|X)] = \sum_x E(Y|X = x) P(X = x)$$

Conditional Density and Expectation

Suppose (X,Y) has a join density $f_{(X,Y)}(x,y)$. Given X = x, how should we compute the conditional probability $P((X,Y) \in A | X = x)$?This cannot be solved by the usual conditional probability formula because $P(X = x) = 0, \forall x$. So we define a conditional density by

$$f_Y(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$$

then

$$P(A|X = x) = \int_A f_Y(y|x) dy$$

Examples

True/False

- If $X \sim Exp(\frac{2}{\lambda})$, then $Var(-2X) = \lambda^2$

$$Var(-2X) = (-2)^2 Var(X) = 4(\frac{1}{(\frac{2}{\lambda})^2}) = \frac{4\lambda^2}{4} = \lambda^2$$

- If $X \sim Uniform[-1, 1]$, then $Var(-3X - 1) = \frac{18}{12}$

$$Var(-3X - 1) = (-3)^2 Var(X) + Var(1) = 9\frac{(1 - (-1))^2}{12} = \frac{36}{12} = 3$$

- If $X \sim Poisson(2)$ and $Y \sim Poisson(1)$, then $X + Y \sim Poisson(3) \cdots$ False, X and Y need to be independent for $X + Y \sim Poisson(3)$
- If X and Y are iid RV's with mean 0, then E(XY) = 0 $\cdots$ True
- A box contains 3 red balls and 2 white balls. Sample 2 balls **without replacement** from the box. Then P(1st ball is red and 2nd ball is white) = P(1st ball is white and 2 ball is red) $\cdots$ True
- If $X \sim N(0, \sigma^2)$, then $E(X^3) = 0 \cdots$ True
- $I_{A \cap B} = I_A I_B$ always holds. $\cdots$ True
- A box contains 4 red balls (R), 3 white balls (W), and 2 black balls (B). Sample 2 balls from the box **with replacement**. Let X and Y denote the numbers of R's and W's in the sample respectively. Then $Cov(X,Y) = \frac{-8}{27} \cdots$ True
- Let X + 2Y = 5. Then $Corr(X,Y) = -1/2 \cdots$ From properties of covariance, if X = -2Y + 5, then Corr(X,Y) = -1

Poisson RV's

Suppose X and Y are iid Poisson(1) random variables.

- For a constant c, calculate mean and variance of $X - cY$

$$E(X - cY) = E(X) - cE(Y) = 1 - c$$

$$Var(X - cY) = Var(X) + c^2 Var(Y) = 1 + c^2$$

- Find all values of c such that $X - cY$ follows a Poisson distribution $\cdots$ Set $1 - c = 1 + c^2$

$$c^2 + c = c(c+1) = 0; c = 0 or c = -1$$

Playoff

Suppose team A and team B meet in playoffs and play against each other in a 7-game series, whichever team withs 4 games 1st moves to next round. Assume outcomes in each game are independent. Winning probability for A is $\frac{3}{5}$, for B is $\frac{2}{5}$

- Probability that A sweeps B in 4 games is equal to: $\cdots (\frac{3}{5})^4$
- Let N be total number of games needed to finish the series and I_A be the indicator for A to win the series. Are N and I_A independent? $\cdots$ For the two to be independent, $P(N = 4 | I_A = 1) = P(N = 4 | I_A = 0)$ which only holds when A and B have same probability to win each game. Therefore, not independent.

High School

A high school has 1000 students: the numbers of students from 9th to 12th grade are 400, 300, 200, and 100 respectively. Randomly sample 20 students with replacement. Let N_k denote the number of kth graders in the sample, where k=9, 10, 11, 12.

- The probability distribution of N_{11} is $\cdots$ Bin(20, 200/1000) = Bin(20, 0.2)
- The probability distribution of $N_9 + N_{10}$ is $\cdots$ Bin(20, 700/1000) = Bin(20, 0.7)
- Correlation between N_9 and N_{12} is $\cdots \frac{-2}{3\sqrt{6}}$
- Covariance between $N_9 + N_{10}$ and $N_{11} + N_{12} \cdots$ If we let $X = N_9 + N_{10}$ and $Y = N_{11} + N_{12}$. Then, because X + Y = 20, the Corr(X,Y) = -1 and Var(X) = Var(Y). Therefore

$$Cov(X,Y) = -\sqrt{VarX * VarY} = -Var(X) = -20(\frac{7}{10})(\frac{3}{10})$$

Poisson Process

Suppose number of phone calls per hour arriving at an answering service follows a Poisson process with the rate $\lambda = 3$ or (interarrival times are iid exponential RV's with mean 20 minutes)

- Given that 6 calls arrive in the first two hours, what is the conditional probability that 4 calls arrive in the second hour? $\cdots \binom{6}{2} \frac{1}{2^6}$
- Let $T(i,j)$ denote the time interval from the ith arrival to the jth arrival. Then the correlation between T(1,3) and T(2,4) is $\cdots \frac{1}{2}$
- Then the correlation beteen T(0,2) and T(0,4) is $\cdots Cov(T(0,2), T(0,4)) = Var[T(0,2)] = \frac{2}{\lambda^2} \cdots Corr(T(0,2), T(0,4)) = \frac{\frac{2}{\lambda^2}}{\sqrt{\frac{2}{\lambda^2} \frac{4}{\lambda^2}}} = \frac{1}{\sqrt{2}} = \sqrt{\frac{1}{2}}$