Math Tricks

$\binom{n}{r} = \frac{n!}{r!(n-r)!}$ (Combination - number of ways to group r of n objects, order not important)

$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$ (a is constant, $0 < r < 1$)

$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

$E(X^2) = E[X(X-1)] + E(X)$

Basic Probability

$f(x) \geq 0$

$\sum_x f(x)$ or $\int_{-\infty}^{\infty} f(x)dx = 1$

$\Pr[X > x] = 1 - \Pr[X \leq x]$

$\Pr[x_1 \leq X \leq x_2] = \Pr[X \leq x_2] - \Pr[X \leq x_1]$

Properties of $E(X)$

$E(X) = \sum_x xf(x)$ or $\int_{-\infty}^{\infty} xf(x)dx$

$E(aX + b) = aE(X) + b$

$E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$

If $\exists\ a$ such that $\Pr[X \geq a] = 1$, then $E(X) > a$

If $\exists\ b$ such that $\Pr[X \leq b] = 1$, then $E(X) < b$

Properties of $Var(X)$

$Var(X) = \sum_x (x-\mu)^2 f(x)$ or $\int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$

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

$Var(X) = 0 \Rightarrow \Pr[X = \mu] = 1$ (i.e., X only takes on one value)

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

$Var(X_1 + X_2 + ... + X_n) = Var(X_1) + Var(X_2) + ... + Var(X_n)$ (if all X_i are independent)

Skewness - $E[(x - \mu)^3]$
Skewness Coefficient - divide by σ^3

Kurtosis - $E[(x - \mu)^4]$

Degree of Excess Kurtosis - $\dfrac{\mu_4}{\sigma^4} - 3$ (Kurtosis of normal distribution is 3)

Central Moment - $\mu_r = E[(x - \mu)^r]$, $r \geq 2$
Raw Moment - $E[x^r]$

Functions of Random Variables

$y = r(x)$ is a function of random variable x (r must be monotonic... increasing or decreasing)
Inverse function is $x = s(y)$

Discrete pdf: $g(y) = \Pr[Y = y] = \Pr[r(x) = y] = \sum_{x:r(x)=y} f(x)$

Continuous pdf: $g(y) = \dfrac{d}{dy}G(y) = \dfrac{dF(s(y))}{ds(y)} \cdot \left|\dfrac{ds(y)}{dy}\right| = f(s(y))\left|\dfrac{ds(y)}{dy}\right|$

Continuous cdf: $G(y) = \Pr[Y \leq y] = \Pr[r(x) \leq y) = \int_{x:r(x)=y} f(x)dx$ = Pr[X ≤ s(y)] = F(s(y))

Multiple Variables

Joint Distribution - $f(x,y) = \Pr(X = x$ and $Y = y)$

$f(x,y) \geq 0$

$$\sum_x \sum_y f(x,y) = 1$$

$\Pr[a \leq x \leq b$ and $c \leq y \leq d] = \int_a^b \int_c^d f(x,y)dydx$

Joint Cumulative Distribution - $F(x,y) = \Pr(X \leq x$ and $Y \leq y)$

$$\sum_{X \leq x} \sum_{Y \leq y} f(x,y)$$

$$\int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)dydx$$

Marginal Distribution - $f_x(x) = \sum_y f(x,y)$

Covariance - measures linear relationship between x and y

$$Cov(X,Y) = E[(x - \mu_x)(y - \mu_y)] = E(xy) - \mu_x\mu_y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \sigma_{xy}$$

X and Y **independent** $\Rightarrow$ $Cov(X,Y) = 0$ (doesn't work the other way around)

$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
$Var(aX + bY + c) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$

Correlation - $\rho = \dfrac{Cov(X,Y)}{\sqrt{Var(X)+Var(Y)}} = \dfrac{\sigma_{xy}}{\sigma_x\sigma_y}$

$\rho = 1 \Rightarrow$ perfect linear relationship between X and Y
$\rho = -1 \Rightarrow$ perfect inverse linear relationship between X and Y (i.e., slope is negative)

Independence - $\Pr(A \,\&\, B) = \Pr(A \cap B) = \Pr(A)\cdot\Pr(B)$
$Cov(X,Y) = 0$
$f(x,y) = f_x(x)\cdot f_y(y)$

Conditional Probabilities - $f_{x|y}(x|y) = \Pr(X = x \mid Y = y) = \dfrac{\Pr(X = x \,\&\, Y = y)}{\Pr(Y = y)} = \dfrac{f_{xy}(x,y)}{f_y(y)}$

x \ y	1	2	3	4	Marginal (x)
1	0.1	0	0.1	0	0.2
2	0.3	0	0.1	0.2	0.6
3	0	0.2	0	0	0.2
Marginal (y)	0.4	0.2	0.1	0.2	1

$\Pr[X = 1 \mid Y = 1] = 0.1/0.4 = 0.25$
$\Pr[X = 2 \mid Y = 1] = 0.3/0.4 = 0.75$

$\Pr[A \cap B] = \Pr[B|A]\cdot\Pr[A]$

Multiple Conditions - $\Pr[A \cap B \cap C] = \Pr[A \cap B]\cdot\Pr[C]$
If A & B are independent... $= \Pr[A|C]\cdot\Pr[B|C]\cdot\Pr[C]$

Conditional Expectation - $E(X|Y) = \int_{-\infty}^{\infty} xf_{x|y}(x \mid y)dx$ or $\sum_x xf_{x|y}(x \mid y)$

e.g., $E(X = 1 \mid Y = 1) = 1(0.25) + 2(0.75) = 1.75$

Law of Iterated Expectations - fix a level of Y and calculate $E(X|Y)$, then take expected value of these to find $E(X)$; use if you have joint distribution and can't get X by itself

$$E(X) = E_y[E(X|Y)] = \int_{-\infty}^{\infty} E(X \mid Y) f_y(y)dy$$

Conditional Variance - $Var(X|Y) = E[(X - E(X|Y))^2] =$

$$\int_{-\infty}^{\infty}[x-E(X\mid Y)]^2 f_{x|y}(x\mid y)dx \text{ or } \sum_x [x-E(X\mid Y)]^2 f_{x|y}(x\mid y)$$

Multivariate Distributions

$$\mathbf{x}=\begin{pmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix} \qquad E(\mathbf{x})=\begin{pmatrix} E(x_1)\\ E(x_2)\\ \vdots\\ E(x_n)\end{pmatrix}=\begin{pmatrix} \mu_1\\ \mu_2\\ \vdots\\ \mu_n\end{pmatrix}=\boldsymbol{\mu}$$

$$Cov(\mathbf{x})=E\Big[\underset{n \times 1}{(\mathbf{x}-\boldsymbol{\mu})^T}\underset{1 \times n}{(\mathbf{x}-\boldsymbol{\mu})}\Big]=\underset{n \times n}{\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn}\end{pmatrix}}=\boldsymbol{\Sigma}$$

Symmetric matrix - $\sigma_{ij}=\sigma_{ji}$

σ_{ij} = $Cov(x_i,x_j)$ $(i \neq j)$ or $Var(x_i)$ $(i = j)$

Independent - $\boldsymbol{\Sigma}$ is diagonal matrix with variances only

Discrete Distributions

Bernoulli - simple yes/no or success/failure; Pr[success] = p; $x \in$ {0, 1}

pdf: $p^x(1-p)^{1-x}$ $\qquad E(X) = p$ $\qquad Var(X) = p(1-p)$

Binomial - x successes in n trials; Pr[success] = p; sum of n *iid* Bernoulli's; $x \in$ {0, 1, 2,…n}

pdf: $\binom{n}{x}p^x(1-p)^{n-x}$ $\qquad E(X) = np$ $\qquad Var(X) = np(1-p)$

De Moivre-Laplace Central Limit Theorem - X ~ Binomial(n,p), the for fixed $a \le b$,

$$\Pr\left[a \le \frac{x-np}{\sqrt{npq}} \le b\right]=\Pr\left[np+a\sqrt{npq} \le x \le np+b\sqrt{npq}\right]=\Phi(b)-\Phi(a) \text{ as } n \to \infty$$

Example: Bin(200,0.05), Pr[$x \le 20$] = Pr[$0 \le x \le 20$]

Find $a=\dfrac{0-np}{\sqrt{npq}}=\dfrac{-200(0.05)}{\sqrt{200(0.05)(0.95)}}=-3.2444$ and $b=\dfrac{20-np}{\sqrt{npq}}=3.2444$

Pr[$0 \le x \le 20$] = Φ(3.2444) - Φ(-3.2444) = 0.998823

Note: Same calculation using Binomial yields 0.99884

Geometric - x trials until first success; Pr[success] = p; $x \in$ {1, 2, 3,…}

pdf: $p(1-p)^{x-1}$ $E(X) = \frac{1}{p}$ $Var(X) = \frac{1-p}{p^2}$

Negative Binomial - x trials until r^{th} success; Pr[success] = p; $x \in \{r, r+1, \ldots\}$

pdf: $\binom{x-1}{r-1} p^r (1-p)^{x-r}$ $E(X) = \frac{r}{p}$ $Var(X) = \frac{r(1-p)}{p^2}$

Poisson - x successes in λ (continuous area, volume, time, etc.); $x \in \{0, 1, 2, \ldots\}$

pdf: $\frac{\lambda^x}{x!} e^{-\lambda}$ $E(X) = \lambda$ $Var(X) = \lambda$

Hypergeometric - binomial with dependent trials; m = # of possible successes; x = # of desired successes; N = total # of trials possible; n = # of trials used

pdf: $\frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$ $E(X) = n\left(\frac{m}{N}\right)$ $Var(X) = n\left(\frac{m}{N}\right)\left(1-\frac{m}{N}\right)\left(\frac{N-n}{N-1}\right)$

Continuous Distributions

Uniform - equal probability of occurrence anywhere in range [a,b]

$f(x) = \begin{cases} \frac{1}{b-a} & , a \le x \le b \\ 0 & , \text{else} \end{cases}$ $E(X) = \frac{a+b}{2}$ $Var(X) = \frac{(b-a)^2}{12}$

Exponential - time until first occurrence; closely related to Poisson

$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0 & , \text{else} \end{cases}$ $E(X) = \frac{1}{\lambda}$ $Var(X) = \frac{1}{\lambda^2}$

Gamma (Erlang) - time until k^{th} occurrence; sum of k *iid* Exponential(λ)

$f(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$ $E(X) = \frac{k}{\lambda} = \alpha\beta$ $Var(X) = \frac{k}{\lambda^2} = \alpha\beta^2$

NOTE: $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x = (\alpha-1)\Gamma(\alpha-1) = (\alpha-1)!$ (if α is integer)

Normal - symmetric about mean

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ $E(X) = \mu$ $Var(X) = \sigma^2$

Bivariate Normal -

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\left(\frac{\varepsilon_x^2+\varepsilon_y^2-2\rho\varepsilon_x\varepsilon_y}{1-\rho^2}\right)\right]$$

$$\varepsilon_x=\frac{x-\mu_x}{\sigma_x}\,,\ \varepsilon_y=\frac{y-\mu_y}{\sigma_y}$$

Marginals are also normal: $f_x(x) \sim N(\mu_x,\ \sigma_x^2)$

Conditional Distribution of X | Y also normal: $f(x|y) \sim N\left(\mu_x+\frac{\rho\sigma_x}{\sigma_y}(y-\mu_y),\sigma_x^2(1-\rho^2)\right)$

Multivariate Normal -

$$f(\mathbf{x})=(2\pi)^{\frac{-n}{2}}|\mathbf{\Sigma}|^{\frac{-1}{2}}\exp\left[\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

Special Cases:

x_i independent: $f(\mathbf{x})=(2\pi)^{\frac{-n}{2}}(\sigma_1\sigma_2\cdots\sigma_n)^{-1}\exp\left[\frac{-1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right]$

x_i iid $N(\mu,\sigma^2)$: $f(\mathbf{x})=(2\pi)^{\frac{-n}{2}}\sigma^{-n}\exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$

Properties:

$\underset{m\times 1}{\mathbf{Y}} = \underset{(m\times n)(n\times 1)}{\mathbf{A}\mathbf{x}} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu}+\mathbf{b},\ \underset{(m\times n)(n\times n)(n\times m)}{\mathbf{A}\mathbf{\Sigma}\mathbf{A}^T})$

Partition into subsets: $\mathbf{x}=\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}\begin{matrix}n\times 1\\ m\times 1\end{matrix}$ $\quad \boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}$ $\quad \mathbf{\Sigma}=\begin{array}{cc} n\times n & n\times m \\ \multicolumn{2}{c}{\begin{pmatrix}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22}\end{pmatrix}} \\ m\times n & m\times m\end{array}$

$\mathbf{\Sigma}_{21} = \mathbf{\Sigma}_{12}^T$

Marginal distributions also normal: $\mathbf{x}_1 \sim N(\boldsymbol{\mu},\mathbf{\Sigma}_{11})$ and $\mathbf{x}_2 \sim N(\boldsymbol{\mu},\mathbf{\Sigma}_{22})$

(Note: Lose info in $\mathbf{\Sigma}_{12}$ unless they're $\mathbf{x}_1$ & $\mathbf{x}_2$ are independent)

Conditional distribution: $\mathbf{x}_1|\mathbf{x}_2 \sim N\left(\boldsymbol{\mu}+\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2),\mathbf{\Sigma}_{11}-\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right)$

Special Distributions

Standard Normal - normal with mean 0 and variance 1

$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}x^2}$ $E(X) = 0$ $Var(X) = 1$

$X \sim N(\mu, \sigma^2) \Rightarrow \boxed{z = \frac{x-\mu}{\sigma}} \sim N(0,1)$

$\phi(z)$ = cdf of standard normal

Chi Squared - squared standard normal is chi squared with 1 degree of freedom

$z \sim N(0,1) \Rightarrow \boxed{X = z^2} \sim \chi_1^2$ (1 df)

$f(x) = \frac{\left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2\Gamma\left(\frac{n}{2}\right)}$, $x > 0$, $n = 1, 2, \ldots$ $E(X) = n$ $Var(X) = 2n$

$X_1 \sim \chi_m^2$ and $X_2 \sim \chi_m^2$; if X_1 and X_2 are independent, then $X_1 + X_2 \sim \chi_{m+n}^2$

F - ratio of chi squared distributions divided by their degrees of freedom

$X_1 \sim \chi_m^2$ and $X_2 \sim \chi_n^2 \Rightarrow \boxed{F = \frac{X_1/m}{X_2/n}} \sim F_{m,n}$ (1 df)

$f(x) = \frac{kx^{(m-2)/2}}{\left(1 + \frac{m}{n}x\right)^{(m+n)/2}}$, $x > 0$, m, $n = 1, 2, \ldots$ $E(X) =$ $Var(X) =$

t - flatter than a standard normal (wider tails)

$z \sim N(0,1)$ and $X \sim \chi_n^2 \Rightarrow \boxed{t = \frac{z}{\sqrt{X/n}}} \sim t_n$

$f(x) = k\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$, $n = 1, 2, \ldots$ $E(X) =$ $Var(X) =$

Birthday Problem

$\Pr$(2 people in group of n people have b-day on same day) =
1 - $\Pr$(Everyone's b-days are different) =

$$1-1\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right)\cdots\left(\frac{365-n+1}{365}\right)=1-\frac{364!}{365^{n-1}}$$

n	$\Pr$(2 w/ same b-day)
20	1 - 0.5886
30	1 - 0.3024
40	1 - 0.1120
50	1 - 0.0305

Sampling Theory

Examples - unemployment, household consumption survey

Random sample - set of *iid* rv's $x_1, x_2, ..., x_n$; x_i's have joint distribution $f(\mathbf{x},\boldsymbol{\theta}) = [f(x,\boldsymbol{\theta})]^n$

$\boldsymbol{\theta}$ is vector of parameters (e.g., $\boldsymbol{\theta} = \begin{pmatrix}\mu \\ \sigma^2\end{pmatrix}$)

Statistic - function that doesn't depend on any of the known parameters; examples:

Sample Mean: $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

Sample Variance: $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Sample Covariance: $\hat{\sigma}_{xy} = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

Note for sample variance and covariance, must replace the mean with a function of the observations; can't have a function on the unknown parameters

Theorem: If random sample is from a population with mean μ and variance σ^2, then the sample mean is a random variable with mean μ and variance σ^2/n.

