

\*\*\*\* Basic Probability \*\*\*\*

Events A,B subset Omega

A : Omega -> [0,1]

Conditional Probability

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

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

Total Probability

Partition Omega = B<sub>1</sub> cup B<sub>2</sub> cup .. cup B<sub>n</sub>

B<sub>i</sub> cap B<sub>j</sub> = empty i != j

$$P(A) = P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + \dots + P(A | B_n)P(B_n)$$

Independence: A,B independent iff

$$P(A | B) = P(A)$$

$$P(B | A) = P(B)$$

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

Bayes Rule:

$$P(B | A) = P(A | B) * P(B) / P(A)$$

explain "model"=B best using "data"=A

\*\*\* Useful tools

samples space diagrams

[ ]

[ [A] ] = Omega

Tree Diagram:

C = {H,T} flip coin,

D = {L,G} roll die: L <= 2, G >= 2.

COIN	DIE	Probability
(1/2) / [H]	(1/3) / [L]	P(H cap L) = (1/2)*(1/3) = 1/6
	(2/3) \ [G]	P(H cap G) = (1/2)*(2/3) = 1/3
*	(1/3) / [L]	P(T cap L) = (1/2)*(1/3) = 1/6
(1/2) \ [T]	(2/3) \ [G]	P(T cap G) = (1/2)*(2/3) = 1/3
	conditional	total probability = 1

\*\*\*\* Random Variables \*\*\*\*

Random Variable X

function  $X : \Omega \rightarrow \mathbb{R}$

$\Omega$

discrete (rolls of dice, flip of coin)

e.g.  $P(X = \text{heads})$

continuous (rain fall, time at store)

e.g.  $P(\text{Rain} < 1 \text{ inch})$

probability density function (pdf)

$p(a) = P(X = a)$

cumulative density function (cdf)

$F(a) = P(X \leq a)$

$f(x)$  is pdf

$P(a \leq X \leq b) = \int_{x=a}^b f(x) dx$

Expectation X, pdf =  $f(x)$  ( discrete =  $P(X = x)$ )

$E[g(X)] = \sum_{i=1} g(a_i) f(a_i)$

$= \int_x g(x) f(x) dx$

Variance X

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

linearity

$E[aX + bY + c] = aE[X] + bE[Y] + c$

$\text{Var}[aX+b] = a^2 * \text{Var}[X]$

$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2*\text{Cov}(X,Y)$

\*\*\*\* Distributions \*\*\*\*

(Discrete)

Bernoulli  $X \sim \text{Ber}(p)$

$$X = \{1 \text{ w.p. } p, 0 \text{ w.p. } 1-p\}$$
$$E[X] = p \quad \text{Var}[X] = p(1-p)$$

Binomial  $X \sim \text{Bin}(n,p)$

$$X = \sum_{i=1}^n X_i \text{ s.t. } X_i \sim \text{Ber}(p)$$
$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$E[X] = np \quad \text{Var}[X] = np(1-p)$$

Geometric  $X \sim \text{Geo}(p)$

$$X = \# \text{ trials until "success" on Ber}(p)$$
$$P(X=k) = (1-p)^{k-1} p$$
$$E[X] = 1/p \quad \text{Var}[X] = (1-p)/p^2$$

(Continuous)

Uniform  $X \sim \text{Unif}(A,B)$

$$X = \{1/(B-A) \text{ if } x \text{ in } [A,B], 0 \text{ o.w.}\}$$
$$E[X] = (B-A)/2 \quad \text{Var}[X] = (1/12)(B-A)^2$$

Exponential  $X \sim \text{Exp}(l)$

$$f(x) = l \exp(-l*x), F(a) = 1 - \exp(-l*a)$$

arrival time until first event,  $l = \text{rate}$

$$E[X] = 1/l \quad \text{Var}[X] = 1/l^2$$

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

$$f(x) = (1/\sqrt{2 \pi} \sigma) * \exp(-(x-\mu)^2 / 2 \sigma^2)$$
$$F(a) = \int_{-\infty}^a f(x) dx$$
$$E[X] = \mu \quad \text{Var}[X] = \sigma^2$$

\*\*\*\* Joint Probability \*\*\*\*  
 $P(X=a, Y=b) = P(\{X=a\} \cap \{Y=b\})$   
 $p_{\{X,Y\}}(a,b) \geq 0$   
 $\sum_i \sum_j p_{\{X,Y\}}(a_i, b_j) = 1$

	A		
	0	1	
	-----		
0	0.3	0.2	0.5
1	0.1	0.4	0.5
	-----		
	0.4	0.6	

TABLE  
 $P(A=1, B=0) = 0.2$   
 $P(A=0) = 0.4 = P(A=0, B=1) + P(A=0, B=0)$  ← marginal  
 $P(A=1 | B=1) = 0.8 = P(B=1, A=1) / P(B=1)$  ← conditional

Independence:  $p_{\{X,Y\}}(a,b) = p_X(a) * p_Y(b)$  \*\*for all\*\* a,b

\*\*\*\* Continuous Joint Probability \*\*\*\*

joint pdf  $f(x,y)$  for RV X,Y  
 $P(a \leq X \leq b, c \leq Y \leq d) = \int_{x=a}^b \int_{y=c}^d f(x,y) dy dx$   
 $f(x,y) \geq 0$   
 $\int_x \int_y f(x,y) dx dy = 1$

Marginal:

$f_X(x) = \int_{y=-\infty}^{\infty} f(x,y) dy$   
 $f_Y(y) = \int_{x=-\infty}^{\infty} f(x,y) dx$

