1 Chapters 1-5: Probability Theory

Bonferoni's Inequality: $P(A \cap B) \geq P(A) + P(B) - 1$

Counting:

	W/out repl.	W/ repl.
Ordered	$\frac{n!}{(n-r)!}$	n^r
Unordered	$\binom{n}{r}$	$\binom{n+r-1}{r}$

where $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

Conditional Probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$, $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
Baye's Rule: $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}$

CDF: $F_X(x) = P_X(X \leq x)$, for all x. If $F_X(x) = F_Y(y)$ then x and y are identically distributed.
$\frac{d}{dx}F_X(x) = f_X(x)$ and $P(X \leq x) = F_X(x) = \int_{-inf}^{x} f_X(t)dt$.
CDF Transformations: $Y = g(X)$, find $F_Y(y)$.

1. Take the first derivative of g(x).
2. Find $g^{-1}(y)$ (get the inverse of g(x) and plug in y). *Watche out for sqrs or discontinuity!*
   (a.) If g is increasing $F_Y(y) = F_X(g^{-1}(y))$
   (b.) If g is increasing $F_Y(y) = F_X(g^{-1}(y))$

PDF Transformations: $Y = g(X)$, find distribution (pdf) of Y. $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$
Probability integral tansform: any $Y = F_X(x) \sim$ Uniform (0,1) such that $P(Y \leq y) = y, 0 < y < 1$.

Expectations and Variance:

1. Definition:
   a. $\int_{-\infty}^{\infty} g(x)f(x)dx$
   b. $\sum_{x \in X} g(x)f(x) = \sum_{x \in X} g(x)P(X = x)$
2. Rules:
   a. $E(ag_1(X) + bg_2(X) + c) = aEg_1(X) + bEg_2(X) + c$
   b. If $g(x) \geq 0$ for all x, then $Eg(x) \geq 0$
   c. If $g_1(x) \geq g_2(x)$ for all x, then $Eg_1(X) \geq Eg_2(X)$
   d. If $a \leq g(x) \leq b$ for all x, then $a \leq Eg(x) \leq b$
   e. Var(X)=$(X - EX)^2 = EX^2 - (EX)^2$
   f. Var(aX+b)=a^2Var(X).
3. Add one, subtract one trick:
   $E(X-b)^2 = E(X - EX + EX - b)^2 = E((X - EX) + (EX - b))^2 = E(X - EX)^2 + (EX - b)^2 + 2E((X - EX)(EX - b))$
   Which reduces to: $E(X - EX)^2 + (EX - b)^2$, the variance and the bias.

Moments: n^{th} moment is $\mu'_n = EX^n$ or central moment is $\mu_n = E(X - EX)^n$

1. Moment Generating Function. The n^{th} moment is equal to the n^{th} derivative of $M_X(t)$, with respect to t, evaluated at t=0.
   (a.) $M_X(t) = \int_{-\infty}^{\infty} e^{tx} fX(x)dx$ or,
   (b.) $M_X(t) = \sum_x e^{tx} P(X = x)$ if X is discrete
   (c.) $M_{aX+b}(t) = e^{bt} M_X(at)$
2. If X and Y have bounded support then $F_X(u) = F_Y(u)$ for all u iff $EX^r = EY^r$, or, if mgf's exist and $M_X(t) = M_Y(t)$ in some neighborhood of 0 then $F_X(u) = F_Y(u)$ for all u.
3. Convergence of MGFs: convergence for $|t| < h$ of mgfs implies convergence of CDFs.

Exponential Families

$$f(x|\theta) = h(x)c(\theta)exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right)$$

Where, $h(x) and c(\theta) \geq 0$, $t_k(x)$ do not depend on θ and $w_k(\theta)$ do not depend on x.
Calculation Shortcut for moments of an exponential family.
$E\left(\sum_{i=1}^{k} \frac{\delta w_i(\theta)}{\delta \theta_j} t_i(X)\right) = -\frac{\delta}{\delta \theta_j} log(c(\theta))$; and
$Var\left(\sum_{i=1}^{k} \frac{\delta w_i(\theta)}{\delta \theta_j} t_i(X)\right) = -\frac{\delta^2}{\delta \theta_j^2} log(c(\theta)) - E\left(\sum_{i=1}^{k} \frac{\delta^2 w_i(\theta)}{\delta \theta_j^2} t_i(X)\right)$
If the natural parameter space is less than the number of terms in the exponent it is a **Curved Exponential**.