Sample Mean: $\bar{x} = \dfrac{1}{n}(x_1 + x_2 + \cdots + x_n)$

$$E(\bar{x}) = \frac{1}{n}\big(E(x_1) + E(x_2) + \cdots + E(x_n)\big) = \frac{1}{n}nE(X) = E(X)$$

Sample Variance: assume $x_i \sim N(\mu,\sigma^2)$

$$Var(\bar{x}) = \left(\frac{1}{n}\right)^2\big(nVar(X)\big) = \frac{1}{n}Var(X)$$ (can sum variances of independent normals)

General case for variance:

$$Var(\bar{x}) = E\left[(\bar{x}-\mu)^2\right] = E\left[\left(\frac{1}{n}x_1 + \frac{1}{n}x_2 + \cdots + \frac{1}{n}x_n - \mu\right)\left(\frac{1}{n}x_1 + \frac{1}{n}x_2 + \cdots + \frac{1}{n}x_n - \mu\right)\right] =$$

$$E\left[\left(\frac{1}{n}(x_1-\mu) + \frac{1}{n}(x_2-\mu) + \cdots + \frac{1}{n}(x_n-\mu)\right)\left(\frac{1}{n}(x_1-\mu) + \frac{1}{n}(x_2-\mu) + \cdots + \frac{1}{n}(x_n-\mu)\right)\right]$$

$$= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)\right)\left(\frac{1}{n}\sum_{j=1}^{n}(x_j-\mu\right)\right] =$$

$$E\left[\frac{1}{n^2}\left(\sum_{i=1}^{n}(x_i-\mu)\right)^2 + \frac{1}{n^2}\left(\sum_{i=1}^{n}\sum_{j\neq i}(x_i-\mu)(x_j-\mu)\right)\right] =$$

$$\frac{1}{n^2}\sum_{i=1}^{n} E(x_i-\mu)^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} E\left[(x_i-\mu)(x_j-\mu)\right] =$$

$$\frac{1}{n^2}\sum_{i=1}^{n} Var(X) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} Cov(x_i, x_j) = \frac{1}{n^2}n\sigma^2 + \frac{1}{n^2}n0 = \frac{\sigma^2}{n}$$

Estimation - want estimators that are unbiased and efficient (small variance)

Unbiased - the mean of the distribution of the estimator equals the parameter we're trying to estimate:

$$E(\hat{\theta}) = \int_{-\infty}^{\infty} \theta f(\theta)d\theta = \theta$$

Example - Normal Distribution

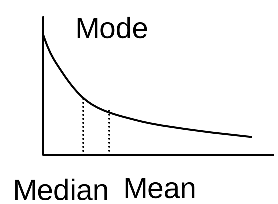
$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ ~ $N(\mu,\sigma^2/n)$ (shown last time) so $\hat{\theta}_1$ is unbiased estimate

$\hat{\theta}_2 =$ median($x_1,x_2,...,x_n$); for $N(\mu,\sigma^2/n)$, $E(\hat{\theta}_2) = \mu$; $\hat{\theta}_2$ is unbiased estimate

$\hat{\theta}_3 = x_5$ ~ $N(\mu,\sigma^2)$, $E(\hat{\theta}_3) = \mu$; $\hat{\theta}_3$ is unbiased estimate

$\hat{\theta}_4 = 1/2(x_1 + x_2)$; $E(\hat{\theta}_4) = 1/2[E(x_1) + E(x_2)] = 1/2[\mu + \mu] = \mu$; $\hat{\theta}_4$ is unbiased estimate

Example - Exponential Distribution



$f(x,\theta) = \frac{1}{\theta}e^{-x/\theta}$, $x > 0$

$F(x,\theta) = 1 - e^{-x/\theta}$, $x > 0$

$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$; $E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}n\theta = \theta$ ... unbiased

$\hat{\theta}_2 =$ median($x_1,x_2,...,x_n$)... $F(x,\theta) = 1 - e^{-x/\theta} = 0.5 \Rightarrow$ median = $\theta\ln(2)$ ... biased

$\hat{\theta}_3 = x_5$; $E(\hat{\theta}_3) = \theta$... unbiased

$\hat{\theta}_4 = 1/2(x_1 + x_2)$; $E(\hat{\theta}_4) = 1/2[E(x_1) + E(x_2)] = 1/2[\theta + \theta] = \theta$... is unbiased

Bias - $b(\hat{\theta}) = E\left|\hat{\theta} - \theta\right|$; unbiased means $b(\hat{\theta}) = 0$

Efficiency - $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$

Example - Normal Distribution

$Var(\hat{\theta}_1) = \sigma^2/n$ ... this is the most efficient

$Var(\hat{\theta}_3) = \sigma^2$

$Var(\hat{\theta}_4) = \sigma^2/2$

Purposely Choose Biased Estimator? - maybe, if you have great gains in efficiency (much lower variance); example: median may be biased, but bias may be small and median has lower variance than mean so the trade-off may be worth while

Mean Squared Error - $MSE(\hat{\theta}) = E\left[(\hat{\theta}-\theta)^2\right] = Var(\hat{\theta}) + \left[b(\hat{\theta})\right]^2$; trades off bias and variance; new standard for estimate is lowest MSE (although sometimes unbiased is more important)

Unbiased - $MSE(\hat{\theta}) = Var(\hat{\theta})$

Relative Efficiency - $\hat{\theta}_1$ is relatively more efficient than $\hat{\theta}_2$ if $\dfrac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)} = \dfrac{E\left[(\hat{\theta}_1-\theta)^2\right]}{E\left[(\hat{\theta}_2-\theta)^2\right]} < 1$

Unbiased - relatively more efficient if $\dfrac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)} < 1$

UMVU - uniformly minimum variance unbiased estimate; unbiased estimator, $\hat{\theta}$, such that for any other unbiased estimator, $\theta *$, $Var(\hat{\theta}) < Var(\theta *)$; if we want an unbiased estimator, this will be the best one to use

Sample Variance: $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ or $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ ... $\hat{\sigma}^2$ has smaller MSE, but s^2 is unbiased

Show $E(s^2) = \sigma^2$:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[x_i - \frac{1}{n}\sum_{j=1}^{n}x_j\right]^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[x_i^2 - 2x_i\sum_{j=1}^{n}x_j + \left(\frac{1}{n}\sum_{j=1}^{n}x_j\right)^2\right] =$$

$$\frac{1}{n-1}\sum_{i=1}^{n}\left[x_i^2 - 2x_i\sum_{j=1}^{n}x_j + \frac{1}{n^2}\sum_{j=1}^{n}x_j\sum_{k=1}^{n}x_k\right] =$$

$$\frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 \overbrace{-\frac{2}{n}\sum_{i=1}^{n}x_i^2 - \frac{2}{n}\sum_{i=1}^{n}x_i\sum_{j\neq i}x_j} + \sum_{i=1}^{n}\left(\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}x_jx_k\right)\right] =$$

$$\frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - \frac{2}{n}\sum_{i=1}^{n}x_i^2 - \frac{2}{n}\sum_{i=1}^{n}x_i\sum_{j\neq i}x_j + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j\right] =$$

$$\frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - \frac{2}{n}\sum_{i=1}^{n}x_i^2 - \frac{2}{n}\sum_{i=1}^{n}x_i\sum_{j\neq i}x_j + \overbrace{\frac{1}{n}\sum_{i=1}^{n}x_i^2 + \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}x_ix_j}\right] =$$

$$\frac{1}{n-1}\left[\left(1 - \frac{2}{n} + \frac{1}{n}\right)\sum_{i=1}^{n}x_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}x_ix_j\right] = \frac{1}{n-1}\left[\left(\frac{n-1}{n}\right)\sum_{i=1}^{n}x_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}x_ix_j\right] =$$

$$\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}x_ix_j$$

$E(x_i^2) = \sigma^2 + \mu^2$ (comes from assumption that x_i iid mean μ and variance σ^2)

Proof: $Var(x_i) = E(x_i^2) - \left[E(x_i)\right]^2 = E(x_i^2) - \mu^2 = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2$

$E(x_ix_j) = \mu^2$

Proof: $Cov(x_i, x_j) = E(x_i x_j) - E(x_i)E(x_j) = E(x_i x_j) - \mu^2$

If x_i's are independent $Cov(x_i, x_j) = 0$ so $E(x_i x_j) = \mu^2$

$$E(s^2) = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} x_i x_j\right] = \frac{1}{n}\sum_{i=1}^{n} E(x_i^2) - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} E(x_i x_j) =$$

$$\frac{1}{n}\sum_{i=1}^{n}(\sigma^2 + \mu^2) - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\mu^2 = \frac{1}{n}n\sigma^2 + \frac{1}{n}n\mu^2 - \frac{1}{n(n-1)}n(n-1)\mu^2 =$$

$\sigma^2 + \mu^2 - \mu^2 = \sigma^2$ ... unbiased

$Var(s^2) = 2\sigma^4/(n - 1)$

$MSE(s^2) = Var(s^2) + [b(s^2)]^2 = 2\sigma^4/(n - 1)$

$$Var(\hat{\sigma}^2) = Var\left(\frac{n-1}{n}s^2\right) = \left(\frac{n-1}{n}\right)^2 Var(s^2) = \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1} = \frac{2(n-1)\sigma^2}{n^2}$$

$$b(\hat{\sigma}^2) = E(\hat{\sigma}^2) - \sigma^2 = \left(\frac{n-1}{n}\right)E(s^2) - \sigma^2 = \left(\frac{n-1}{n}\right)\sigma^2 - \sigma^2 = \frac{-1}{n}\sigma^2$$

$$MSE(\hat{\sigma}^2) = \frac{2(n-1)\sigma^4}{n^2} + \frac{1}{n^2}\sigma^4 = \frac{(2n-1)}{n^2}\sigma^4$$

$$\frac{MSE(\hat{\sigma}^2)}{MSE(s^2)} = \frac{(2n-1)\sigma^4/n^2}{2\sigma^4/(n-1)} = \frac{(2n-1)(n-1)}{2n^2} = \frac{(n-1)}{n}\cdot\frac{(2n-1)}{2n} < 1 \;\forall\; n \geq 1$$

$\therefore$ $\hat{\sigma}^2$ is a relatively more efficient estimator of σ^2 than s^2 (but s^2 is unbiased and $\hat{\sigma}^2$ is biased... people still debate which is better to use)

Asymptotic Theory

What happens to our statistic as the size of our sample increases? Does our statistic converge to the correct value... do s^2 and $\hat{\sigma}^2$ converge to σ^2 as $n \to \infty$?

Non-Stochastic Convergence - let $\{b_n\}$ be a sequence of real numbers. If there exists a real number b, and if for every $\delta > 0$ there exists an integer $N(\delta)$ such that for all $n \geq N(\delta)$ $|b_n - b| < \delta$ then b is the limit of the sequence b_n and write $b_n \to b$, read "b_n converges to b"

Example -

$\{bn\} = \sum_{i=0}^{n}\left(\frac{1}{2}\right)^n = 1 + {}^1/_2 + {}^1/_4 + {}^1/_8 + \ldots$

$b_1 = 1$
$b_2 = 1 + {}^1/_2 = {}^3/_2$
$b_3 = 1 + {}^1/_2 + {}^1/_4 = {}^7/_4$

Converges to $b = 2$

Statistical Convergence - different types; some stronger than others; 1⇒3; 2⇒3; 3⇒4

1. Almost Sure Convergence
2. Convergence in r^{th} mean
3. Convergence in Probability
4. Convergence in Distribution

Let b_n be a statistic based on a random sample of observations

e.g., $b_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ (sample mean) or $b_n = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})$ (sample variance)

use n observations of x_i to construct the statistic b_n

Almost Sure Convergence - $\{b_n\}$ converges almost surely to b *iff* there exists a real number b such that $\Pr[b_n \to b] = 1$; written $b_n \xrightarrow{as} b$

Example - $\{x_i\}$ is sequence of *iid* rv's with $E(x_i) = \mu < \infty$ (distribution type doesn't matter)

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \overset{as}{\to} \mu$$

Convergence in Probability - $\{b_n\}$ converges in probability if there exists a real number b such that for every $\varepsilon > 0$, $\Pr[|b_n - b| < \varepsilon] \to 1$ as $n \to \infty$; written $b_n \xrightarrow{P} b$

Example - $\{x_i\}$ is sequence of rv's with $E(x_i) = \mu$, $Var(x_i) = \sigma^2 < \infty$, and $Cov(x_i, x_j) = 0$; assumptions are much weaker (easier to verify) than almost sure convergence; no assumptions about identical distribution or independence ($Cov = 0 \not\Rightarrow$ independent; only linear independence)

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \overset{P}{\to} \mu$$

Convergence in r^{th} Mean - used in time series; $\{b_n\}$ is sequence of real valued rv's; if there exists a real number b such that $E[|b_n - b|^r] \to 0$ as $n \to \infty$ for some $r > 0$, then b_n converges in r^{th} mean; written $b_n \xrightarrow{(r)} b$;

Quadratic Mean - use $r = 2$ (which we usually do); $b_n \xrightarrow{q.m.} b$; note $|b_n - b|^2 = (b_n - b)^2$ so $E[|b_n - b|^2] = E[(b_n - b)^2] = Var(b_n)$; converges if variance of statistic converges to zero

Lower r - $b_n \xrightarrow{(r)} b$ for some $r \geq 1$, then $b_n \xrightarrow{(s)} b$ for $0 < s < r$ (i.e. quadratic mean convergence ⇒ mean convergence, $r = 1$)

Convergence in Distribution - $\{b_n\}$ is sequence of real valued rv's with distribution functions $\{F_n\}$; if $F_n(x) \to F(x)$ as $n \to \infty$ for every continuity point in x, where $F(x)$ is the distribution function of a rv X, then bn converges in distribution to the rv X; written $b_n \xrightarrow{d} X$

Example - $t_n \overset{d}{\to} N(0,1)$

Consistent Estimator - $\hat{\theta}$ is consistent estimator of θ *iff* $\hat{\theta} \xrightarrow{P} \theta$

Examples:

$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ is consistent estimator of θ (mean)

$\hat{\theta}_2 =$ median; if $Cov(x_i,x_j) = 0$ and $E(x_i) = \mu$ and $Var(x_i) = \sigma^2 < \infty$, and x_i come from a symmetric distribution, then median = mean and $\hat{\theta}_2$ is consistent estimator of θ

Exponential Distribution - use median/$\ln(2)$

$\hat{\theta}_3 = x_5$ and $\hat{\theta}_4 = \frac{1}{2}(x_1 + x_2)$ are not consistent estimators; they don't change at all as sample size gets bigger

Weak Law of Large Numbers (WLLN) - as sample size increases, the sample mean converges in probability to population mean; 4 sets of conditions under which WLLN holds

Khinchin - $\{x_n\}$ is a sequence of *iid* rv's with finite mean μ

$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{P} \mu$

Chebychev - $\{x_n\}$ is a sequence of independent rv's with means μ_n and variances σ_n^2; if variances are bounded above (i.e., $\sigma_n^2 < c < \infty$) and

$\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i$, then $(\bar{x}_n - \bar{\mu}_n) \xrightarrow{P} 0$ (sample mean minus mean of the means)

Markov - $\{x_n\}$ is a sequence of rv's with means μ_n , if $Var(\bar{x}_n) \rightarrow 0$ as n $\rightarrow \infty$, then

$(\bar{x}_n - \bar{\mu}_n) \xrightarrow{P} 0$ (doesn't assume independence)

Kolmogorov - $\{x_n\}$ is a sequence of independent rv's; $z_n = \bar{x}_n - \bar{\mu}_n$

if $\lim_{n\to\infty} E\left[\frac{z_n^2}{(1+z_n^2)}\right] = 0$ then $z_n \xrightarrow{P} 0$

Strong Law of Large Numbers - interested in almost sure convergence;

Theorem - $\{x_n\}$ is a sequence of *iid* rv's, then $\bar{x}_n \xrightarrow{as} \mu$ iff $E[|x_n|] < \infty$ (i.e., finite mean)