Conditional:

$f(x | Y=y) = f(x,y) / f_Y(y)$  (or 0 if  $f_Y(y) = 0$ )

Independence:  $f(x,y) = f_X(x) * f_Y(y)$

Expectation

$E[g(X,Y)] = \sum_i \sum_j g(a_i, b_j) P(X=a_i, Y=b_j)$   
 $\int_x \int_y g(x,y) f(x,y) dx dy$   
 $E[X | Y=y] \dots \text{let } g(X) = \{X | Y=y\} = f(x | Y=y)$

Covariance

$\text{Cov}(X,Y) = E[(X-E[X]) * (Y-E[Y])] = E[XY] - E[X] * E[Y]$   
 $\text{Cov}(X,X) = \text{Var}[X]$

X,Y independent →  $\text{Cov}(X,Y) = 0$

$\text{Cov}(X,Y) \neq 0 \rightarrow$  X,Y independent (maybe, maybe not)

$\text{Cov}(X,Y) > 0 \Rightarrow$  X increases, Y also increases

$\text{Cov}(X,Y) < 0 \Rightarrow$  X increases, Y decreases

correlation  $\rho(X,Y)$  in  $[-1,1]$

$= \text{Cov}(X,Y) / \sqrt{\text{Var}(X) * \text{Var}(Y)}$

\*\*\*\* Estimation \*\*\*\*

Goal: estimate parameter of distribution

$\mu$  or  $\sigma$  in  $N(\mu, \sigma)$

$p$  in  $Ber(p)$

$p$  or  $n$  in  $Bin(n, p)$

$\lambda$  in  $Exp(\lambda)$

$X_1, X_2, \dots, X_n \sim D(\theta)$  (iid)

$\theta$  is generic parameter

iid: independent identically distributed

$x = (X_1, X_2, \dots, X_n)$       $\leftarrow$  for R commands

$\hat{\theta} = T(X_1, X_2, \dots, X_n)$

estimator

$T(\cdot)$  is an algorithm on data

Bias

$\text{bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$

unbiased  $\iff$  (bias = 0)   or   ( $E[\hat{\theta}] = \theta$ )

sample mean

$\bar{X}_n = (1/n) \sum_{i=1}^n X_i$

$E[\bar{X}_n] = E[X_i] = \mu$     $\leftarrow$  mean of distribution

$\text{mean}(x)$       $\leftarrow$  in R

sample variance

$S_n^2 = (1/(n-1)) \sum_{i=1}^n (X_i - \bar{X}_n)^2$

$E[S_n^2] = \text{Var}[X_i] = \sigma^2$

$\text{sd}(x)$       $\leftarrow$  in R

\*\*\*\* Confidence Intervals \*\*\*\*  
100(1-alpha)% confidence interval  
 $P(L < \theta < R) = 1 - \alpha$   
L,R is based on data, theta is unknown

$Z = (\bar{X}_n - \mu) / (\sigma / \sqrt{n})$   
if  $X_i \sim N(\mu, \sigma^2)$   
then  $Z \sim N(0, 1)$   
 $\Delta_{\alpha} = z_{\alpha/2} \sigma / \sqrt{n}$   
 $L = \bar{X}_n - \Delta_{\alpha}$   
 $R = \bar{X}_n + \Delta_{\alpha}$   
 $z_{\alpha/2} = \text{qnorm}(1 - \alpha/2)$       <- in R  
 $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$

$X_i \sim \text{Ber}(p)$   
 $\Delta_{\alpha} \leq z_{\alpha/2} \sqrt{p(1-p)} / \sqrt{n}$   
gives upper bound  $p(1-p) \leq 1/4$  on sigma

If sigma unknown -> use t-distribution  
 $T = (\bar{X}_n - \mu) / (S_n / \sqrt{n})$   
 $T \sim t\text{-dist}(df=n-1)$

$t_{\alpha/2} = \text{qt}(1 - \alpha/2, df=n-1)$  <- in R  
 $P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$   
 $\Delta_{\alpha} = t_{\alpha/2} S_n / \sqrt{n}$   
 $L = \bar{X}_n - \Delta_{\alpha}$   
 $R = \bar{X}_n + \Delta_{\alpha}$

\*\*\*\* Hypothesis Testing \*\*\*\*

Null Hypothesis :  $H_0$  | guess where data ( $X_1, X_2, \dots$ ) comes from

Alternative Hypothesis :  $H_1$  guess of how  $H_0$  is broken

e.g.  $H_0: \mu = 10$  ,  $H_1: \mu > 10$

convert to t-distribution

$T = (\bar{X}_n - \mu) / (S_n / \sqrt{n})$

$t \leftarrow$  realization of actual data

critical value at alpha

$H_1: \mu > 0$

$t_{\alpha} = qt(1-\alpha, df=n-1)$   $\leftarrow$  in R

$P(T < t_{\alpha}) = 1-\alpha$

$H_1: \mu < 0$

$t_{\alpha} = qt(\alpha, df=n-1)$

$P(T < t_{\alpha}) = \alpha$   $\leftarrow$  in R

(not based on data, but then compare  $t$  (from data) vs.  $t_{\alpha}$ )

p-value

$H_1: \mu > 0$

$p = 1-pt(t, df = n-1)$

$P(T < t) = 1-p$

$H_1: \mu < 0$

$p = pt(t, df = n-1)$

$P(T < t) = p$

(based on data, but then compare versus some alpha threshold probability)