Location and Scale Families
$g(x|\mu, \sigma) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$ is a pdf if σ is a constant > 0
If $f(x)$ is a pdf then $f(x - \mu)$ is a location family with location parameter μ.
If $f(x)$ is a pdf then $(1/\sigma)f(x/\sigma)$ is a scale family with scale parameter σ.
If $f(x)$ is a pdf then $(1/\sigma)f((x - \mu)/\sigma)$ is a location-scale family with parameter (μ, σ)
Given $(1/\sigma)f((x - \mu)/\sigma)$ is pdf for the random variable X then there exists a random variable Z, with pdf f(z), such that $X = \sigma Z + \mu$. This leads to some useful properties.

a. $EX = \sigma EZ + \mu$
b. $Var(X) = \sigma^2 Var(Z)$

Inequalities and Identities:

a. Chebychev's Inequality: $P(g(X) \geq r) \leq \frac{E(g(x))}{r}$, for any non-negative function $g(x)$.
b. Stein's Lemma: If $X \sim n(\theta, \sigma^2)$, $E\left[g(x)(X - \theta)\right] = \sigma^2 E(g'(X))$ (useful for calculating higher order moments)
c. For any function $h(x)$, $E\left[h(\chi_p^2)\right] = pE\left(\frac{h(\chi_{p+2}^2)}{\chi_{p+2}^2}\right)$

d. HWANG: If g(X) has finite expectation and g(-1) is finite then:

i If $X \sim$Poisson, $E(\lambda g(X)) = E(Xg(X-1))$ (Useful for moment calcs)

ii If $X \sim$Neg. Bin(r,p), $E((1-p)g(X)) = E\left(\frac{X}{r+X-1}g(X-1)\right)$

For $\frac{1}{p}+\frac{1}{q}=1$, then $\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab$ with equality only if $a^p = b^q$.
Holders Inequality: $|EXY| \leq E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}$ Cauchy-Schwartz: $|EXY| \leq E|XY| \leq (E|X|^2)^{1/2}(E|Y|^2)^{1/2}$ (special case of above which proves $Cov(X,Y)^2 \leq \sigma_X^2\sigma_Y^2$)
Liapounov's: $(E|X|^r)^{1/r} \leq (E|X|^s)^{1/s}$ for $1 < r < s < \infty$
Minkowski's: $[E|X+Y|^p]^{1/p} \leq [E|X|^p]^{1/p} + [E|Y|^p]^{1/p}$
Application to sums: $\sum_{i=1}^n |a_i b_i| \leq (\sum_{i=1}^n a_i^p)^{1/p}(\sum_{i=1}^n b_i^q)^{1/q}$, $\frac{1}{p}+\frac{1}{q}=1$
Special Case: $\frac{1}{n}\left(\sum_{i=1}^n |a_i|\right)^2 \leq \sum_{i=1}^n a_i^2$

Functional Inequalities

g(x) is *convex* if $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$ for all x and y and $0 < \lambda < 1$.
Jensen's Inequality: IF $g(x)$ is a convex function, $Eg(x) \geq g(EX)$
:

a. A function is convex if $g(\lambda x + (1-\lambda)y)) \leq \lambda g(x) + (1-\lambda)g(y)$ for all x and y and $0 < \lambda < 1$

b. A function is concave if $-g(x)$ is convex

Covariance Inequality: if g(x) is non-decreasing and h(x) is non-increasing $E(g(X)h(x)) \leq Eg(X)Eh(X)$
If g(x), h(x) are both non-increasing or non-decreasing then, $E(g(X)h(x)) \geq Eg(X)Eh(X)$

Joint and Marginal Distributions:
Let (X,Y) be random vector:

a. $f_X(x) = \sum_y f_{X,Y}(x,y)$ and vice versa if discrete

b. $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$ and vice versa if continuous