Kolmogorov II - $\{x_n\}$ is a sequence of independent rv's with finite variances

if $\frac{1}{n^2}\sum_{i=1}^{n} Var(x_n) < \infty$, then $(\bar{x}_n - \bar{\mu}_n) \xrightarrow{as} 0$ (allows mean to change over time, but mean of sample means approaches mean of means)

Central Limit Theorem - for a large sample from any distribution, we can approximate the distribution of the sample mean with a normal distribution; $\{x_n\}$ be a sequence of rv's with

$s_n = \sum_{i=1}^{n} x_i$ and $\bar{x} = \frac{1}{n} s_n$; the standardized mean $z_n = \frac{\bar{x}_n - E(\bar{x}_n)}{\sqrt{Var(\bar{x}_n)}} = \frac{s_n - E(s_n)}{\sqrt{Var(s_n)}}$, where

$Var(s_n) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2$ and $Var(\bar{x}_n) = \frac{1}{n}\sigma^2$; z_n ~ N(0,1)

$$\bar{x}_n \to N(E(\bar{x}_n), Var(\bar{x}_n)) = N(\mu, \frac{1}{n}\sigma^2)$$

De Moivre - proved CLT where x_i's are independent Bernoulli rv's
Lindberg-Levy - proved CLT where x_i's are *iid* with $Var(x_i) = \sigma^2 < \infty$
Lindberg-Feller - proved CLT where x_i's are independent with $E(X_i) = \mu_i$ and $Var(x_i) = \sigma_i^2$,

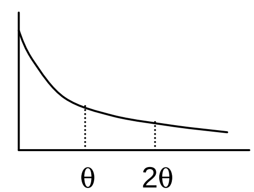
$s_n^2 \equiv \sum_{i=1}^{n} \sigma_i^2$, $S_n \equiv \sum_{i=1}^{n} x_i$; for some ε > 0, $\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \int_{|x-\mu_i|>\varepsilon S_n} (x-\mu_i)^2 \cdot f_i(x)dx = 0$

Only looking at portion of variance that is far away from mean; if that is converging to zero (i.e., tails aren't too big), the CLT holds

Multivariate CLT - $\{\mathbf{x}_n\}$ is sequence of k-variate iid rv's ($\mathbf{x}_n$ is k x 1 vector) with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, then $\mathbf{z}_n = \sqrt{n}(\bar{\mathbf{x}}_n - \boldsymbol{\mu}) \xrightarrow{d} N_k(\mathbf{0}, \boldsymbol{\Sigma})$

Chebychev's Inequality - for constant $k > 0$, $\Pr(|x - E(X)| \geq k) \leq Var(X)/k^2$; probability that you're more than k away from the mean is less than or equal to variance over k^2; provides an upper bound on $\Pr(|x - E(X)| \geq k)$... usually a very generous upper bound (much higher than the actual probability will be... will see on HW3)
As Lower Bound - $\Pr(|x - E(X)| < k) > 1 - Var(X)/k^2$

Markov's Inequality - $\Pr(x \geq \lambda E(X)) \leq 1/\lambda$ for a <u>positive</u> rv x and $\lambda > 0$;
Example - look at exponential: $f(x) = e^{-x/\theta}$, $E(X) = \theta$... $\Pr(x \geq 2\theta) \leq 1/2$... yep



Likelihood Functions - how to estimate parameters; let $x_1, x_2,..., x_n$ be a random sample from a density function $f(x;\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is vector of distribution parameters; likelihood function

$L(\boldsymbol{\theta};x) = f_x(x_1,x_2,...,x_n,\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta})$; looks just like a joint pdf except now the x_i's are given and the parameters are the unknowns

Example - $x_1, x_2,..., x_n$ are iid N(μ,σ2) so $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$

$$L(\mu,\sigma^2;x) = (2\pi\sigma)^{=n/2} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

Log-Likelihood Function - natural logarithm of the likelihood function; will be negative because you're taking $\ln$ of probabilities (< 1)

Example - $\ln[L(\mu,\sigma^2;x)] = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$

Information Matrix - estimate parameters for distribution from sample (e.g., μ and σ^2 for normal distribution); need to estimate covariance matrix for unbiased estimators

Single Parameter - $I(\theta) = -E\left[\frac{\partial^2 \ln L(\theta;\mathbf{x})}{\partial\theta^2}\right] = E\left[\left\{\frac{\partial \ln L(\theta;\mathbf{x})}{\partial\theta}\right\}^2\right]$

2 Or More Parameters - $I_{i,j} = -E\left[\frac{\partial^2 \ln L(x;\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right] = E\left[\frac{\partial \ln L(x;\boldsymbol{\theta})}{\partial\theta_i}\cdot\frac{\partial \ln L(x;\boldsymbol{\theta})}{\partial\theta_j}\right]$

Information Matrix - $\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} I_{11} & I_{12} & \cdots & I_{1k} \\ I_{21} & I_{22} & \cdots & I_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ I_{k1} & I_{k2} & \cdots & I_{kk} \end{bmatrix}$; $I_{ij} = I_{ji}$ so it's symmetric matrix

Example - $x_1, x_2,..., x_n$ are iid N(μ,σ2) so

$$\ln\left[L(\mu,\sigma^2;x)\right] = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial \ln L}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \qquad \frac{\partial^2 \ln L}{\partial\mu^2} = \frac{-n}{\sigma^2}$$

$$\frac{\partial \ln L}{\partial\sigma^2} = -\frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 \qquad \frac{\partial^2 \ln L}{\partial(\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} + \frac{-1}{(\sigma^2)^3}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial^2 \ln L}{\partial\mu\partial\sigma^2} = \frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$E\left[\frac{\partial^2 \ln L}{\partial\mu^2}\right] = E\left[\frac{-n}{\sigma^2}\right] = \frac{-n}{\sigma^2}$$

$$E\left[\frac{\partial^2 \ln L}{\partial\mu\partial\sigma^2}\right] = E\left[\frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)\right] = \frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(E(x_i)-\mu) = \frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(\mu-\mu) = 0$$

$$E\left[\frac{\partial^2 \ln L}{\partial(\sigma^2)^2}\right] = E\left[\frac{n}{2(\sigma^2)^2} + \frac{-1}{(\sigma^2)^3}\sum_{i=1}^{n}(x_i-\mu)^2\right] = \frac{n}{2(\sigma^2)^2} + \frac{-1}{(\sigma^2)^3}\sum_{i=1}^{n}E\left[(x_i-\mu)^2\right] =$$

$$\frac{n}{2(\sigma^2)^2} + \frac{-1}{(\sigma^2)^3}n\sigma^2 = \frac{n}{2(\sigma^2)^2} + \frac{-n}{(\sigma^2)^2} = \frac{-n}{2(\sigma^2)^2}$$

$$\mathbf{I}(\mu,\sigma^2) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{bmatrix}$$

Regularity Conditions -

(1) $\frac{\partial \ln L}{\partial\theta}$ exists for all $\mathbf{x}$ and θ

(2) $\frac{\partial}{\partial\theta}\int \ln L(\theta,\mathbf{x})d\mathbf{x} = \int\frac{\partial}{\partial\theta}\ln L(\theta,\mathbf{x})d\mathbf{x}$ (i.e., can switch order of integration an differentiation

(3) $\frac{\partial}{\partial\theta}\int\hat{\theta}(\mathbf{x})\ln L(\theta,\mathbf{x})d\mathbf{x} = \int\hat{\theta}(\mathbf{x})\frac{\partial \ln L(\theta,\mathbf{x})}{\partial\theta}d\mathbf{x}$

(4) $0 < E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right] < \infty$ for all θ

(5) $\frac{\partial \ln L}{\partial \theta}$ and $\frac{\partial^2 \ln L}{\partial \theta^2}$ exist and are continuous for all θ

Cramer-Rao Lower Bound (CRLB) -

Single Parameter - if regularity conditions hold, if $\hat{\theta}$ is an unbiased estimator of the parameter θ then $Var(\hat{\theta}) \geq \frac{1}{I(\theta)}$

Multiple Parameters - If $\hat{\theta}_j(\mathbf{x})$ is an unbiased estimator of θj for j = 1,...,k (k parameters). Assuming regularity conditions hold $Cov(\hat{\boldsymbol{\theta}}) - [\mathbf{I}(\boldsymbol{\theta})]^{-1}$ is a positive semi-definite matrix (i.e., $[\mathbf{I}(\boldsymbol{\theta})]^{-1}$ is a lower bound for the covariance matrix of an unbiased estimator)

Significance - if you have an unbiased estimator that attains the CRLB of the variance, then you know that this is the most efficient unbiased estimator (UMVU); it's not always possible to find an unbiased estimator that attains the CRLB

Example - consider a random sample from a normal distribution $N(\mu,\sigma^2)$

Unbiased estimators:

$$\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Information matrix (did this last time)

$$\mathbf{I}(\mu,\sigma^2) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{bmatrix}$$

Take inverse of this

$$[\mathbf{I}(\mu,\sigma^2)]^{-1} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

so the CRLB for the variance of $\hat{\mu}$ is σ^2/n

$$Var(\hat{\mu}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$ ... so $\hat{\mu}$ is the CRLB

$$Var(s^2)? \quad \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2} \sim \chi^2_{(n-1)} \Rightarrow \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$Var(s^2) = \frac{(\sigma^2)^2}{(n-1)^2}2(n-1) = \frac{2\sigma^4}{n-1}$$ ... so s^2 is NOT the CRLB = $\frac{2\sigma^4}{n}$

Cramer-Rao Inequality - let $x_i \sim f(x;\theta)$ and $T = T(x_1,...,x_n)$ be a statistic such that $E(T) = u(\theta)$ (some function of θ). Assume regularity conditions. Then $Var(T) \geq [u'(\theta)]^2/I(\theta)$

Maximum Likelihood Estimator (MLE) - choose values of the parameters that maximize the likelihood function (or the log-likelihood function). Take all the partial derivatives of $\ln L(\boldsymbol{\theta};\mathbf{x})$ and set them equal to zero and solve for $\tilde{\boldsymbol{\theta}}$

Score - derivative of $\ln L(\theta,\mathbf{x})$

Score Vector $\mathbf{S}(\hat{\boldsymbol{\theta}}_n)$ - comprised of all partial derivatives of $\ln L(\boldsymbol{\theta};\mathbf{x})$

Example - Let $x_1,...,x_n$ be a random sample form a $N(0,\theta)$ distribution

$$L(\theta;\mathbf{x}) = (2\pi\theta)^{-n/2}\prod_{i=1}^{n}\exp\left\{\frac{-x_i^2}{2\theta}\right\} = (2\pi\theta)^{-n/2}\exp\left\{\sum_{i=1}^{n}\frac{-x_i^2}{2\theta}\right\}$$

$$\ln L(\theta;\mathbf{x}) = \frac{-n}{2}\ln(2\pi) - \frac{n}{2}\ln(\theta) - \frac{1}{2\theta}\sum_{i=1}^{n}x_i^2$$

$$\frac{\partial \ln L}{\partial\theta} = \frac{-n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}x_i^2 = 0$$

$$-n + \frac{1}{\theta}\sum_{i=1}^{n}x_i^2 = 0$$

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}x_i^2 \text{ is the MLE of } \theta$$

Example - Let $x_1,...,x_n$ be a random sample from a Pareto distribution

$$f(x;\theta) = \theta x^{-\theta-1} \text{ for } 1 \le x < \infty$$

$$L(\theta;\mathbf{x}) = \prod_{i=1}^{n}\theta x_i^{-\theta-1} = \theta^n\prod_{i=1}^{n}x_i^{-\theta-1}$$

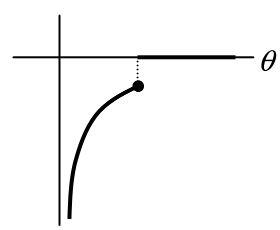
$$\ln L(\theta;\mathbf{x}) = n\ln(\theta) - (\theta+1)\sum_{i=1}^{n}\ln(x_i)$$

$$\frac{\partial \ln L}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^{n}\ln(x_i) = 0$$

$$\frac{n}{\theta} = \sum_{i=1}^{n}\ln(x_i)$$

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n}\ln(x_i)}$$

Example - Let $x_1,...,x_n$ be a random sample from a distribution with density



$$f(x;\theta) = \frac{1}{\theta} \text{ for } 0 \le x \le \theta\text{, 0 otherwise}$$

$$L(\theta;\mathbf{x}) = \prod_{i=1}^{n}\frac{1}{\theta} = \theta^{-n} \text{ for } 0 \le x_i \le \theta,\ i = 1,...n\text{, 0 otherwise}$$

$$\ln L(\theta;\mathbf{x}) = -n\ln(\theta) \text{ for } 0 \le x_i \le \theta,\ i = 1,...n\text{, 0 otherwise}$$

$$\frac{\partial \ln L}{\partial\theta} = \frac{-n}{\theta} = s(\theta;\mathbf{x}) - n\ln(\theta) \text{ for } 0 \le x_i \le \theta,\ i = 1,...n\text{, 0 otherwise}$$

Problem - there is a discontinuity at θ, $s(\theta;\mathbf{x})$ is not a function of x; can't set $s(\theta;\mathbf{x}) = 0$

Solution - want to make $s(\theta;\mathbf{x})$ as close to zero as possible so we want to make θ as large as possible; we know that $\theta \geq x_i \therefore$ min possible value of $s(\theta;\mathbf{x})$ is at $\hat{\theta} = \max(x_1,\ldots,x_n)$

Multiple Parameter Example - Let $x_1,...,x_n$ be a random sample from N(μ,σ 2)

$$L(\mu,\sigma^2;\mathbf{x}) = \prod_{i=1}^{n}(2\pi)^{-1/2}(\sigma^2)^{-1/2}\exp\left[\frac{-1}{2\sigma^2}(x_i-\mu)^2\right] =$$

$$(2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\ln L(\mu,\sigma^2;\mathbf{x}) = -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(x_i-\mu)(-1) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\partial \ln L}{\partial (\sigma^2)} = -\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$s(\mu,\sigma^2;\mathbf{x}) = \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \\ -\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0 \Rightarrow \left(\sum_{i=1}^{n}(x_i)\right)-n\mu = 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x}$$

Substitute into the second equation

$$-\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 = 0 \Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \text{n} \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

NOTE: This is a biased estimator

Properties of MLEs

Invariance - Let $\hat{\theta}$ be a MLE of θ. if $g(\bullet)$ is a function of θ then the MLE of $g(\theta)$ exists and is given by $g(\hat{\theta})$

Consistency and Uniqueness - under regularity conditions for Cramer-Rao lower bound, there exists a solution vector to the likelihood equations that is consistent $\lim_{n\to\infty}\Pr\left[\mathbf{S}(\hat{\boldsymbol{\theta}}_n)=\mathbf{0}\right]=1$ (existence) and $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}$ (consistency)

Asymptotic Normality - $\lim_{n\to\infty}\mathbf{I}(\boldsymbol{\theta})/n = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ under the regularity conditions of the CRLB

$\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}))$; i.e., MLEs have normal distribution at the limit

Asymptotic Efficiency - asymptotic variance of $\hat{\boldsymbol{\theta}}_n$ equals the limit of the CRLB

Inverse of 2x2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - cb}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Hypothesis Testing

MLE's are point estimates of parameters/coefficients; really have a distribution

Basic Concept - develop region in which we accept the hypothesis and one where we reject it

- **H** - represents all possible values of θ (point estimate)
- **Two regions** -
  - H_0 - null hypothesis; region where we accept the hypothesis
  - H_a - alternative hypothesis; region where we reject
- **Rules** -
  - **No Overlap** - $H_0 \cap H_a = \varnothing$
  - **Covers Everything** - $H_0 \cup H_a = H$
- **Example** - from midterm question 4; want to test the hypothesis that $\theta_1 = 1$
  - H_0: $\theta_1 = 1$
  - H_a: $\theta_1 \neq 1$ (but > 0... required in distribution)

