DSCI 551 : Stats & Prob

Lecture 1

$1 = P(A) + P(A^c)$ — complement

- Inclusion - exclusion principle

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

for 3 events...

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

- Independent Events

$P(A \cup B) = P(A) \cdot P(B)$

for cont. $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

- Entropy

$H(Y) = -\sum_x P(x) \ln[P(x)]$ — log = $\log_e$

- Mean

$E(X) = \sum_x x \cdot P(X=x)$ if cont. $E(X) = \int_x x f_X(x) dx$

- Variance

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

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

Lecture 2

Linearity of Expectations — X, Y any random var

$E(aX) = a E(X)$

$E(X+Y) = E(X) + E(Y)$

$E(aX + bY) = a E(X) + b E(Y)$

$E(XY) \neq E(X) \cdot E(Y)$

Linearity of Variance

X, Y are independent random var.

$Var(aX) = a^2 Var(X)$

$Var(X \pm Y) = Var(X) + Var(Y)$ $\pm \underbrace{2 Cov(X,Y)}_{=0 \text{ if ind.}}$ — for dependent

$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$

RANDOM VARIABLES

1. Bernoulli, $X \sim Ber(p)$

- pmf $\begin{cases} \text{success} : p(1) = p \\ \text{fail} : p(0) = 1-p \end{cases}$
- cdf : $F_X(a) = P(x \le a)$
- expect : $E(X) = p$
  $Var(X) = p(1-p)$

2. Binomial, $X \sim Bin(n,p)$
   - # of successes after n^{th} trials (n independent bernoulli trials)

- pmf : $p(i) = p(X=i) = \binom{n}{i} p^i (1-p)^{n-i}$
  $\sum_{i=0}^{n} p(i) =$
- expect : $E(X) = np$
- Var : $Var(X) = np(1-p)$

independent prop:
$\left.\begin{matrix} X \sim Bin(n,p) \\ Y \sim Bin(m,p) \end{matrix}\right\} X+Y \sim Bin(n+m, p)$

3. Geometric, $X \sim Geom(p)$
   - # of fails, until first success

- pmf : $p(x) = P(x \text{ fails, then success})$
  $= (1-p)^x \cdot p$ for $x = 0, 1, \ldots$
- expect : $E(X) = \frac{1-p}{p}$
  $Var(X) = \frac{1-p}{p^2}$

4. Pascal (Negative Binomial), $X \sim$ Negative Bin (k,p)
   - # losses before experiencing k wins

- pmf : $P(X) = P(X=x \mid k,p) = \binom{k-1+x}{x} p^k (1-p)^x$ for $x = 0, 1, \ldots$
- expect : $E(X) = \frac{k(1-p)}{p}$
  $Var(X) = \frac{k(1-p)}{p^2}$

5. Poisson, $X \sim Pois(\lambda)$ — rate of occurrence
   - Ave # of successes in n^{th} trials / fixed interval
   - approx to binom (when n large, p small)

$\lambda = np$

- pmf : $p(i) = P(X=i) = \frac{\lambda^i}{i!} e^{-\lambda}$
- expect : $E(X) = \lambda$
- Var : $Var(X) = \lambda$
- Ind prop

$\left.\begin{matrix} X \sim Pois(\lambda_1) \\ Y \sim Pois(\lambda_2) \end{matrix}\right\} X+Y = Pois(\lambda_1 + \lambda_2)$

Lecture 3

Marginal dist : In a system w/ > 1 random var, the dist of standalone is marginal dist.

i.e.

X\Y	1	2
1	0.2	0.1
2	0.4	0.3

=>

X	P
1	0.3
2	0.7

OR

Y	P
1	0.6
2	0.4

Covariance (linear)

$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$

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

Pearson's Correlation (linear)

$Corr(X,Y) = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)\left(\frac{Y - \mu_Y}{\sigma_Y}\right)\right]$

$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$ $= [0, 1]$

Kendall's τ_K (non-linear)

$(x_i, y_i), (x_j, y_j)$ where $i \neq j$

Concordant	discordant
both $x_i < x_j$ and $y_i < y_j$ and vice versa	$x_i < x_j$ and $y_i > x_j$ & vice versa

Lecture 4

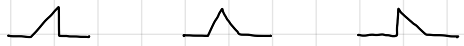
prob A given B : $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$