$Eg(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$
To calculate joint probabilities you have to first determine the limits of integration (see example):
$P((X,Y) \in A) = \int_A \int f(x,y)dxdy$

Conditional Distributions and Independence

For discrete variables $f(x|y) = P(X = x|Y = y) = \frac{f(x,y)}{f_Y(y)}$ Which is equivalent to the form for continuous vars.
$E(g(Y)|x) = \int_{-\infty}^{\infty} g(y)f(y|x)dy$ (Replace int with sum over y for discrete case)
$Var(Y|x) = E(Y^2|X) - E(Y|x)^2$
IF $f(x,y) = f_X(x)f_Y(y)$ then X and Y are independent.
Lemma: X and Y are independent iff there exists funcs g(x) and h(y) such that (for all x, y) f(x,y)=g(x)h(y). I.E. if f(x,y) is not a cross-product over the defined region then they are not indpendent.
If X and Y are independent with mgfs $M_X(t)$ and $M_Y(t)$, Z=X+Y has mgf $M_Z(t) = M_X(t)M_Y(t)$
If X and Y are independent and U=g(x) is a function of only x and V=h(y) is afunction of only y, then U, V are independent.

If X and Y are two r.v.'s then, $EX = E(E(X|Y))$
$Var(X) = E(Var(X|Y)) + Var(E(X|Y))$
$Var(aX + bY) = a^2 VarX + b^2 Var(Y) + 2abCov(X, Y)$ (Note that if $\rho > 0$ then variance is higher but if $\rho < 0$ variance is lower.

Bivariate Transformations:
$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))|J|$, where h_1 and h_2 are the *inverse* of the transformations and J is the jacobian:
Bivariate jacobian is:

$$J = \begin{vmatrix} \frac{\delta x}{\delta u} & \frac{\delta x}{\delta v} \\ \frac{\delta y}{\delta u} & \frac{\delta y}{\delta v} \end{vmatrix} = \frac{\delta x}{\delta u}\frac{\delta y}{\delta v} - \frac{\delta y}{\delta u}\frac{\delta x}{\delta v}$$

where, $\frac{\delta x}{\delta u} = \frac{\delta h_1(u,v)}{\delta u}$, $\frac{\delta x}{\delta u} = \frac{\delta h_1(u,v)}{\delta v}$, $\frac{\delta y}{\delta u} = \frac{\delta h_2(u,v)}{\delta u}$, and $\frac{\delta y}{\delta v} = \frac{\delta h_2(u,v)}{\delta v}$.
If the transorfmation is only piecewise 1:1 onto, then the joint pdf is represented by:
$f_{U,V}(u, v) = \sum_{i=1}^{k} f_{X,Y}(h_{1i}(u, v), h_{2i}(u, v))|J_i|$
If X and Y are indep. r.v.'s with pdf's $f_X(x)$ and $f_Y(y)$, then the pdf of $X = X + Y$ is:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z - w) dw, \text{ with } X = W$$

$Z = X - Y$ is:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(w - z) dw, \text{ with } X = W$$

$Z = XY$ is:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z/w)|1/w| dw, \text{ with } X = W$$

$Z = X/Y$ is:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(w/z)|w/z^2| dw, \text{ with } X = W$$

Hierarchical Models and Mixture Distributions. If X and Y are two r.v.'s then, $EX = E(E(X|Y))$
A random variable X is said to have a *mixture distribution* if it's parameter(s) also has a distribution.
$Var(X) = E(Var(X|Y)) + Var(E(X|Y))$

Covariance and Correlation
$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = EXY - \mu_X \mu_Y$
$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$
The bivariate normal:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} exp\left[\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right]$$

Properties of the bivariate normal: (Note: marginal normality does not imply joint normality!)

a. The marginal dist. of X is $n(\mu_X, \sigma_X^2)$ and vice versa for Y

b. The correlation between X an Y is ρ.

c. For any constants a and b, the distribution of aX+bY is: $n(a\mu_x + b\mu_Y, a^2\sigma_X + b^2\sigma_Y + 2ab\rho\sigma_X\sigma_Y)$