Simple Hypothesis - only checks 1 value

- H_0: $\theta_1 = 1$
- H_a: $\theta_1 \neq 1$ (but > 0... required in distribution)... this is composite alternative)

Composite Hypothesis -

- H_0: $\theta_1 \geq 1$
- H_a: $\theta_1 < 1$

Accepting Null Hypothesis - if $\hat{\theta} \in H_0$; reject if $\hat{\theta} \in H_a$

Errors - statistically speaking we can never be 100% sure

- **Type 1 Error** - chance that we observe $\hat{\theta} \in H_a$ even though H_0 is true;
  - $\alpha = \Pr[\text{reject } H_0 | H_0 \text{ true}]$
- **Type 2 Error** - chance that we observe $\hat{\theta} \in H_0$ even though H_a is true
  - $\beta = \Pr[\text{fail to reject } H_0 | H_0 \text{ false}]$
- **Example** - suppose that the life of a light bulb is drawn from an exponential distribution
  - $f(x;\theta) = \exp(-x/\theta)/\theta$ ($x > 0$, $\theta > 0$); $F(x;\theta) = 1 - \exp(-x/\theta)$ ($x > 0$; $\theta > 0$); there are two types of light bulbs (ordinary and long life)
  - H_0: $\theta = 2{,}000$ hours
  - H_a: $\theta = 10{,}000$ hours
  - Accept H_0 if $x < d$, where x is the lifetime in hours of the light bulb that you test; reject H_0 (accept H_a) if $x > d$, where d is some set number of hours (e.g., $d = 6{,}000$). Can't be 100% sure because exponential is infinite distribution so there's always a chance that an ordinary bulb lasts more than 6,000 hours (Type I error) and there's a chance that a long life bulb lasts less than 6,000 hours (Type II error)
  - $\alpha = 1 - F(6000{,}2000) = 1 - (1 - \exp(-6000/2000)) = 0.049$... 5% chance that we say it's a long life bulb when it's really an ordinary bulb
  - $\beta = F(6000{,}10000) = 1 - \exp(-6000/10000) = 0.45$... so 45% of the time we'll wrongly label long life bulbs as ordinary bulbs

Critical Region - region where we reject H_0

- **Example** - from previous example, $x > 6{,}000$ is critical region

Level of Significance - probability of making a type 1 error; should choose what you're willing to accept before crunching the data

- **Trade-off Problem** - if we lower the chance of type 1 error, we increase the chance of type II error (i.e. lower power of test)

Long Run - level of significance tells you that if you conduct the test many times, that percentage of them will have a type I error (e.g., 5% significance, for every 20 tests, you'll expect to get 1 false rejection)

Power of a Test - 1 - Pr[Type 2 error] = Pr[reject H_0|H_0 false]; want to be as high as possible

Example - from light bulb example, Power = 1 - 0.45 = 0.55... extremely low

Trade-off Problem - would like to maximize power and minimize level of significance, but improving one is done at the expense of the other; from light bulb example, to increase significance, we need to increase d; to increase power, we need to decrease d; other solution is to use more light bulbs

Best Test - within a class of tests, on is said to be "best" if it has maximum power amongst all tests with level of significance (size) less than or equal to some given value

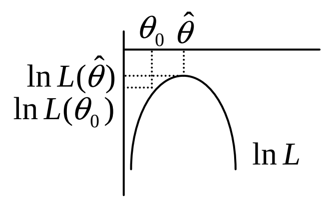
Consistent Test - power goes to one as sample size grows to infinity

Specific Tests

Assumptions for LRT, LMT, & WT - standard regularity conditions; information matrix is non-singular; know distribution (or density) that generates the data (i.e., $f(\mathbf{x};\boldsymbol{\theta}_0)$)

General - want to test whether $\theta = \theta_0$ in the known distribution; we know what the likelihood and log-likelihood functions are; just want to test whether or not the parameter value is θ_0

Likelihood Ratio Test - based on $\dfrac{L(\mathbf{x};\theta_0)}{L(\mathbf{x};\hat{\theta})}$



Look at difference between $\ln L(\mathbf{x};\theta_0) - \ln(\mathbf{x};\hat{\theta})$

H_0: $\theta = \theta_0$ (simple hypothesis)

H_a: $\theta \neq \theta_0$ (composite alternative)

Reject H_0 if this difference is "big"; accept H_0 if the difference is "small"

Under general conditions: $\boxed{\xi_{LR} = -2\left(\ln L(\mathbf{x};\theta_0) - \ln(\mathbf{x};\hat{\theta})\right) \sim \chi^2}$ (df equals number of parameters)

Example - consider T observations form a Bernoulli rv (x_t = 1 with probability θ and 0 with probability (1 - θ)); we wish to test H_0: $\theta = \theta_0$ (e.g., 1/2); H_a: $\theta \neq \theta_0$

$$\ln L(x;\theta) = \sum_{t=1}^{T}\left[x_t \ln\theta + (1-x_t)\ln(1-\theta)\right] = \ln\theta\sum_{t=1}^{T} x_t + \ln(1-\theta)\sum_{t=1}^{T}(1-x_t) =$$
$$T\bar{x}\ln\theta + T(1-\bar{x})\ln(1-\theta)$$

MLE to get $\hat{\theta}$

$$\frac{\partial \ln L(x;\theta)}{\partial\theta} = \frac{1}{\theta}\sum_{t=1}^{T} x_t + \frac{-1}{1-\theta}\sum_{t=1}^{T}(1-x_t) = 0$$

$$\frac{1}{\theta}T\bar{x} - \frac{T(1-\bar{x})}{1-\theta} = 0 \;\Rightarrow\; \frac{1}{\theta}T\bar{x} = \frac{T(1-\bar{x})}{1-\theta} \;\Rightarrow\; \frac{1}{\theta}\bar{x} = \frac{(1-\bar{x})}{1-\theta} \;\Rightarrow\; \bar{x}(1-\theta) = \theta(1-\bar{x}) \;\Rightarrow$$
$$\bar{x} - \bar{x}\theta = \theta - \bar{x}\theta \;\Rightarrow\; \hat{\theta} = \bar{x}$$

$$\xi_{LR} = -2\left(\ln L(\mathbf{x};\theta_0) - \ln(\mathbf{x};\hat{\theta})\right) =$$

$$-2\left[T\bar{x}\ln\theta_0 + T(1-\bar{x})\ln(1-\theta_0) - T\bar{x}\ln\bar{x} + T(1-\bar{x})\ln(1-\bar{x})\right] =$$

$$2T\left[\bar{x}\ln\left(\frac{\bar{x}}{\theta_0}\right) + (1-\bar{x})\ln\left(\frac{1-\bar{x}}{1-\theta_0}\right)\right] \sim \chi_1^2$$

Suppose you toss a coin 10 times and observe 8 heads. What is the test statistic for H_0: θ = 1/2 (i.e., have a fair coin)?

T = 10, $\bar{x}$ = 0.8, θ_0 = 0.5

$$\xi_{LR} = 2(10)\left[0.8\ln\left(\frac{0.8}{0.5}\right) + (0.2)\ln\left(\frac{0.2}{0.5}\right)\right] = 3.85$$

Critical value χ_1^2 for 0.95 (one tailed) is 3.84 ∴ reject H_0

Lagrange Multiplier Test - looks at slope of log likelihood at θ_0

H_0: $\theta = \theta_0$ (simple hypothesis)

H_a: $\theta \neq \theta_0$ (composite alternative)

If slope is "small" accept H_0; if it's "big" reject H_0; based on fact that at $\hat{\theta}$, slope of $\ln L$ is 0 (i.e., score = 0)

Set up as optimization: Max $\ln L(\mathbf{x};\theta)$ subject to $\theta = \theta_0$

Lagrangian: $\ell = \ln L(\mathbf{x};\theta) - \boldsymbol{\lambda}'(\theta - \theta_0)$, where $\boldsymbol{\lambda}$ is vector of shadow prices

First Order Conditions:

$$\frac{\partial \ell}{\partial \theta} = \frac{\partial \ln L(\mathbf{x};\theta)}{\partial \theta} - \boldsymbol{\lambda} = 0$$

$$-\frac{\partial \ell}{\partial \boldsymbol{\lambda}} = \theta - \theta_0 = 0$$

From first condition: $\boldsymbol{\lambda} = \dfrac{\partial \ln L(\mathbf{x};\theta)}{\partial \theta} = S(\mathbf{x};\theta)$ (score)

Distribution of the score function has mean zero and variance $I(\theta)$

$\dfrac{1}{\sqrt{n}} S(\mathbf{x};\theta) \xrightarrow{d} N(0, \boldsymbol{\Sigma}(0))$, $\boldsymbol{\Sigma}(0) = \lim\limits_{n\to\infty} \dfrac{I(\theta)}{n}$ ... (wave magic wand)

Test Statistic: $\boxed{\xi_{LM} = S^T(\mathbf{x};\theta_0) I^{-1}(\theta_0) S(\mathbf{x};\theta_0) \sim \chi_k^2}$ (k = number of parameters)

Example - Consider the previous example of Bernoulli trials

Remember that $\ln L(x;\theta) = T\bar{x}\ln\theta + T(1-\bar{x})\ln(1-\theta)$ and MLE is $\hat{\theta} = \bar{x}$

$$S(x;\theta) = \frac{\partial \ln L(x;\theta)}{\partial \theta} = \frac{T\bar{x}}{\theta} - \frac{T(1-\bar{x})}{1-\theta} = \frac{T(\bar{x}-\theta)}{\theta(1-\theta)}$$

$$\frac{\partial^2 \ln L(x;\theta)}{\partial \theta^2} = -\frac{T\bar{x}}{\theta^2} - \frac{T(1-\bar{x})}{(1-\theta)^2}$$

$$I(\theta) = -E\left(\frac{\partial^2 \ln L(x;\theta)}{\partial \theta^2}\right) = -E\left(-\frac{T\bar{x}}{\theta^2} - \frac{T(1-\bar{x})}{(1-\theta)^2}\right) = \frac{TE(\bar{x})}{\theta^2} + \frac{T(1-E(\bar{x}))}{(1-\theta)^2}$$

For T Bernoulli trials $E(\bar{x}) = \theta \therefore I(\theta) = \frac{T\theta}{\theta^2} + \frac{T(1-\theta)}{(1-\theta)^2} = T\left(\frac{1}{\theta} + \frac{1}{(1-\theta)}\right) = \frac{T}{\theta(1-\theta)}$

$$\xi_{LM} = S^T(\mathbf{x};\theta_0)I^{-1}(\theta_0)S(\mathbf{x};\theta_0) = \frac{T(\bar{x}-\theta_0)}{\theta_0(1-\theta_0)}\frac{\theta_0(1-\theta_0)}{T}\frac{T(\bar{x}-\theta_0)}{\theta_0(1-\theta_0)} = \frac{T(\bar{x}-\theta_0)^2}{\theta_0(1-\theta_0)}$$

Suppose T = 10 and $\bar{x}$ = 0.8 and θ_0 = 0.5 (assuming fair coin)

$$\xi_{LM} = \frac{10(0.8-0.5)^2}{0.5(1-0.5)} = \frac{0.9}{0.25} = 3.6$$

Critical value χ_1^2 for 0.95 (one tailed) is 3.84 $\therefore$ don't reject H_0

Wald Test - looks at distance between θ_0 and $\hat{\theta}$; if distance is "small" accept H_0; if it's "big" reject H_0; based on asymptotic normality of the MLE

$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0, \mathbf{\Sigma}^{-1}(0))$, $\mathbf{\Sigma}(0) = \lim_{n\to\infty}\frac{I(\theta)}{n}$ ... (wave magic wand)

Test Statistic: $\boxed{\xi_W = (\hat{\theta}-\theta_0)^T I(\hat{\theta})(\hat{\theta}-\theta_0) \sim \chi_k^2}$ (k = number of parameters)

Example - same as before

$\ln L(x;\theta) = T\bar{x}\ln\theta + T(1-\bar{x})\ln(1-\theta)$ and MLE is $\hat{\theta} = \bar{x}$

$S(x;\theta) = \frac{T(\bar{x}-\theta)}{\theta(1-\theta)}$ and $I(\theta) = \frac{T}{\theta(1-\theta)}$

$$\xi_W = (\hat{\theta}-\theta_0)^T I(\hat{\theta})(\hat{\theta}-\theta_0) = (\bar{x}-\theta_0)\frac{T}{\bar{x}(1-\bar{x})}(\bar{x}-\theta_0) = \frac{T(\bar{x}-\theta_0)^2}{\bar{x}(1-\bar{x})}$$

Suppose T = 10 and $\bar{x}$ = 0.8 and θ_0 = 0.5 (assuming fair coin)

$$\xi_W = \frac{10(0.8-0.5)^2}{0.8(1-0.8)} = \frac{0.9}{0.16} = 5.625$$

Critical value χ_1^2 for 0.95 (one tailed) is 3.84 $\therefore$ reject H_0

Note - Likelihood ratio test doesn't account for variance (or information matrix); Lagrange multiplier and Wald tests do; but LRT is generally easiest to do

Testing Restrictions on Subset of Parameters

Under H_0: $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^o$

$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix}_{kx1}$, $\boldsymbol{\theta}_1$ (k_1x1) restricted under H_0; $\boldsymbol{\theta}_2$ (k_2x1) unrestricted under H_0

Let the MLE of $\boldsymbol{\theta}_2$ under H_0 be $\tilde{\boldsymbol{\theta}}_2$ so $\tilde{\boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{\theta}_1^o \\ \tilde{\boldsymbol{\theta}}_2 \end{bmatrix}_{kx1}$, whereas $\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_1 \\ \hat{\boldsymbol{\theta}}_2 \end{bmatrix}_{kx1}$ (use unrestricted MLE for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$)

Likelihood Ratio Test - $\xi_{LR} = -2\left(\ln L(\mathbf{x};\tilde{\boldsymbol{\theta}}) - \ln(\mathbf{x};\hat{\boldsymbol{\theta}})\right) \sim \chi^2_{k_1}$

Lagrange Multiplier Test -

$$\xi_{LM} = S^T(\mathbf{x};\tilde{\boldsymbol{\theta}})\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})S(\mathbf{x};\tilde{\boldsymbol{\theta}}) \sim \chi^2_{k_1}$$

$$\left[S_1(\mathbf{x};\tilde{\boldsymbol{\theta}})^T \quad \mathbf{0}^T\right]\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}})\begin{bmatrix} S_1(\mathbf{x};\tilde{\boldsymbol{\theta}}) \\ \mathbf{0}^T \end{bmatrix}$$

Choose the upper left-hand corner of the inverse of the information matrix

$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \mathbf{I}_{21} & \mathbf{I}_{22} \end{bmatrix}$, upper left-hand corner of inverse is $\left[\mathbf{I}_{11} - \mathbf{I}_{12}\mathbf{I}_{22}^{-1}\mathbf{I}_{21}\right]^{-1}$

$\therefore\ \xi_{LM} = S^T(\mathbf{x};\tilde{\boldsymbol{\theta}})\left[\mathbf{I}_{11} - \mathbf{I}_{12}\mathbf{I}_{22}^{-1}\mathbf{I}_{21}\right]^{-1}S(\mathbf{x};\tilde{\boldsymbol{\theta}}) \sim \chi^2_{k_1}$, with $\mathbf{I}_{ij}$ evaluated at $\tilde{\boldsymbol{\theta}}$

Wald Test -

$$\xi_W = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o)^T\left[\mathbf{I}_{11}(\hat{\boldsymbol{\theta}}) - \mathbf{I}_{12}(\hat{\boldsymbol{\theta}})\mathbf{I}_{22}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{I}_{21}(\hat{\boldsymbol{\theta}})\right](\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^o) \sim \chi^2_{k_1}$$

Example - from HW5 problem 2, test H_0: $p_1 = 0.6$ (H_1: $p_1 \neq 0.6$)