Lecture 5 : Continous distributions

- pdf: $f_X(x)$, cond: $f(x) \ge 0 \quad \forall x$, Note: $P(X=a)=0$
- Mode: $\text{Mode}(X) = \arg\max_x f_X(x)$
- ~~Entropy~~ $\cancel{H(x) = -\int_x f_X(x) \log[f_X(x)]\, dx}$
- cdf: $F(a) = P[X \le a] = \int_{-\infty}^{a} f(x)\, dx$ — probability
- mean: $E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$

Law of unconcious statistician
cts: $E(g(X)) = \int_{-\infty} g(x) f(x)\, dx$
dis: $E(g(X)) = \sum g(x) P(x)$

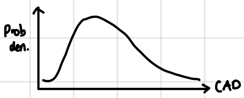
- Variance: $\text{Var}(X) = E[(X-\mu_x)^2] = \int_x (x-\mu_x)^2 f_x(x)\, dx$
  $\text{Var}(X) = E(X^2) - [E(X)]^2$
- p-quantile / $Q(p)$: prob p getting a smaller outcome
  $P[X \le Q(p)] = p$
- median: 0.5 quantile / 50-50 to see greater/less value
  $P[X \le M(X)] = 0.5$
- prediction interval: $p \times 100\%$ prediction interval.
  $\frac{1-p}{2}$ of overshooting, $\frac{1-p}{2}$ of undershooting
  e.g. 90% prediction interval $\Rightarrow$ 0.05 and 0.95 quantiles
- skewness:



$$\text{Skewness}(X) = E\left[\left(\frac{X-\mu_x}{\sigma_x}\right)^3\right] = \begin{cases} \sum_x \left(\frac{x-\mu_x}{\sigma_x}\right)^3 \cdot P(X=x) & \text{discrete} \\ \int_x \left(\frac{x-\mu_x}{\sigma_x}\right)^3 \cdot f_x(x)\, dx & \text{cont.} \end{cases}$$

Other Distributions

- pdf:

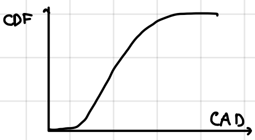


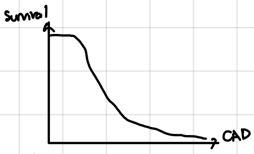
$f_X(x) = \frac{dF_X(x)}{dx}$ — cdf

- CDF (cumulative dist func) (unitless (is a prob.))
  - must never decrease
  - must be $[0,1]$
  - as $x \to -\infty$, $F_X(x) = 0$
  - as $x \to \infty$, $F_X(x) = 1$

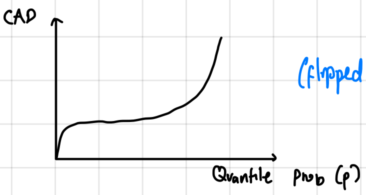
$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$ (cdf, pdf)

$P(a \le X \le b) = \int_a^b f_X(x)\, dx = F_X(b) - F_X(a)$





- Survival Function
  $S_X(x) = P(X > x) = 1 - F_X(x)$
- Quantile Function: takes prob p and maps to p-quantile
  $Q(p) = F^{-1}(p)$, only exist in $0 \le p \le 1$

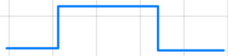


(flipped cdf)

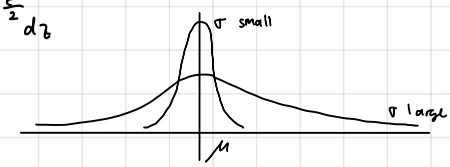
Lecture 6

CONTINOUS VARS

1. Uniform, $X \sim \text{Unif}(a,b)$
   - pdf: $f_X(x) = \frac{1}{b-a}$ for $a \le x \le b$
   - mean: $E(X) = \frac{a+b}{2}$
   - Var: $\text{Var}(X) = \frac{(b-a)^2}{12}$



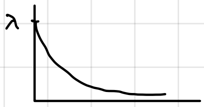
2. Gaussian/Normal, $X \sim \mathcal{N}(\mu, \sigma^2)$
   - pdf: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$ for $-\infty < x < \infty$
   - cdf: $\Phi(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}\, dz$
   - mean: $E(X) = \mu$
   - Var: $\text{Var}(X) = \sigma^2$