Multivariate Distributions

The marginal distribution of the first k variables is:
$f(x_1, ..., x_k) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, ..., x_n)dx_{k+1}...dx_n$
$f(x_1, ..., x_k) = \sum_{(x_{k+1},...x_n)\in\Re^{n-k}} f(x_1, ..., x_n)$
$f(x_{k+1}, ..., x_n|x_1, ..., x_k) = \frac{f(x_1,...,x_n)}{f(x_1,...,x_k)}$
The multinomial distribution:

$$f(x_1, ..., x_n) = \frac{m!}{x_1! \cdot ... \cdot x_n!} p_1^{X_1} \cdot ... \cdot p_n^{x_n} = m! \prod_{i=1}^{n} \frac{p_i^{x_i}}{x_i!}$$

The Multinomial Theorem: (A is the set of vectors such that each $x_i > 0$ and $\sum_{i=1}^{m} x_i = m$)

$$(p_1 + ... + p_n)^m = \sum_{x\in A} \frac{m!}{x_1! \cdot ... \cdot x_n!} p_1^{X_1} \cdot ... \cdot p_n^{x_n}$$

Note that all of the pairwise Cov are (-): $Cov(X_i, X_j) = E[(X_i - p_i)(X_j - p_j)] = -mp_ip_j$.
If $f(x_1, ..., x_n) = f_{X_1}(x_1) \cdot ... \cdot f_{X_n}(x_n) = \prod_{i-1}^{n} f_{X_i}(x_i)$, then $X_1, ..., X_n$ are mutually indep. vectors or variables.
The following holds for mutually independent random variables or vectors.

a. $E[g_1(X_1) \cdot ... \cdot g_n(X_n)] = Eg_1(X_1) \cdot ... \cdot Eg_n(X_n)$

b. $Z = \sum_{i=1}^{n} X_i$ has mgf $M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t)$ if identically distributed then $= (M_X(t))^n$.

c. With $a_1, ...a_n$ and $b_1, ...b_n$ fixed constants and $Z = \sum_{i=1}^{n} a_iX_i+b_i$, $M_Z(t) = e^{t(\sum b_i)} \prod_{i=1}^{n} M_{X_i}(a_it)$

d. If $X_i \sim n(\mu_i, \sigma_i^2)$ then, $Z = \sum_{i=1}^{n}(a_iX_i + b_i) \sim n(\sum_{i=1}^{n}(a_i\mu_i + b_i), \sum_{i=1}^{n}(a_i^2\sigma_i^2))$

If the joint pdf or pmf of random vectors $X_1, ..., X_n$ can be factored into $\prod_{i=1}^{n} g_i(x_i)$ then they are mutually independent.
For mutually independent random vectors if $U_i = g_i(X_i)$ is a function of only X_i then all U_i are mutually independent.
Finding the distribution of a transformation of a random vector:

$$f_U(u_1, ..., u_n) = \sum_{i=1}^{k} f_X(h_{1i}(u_1, ..., u_n), ..., h_{ni}(u_1, ..., u_n))|J_i|$$

Where the Jacobian computed from the ith inverse is:

$$J_i = \begin{vmatrix} \frac{\delta x_1}{\delta u_1} & \frac{\delta x_1}{\delta u_2} & \cdots & \frac{\delta x_1}{\delta u_n} \\ \frac{\delta x_2}{\delta u_1} & \frac{\delta x_2}{\delta u_2} & \cdots & \frac{\delta x_2}{\delta u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta x_n}{\delta u_1} & \frac{\delta x_n}{\delta u_2} & \cdots & \frac{\delta x_n}{\delta u_n} \end{vmatrix} = \begin{vmatrix} \frac{\delta h_{1i}(u)}{\delta u_1} & \frac{\delta h_{1i}(u)}{\delta u_2} & \cdots & \frac{\delta h_{1i}(u)}{\delta u_n} \\ \frac{\delta h_{2i}(u)}{\delta u_1} & \frac{\delta h_{2i}(u)}{\delta u_2} & \cdots & \frac{\delta h_{2i}(u)}{\delta u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta h_{ni}(u)}{\delta u_1} & \frac{\delta h_{ni}(u)}{\delta u_2} & \cdots & \frac{\delta h_{ni}(u)}{\delta u_n} \end{vmatrix}$$

The determinant is calculated by finding minors and cofactors (matrix algebra review).