$f(\mathbf{x};\boldsymbol{\theta}) = p_1^{x_1}p_2^{x_2}p_3^{x_3}(1-p_1-p_2-p_3)^{x_4}$, $x_1+x_2+x_3+x_4$,

where $\mathbf{x} = [x_1 \quad x_2 \quad x_3 \quad x_4]^T$ and $\boldsymbol{\theta} = [p_1 \quad p_2 \quad p_3]^T$

$$L(\mathbf{x};\boldsymbol{\theta}) = \prod_{i=1}^{N} f(\mathbf{x}_i;\boldsymbol{\theta}) = \prod_{i=1}^{N} p_1^{x_{1i}}p_2^{x_{2i}}p_3^{x_{3i}}(1-p_1-p_2-p_3)^{x_{4i}} =$$

$p_1^{n_1}p_2^{n_2}p_3^{n_3}(1-p_1-p_2-p_3)^{N-n_1-n_2-n_3}$, where $n_j = \sum_{i=1}^{N} x_{ji}$, $j = 1,2,3$

Note: $E(n_j) = \sum_{i=1}^{N}\frac{n_j}{N} = \sum_{i=1}^{N} p_j = Np_j$

Note: for simplicity substitute $n_4 = N - n_1 - n_2 - n_3$

$$L(\mathbf{x};\boldsymbol{\theta}) = p_1^{n_1}p_2^{n_2}p_3^{n_3}(1-p_1-p_2-p_3)^{n_4}$$

$$\ln L(\mathbf{x};\boldsymbol{\theta}) = n_1\ln(p_1) + n_2\ln(p_2) + n_3\ln(p_3) + n_4\ln(1-p_1-p_2-p_3)$$

Score $S(\mathbf{x},\boldsymbol{\theta})$ = $\begin{bmatrix} \dfrac{\partial \ln L(\mathbf{x};\boldsymbol{\theta})}{\partial p_1} \\ \dfrac{\partial \ln L(\mathbf{x};\boldsymbol{\theta})}{\partial p_2} \\ \dfrac{\partial \ln L(\mathbf{x};\boldsymbol{\theta})}{\partial p_3} \end{bmatrix}$ = $\begin{bmatrix} n_1 \dfrac{1}{p_1} - n_4 \dfrac{1}{1-p_1-p_2-p_3} \\ n_2 \dfrac{1}{p_2} - n_4 \dfrac{1}{1-p_1-p_2-p_3} \\ n_3 \dfrac{1}{p_3} - n_4 \dfrac{1}{1-p_1-p_2-p_3} \end{bmatrix}$

Standard MLEs: solve $S(\mathbf{x},\boldsymbol{\theta})$ = **0** for $\boldsymbol{\theta}$; lots of complicated algebra will show that

$$\hat{p}_j = \frac{n_j}{N},\; j = 1,2,3$$

Restricted MLEs: plug in 0.6 for p_1 and resolve for non-restricted parameters (p_2 and p_3)

$$\frac{\partial \ln L(\mathbf{x};\boldsymbol{\theta})}{\partial p_2} = n_2 \frac{1}{p_2} - n_4 \frac{1}{1-0.6-p_2-p_3} = 0$$

$$\frac{\partial \ln L(\mathbf{x};\boldsymbol{\theta})}{\partial p_3} = n_3 \frac{1}{p_3} - n_4 \frac{1}{1-0.6-p_2-p_3} = 0$$

Solve first one for p_2

$$\frac{0.4n_2 - p_2 n_2 - p_3 n_2}{p_2} = n_4 \;\Rightarrow\; \frac{0.4n_2 - p_3 n_2}{p_2} = n_4 + n_2 \;\Rightarrow\; p_2 = \frac{0.4n_2 - p_3 n_2}{n_4 + n_2}$$

Substitute into the second one and solve for p_3

$$n_3 \frac{1}{p_3} - n_4 \frac{1}{1-0.6-\left(\dfrac{0.4n_2 - p_3 n_2}{n_4+n_2}\right) - p_3} = 0 \;\Rightarrow$$

$$\frac{n_3}{p_3} - \frac{n_4(n_4+n_2)}{0.4n_4 + 0.4n_2 - 0.4n_2 + p_3 n_2 - p_3 n_4 - p_3 n_2} = 0 \;\Rightarrow\; \frac{0.4n_3 - p_3 n_3}{p_3} = n_4 + n_2$$

$$\Rightarrow\; \frac{0.4n_3}{p_3} = n_2 + n_3 + n_4 \;\Rightarrow\; p_3 = \frac{0.4n_3}{n_2+n_3+n_4} = \frac{0.4n_3}{N-n_1}$$

Go back and solve p_2

$$p_2 = \frac{0.4n_2 - \dfrac{0.4n_3}{n_2+n_3+n_4} n_2}{n_4+n_2} \;\Rightarrow\; p_2 = \frac{\dfrac{0.4n_2^2 + 0.4n_2n_3 + 0.4n_2n_4 - 0.4n_2n_3}{n_2+n_3+n_4}}{n_4+n_2} \;\Rightarrow$$

$$p_2 = \frac{\dfrac{0.4n_2(n_2+n_4)}{n_2+n_3+n_4}}{n_4+n_2} \Rightarrow p_2 = \frac{0.4n_2}{n_2+n_3+n_4} = \frac{0.4n_2}{N-n_1}$$

Compute all the values we'll need for the test statistics

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} n_1/N \\ n_2/N \\ n_3/N \end{bmatrix} = \begin{bmatrix} 315/556 \\ 108/556 \\ 101/556 \end{bmatrix} = \begin{bmatrix} 0.5665 \\ 0.1942 \\ 0.1817 \end{bmatrix}$$

$$\tilde{\boldsymbol{\theta}} = \begin{bmatrix} 0.6 \\ 0.4n_2/(N-n_1) \\ 0.4n_3/(N-n_1) \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.4(108)/(556-315) \\ 0.4(101)/(556-315) \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.1793 \\ 0.1676 \end{bmatrix}$$

$$S(\mathbf{x};\boldsymbol{\theta}_0) = \begin{bmatrix} n_1\dfrac{1}{p_1} - n_4\dfrac{1}{1-p_1-p_2-p_3} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \dfrac{315}{0.6} - \dfrac{32}{1-0.6-0.1793-0.1676} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -77.5 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{I}(\boldsymbol{\theta}_0) = \left[\begin{array}{c|cc} 11395.1 & 10468.44 & 10468.44 \\ \hline 10468.44 & 13570.2 & 10468.44 \\ 10468.44 & 10468.44 & 13785.17 \end{array}\right]$$

$$\left[\mathbf{I}_{11} - \mathbf{I}_{12}\mathbf{I}_{22}^{-1}\mathbf{I}_{21}\right]^{-1} =$$

$$\left[11395.1 - \begin{bmatrix} 10468.44 & 10468.44 \end{bmatrix} \begin{bmatrix} 13570.2 & 10468.44 \\ 10468.44 & 13785.17 \end{bmatrix}^{-1} \begin{bmatrix} 10468.44 \\ 10468.44 \end{bmatrix}\right]^{-1} =$$

$$\left[11395.1 - \begin{bmatrix} 10468.44 & 10468.44 \end{bmatrix} \begin{bmatrix} 0.000178 & 0.000135 \\ 0.000135 & 0.000175 \end{bmatrix} \begin{bmatrix} 10468.44 \\ 10468.44 \end{bmatrix}\right]^{-1} =$$

$$\begin{bmatrix} 0.000432 \end{bmatrix}$$

$$\mathbf{I}(\hat{\boldsymbol{\theta}}) = \left[\begin{array}{c|cc} 10641.88 & 9660.5 & 9660.5 \\ \hline 9660.5 & 12522.87 & 9660.5 \\ 9660.5 & 9660.5 & 12721.25 \end{array}\right]$$

$$\left[\mathbf{I}_{11} - \mathbf{I}_{12}\mathbf{I}_{22}^{-1}\mathbf{I}_{21}\right] = \left[10641.88 - \begin{bmatrix} 9660.5 & 9660.5 \end{bmatrix} \begin{bmatrix} 12522.87 & 9660.5 \\ 9660.5 & 12721.25 \end{bmatrix}^{-1} \begin{bmatrix} 9660.5 \\ 9660.5 \end{bmatrix}\right] =$$

$$\left[10641.88 - \begin{bmatrix} 9660.5 & 9660.5 \end{bmatrix} \begin{bmatrix} 0.000193 & -0.00015 \\ -0.00015 & 0.00019 \end{bmatrix}^{-1} \begin{bmatrix} 9660.5 \\ 9660.5 \end{bmatrix}\right] = \begin{bmatrix} 2264.1 \end{bmatrix}$$

Critical Value = χ_1^2 = 3.84

Likelihood Ratio Test -

$\xi_{LR} = -2\left(\ln L(\mathbf{x};\tilde{\boldsymbol{\theta}}) - \ln(\mathbf{x};\hat{\boldsymbol{\theta}})\right) =$

$-2\big(315\ln(0.6) + 108\ln(0.1793) + 101\ln(0.1676) + 32\ln(0.0531) -$

$315\ln(0.5665) - 108\ln(0.1942) - 101\ln(0.1817) - 32\ln(0.0576)\big)$ = **2.571** < 3.84 $\therefore$ don't reject H_0. The data seems to agree with p_1 = 0.6

Lagrange Multiplier Test -

$\xi_{LM} = S_1^T(\mathbf{x};\tilde{\boldsymbol{\theta}})\left[\mathbf{I}_{11} - \mathbf{I}_{12}\mathbf{I}_{22}^{-1}\mathbf{I}_{21}\right]^{-1} S_1(\mathbf{x};\tilde{\boldsymbol{\theta}}) = [77.5][0.000432][77.5]$ = **2.592** < 3.84 $\therefore$ don't reject H_0. The data seems to agree with p_1 = 0.6

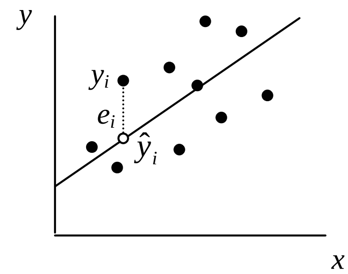
Wald Test -

$\xi_W = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{\text{o}})^T \left[\mathbf{I}_{11}(\hat{\boldsymbol{\theta}}) - \mathbf{I}_{12}(\hat{\boldsymbol{\theta}})\mathbf{I}_{22}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{I}_{21}(\hat{\boldsymbol{\theta}})\right](\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{\text{o}})$

$(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^{\text{o}}) = [0.5665] - [0.6] = [-0.0335]$

$\xi_W = [-0.335][2264.1][-0.0335]$ = **2.533** < 3.84 $\therefore$ don't reject H_0. The data seems to agree with $p_1 = 0.6$

Regression Analysis



Linear Regression with Constant and Slope

Idea - minimize $\sum_{i=1}^{n} e_i^2$ (i.e., want to minimize the sum of square of e_i's)

$e_i = y_i - \hat{y}_i$ (i.e., the vertical distance between observation and fitted line), where $\hat{y}_i$ is the fitted value

If $\sum_{i=1}^{n} e_i^2$ = 0, the line is a perfect fit

Fit Line - $f(x_i) = a + bx_i$

Solve optimization problem: Min $Q = \sum_{i=1}^{n} \left(y_i - a - bx_i\right)^2$

First Order Conditions - $\dfrac{\partial Q}{\partial a} = -2\sum_{i=1}^{n} \left(y_i - a - bx_i\right)$ = 0 (1)

$\dfrac{\partial Q}{\partial b} = -2\sum_{i=1}^{n} x_i\left(y_i - a - bx_i\right)$ = 0 (2)

From (1): $\dfrac{1}{n}\sum_{i=1}^{n} y_i = \dfrac{1}{n}\left(a + bx_i\right) \Rightarrow \bar{y} = a + b\bar{x} \Rightarrow a = \bar{y} - b\bar{x}$

Substitute this into (2): $\sum_{i=1}^{n} x_i y_i - (\bar{y} - b\bar{x})x_i - bx_i^2$ = 0

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i - (\bar{y} - b\bar{x})\frac{1}{n}\sum_{i=1}^{n} x_i - b\frac{1}{n}\sum_{i=1}^{n} x_i^2 = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i - (\bar{y} - b\bar{x})\bar{x} = b\frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy} = b\frac{1}{n}\sum_{i=1}^{n} x_i^2 - b\bar{x}^2$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = b\sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\boxed{\hat{b} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} = \frac{S_{xy}}{S_x^2}$$

$$\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) =$$

$$\frac{1}{n}\sum_{i=1}^{n} (x_i y_i - \bar{x}y_i - x_i\bar{y} + \overline{xy}) =$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\frac{1}{n}\sum_{i=1}^{n} y_i - \bar{y}\frac{1}{n}\sum_{i=1}^{n} x_i + \overline{xy} =$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy} - \overline{xy} + \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}$$

where $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$ and $S_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$

Substitute into equation for a:

$$\boxed{\hat{a} = \bar{y} - \hat{b}\bar{x}}$$

Goodness of Fit

$\sum_{i=1}^{n} e_i$ = 0

$e_i = y_i - a - bx_i$

$\frac{1}{n} e_i = \frac{1}{n}\sum_{i=1}^{n} y_i - a - b\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{y} - a - b\bar{x}$

Substitute $\hat{a}$ and $\hat{b}$

$\frac{1}{n} e_i = \bar{y} - (\bar{y} - \hat{b}\bar{x}) - \hat{b}\bar{x} = \bar{y} - \bar{y} + \hat{b}\bar{x} - \hat{b}\bar{x}$ = 0

$\therefore \sum_{i=1}^{n} e_i$ tells us nothing about how well the estimates fit the data. Are the actual y_i's exactly on the line? or widely scattered about it?

Multiple R^2 - measure of goodness of fit; tells us the percent of variation in the data (y) that the variable(s) (x) can explain; want R^2 as high as possible, but is possible to have high R^2 without explaining anything (e.g., time series); as you add variables R^2 increases

Residual Sum of Squares - RSS = $\sum_{i=1}^{n} e_i^2$ (fit constant plus slope term)

Total Sum of Squares - TSS = $\sum_{i=1}^{n} (y_i - \bar{y})^2$ (equivalent to fitting a constant [$\bar{y}$] to the data)

$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$; RSS $\le$ TSS $\Rightarrow 0 \le R^2 \le 1$

Classical Linear Regression Model

Assumptions

(1) The model is known to be $\overset{n \times 1\ \text{vector}}{\mathbf{y}} = \underset{n \times k\ \text{matrix}}{\mathbf{x}}\overset{n \times 1\ \text{vector}}{\boldsymbol{\beta}} + \underset{n \times 1\ \text{vector}}{\boldsymbol{\varepsilon}}$

Assume $\boldsymbol{\beta}$ is finite

Note $\mathbf{x}$ will have a column of 1's for the constant term

(2) $\mathbf{x}$ is a non-stochastic, finite, n x k matrix; i.e., x's are not random variables

(3) $\mathbf{x}^T\mathbf{x}$ is non-singular for all $n > k$, i.e., exists

(4) $E(\boldsymbol{\varepsilon}) = \mathbf{0}$

(5) $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, $\sigma^2 < \infty$

Estimators

$$\boxed{\hat{\boldsymbol{\beta}} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y}}$$

To show this minimize sum of squared errors SSE = $(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})$

First order condition: $\frac{\partial \text{SSE}}{\partial \boldsymbol{\beta}^T} = -2\mathbf{x}^T(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) = 0$

$\mathbf{x}^T\mathbf{y} = \mathbf{x}^T\mathbf{x}\boldsymbol{\beta}$

$(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{x}\boldsymbol{\beta} = \boldsymbol{\beta}$

Properties

(1) **Existence** - given assumptions 1-3, $\hat{\boldsymbol{\beta}}_{OLS}$ exists for all $n > k$ and is unique

i.e., $(\mathbf{x}^T\mathbf{x})^{-1}$ exists then $(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y}$ exists and will be unique (since there is only one $(\mathbf{x}^T\mathbf{x})^{-1}$)

(2) **Unbiased** - given assumptions 1-4, $E(\hat{\boldsymbol{\beta}}_{OLS}) = \boldsymbol{\beta}$

$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y}$

From assumption 1: $(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T(\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\varepsilon})$

Multiply it out: $(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{x}\boldsymbol{\beta}+(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\boldsymbol{\varepsilon} = \boldsymbol{\beta}+(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\boldsymbol{\varepsilon}$