   - scaling: if $X \sim N(\mu, \sigma^2)$, $Y = \frac{X-\mu}{\sigma}$ then $Y \sim N(0,1)$

3. Log-Normal, $X \sim \text{Log-Normal}(\mu, \sigma^2)$
   - pdf: $f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{|\log(x) - \mu|^2}{2\sigma^2}\right\}$ for $x \ge 0$
   - mean: $E(X) = \exp\left[\mu + \frac{\sigma^2}{2}\right]$
   - Var: $\text{Var}(X) = \exp[2(\mu+\sigma^2)] - \exp(2\mu+\sigma^2)$

4. Exponential, $X \sim \exp(\lambda)$ or $\beta = \frac{1}{\lambda}$
   - pdf: $f(x) = \lambda \exp(-\lambda x)$ for $x \ge 0$, 0 otherwise
   - cdf: $F(a) = P(X \le a) = 1 - e^{-\lambda a}$
   - mean: $E(X) = \frac{1}{\lambda}$
   - Var: $\text{Var}(X) = \frac{1}{\lambda^2}$



5. Beta, $X \sim \text{Beta}(\alpha, \beta)$
   - pdf: $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ for $0 \le x \le 1$ — gamma func
   - mean: $E(X) = \frac{\alpha}{\alpha+\beta}$
   - Var: $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

6. Weibull, $X \sim \text{Weibull}(\lambda, k)$
   - pdf: $f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{x}{\lambda}\right)^k\right]$ for $x \ge 0$
   - mean: $E(X) = \lambda^{\frac{1}{k}}\, \Gamma\left(1 + \frac{1}{k}\right)$
   - Var: $\text{Var}(X) = \lambda^{\frac{2}{k}}\left[\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right]$

7. Gamma, $X \sim \text{Gamma}(k, \theta)$
   - pdf: $f(x) = \frac{1}{\Gamma(k) \cdot \theta^k} x^{k-1} \exp\left(-\frac{x}{\theta}\right)$ for $x \ge 0$
   - mean: $E(X) = k\theta$
   - Var: $\text{Var}(X) = k\theta^2$

Relevant R functions

→ prefixes:
- d : density func / $f(x)$
- p : cdf / $F(x)$
- q : quantile func / inverse cdf
- r : RNG

→ distributions:
- unif
- norm (q = , mean = , sd =) sd not var!
- lnorm
- geom
- pois
- binom

Continuous Multivariate Dist

$$f_{X|Y\geq b} = \begin{cases} \dfrac{f_X(x)}{P(Y\geq b)} & \text{for } y\geq b \\ 0 & y<b \end{cases}$$

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

Lecture 7

Bivariate Gaussian Family

- Conditions:

$$-\infty < \mu_{X/Y/\dots} < \infty$$
$$\sigma^2_{X/Y/\dots} > 0$$
$$-1 \leq \rho_{XY} \leq 1$$

- Covariance matrix, Σ

d=2

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

d=3

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{XY} & \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{XZ} & \sigma_{YZ} & \sigma_Z^2 \end{bmatrix}$$

$$\sigma_{XY} = Cov(X,Y) = \rho_{XY}\cdot\sigma_X\cdot\sigma_Y$$

- Correlation matrix, P

d=2

$$P = \begin{bmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{bmatrix}$$

d=3

$$P = \begin{bmatrix} 1 & \rho_{XY} & \rho_{XZ} \\ \rho_{XY} & 1 & \rho_{YZ} \\ \rho_{XZ} & \rho_{YZ} & 1 \end{bmatrix}$$

where $$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\,\sigma_Y^2}}$$

Gaussian dist

$$(Y|X=x) \sim \mathcal{N}\left(\mu = \mu_Y + \frac{\sigma_Y}{\sigma_X}\rho_{XY}(x-\mu_X),\ \sigma^2 = (1-\rho_{XY}^2)\sigma_Y^2\right)$$

Linear combination of Gaussian

e.g. $Z = aX + bY + c$ where $X,Y \sim Norm(\mu_{X,Y}, \sigma^2_{X,Y})$

$$E(Z) = aE(X) + bE(Y) + c$$
$$= a\mu_X + b\mu_Y + c$$

$$Var(Z) = Var(aX) + Var(bY) + 0 + Cov(aX,bY)$$
$$= a^2Var(X) + b^2Var(Y) + 0 + 2ab\,Cov(X,Y)$$
$$= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$$