1. Find the minor of each 2 by 2 sub-matrix (the determinant of each sub-matrix)
2. The cofactor is the minor if the addition of the column and row number is even, if odd then it is -1xminor.
3. Create a matrix of cofactors, C
4. Pick a row from the original matrix, r_i, and find $r_i C_i^T$. (pick the row or column with the most 0's)
5. For upper or lower or diagonal matrices the determinant is the product of the diagonal.

Properties of a Random Sample
iid implies variables are mutually independent.
The joint pdf/pmf of iid variables is: $f(x_i, ...x_n|\theta) = \prod_{i=1}^n f(x_i|\theta)$
The probability distribution of a statistic Y is called the sampling distribution of Y.
Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, $(n-1)s^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$
$E\left(\sum_{i=1}^n g(X_i)\right) = nEg(X_1)$
$Var\left(\sum_{i=1}^n g(X_i)\right) = nVar(g(X_1))$
For any random sample with mean μ and variance σ^2

a. $E\bar{X} = \mu$
b. $Var\bar{X} = \sigma^2/n$
c. $ES^2 = \sigma^2$
d. $M_{\bar{X}}(t) = [M_X(t/n)]^n$

If X and Y are indep. then Z=X+Y has pdf: $f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)dw$ Where w is either X or Y
For an exponential family, the statistics defined as $\sum_{j=1}^n t_i(X_j)$, $i = 1, ..., k$ have an exponential distribution of the form: $f_T(u_1, ..., u_k|\theta) = H(u_1, ...u_k)[c(\theta)]^n exp\left(\sum_{i=1}^k w_i(\theta)u_i\right)$ if the set $\{(w_1(\theta), ...w_k(\theta)), \theta \in \Theta\}$ contains an open set in $\Re^k$.

Order Statistics Let $X_1, ..., X_n$ be a random sample from a discrete distribution with pmf $f_X(X_i) = p_i$. Define:
$P_0 = 0$, $P_1 = p_i$, $P_2 = p_1 + p_2, ..., P_i = \sum_{i=1}^n p_i$
$P(X_{(j)} \leq x_i) = \sum_{k=j}^n \binom{n}{k} P_i^k (1-P_i)^{n-k}$, and
$P(X_{(j)} = x_i) = \sum_{k=j}^n \binom{n}{k} \left[P_i^k(1-P_i)^{n-k} - P_{i-1}^k(1-P_{i-1})^{n-k}\right]$.
For the continuous case the pdf of $X_{(j)}$ is:

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)[F_X(x)]^{j-1}[1-F_X(x)]^{n-j}$$

The joint pdf of $X_{(i)}, X_{(j)}$ with $1 \leq i < j \leq n$ is:

$$f_{X_{(j)}}(u,v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v)[F_X(u)]^{i-1} \times [F_X(v) - F_X(u)]^{j-1-i}[1-F_X(v)]^{n-j}$$

Convergence Concepts

Convergence in Probability: $X_1, X_2, ...$ converge in prob to X if for every $\epsilon > 0$

$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$, or $\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$

Weak Law of Large Numbers:

For iid r.v.'s with $EX_i = \mu$ and $VarX_i = \sigma^2$, then for every $\epsilon > 0$, $\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$

If $X_1, X_2, ...$ converge in prob to X then $h(X_1), h(X_2), ...$ converge in prob to $h(X)$.

Almost Surely Convergence: for every $\epsilon > 0$, $P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1$

Strong Law of Large Numbers: For iid r.v.'s with $EX_i = \mu$ and $VarX_i = \sigma^2$, then for every $\epsilon > 0$,

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1, \bar{X} \text{ converges almost surely to } \mu$$

Convergence in Distribution: $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$

Convergence in probability means convergence in distribution

Convergence in probability to a constant is only possible if the sequence converges in distribution to a constant.