$E(\hat{\boldsymbol{\beta}}_{OLS}) = E\left[\boldsymbol{\beta}+(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\boldsymbol{\varepsilon}\right]$

From assumption 2 ($\mathbf{x}$ not random variable): $E\left[\boldsymbol{\beta}+(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\boldsymbol{\varepsilon}\right] = \boldsymbol{\beta}+(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T E(\boldsymbol{\varepsilon})$

From assumption 4 ($E(\boldsymbol{\varepsilon}) = \mathbf{0}$): $\boldsymbol{\beta}+(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T E(\boldsymbol{\varepsilon}) = \boldsymbol{\beta}$

$\therefore\ \hat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{\beta}$ so $\hat{\boldsymbol{\beta}}_{OLS}$ is unbiased estimator

(3) **Normal** - given assumptions 1-5, $\hat{\boldsymbol{\beta}}_{OLS} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{x}^T\mathbf{x})^{-1})$

(4) **Efficiency** - given assumptions 1-5, $\hat{\boldsymbol{\beta}}_{OLS}$ is the maximum likelihood estimator (MLE) and best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$

Proof: need to show that $\hat{\boldsymbol{\beta}}_{OLS}$ is MLE of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}_{OLS}$ attains the Cramer-Rao Lower Bound (CRLB)

$y_i = \mathbf{x}_i\boldsymbol{\beta}+\varepsilon_i$ ($\mathbf{x}_i$ is row vector of single observation; non-stochastic; $\varepsilon_i \sim N(0,\sigma^2)$)

$$\therefore\ f(y_i,\mathbf{x}_i,\boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}(y_i-\mathbf{x}_i\boldsymbol{\beta})^2\right]$$

$$\ln L(y_i,\mathbf{x}_i,\boldsymbol{\beta}) = -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})$$

$\frac{\partial \ln L(y_i,\mathbf{x}_i,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}$ = (same thing as for minimizing least squares) =

$-\frac{1}{2\sigma^2}\left[-2\mathbf{x}^T(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})\right] = \mathbf{0} \Rightarrow \mathbf{x}^T\mathbf{y}-\mathbf{x}\boldsymbol{\beta} \Rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y}\ \therefore \hat{\boldsymbol{\beta}}_{OLS}$ is MLE

$\frac{\partial \ln L(y_i,\mathbf{x}_i,\boldsymbol{\beta})}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2}+\frac{1}{2(\sigma^2)^2}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) = \mathbf{0} \Rightarrow$ (substitute in the MLE estimate for $\hat{\boldsymbol{\varepsilon}} = \mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}$) $\Rightarrow\ n = \frac{1}{\sigma^2}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}$ (MLE, but biased)

Note: most statistical packages use $s^2 = \frac{1}{n-k}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}$ (not the MLE, but it's unbiased)

To get CRLB need $\frac{\partial^2 \ln L(y_i, \mathbf{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} = -\frac{1}{\sigma^2}(\mathbf{x}^T\mathbf{x})$... wave magic wand... CRLB of $Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{x}^T\mathbf{x})^{-1}$ (assuming σ^2 is given) $\therefore$ $\hat{\boldsymbol{\beta}}_{OLS}$ attains CRLB

Test Linear Restrictions on Coefficients

Consider model: $\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \end{bmatrix}$ (as long as # parameters k < # of observations n)

Want to test hypothesis of form:

H$_0$: $\mathbf{R}\cdot\boldsymbol{\beta}$ - $\mathbf{q}$ = $\mathbf{0}$, where $\mathbf{R}$ is matrix describing the combinations of coefficients

H$_a$: $\mathbf{R}\cdot\boldsymbol{\beta}$ - $\mathbf{q} \neq \mathbf{0}$,

We are interested in testing restrictions such as: (assume only 3 parameters)

(a) $\beta_1 = 0 \Rightarrow \mathbf{R} = [1 \quad 0 \quad 0]$; $\mathbf{q} = [0]$

(b) $\beta_2 = \beta_3 \Rightarrow \mathbf{R} = [0 \quad 1 \quad -1]$; $\mathbf{q} = [0]$

(c) $\beta_1 + \beta_2 = 1 \Rightarrow \mathbf{R} = [1 \quad 1 \quad 0]$; $\mathbf{q} = [1]$

If we want to jointly test all 3 restrictions at once: $\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{bmatrix}$; $\mathbf{q} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Note: testing too much at once can burry insignificant results

Note: need to have rank($\mathbf{R}$) = # of restrictions (i.e., no redundant restrictions)

Under H$_0$: $\mathbf{R}\cdot\boldsymbol{\beta}$ - $\mathbf{q}$ ~ $N(\mathbf{0}, \sigma^2\mathbf{R}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^T) \Rightarrow (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})^T\left[\sigma^2\mathbf{R}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^T\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) \sim \chi^2_m$

(m = # linearly independent restrictions = rank($\mathbf{R}$))

But don't know σ^2 so we have to estimate it with $s^2 = \frac{1}{n-k}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} \sim \chi^2_{n-k}$

(n = # of observations; k = # of parameters)

So use new test statistic: $\frac{\left\{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})^T\left[\mathbf{R}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^T\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})\right\}/m}{\underbrace{\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}/(n-k)}_{= s^2}} \sim F_{m,n-k}$

Using Stata

Welcome to Stata, one of the most un-user friendly programs ever created. Although later versions have some features to make it easier to use, they more than make up for it by not being 100% compatible with previous versions... and some even generate code through the "helpful" dialog boxes that cause errors. This tutorial is intended to give you the very basic tools needed to get things done in Stata.

How this document is organized:

- `Courier font` is what you type or see in Stata.
- Things in `[brackets]` are optional arguments.
- `Blue text` are reserved commands in Stata; the underlined part of the next is how you can abbreviate the command.
- `Red text` are Stata error messages.
- File extensions are listed (e.g., `.do`, `.dta`, `.log`) to let you know what Stata is expecting. Stata doesn't not require the extensions to be used if the file is of the type expected.

How Stata Works

Stata uses a combination of command-line and menu-driven inputs along with a very basic spreadsheet-style data editor. There are several types of files you can work with in Stata, but the basic one is the Stata data set file (`.dta`). This file contains all your data as well as variable names.

Extension	File Type
.dta	Stata data file
.log	Log file (explained below)
.raw	ASCII (text) file

PgUp and PgDn buttons scroll through commands in the Review window (i.e., writes them in the Stata Command window for you)

Data Sets

You can view the data Stata is working with by opening the data editor or data browser. Both of these work similar to a spreadsheet with the variables listed in columns. The only difference is that the editor lets you change values and the browser doesn't. If you insist on using the command line, you can use the `list` command. Although it's old school, list could help find your problem areas when used in conjunction with `if`. For example:

```
list varname if varname > 5000
```

will list all observations of the variable `varname` that are greater then 5000. The good thing is that each one has the observation number. You can jot that down and look up the data points in the data editor or you can go back to your original data source to track down potential problems.

Save Data

```
save filename.dta [,*]
```

Options:

- `nolabel` - omits value labels; still saves associations between variables and value label names (just not the labels themselves).
- `replace` - allows you to overwrite the existing file; prevents `"file filename.dta already exists"` error.
- `orphans` - saves all value labels, including those not attached to any variables
- `emptyok` - allows you to save an empty data set to prevent `"no variables defined"` error. (Used for programming.)
- `intercooled` - makes Stata/SE save in Intercooled Stata format.

Load Data (`.dta` Files)

```
use filename.dta [,*]
```

Options:

- `clear` - Stata will not let you load a data file if you already have a data in memory. Using the `clear` option removes any data from memory (even if it hasn't been saved) to allow you to load the data file.

Load Data (Other Sources)

Formatted Text File - one observation per line; values are tab or comma delimited; can have variable names in the first line (optional); if you don't include file extension, `.raw` is assumed.

```
insheet [varlist] using filename.raw [,*]
```

Options:

- `varlist` - list variables names separated by spaces (not commas)
- `double` - forces Stata to store variables as doubles (rather than floats)
- `no[names]` - informs Stata whether variable names are included; Stata will figure it out on its own, but this option will allow the file to open faster
- `comma` - specify comma delimited (not required)
- `tab` - specify tab delimited (not required)

`delimiter("*")` - specify a different delimiter in the data (e.g., `delimiter(";")`)

`clear` - removes any data from memory (even if it hasn't been saved) to allow you to load the data file.

Examples:

```
insheet using newdata
insheet using newerdata.txt, clear
insheet using weirddata.txt, clear delimiter("&")
insheet height gender mom dad using heights.dat
```

Log Files

These keep track of everything that happens during your Stata session by recording everything that appears in the Stata Results window (the one with the black background). A log file can be handy for tracking down errors when you're running a `.do` file. If you specify the `.log` extension , the file is saved in ASCII (text) format which means the colors are not saved, but it's pretty easy to tell the difference between your commands and Stata output because commands are preceded by a period (`.`). There's a different format for the Stata viewer, but it's not really any better than a text file. There are also other options for a log file than aren't covered here, but this section should give you all you need to know. The only tricky part is deciding where (or if) to turn the log file on or off during execution of your .do file or Stata session. You don't really need to close the log file to be able to read it.

```
log using filename.log [,*]
log off
log on
log close
```

Options:

`replace` - overwrites current log file

`append` - adds this session to the end of the log file

text |

Examples:

```
log using newlog.log, replace
log using "file with spaces.log"
log close
```

Commands

Stata commands are the things that get things done in Stata. They are how you tell Stata to do what it is you want done.

Note: `exp` refers to any expression, logical or mathematical; the type should be clear in the context; if `exp` is written twice in a single line, it does not imply that it is the same expression. Expressions use the following operators:

Arithmetic		Logical		Relational	
+	addition	~	not	>	greater than
-	subtraction	!	not	<	less than
*	multiplication	|	or (shift \\)	>=	> or equal
/	division	&	and	<=	< or equal
^	power			==	equal
+	string concatenation			~=	not equal
				!=	not equal

Generate - creates a new variable based on `exp`

```
generate [type] newvar[:lblname] = exp [if exp]
```

Options:
`type` - specifies the variable type; if none is specified, Stata will automatically select `float` for numeric data and `str` for text

Examples:

```
generate age2 = age*age
generate biginc = income>100000 & income!=.
gen double unitpr = cost/quantity
gen byte biginc = income>100000 & income!=.
gen xlag = x[_n-1]
```

List - prints data on the screen

```
list [varlist] [if exp] [, *]
```

Options:
`table` - lists variables vertically, one observation per row

	make	price	mpg	rep78
1.	AMC Concord	4,099	22	3
2.	AMC Pacer	4,749	17	3
3.	AMC Spirit	3,799	22	.
4.	Buick Century	4,816	20	3

display - lists observations together; useful if there are a lot of variables to keep it from wrapping around the screen

1.

make	price	mpg	rep78	headroom	trunk
AMC Concord	4,099	22	3	2.5	11

weight	length	turn	displa~t	gear_r~o	foreign
2,930	186	40	121	3.58	Domestic

2.

make	price	mpg	rep78	headroom	trunk
AMC Pacer	4,749	17	3	3.0	11

weight	length	turn	displa~t	gear_r~o	foreign
3,350	173	40	258	2.53	Domestic

Replace - changes the contents of an existing variable

```
replace oldvar = expression1 [if expression] [, nopromote]
```

Options:
  oldvar - name of a variable that already exists in the data set
  nopromote - prevents replace from promoting the variable type to accommodate the change (e.g., if you replace an integer variable with data containing 3.14 and prevent the type to promote, you'll end up with 3)

Examples:

```
replace income=. if income<=0
replace age = 25 in 1007
```

Set Memory - specifies how much system memory you want to be dedicated to Stata ; **Note:** typing memory without set before it will display a report of Stata's memory usage

```
set memory #[b|k|m|g] [, permanently ]
```

Options:
  # - amount of memory to set; specified in terms of bytes (b), kilobytes (k), megabytes (m), or gigabytes (g)
  permanently - specifies that in addition to making the change right now, Stata will remember the new limit and use it in the future when you open Stata

Examples:

```
set memory 5m
```

Set Type - specifies the default data type assigned to new variables (such as by generate) when the storage type is not explicitly specified

```
set type *
```

where * is either a numeric storage type listed here or a string explained below the table

Numeric Storage Type	Bytes	Minimum	Maximum	Closets to 0 without being 0
`byte`	1	-127	100	+/-1
`int`	2	-32,767	32,740	+/-1
`long`	4	-2,147,483,647	2,147,483,620	+/-1
`float`	4	-1.70141173319*10^38	1.70141173319*10^36	+/-10^-36
`double`	8	-8.9884656743*10^307	8.9884656743*10^308	+/-10^-323

Precision for float is 3.795x10^-8
Precision for double is 1.414x10^-16

Character strings are specified by `str#`, where # gives the maximum length of the string (ranges from 1 to 80). Each character reserved by a string takes one byte regardless of the data stored in the string (e.g., "`it`" stored in a variable of type `str80`, still takes up 80 bytes).

Summarize

```
summarize [varlist] [if expression] [, detail]
```

Options:

`varlist` - list of variables, separated by spaces (not commas); if you don't indicate a variable list, Stata will summarize all the variables in the data set

`if expression` - allows you to specify a subset of the data to be summarized

`detail` - standard summarize command lists number of observations, mean, standard deviation, minimum and maximum; specifying `detail` adds 1, 5, 10, 15, 75, 90, 95, 99th percentiles, variation, skewness, and kurtosis

Functions

Functions are actually series of embedded commands designed to accomplish a specific task. They make working with Stata a little easier because you don't have to program them in yourself. This is just a subset of frequently used functions. You can get more functions by using the online help in Stata and searching for these.

Type of function	See help
Mathematical Functions	`mathfun`
Probability Functions	`probfun`
Random Numbers	`random`
String Functions	`strfun`
Programming Functions	`progfun`
Date Functions	`datefun`
Time-series Functions	`tsfun`

Mathematical Functions

`abs(x)`	returns the absolute value of x
`exp(x)`	returns the ex
`int(x)`	returns the integer obtained by truncated x towards zero
`ln(x)` or `log(x)`	returns the natural logarithm of x
`log10(x)`	returns the base 10 logarithm of x
`max(x1,x2,...,xn)`	returns the maximum of x1, x2, ..., xn (missing values are ignored)
`min(x1,x2,...,xn)`	returns the minimum of x1, x2, ..., xn (missing values are ignored)
`round(x,y)`	returns x rounded off to units of y
`sqrt(x)`	returns the square root of x

Probability Functions

`binomial(n,k,p)`	returns the probability of k or more successes in n trials when the probability of a success on a single trial is p
`chi2(n,x)`	returns the cumulative chi-squared distribution with n degrees of freedom
`chi2tail(n,x)`	returns the reverse cumulative (upper-tail) chi-squared distribution with n degrees of freedom; `chi2tail(n,x) = 1 - chi2(n,x)`
`F(n1,n2,f)`	returns the cumulative F distribution with n1 numerator and n2 denominator degrees of freedom
`Fden(n1,n2,f)`	returns the probability density function for the F distribution with n1 numerator and n2 denominator degrees of freedom
`Ftail(n1,n2,f)`	returns the reverse cumulative (upper-tail) F distribution with n1 numerator and n2 denominator degrees of freedom; `Ftail(n1,n2,f) = 1 - F(n1,n2,f)`
`invbinomial(n,k,P)`	returns the inverse binomial: for P<=0.5, probability p such that the probability of observing k or more successes in n trials is P; for P>0.5, probability p such that the probability of observing k or fewer successes in n trials is 1-P.
`invchi2(n,p)`	returns the inverse of `chi2()`; if `chi2(n,x) = p`, then `invchi2(n,p) = x`
`invF(n1,n2,p)`	returns the inverse cumulative F distribution; if `F(n1,n2,f) = p`, then `invF(n1,n2,p) = f`
`invnorm(p)`	returns the inverse cumulative standard normal distribution; if `norm(z) = p`, then `invnorm(p) = z`
`norm(z)`	returns the cumulative standard normal distribution
`normden(z)`	returns the standard normal density
`normden(x,m,s)`	returns the normal density with mean m and standard deviation s; `normden(x,m,s) = normden((x-m)/s)/s`
`tden(n,t)`	returns the probability density function of Student's t distribution with n > 0 degrees of freedom
`ttail(n,t)`	returns the reverse cumulative (upper-tail) Student's t distribution with n > 0 degrees of freedom

Random Numbers

`uniform()`	returns uniformly distributed pseudo-random numbers on the interval [0,1)
`invnorm(uniform())`	returns normally distributed random numbers with mean zero and standard deviation one

String Functions

Programming

Data Functions

Time-series Functions

Matrix Functions

set seed #
uniform()
invnorm(uniform())

sum(x)
sum(x!=.)

