**** 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_1 cup B_2 cup .. cup B_n
      B_i cap B_j = empty i != j
  P(A) = P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + .. + 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
               (1/3) / [L]   P(H cap L) = (1/2)*(1/3) = 1/6
  (1/2) / [H]
               (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 -> R

Omega
  discrete (rolls of dice, flip of coin)
   e.g.  P(X = heads)
  continuous (rain fall, time at store)
   e.g.  P(Rain < 1 inch)


probability density function  (pdf)
  p(a) = P(X = a)
cumulative density function (cdf)
  F(a) = P(X <= a)

f(x) is pdf
P(a <= X <= 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
  Var[X] =  E[(X-E[X])^2] = E[X^2] - E[X]^2

linearity
  E[aX + bY+ c] = a*E[X] + b*E[Y] + c
  Var[aX+b] = a^2 * Var[X]
  Var[X+Y] = Var[X] + Var[Y] + 2*Cov(X,Y)
```

```
  **** Distributions  ****
(Discrete)
 Bernoulli X~Ber(p)          X = {1 wp p,  0 wp 1-p}
                             E[X] = p    Var[X] = p(1-p)

 Binomial  X~Bin(n,p)        X = sum_{i=1}^n X_i s.t. X_i~Ber(p)
                             P(X=k) = (n choose k) p^k (1-p)^{n-k}
                             E[X] = np   Var[X] = np(1-p)

 Geometric X~Geo(p)         X = # trials until “success” on Ber(p)
                            P(X=k) = (1-p)^{k-1} p
                            E[X] = 1/p  Var[X] = (1-p)/p^2



(Continuous)
 Uniform  X~Unif(A,B)       X = {1/(B-A) if x in [A,B], 0 o.w.}
                            E[X] = (B-A)/2   Var[X] = (1/12)(B-A)^2

 Exponential X~Exp(l)       f(x) = l exp(-l*x), F(a) = 1-exp(-l*a)
                            arrival time until first event, l = rate
                            E[X] = 1/l   Var[X] = 1/l^2

 Normal X~N(mu,sigma^2)     f(x) = (1/sqrt{2 pi} sigma) *
                                      exp(-(x-mu)^2 / 2 sigma^2)
                            F(a) = int_{x=-infty}^a f(x) dx
                            E[X] = mu   Var[X] = sigma^2
```

```
  **** Joint Probability ****
P(X=a,Y=b) = P({X=a} cup {Y=b})
p_{X,Y}(a,b) >= 0
sum_i sum_j p_{X,Y}(a_i,b_j) = 1

       A
      0      1
   ------------
 0 | 0.3 | 0.2 | 0.5      TABLE
B   ------------
 1 | 0.1 | 0.4 | 0.5
    -----------
    0.4    0.6
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 <= X <= b,  c <= Y <= d)  = int_{x=a}^b int_{y=c}^d f(x,y) dy dx
  f(x,y) >= 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_i) P(X=a_i, Y=b_j)
             int_x int_y g(x,y) f(x,y) dx dy
 E[X | Y=y] .. let g(X) = {X | Y=y} = f(x | Y=y)

Covariance
 Cov(X,Y) = E[(X-E[X])*(Y-E[Y])] = E[XY] - E[X]*E[Y]
 Cov(X,X) = Var[X]

 X,Y independent -> Cov(X,Y) = 0
 Cov(X,Y)  -/not/->  X,Y independent (maybe, maybe not)

 Cov(X,Y) > 0 == X increases, Y also increases
 Cov(X,Y) < 0 == X increases, Y decreases

 correlation rho(X,Y) in [-1,1]
  = Cov(X,Y) / sqrt(Var(X)*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)
  l in Exp(l)


 X_1, X_2, .., X_n ~ D(theta)  (iid)
   theta is generic parameter
   iid: independent identically distributed
   x = (X_1, X_2, .., X_n)     <- for R commands

  hat{theta} = T(X_1, X_2, .., X_n)
   estimator
   T(..) is an algorithm on data


Bias
 bias(hat{theta}) = E[hat{theta}] - theta
 unbiased == (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   <- mean of distribution
 mean(x)       <- in R

sample variance
S_n^2 = (1/(n-1)) sum_{i=1}^n (X_i - bar{X_n})^2
 E[S_n^2] = Var[X_i] = sigma^2
 var(x)        <- in R


 *** Central Limit Theorem ***
Random Sample X_1, ..., X_n ~iid f(theta)
  iid :=  independent and identically distributed
  bar{X}_n = (1/n) sum_i X_i
  (1) E[bar{X}_n] = E[X_i] = mu
  (2) Var[bar{X}_n] = Var[X_i]/n = sigma^2/n
  (3) (\bar{X}_n - mu)/(sigma/sqrt(n)) ~[lim n-> infty]~ N(0,1)
```

```
 **** Confidence Intervals ****
 100(1-alpha)% confidence interval
 P(L < theta < R) = 1-alpha
  L,R are R.V. base on sample, theta is unknown true value

Z = (bar{X_n} - mu) / (sigma/sqrt{n})
  if X_i ~ N(mu,sigma^2)
  then Z ~ 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} = qnorm(1-alpha/2)       <- in R
    P(-z_{alpha/2} < Z < z_{alpha/2}) = 1-alpha

X_i ~ Ber(p)
  Delta_alpha <= z_{alpha/2} sqrt{p(1-p)} / sqrt{n}
   gives upper bound p(1-p) <= 1/4 on sigma


If sigma unknown -> use t-distribution
  T = (bar{X_n} - mu) / (S_n / sqrt{n})
  T ~ t-dist(df=n-1)

  t_{alpha/2} = 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
  P(L < mu < R) = 1-alpha
```

```
 **** Hypothesis Testing ****
Null Hypothesis : H_0  |guess where data (X_1, X_2, ..) comes from
Alterantive Hypothesis : H_1  guess of how H_0 is broken
  e.g. H_0:  mu = 10, X_i~N(mu,sigma^2)   , H_1: mu > 10

convert to t-distribution
  T = (bar{X_n} - mu)/(S_n/sqrt{n})
  t <- realization of actual data


critical value at alpha
H1: mu > 0
 t_alpha = qt(1-alpha, df=n-1)   <- in R
 P(T < t_alpha) = 1-alpha
H1: mu < 0
 t_alpha = qt(alpha, df=n-1)
 P(T < t_alpha) = alpha          <- in R

(not based on data, but then compare t (from data) vs. t_alpha)

p-value
H1: mu > 0
 p = 1-pt(t, df = n-1)
 P(T < t) = 1-p
H1: mu < 0
 p = pt(t, df = n-1)
 P(T < t) = p

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