Central Limit Theorem:

Requirements:

a. $X_1, X_2, ...$ iid

b. MGF's exist in a neighborhood of 0 (or finite variance greater than 0)

c. $EX_i = \mu$ and $VarX_i = \sigma^2$

$$\lim_{n\to\infty} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} d^{-y^2/2} dy$$

Slutsky's Theorem: If $X_n \to X$ in dist. and $Y_n \to a$ (a constant) then, $Y_n X_n \to aX$ and $X_n + Y_n \to X + a$ in distribution.

The Delta Method

Taylor Series: The Taylor polynomial of order r about a is $T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i$

The remainder of a taylor series polynomial: $g(x) - T_r(x) = \int_a^x \frac{g^{(r+1)}(t)}{r!}(x-t)^r dt$

Given r.v.'s $T_1, ..., TK$ with means $\theta_1, ..., \theta_k$ and a differentiable function $g(\mathbf{T})$). Define:

$$g'(\theta) = \frac{\delta}{\delta t_i} g(\mathbf{t}) \mid_{t_1=\theta_1,...,t_k=\theta_k}$$

then,

$$g(\mathbf{t}) \approx g(\theta) + \sum_{i=1}^{k} g_i(\theta)(t_i - \theta_i)$$

and, $E_\theta g(\mathbf{T}) \approx g(\theta)$ and

$$Var_\theta g(\mathbf{T}) \approx \sum_{i=1}^{k} \left[g_i'(\theta)\right]^2 Var_\theta(T_i) + 2\sum_{i>j} g_i'(\theta) g_j'(\theta) Cov_\theta(T_i, T_j)$$

Delta method CLT: If Y_n satisfies CLT then $\sqrt{n}[g(Y_n) - g(\theta)] \to n(0, \sigma^2[g'(\theta)]^2)$ in distribution.
Second Order Delta Method: If $g'(\theta) = 0$ and $g''(\theta) \neq 0$ then,

$$n\left[g(Y_n) - g(\theta)\right] \to \sigma^2 \frac{g''(\theta)}{2} \chi_1^2 \text{ in distribution.}$$

Multivariate Delta Method: For $\mathbf{X_i}, ..., \mathbf{X_n}$ with $E(X_{ij}) = \mu_i$ and $Cov(X_{ik}, X_{jk}) = \sigma_{ij}$:

$$\sqrt{n}\left[g(\bar{X}_1, ..., \bar{X}_s) - g(\mu_i, ..., \mu_p)\right] \to n(0.\tau^2) \text{ in distribution}$$

where $\tau^2 = \sum\sum \sigma_{ij} \frac{\delta g(\mu)}{\delta \mu_i} \cdot \frac{\delta g(\mu)}{\delta \mu_j} > 0$.

2 Chapters 6-9: Statistical Theory

2.1 Sufficiency

Order statistics are always sufficient for θ and reduce $n!$ different values for the data to one value for T. To find Sufficiency:

1. If $\frac{p(x|\theta)}{g(T(x)I\theta)}$, where p is the pdf/pmf of X and q is the pdf/pmf of T, is free of θ then T is sufficient. Note that this may not be the easiest method.
2. Neyman-Fisher Factorization theorem: If $f(X|\theta)$ can be factorized to $f(x|\theta) = g(T(x)|\theta)h(x)$ then $T(x)$ is sufficient.
3. For Exponential Families: $f(x|\theta) = h(x)c(\theta)exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right), T(\mathbf{X}) = (\sum t_1(x_i), ..., \sum t_k(x_i))$ is sufficient.

If T is sufficient and $T = c(U)$ where U is some other statistic then U is also sufficient.
If T is sufficient and $U = G(T)$ with G being 1:1, then U is also sufficient.
Minimal Sufficiency: (If you can show minimal sufficiency than you have also shown sufficiency)

1. Lehman-Scheffe: If $f(x|\theta)/f(y|\theta)$ does not depend on theta iff T(x)=T(y).
2. Exponential Families: If full rank then T(X) as defined above is always minimal sufficient. For curved exponential families the $w_j(\theta)$'s must be linearly independent.