Regression

Basic Regression

```
regress depvar [varlist] [,*]
```

Options:

- `depvar` - name of dependent variable
- `varlist` - list of independent variables, separated by spaces (not commas)
- `level(#)` - specifies the confidence level (e.g., 95) for confidence intervals of the coefficients
- `noconstant` - suppresses the constant (intercept) term
- `robust` - uses the White Heteroskedasticity Consistent Covariance Estimator; results in higher standard errors and lower t-ratios

Examples:

```
regress y x1 x2
reg height gender mom dad, level(95)
reg consumption output, noconstant
```

Using Results

Parameter Estimates - returns the estimated coefficient for `regressorname`

```
_b[regressorname]
```

Predict - generates a new variable that stores the designated prediciton based on the last regression run by Stata

```
predict newvarname [,statistic]
```

Statistic:

- `xb` - fitted values; sample point estimate; this is the default so you don't need to include it
- `residuals` - residuals (dependent variable minus ybar)
- `rstandard` - standardized residuals
- `stdp` - standard error of each predicted value (i.e., $Stdev(\hat{y}_i)$)
- `stdf` - standard error of each forecasted value
- `stdr` - standard error of each residual

Variance - displays the variance-covariance matrix (i.e., $Var(\hat{\boldsymbol{\beta}})$)

```
vce
```

Testing Linear Hypotheses After Estimation

`test coeflist` - test that coefficients are equal 0; list coefficients separated by spaces
`test exp = exp [= ...]` - test that linear expressions are equal

Options:
`accumulate` - adds test to previous test(s) in memory making a joint test

Note: This performs the Wald Test... approximated with an F distribution instead of chi-square

***F*-Test** - to do a real F-test of m restrictions:

1. Run the unrestricted regression: $y_i = x_{1i}\hat{\beta}_1 + x_{2i}\hat{\beta}_2 + x_{3i}\hat{\beta}_3 + \cdots + x_{ki}\hat{\beta}_k + \hat{u}_i$
2. Record SSR (just on paper if you want)
3. Run the restricted regression: $\tilde{y}_i = \tilde{x}_{1i}\tilde{\beta}_1 + \cdots + \tilde{x}_{ki}\tilde{\beta}_k + \tilde{u}_i$
4. `generate F = ((RstctdSSR - UnrstctdSSR)/m)/(UnrstctdSSR/(N-k))`
5. Compare that to an F(2,N-k)... display `Ftail(m,N-k,F)`

Example -

```
regress lwage educ huswage city unem exper expersq
```

Using Wald Test:

```
test uduc-expr = 0
test city + unem = 0, accumulate
```

Returns 3.96... p-value 0.0199

Using F-Test:

```
generate edex = educ - exper
generate ctun = city + unem
regress lwage edex huswage ctun expersq
generate F = ((190.12475-186.55)/2)/(186.55/(428-7))
```

Returns 4.022... p-value 0.0186

Advanced Regression Techniques

Options:

`beta` - requests that normalized beta coefficients be reported instead of confidence intervals, if the original model is $y = x_1\beta_1 + x_2\beta_2 + u$, beta alters the model to be

$$\frac{y-\bar{y}}{Stdev(y)} = \frac{x_1-\bar{x}_1}{Stdev(x_1)}\tilde{\beta}_1 + \frac{x_2-\bar{x}_2}{Stdev(x_2)}\tilde{\beta}_2 + \tilde{u}$$, where $\tilde{\beta}_i = \beta_i Stdev(x_i)$

`cluster [varname]` - `varname` describes ID variable to allow correlation between errors within a cluster

Heterskedasticity - here's a series of commands to deal with heteroskedasticity; assume only x2 and x3 are correlated to the error terms

```
regress y x1 x2 x3
predict e, residuals
generate e2 = e^2
regress e2 x2 x3
predict sigma2
```

Method 1 - Transform Model

```
generate newy = y/sqrt(sigma2)
generate newx1 = x1/sqrt(sigma2)
etc.
regress newy newx1 newx2 newx3
```

Method 2 - Weights

```
regress y x1 x2 x3 [weight = sigma2]
```

generating lagged variables - `generate lagy = y[_n-1]`

Regressors Correlated with Error Terms - use instrumental variable estimation and the Hausman test

```
ivreg depvar [varlist] (varlist2 = varlist_iv) [,*]
```

Options:

- `varlist2` - list of independent variables that are correlated with the error term
- `varlist_iv` - list of instrument variables used in place of the variables in `varlist2`
- Other options are same are `regress` command

`hushrs` (husband hours) is probably a joint decision when deciding the wife's hours (`hours`), so it's probably correlated with the error term; suppose `huseduc` is known to be a good instrument; test if `huswage` is also a good instrument:

```
ivreg hours kidslt6 educ wage famine unem (hushrs = huseduc),
   robust
hausman, save
ivreg hours kidslt6 educ wage famine unem (hushrs = huswage
   huseduc), robust
hausman
```

Seemingly Unrelated Regression (SUR) - simultaneous equations using pooled data (i.e., cross-section data over time that may not necessarily be from same source)

```
sureg (depvar1 varlist1 [,noconstant]) (depvar2 varlist2) ...
```

Options:

- `noconstant` - omits constant term for specified equations `isure` - iterate over the estimated disturbance covariance matrix and parameter estimates until the parameter estimates converge; better finite sample properties

`dfk` - use alternate divisor in computing the covariance matrix for the equation errors; better estimates for small samples

`small` - specifies that small sample statistics are to be computed; shifts test statistics from chi-squared and Z statistics to F statistics and t-statistics

Examples:

3 simultaneous equations:

```
sureg (price foreign weight length) (mpg foreign weight)
   (displ foreign weight)
```

Test if coefficient for `foreign` is zero is all equations:

```
test foreign
```

Test across equations

```
test [price] foreign = [mpg] foreign
```

Problem with Heteroskedasticity or Serial Correlation -

Run simple OLS on stacked data (use $n = \min(n_1 n_2)$; drop extra data)

Create new variable to account for pairs

```
regress y x1 x2, cluster[d] robust
```

$$\mathbf{d} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ n \\ 1 \\ 2 \\ \vdots \\ n \end{bmatrix}$$

Fixed Effect Regression with Panel Data -

```
xtreg depvar [varlist], type i(varname)
```

Options:

Type is one of the followed depending on which estimation technique is used:

`be` - between-effects estimator: $y_{i\bullet} = \beta_{0i} + \mathbf{x}_{i\bullet}'\boldsymbol{\beta} + u_{i\bullet}$

`fe` - fixed-effects estimator: $(y_{it} - y_{i\bullet}) = (\mathbf{x}_{it} - \mathbf{x}_{i\bullet})'\boldsymbol{\beta} + (u_{it} - u_{i\bullet})$

where $y_{i\bullet} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $x_{mi\bullet} = \frac{1}{T}\sum_{t=1}^{T} x_{mit}$

`re` - GLS random-effects estimator

`pa` - GEE population-averaged estimator

`mle` - maximum-likelihood random-effects estimator

`i(varname)` - specifies the variable corresponding to an independent unit (e.g., a subject id); this variable represents the i in x_{it} (similar to `cluster`)

Output:

Reports # Observations, # Groups (individuals), min, max and avg Obs/Group

`R-Sq`... only care about overall... that's the one based on the original model:

$$y_{it} = \sum_{j=1}^{N} \beta_{0j} d_{jit} + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}$$

`F(##, ###)` (above table) testing H$_0$: $\beta_i = 0$ (i.e., all parameters are zero)... this is the standard F-test checking all parameters simultaneously for a regression

`Const` = $\hat{\beta}_0 = \frac{1}{N}\sum_{i=1}^{n} \hat{\beta}_{0i}$

`Corr(u_i,xb)` = Corr($\hat{\beta}_{oi}, \mathbf{x}_{it}'\hat{\boldsymbol{\beta}}$)... this is to check the assumption of the random effect model which assumes $E(\beta_{0i}\mathbf{x}_{it}) = \mathbf{0}$

`Sigma_u` = standard deviation of u_{it}

`Sigma_e` = standard deviation of $\hat{\beta}_{0i}$

`F(##, ###)` (below table) testing H_0: $Var(\beta_{0i}) = 0$; i.e., whether individual effect is correlated with regressors (or all the same); numerator degrees of freedom is N + k; denominator is NT - (N + k) (assuming same number of time periods per individual)... another way to think of this test is a test on whether N - 1 dummy variables are simultaneously equal to zero (1 dummy is left out and captured with the constant term in the regression)

Programming

General Program

Specify Stata Version - some commands and formats are specific to the version of Stata (so copying someone else's code made not work in your version of Stata). If the person wrote it in a previous version, you may be able to get away with a simple command that allows the older code to work. Type the version number you want to emulate at the beginning of the file:

```
version 8
```

Comments

`*`	Used at the beginning of a line; the line is ignored.
`/* */`	Used in the middle of a line; everything between `/*` and `*/` is ignored.
`//`	Used at the beginning or end of a line (must be preceded by one or more blanks if at the end); everything on the line after `//` is ignored.
`///`	Instructs Stata to view from `///` to the end of a line as a comment, and to join the next line with the current line; must be preceded by one or more blanks; used make long lines more readable.

1. If X is normally distributed with mean 2 and variance 1, find $\Pr[|X-2|<1]$.

$\Pr[|X-2|<1] = \Pr[-1<X-2<1] = \Pr[1<X<3] = \Pr[X<3]-\Pr[X<1] =$
$\Pr\left[\frac{X-2}{1}<\frac{3-2}{1}\right]-\Pr\left[\frac{X-2}{1}<\frac{1-2}{1}\right] = \Pr[z<1]-\Pr[z<-1] =$
0.8413 - 0.1586 = **0.6827**

2. A distributor of bean seeds determines from extensive testing that 5% of a large batch of seeds will not germinate. He sells seeds in packages of 200 and guarantees 90% germination. What is the probability that a given package will violate the guarantee?

X = number of seeds out of 200 that do not germinate; probability of single seed not germinating is 0.05... X ~ Binomial (p = 0.05, n = 200)
Violate guarantee if more than 90% of 200 (i.e., 20) seeds do not germinate.
$\Pr[X>20] = 1-\Pr[X\le 20]$ = 1 - 0.9988 = **0.0012**

3. A fair coin is tossed until a head appears. Let X denote the number of tosses required.
(a) Find the density of X.
(b) Find the mean and variance of X.

Must toss tails (X - 1) times prior to a head.
Pr[Tail] = Pr[Head] = 0.5.
X ~ Geometric with p = 0.5

(a) PDF: $f(x) = p(1-p)^{x-1},\ x\in\{1,2,3,\ldots\}$

(b) Mean: $E(X) = \frac{1}{p} = \frac{1}{0.5} =$ **2**

Variance: $Var(X) = \frac{1-p}{p^2} = \frac{1-0.5}{0.5^2} = \frac{1}{0.5} =$ **2**

4. (*Hyghens problem*) A and B alternately throw a fair pair of six sided dice. A wins if he scores 6 points before B gets 7 points, in which case B wins. If A starts the game what is the probability she wins?

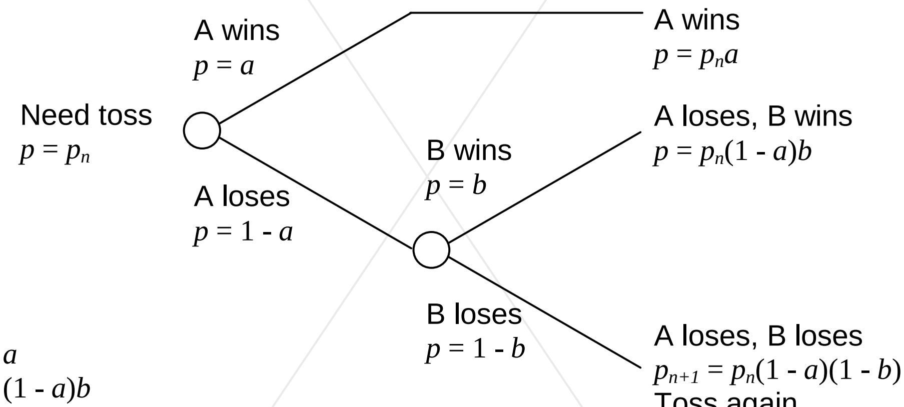
Tossing pair of dice has following results:

d1 \ d2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Distribution of sum of pair of dice:

2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

A wins with probability a = 5/36
B wins with probability b = 6/36



Toss 1:
A wins: a
B wins: $(1 - a)b$
Next toss: $(1 - a)(1 - b)$

Toss 2:
A wins: $a(1 - a)(1 - b)$
B wins: $(1 - a)^2(1 - b)b$
Next toss: $(1 - a)^2(1 - b)^2$

Toss n:
A wins: $a(1 - a)^{n-1}(1 - b)^{n-1}$
B wins: $(1 - a)^n(1 - b)^{n-1}b$
Next toss: $(1 - a)^n(1 - b)^n$

A wins: $\sum_{n=1}^{\infty} a\left[(1-a)(1-b)\right]^{n-1} = a\sum_{n=0}^{\infty}\left[(1-a)(1-b)\right]^{n} = \frac{a}{1-(1-a)(1-b)} = \frac{\frac{5}{36}}{1-\frac{31}{36}\left(\frac{30}{36}\right)} =$

$\frac{180}{366} = \frac{30}{61} =$ **0.4918**

B wins: $\sum_{n=1}^{\infty}(1-a)^n(1-b)^{n-1}b = (1-a)b\sum_{n=0}^{\infty}\left[(1-a)(1-b)\right]^{n} = \frac{(1-a)b}{1-(1-a)(1-b)} = \frac{\left(1-\frac{5}{36}\right)\frac{6}{36}}{1-\frac{31}{36}\left(\frac{30}{36}\right)} =$

$= \frac{186}{366} = \frac{31}{61} = 0.5082$

5. A student in the Department of Economics takes a multiple-choice test consisting of twenty problems, each of which have four answers to choose from. As he missed most of his classes, he only knows the answers to eight of the questions, can narrow the choice to one of two answers for four of the questions, and has no idea as to the correct answers to the remaining eight questions. Calculate:
(a) The expected number of answers that the student will get right.
(b) The probability that the student will answer 70% of the questions correctly.
(c) The probability that the student will answer 50% of the questions correctly.