Minimal sufficient statistics are not unique.
Ancillary Statistic: Distribution of T does not depend on θ.
For location-scale families, find $Z \sim f(z)$ such that $X = \mu + \sigma Z$. Substitute Z into statistic to cancel out θ
For Location Families:

1. If T is location invariant then T is ancillary
2. Sample sd (S) is ancillary, as are other estimates of scale.

For Location-Scale Families:

1. If T is a location-scale invariant statistic, then T is ancillary
2. If T_1, T_2 are such that both $T_{1or2}(ax_1 + b, ..., ax_n + b) = aT_{1or2}$, then T_1/T_2 is ancillary.

Complete Statistics
To find complete sufficient statistic:

1. Find a minimum sufficient statistic
2. Show it is complete (see notes pp26-34)
3. If it is not complete then there is no complete statitic for the family
4. If T is complete and sufficient, then T is also minimal sufficient.
5. If T is complete, then only 1 unbiased estimator based on T is possible
6. If T is complete and suff. then $U = h(T)$ is the UMVUE of its expectation.
7. A statistic is called necessary if for every suff. stat. T , $S = g(T)$

Non-zero constant statistics are complete.
Non-trival (non-constant) ancillary statistic cannot be complete.
A *first order ancillary* statistic is defined to have its expectation free of θ. If a non-trivial function of statistic T is first order ancillary, then T cannot be complete.
Basu's Theorem: A complete suff. stat. is independent of all ancillary statistics.

2.2 Point Estimation

see notes p(45-62)
Method of moments: Easy to compute, reasonable starting estimate, generally consistent (since sample moments are consistent with pop moments.). Not necessarily the best or most efficient, estimates may fall out of the range (work better with large n).
Maximum Likelihood estimators:

a. Find the log-likelihood equation
b. Take the first derivative to find the extreme points
c. Take the second derviative fo the likelihood equation evaluated at $\theta = \hat{\theta}$. If < 0 then max.
d. Remember to compare likelihood values with those at the boundaries of Θ.
e. If $\hat{\theta}$ is a unique solution from the $L(\theta|X)$ then it is an MLE.
f. If non-differentiable check for monotone increasing or decreasing and use boundary.

Two methods for finding the MLE in the two parameter case:

1. Simultaneously maximize and check the negative definiteness of the Hessian matrix (p 53)

2. Profile method, 2-stage maximization. Fix 1 parameter and find the max. to get estimate for non-fixed parameter. Substitute estimate into likelihood and maximize for the other parameter.

Bayes Estimators:
Prior: $\theta \sim \pi(\theta)$ (plug parameter into pdf)
Sampling dist. of $x|\theta$: $x|\theta \sim f(x|\theta)$
Posterior Dist. of θ: $\pi(\theta|X) = f(x|\theta)\pi(\theta)/m(x)$
Marginial dist. of x: $m(x) = \int f(x|\theta)\pi(\theta)d\theta$ (integrate out theta to get function of x)
Posterior mean of θ: $E(\theta|X) = \int \theta\pi(\theta|x)d\theta$ (Bayes Estimator of θ)

2.3 Evaluating Estimators

$MSE(\hat{\theta}) = E_\theta(\hat{\theta} - \theta)^2 = Var(\hat{\theta}) + (Bias(\hat{\theta}))^2$
Score Function (not a statistic since a function of x and θ):

$$s(X,\theta) = \frac{\delta}{\delta\theta}\log f(X|\theta) = \frac{1}{f(X|\theta)}\frac{\delta f(X|\theta)}{\delta\theta}, \text{ for } X_1,...,X_n,\ s_n(\mathbf{X},\theta) = \sum_{i=1}^{n} s(X_i,\theta)$$

Fisher Information: The second order moment of the score function- $E[s(X,\theta)^2]$ (with respect to X given θ). See P.71 for a table of information numbers for comon exponential families.
If the derivative with respect to θ can be done under the integral, $\int_{\mathcal{X}} f(x|\theta)dx = 1$, then $E[s(X,\theta)] = 0$ and we just have the first term in the variance formula.
Fisher information number: $\mathcal{I}(\theta) = Var[s(X,\theta)]$, is the information number that X contains about θ.

$$\mathcal{I}(\theta) = E_\theta\left([\frac{\delta}{\delta\theta}\log f(X|\theta)]^2\right) = E(s^2(X,\theta))$$

If X and Y are independent then $\mathcal{I}_{X,Y}(\theta) = \mathcal{I}_X(\theta) + \mathcal{I}_Y(\theta)$
The Fisher information of a random sample is $n\mathcal{I}(\theta)$
Can substitute sufficient statistic for sample.
To simplify calculation when the interchangability of integration and differentiation holds:

$$\mathcal{I}(\theta) = -E\left[\frac{\delta}{\delta\theta}s(X,\theta)\right]$$

Which is always true for exponential families.

Best Unbiased Estimators (UMVUE)
If W is a BUE of $\tau(\theta)$ then W is unique.
To find the BUE (3 methods):

1. Find the lowest variance bound (Cramer-Rao bound p.73) and show the estimator can acheive the bound

(i) For iid case, the C-R lower bound for unbiased estimators is $1/[n\mathcal{I}(\theta)]$

(ii) C-R lower bound depends only on $\tau(\theta)$ and $f(x|\theta)$ and is a uniform lower bound on the variance

(iii) If $W(\mathbf{X})$ is unbiased for $\tau(\theta)$ then it obtains the C-R lower bound IFF:

$$a(\theta)\left[W(x)-\tau(\theta)\right]=s(x,\theta), \text{ for some function } a(\theta)$$

(iv) For one parameter exponential families, assume $E[T(X)]=\tau(\theta)$, then $\frac{1}{n}\sum T(X_i)$ is unbiased and attains C-R.

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}T(X_i)\right)=\frac{[\tau'(\theta)]^2}{\mathcal{I}_n(\theta)}$$

2. Based on the Rao-Blackwell, construct an estimator using a function of a complete sufficient statistic (p.80)

   (i) Find a complete sufficient statistic T for parameter θ

   (ii) construct a function of T such that $E[\phi(T)]=\tau(\theta)$ for all $\theta\in\Theta\rightarrow\phi(T)$ is UMVUE

3. Based on the Rao-Blackwell method, compute the conditional expectation of an unbiased estimator given a complete sufficient statistic T.

   (i) Find a complete sufficient statistic T for θ

   (ii) Find an unbiased estimator $W(X)$ of $\tau(\theta)$

   (iii) Compute $\phi(T)=E[W(X)|T]$ ($\phi(T)$ is UMVUE) NOTE: use definitions of conditional expectations

2.4 Hypothesis Testing

Power Function: $\beta(\theta)=P_\theta(\mathbf{X}\in R)=P_\theta(\text{reject } H_0)$=P(making type I error)
The power function comes from the distribution of the data
P(Type II error)=$1-\beta$

Likelihood Ratio Tests

$$\lambda(x)=\frac{sup_{\theta\in\Theta_0}L(\theta|x)}{sup_{\theta\in\Theta}L(\theta|x)}=\frac{L(\hat{\theta}_0|x)}{L(\hat{\theta}|x)}$$

where, $\hat{\theta}_0$ is the most likely value of θ resticted to the null parameter space, and $\hat{\theta}$ is the MLE.
To find LRT:

1. Construct LRT test with $\max_{\theta\in\theta_0}L(\theta)$

2. Simplify and set $\leq c$ moving as many constants into c as possible so that the dist. of the LRT is identifiable

LRT based on sufficient statistics: can use the distribution of the sufficient stat. instead of the distribution of the data.
Only need to focus on the $g(T(x);\theta)$ since all of the other stuff should cancel in the ratio.

Unbiased Test:
If: Prob(reject H_0 when H_0 is false)≥Prob(reject H_0 when H_0 is true), then test is *unbiased*
Equivalently: $\beta(\theta') \geq \beta(\theta'')$, for every $\theta' \in \Theta_0^c$ and $\theta'' \in \Theta_0$ (i.e. θ' is the values of θ not in H_0)

Uniformly Most Powerful Test:
For *simple* hypotheses, $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, the UMP level α test always exists.
Neyman-Pearson Theorem: See notes page 112
Monotone likelihood Ratio (MLR) property helps determine if composite hypotheses have a UMP test.
Karlin-Rubin: Sufficient statistic can be used with an MLR family to find level-α UMP tests. (p117)

2.5 Interval Estimators

Two Methods to construct Confidence intervals:

1.