X_1 = # correct for 8 certain questions... $E(X_1) = 8$
X_2 = # correct for 4 50-50 questions; $X_2 \sim$ Binomial ($p = 0.5$, $n = 4$); $E(X_2) = 4(0.5) = 2$
X_3 = # correct for 8 clueless questions; $X_3 \sim$ Binomial ($p = 0.25$, $n = 8$); $E(X_3) = 8(0.25) = 2$

(a) $E(X) = 8 + 2 + 2 =$ **12**

(b) 70% = 14 questions correct... needs 6 more

$\Pr[X_3 = 0] =$
$\Pr[\text{Bin}(p = 0.25, n = 8) = 0]$

$X_3 \setminus X_2$	0	1	2	3	4	
0	0.00626	0.02503	0.03754	0.02503	0.00626	0.10011
1	0.01669	0.06674	0.10011	0.06674	0.01669	0.26697
2	0.01947	0.07787	0.11680	0.07787	0.01947	0.31146
3	0.01298	0.05191	0.07787	0.05191	0.01298	0.20764
4	0.00541	0.02163	0.03244	0.02163	0.00541	0.08652
5	0.00144	0.00577	0.00865	0.00577	0.00144	0.02307
6	0.00024	0.00096	0.00144	0.00096	0.00024	0.00385
7	0.00002	0.00009	0.00014	0.00009	0.00002	0.00037
8	0.00000	0.00000	0.00001	0.00000	0.00000	0.00002
	0.06250	0.25000	0.37500	0.25000	0.06250	1

$\Pr[X_2 = 2] \cdot \Pr[X_3 = 7]$

$\Pr[X_2 = 4]$

Pr(X = 70%) = **0.10983** (add blue cells)
Pr(X ≥ 70%) = 0.16969 (add blue cells and all cells below them in table)

(c) 50% = 10 questions correct... needs 2 more

Pr(X = 50%) = **0.12375** (add blue cells)
Pr(X ≥ 50%) = 0.95203 (add blue cells and all cells below them in table)

6. If $E(X) = E(X^2) = 0$, show that $\Pr(X = 0) = 1$.

$$Var(X) = E(X^2) - [E(X)]^2 = 0$$

$$Var(X) = \begin{cases} \sum_x (x - E(x))^2 f(x) \\ \int_x (x - E(x))^2 f(x)dx \end{cases} \Rightarrow \text{all } x = 0 \ \ \text{(or } f(x) = 0 \text{ which is the same thing)}$$

Either way, $\Pr(X = 0) = 1$

7. Let X_i be random variables distributed N(i,i^2), $i = 1,2,3$. Assume that X_1, X_2, and X_3 are independent. Using only these three random variables:
(a) Give an example of a statistic that has a chi-square distribution with three degrees of freedom.
(b) Give an example of a statistic that has an F distribution with one, two degrees of freedom.
(c) Give an example of a statistic that has a t distribution with two degrees of freedom.

(a) $z_i = \dfrac{X_i - i}{i} \sim$ N(0,1)

$z_i^2 \sim \chi_1^2$

$X = z_1^2 + z_2^2 + z_3^2 \sim \chi_3^2$

(b) $F = \dfrac{z_1^2/1}{\left(z_2^2 + z_3^2\right)/2} \sim F_{1,2}$

(c) $t = \dfrac{z_1}{\sqrt{\left(z_2^2 + z_3^2\right)/2}} \sim t_2$

Documentation

Most of the work was done using course notes. Prof Werner answered various questions in class. The two I remember were those confirming problem 2 used a binomial distribution and problem 3 used a geometric distribution.

I checked my answers with Josh Kneifel, Jon Parker, and Guille Sabbioni.

Formulas for $E(X)$ and $Var(X)$ in problem 3 were taken from *Introduction to Probability and Its Applications* by Richard L. Scheaffer, p.86 (Duxbory, 1990)

1. Calculate the probability density function, expected value and variance of the following:
(a) $Y = 2X^2 + 1$, where X follows a *uniform distribution* on the interval [0,1]
(b) $Y = X^2$, where X has a *standard normal distribution*.

(a) $f(x) = \frac{1}{1-0} = 1,\ 0 \le x \le 1$ (0 otherwise)

$E(X) = \frac{0+1}{2} = \frac{1}{2} \qquad Var(X) = \frac{1-0}{12} = \frac{1}{12}$

$X^2 = \frac{Y-1}{2} \Rightarrow X = \sqrt{\frac{Y-1}{2}} = s(y)$

$0 \le x \le 1 \Rightarrow 0 \le \sqrt{\frac{y-1}{2}} \le 1 \Rightarrow 0 \le \frac{y-1}{2} \le 1 \Rightarrow 0 \le y-1 \le 2 \Rightarrow 1 \le y \le 3$

$\frac{ds(y)}{dy} = \frac{1}{2}\left(\frac{y-1}{2}\right)^{-1/2}\frac{1}{2} = \frac{1}{4}\left(\frac{y-1}{2}\right)^{-1/2}$

$g(y) = f(s(y))\left|\frac{ds(y)}{dy}\right| = \frac{1}{4}\left(\frac{y-1}{2}\right)^{-1/2},\ 1 \le y \le 3$ (0 otherwise)

$E(Y) = E(2X^2 + 1) = 2E(X^2) + 1 = 2(1/3) + 1 =$ **5/3**

$E(X^2) = \int_0^1 x^2 f(x)dx = \int_0^1 x^2 dx = \left.\frac{x^3}{3}\right]_0^1 = 1/3$

Or, $E(X^2) = Var(X) + [E(X)]^2 = 1/12 + (1/2)^2 = 1/12 + 1/4 = 4/12 = 1/3$

$Var(Y) = Var(2X^2) + 1 = 4Var(X^2) = 4(4/45) =$ **16/45**

$Var(X^2) = \int_0^1 (x^2 - E(X^2))^2 f(x)dx = \int_0^1 (x^2 - 1/3)^2 dx = \int_0^1 x^4 - 2/3x^2 + 1/9 dx =$

$\left.\frac{x^5}{5} - \frac{2x^3}{9} + \frac{x}{9}\right]_0^1 = 1/5 - 2/9 + 1/9 = 4/45$

Or, $Var(Y) = E(Y^2) + [E(Y)]^2 = 47/15 + (5/3)^2 = 47/15 + 25/9 = 16/45$

$E(Y^2) = E[(2X^2 + 1)^2] = E(4X^4 + 4X^2 + 1)] = \int_0^1 (4x^4 + 4x^2 + 1) f(x)dx =$

$\int_0^1 (4x^4 + 4x^2 + 1)dx = \left.\frac{4x^5}{5} + \frac{4x^3}{3} + x\right]_0^1 = 4/5 + 4/3 + 1 = 47/15$

$$\boxed{\begin{gathered} g(y) = \frac{1}{4}\left(\frac{y-1}{2}\right)^{-1/2},\ 1 \le Y \le 3 \text{ (0 otherwise)} \\ E(Y) = 5/3 \qquad Var(Y) = 16/45 \end{gathered}}$$

(b) $f(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

$X^2 = Y \Rightarrow X = \pm\sqrt{Y} = s(Y)$; this is symmetric so consider case of $+\sqrt{Y}$ first; then multiply result by 2

$$\frac{ds(Y)}{dY} = \frac{1}{2} Y^{-1/2}$$

$$g(Y) = 2f(s(Y))\left|\frac{ds(Y)}{dY}\right| = 2\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sqrt{Y}^2}\left|\frac{1}{2}Y^{-1/2}\right| = \frac{\sqrt{2}}{2\sqrt{\pi}} e^{-\frac{1}{2}Y} Y^{-1/2} = \frac{\left(\frac{Y}{2}\right)^{-\frac{1}{2}} e^{-\frac{Y}{2}}}{2\sqrt{\pi}}$$

Note: This is the same as the pdf for a chi-square with 1 df, which is expected since we're squaring a standard normal. ($\Gamma(1/2) = \sqrt{\pi}$); mean should be n and variance $2n$ where n = df.

$E(Y) = E(X^2) = Var(X) + [E(X)]^2 = 1 + 0 =$ **1**

$Var(Y) = Var(X^2) = E(X^4) - [E(X^2)]^2 = 3 - 1 =$ **2**

$E(X^4)$ = kurtosis which is 3 for a normal distribution

$$\boxed{g(Y) = \frac{\left(\frac{Y}{2}\right)^{-\frac{1}{2}} e^{-\frac{Y}{2}}}{2\sqrt{\pi}} \qquad E(Y) = 1 \qquad Var(Y) = 2}$$

2. It is well known that many employees of the City of New York "shirk" on the job - i.e., they only work for part of each hour that they are paid to work. Let X be a random variable which measures the ratio of hours a female employee actually works to hours for which she is paid, and let Y be a random variable which measures the ratio of hours a male employee actually works to hours for which he is paid, and suppose that the joint probability density of X and Y is given by:

$$f(x,y) = 0.4(2x + 3y) \text{ for } 0 \le x \le 1,\ 0 \le y \le 1$$

(i) Verify that $f(x,y)$ is a valid probability density.
(ii) Suppose that Mary and Joe Camilleri are employees of the City of New York. Calculate the probability that neither of them actually works more than half the time they are paid to work.
(iii) Calculate the marginal distributions for X and Y, and hence deduce $E(X)$ and $E(Y)$.
(iv) Is shirking by male employees independent of shirking by female employees?
(v) Given that Mary and Joe are lazy employees and never work more than half the time they are paid to work for, what is the probability that Mary works for less than a quarter of the time she is paid to work?
(vi) The hourly shirking cost to the City of new York is a random variable S, where

$$S = 10000(1 - X^2 - Y/3)$$

What is the expected hourly shirking cost? (Hint: Do not work out the distribution of S.)

(i) $0 \le f(x,y) \le 1 \ \forall \ x \in$ [0,1] and $y \in$ [0,1]

$0.4\int_0^1\int_0^1 (2x+3y)dydx = 0.4\int_0^1 (2xy+3/2y^2)\big]_0^1 dx = 0.4\int_0^1 (2x+3/2)dx =$

$0.4(x^2+3/2x)\big]_0^1 = 0.4(1+3/2)$ = **1** $\therefore$ $f(x,y)$ is a probability distribution

(ii) $\Pr[X \le 0.5$ & $Y \le 0.5]$ = $0.4\int_0^{0.5}\int_0^{0.5} (2x+3y)dydx = 0.4\int_0^{0.5} (2xy+3/2y^2)\big]_0^{0.5} dx =$

$0.4\int_0^{0.5} (x+3/8)dx = 0.4(1/2x^2+3/8x)\big]_0^{0.5}$ = 0.4(1/2·1/4 + 3/8·1/2) = 0.4(1/8 + 3/16) =

2/5(5/16) = 2/16 = **1/8 = 0.125**

(iii) $f_X(x) = 0.4\int_0^1 (2x+3y)dy = 0.4(2xy+3/2y^2)\big]_0^1 = 0.4(2x+3/2)$ = **0.8*x* + 0.6**

$E(X) = \int_0^1 xf_X(x)dx = \int_0^1 x(0.8x+0.6)dx = \int_0^1 (0.8x^2+0.6x)dx = (0.8/3x^3+0.6/2x^2)\big]_0^1 =$

$(0.8/3+0.6/2)$ = 4/15 + 3/10 = 8/30 + 9/30 = **17/30 = 0.5667**

$f_Y(y) = 0.4\int_0^1 (2x+3y)dx = 0.4(x^2+3xy)\big]_0^1 = 0.4(1+3y)$ = **0.4 + 1.2*y***

$E(Y) = \int_0^1 yf_Y(y)dy = \int_0^1 y(0.4+1.2y)dy = \int_0^1 (0.4y+1.2y^2)dy = (0.2y^2+0.4y^3)\big]_0^1 =$

$(0.2+0.4)$ = **0.6**

(iv) $f_X(x)\cdot f_Y(y)$ = (0.8x + 0.6)(0.4 + 1.2y) = 0.32x + 0.96xy + 0.24 + 0.72y $\ne$ $f(x,y)$ $\therefore$
X and Y are **NOT independent**

(v) $\Pr[X \le 1/4 \mid X$ & $Y \le 1/2)$ = $\dfrac{\Pr[X \le 1/4 \,\&\, Y \le 1/2]}{\Pr[X \le 1/2 \,\&\, Y \le 1/2]}$ = 0.05/0.125 = **0.4**

$\Pr[X \le 1/4 \,\&\, Y \le 1/2] = 0.4\int_0^{0.25}\int_0^{0.5} (2x+3y)dydx = 0.4\int_0^{0.25} (2xy+3/2y^2)\big]_0^{0.5} dx =$

$0.4\int_0^{0.25} (x+3/8)dx = 0.4(1/2x^2+3/8x)\big]_0^{0.25}$ = 0.4(1/32 + 3/32) = 2/5(4/32) = 1/20
= 0.05

$\Pr[X \le 1/2 \,\&\, Y \le 1/2]$ = 0.125 (from part ii)

(vi) $E(S) = E[10000(1 - X^2 - Y/3)] = 10000(1 - E(X^2) - E(Y)/3) = 10000(1 - 0.4 - 0.6/3) =$ **4000**

$$E(X^2) = \int_0^1 x^2 f_X(x)dx = \int_0^1 x^2(0.8x+0.6)dx = \int_0^1 (0.8x^3+0.6x^2)dx =$$

$$(0.2x^4+0.2x^3)\Big]_0^1 = 0.2 + 0.2 = 0.4$$

3. Suppose that we roll two (six-sided) dice, and let X and Y be the two numbers which appear.
(i) Find the probability density $f(z)$ for $Z = |X - Y|$
(ii) Find the cumulative density $F(z)$
(iii) What is the probability that Z is an odd number?

(i) **$|X - Y|$**

X \ Y	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Distribution of $|X - Y|$:

0	**1**	**2**	**3**	**4**	**5**
6/36	10/36	8/36	6/36	4/36	2/36

(ii) Cumulative distribution of $|X - Y|$:

$Z \geq$		**0**	**1**	**2**	**3**	**4**	**5**
$Z <$	**0**	**1**	**2**	**3**	**4**	**5**	
	0	6/36	16/36	24/36	30/36	34/36	1

(iii) Pr[Z is odd] = Pr[$|X - Y|$] = Pr[1, 3, or 5] = 10/36 + 6/36 + 2/36 = **18/36 = 0.5**

4. If $F_{XY}(x,y)$ is the joint density of X and Y, derive the density function of $Z = X - Y$

DON'T HAVE TO DO

5. Ten percent of certain population suffer form a serious genetic disease. All newborns are given two independent tests for the disease. Each test makes a correct diagnosis 90% of the time. Find the probability that the person really has the illness:
(a) given that both tests are positive
(b) given that only one test is positive

Symbology:

+ = test positive	D = has disease
- = test positive	ND = doesn't have disease
++ = both tests are positive	+- = one test positive and one test negative

Given:

Pr[D] = 0.1 $\Rightarrow$ Pr[ND] = 0.9
Pr[+ | D] = Pr[- | ND] = 0.9 $\Rightarrow$ Pr[- | D] = Pr[+ | ND] = 0.1

(a) Pr[D | ++] = Pr[D & ++]/Pr[++] = 0.081/0.09 = **0.9**
Pr[D & ++] = Pr[D]·Pr[+ | D]·Pr[+ | D] = 0.1(0.9)(0.9) = 0.081
Pr[ND & ++] = Pr[ND] ·Pr[+ | ND]·Pr[+ | ND] = 0.9(0.1)(0.1) = 0.009
Pr[++] = Pr[D & ++] + Pr[ND & ++] = 0.081 + 0.009 = 0.09

(a) Pr[D | +-] = Pr[D & +-]/Pr[+-] = 0.009/0.09 = **0.1**
Pr[D & +-] = Pr[D]·Pr[+ | D]·Pr[- | D] = 0.1(0.9)(0.1) = 0.009
Pr[ND & +-] = Pr[ND] ·Pr[+ | ND]·Pr[- | ND] = 0.9(0.9)(0.1) = 0.081
Pr[+-] = Pr[D & +-] + Pr[ND & +-] = 0.009 + 0.081 = 0.09

Documentation

Prof Werner reviewed most of the problem set with me before I started working on it. On problem 1, she referred me to my notes on functions of random variables. She helped me identify the $f(x)$ and $s(y)$ functions and told me that part b had to be multiplied by 2 to get the distribution to be the same as a chi-squared with 1 df. She also caught an error where I used $E(X)$ rather than $E(X^2)$ in part (a), verified my use of $E(X^2) = Var(X) + [E(X)]^2$, and pointed out that $E(X^4)$ = kurtosis = 3 for a normal distribution for part (b). On problem 2 parts (v) and (vi), Prof Werner verified how I setup the solutions. Prof Werner worked out problem 5 in class.

I checked my answers to problem 1 and with Jon Parker and Josh Kneifel and problem 5 with JC Zannis.

1. The following data show the number of A's obtained in a population of 5 students enrolled in an Accounting PhD program at the Micanopy Institute of Technology.

Student	Number of A's
Alma	3
Bud	0
Carmella	1
Dora	3
Elmer	2

(i) Define the random variable to be the number of A's that a student obtains. Write down the probability distribution of X and then calculate $E(X)$ and $Var(X)$.
(ii) List the ten possible samples of size 3 which we could take from this population if we were to sample without replacement (i.e., name the students who would be in each of these ten samples). Then calculate the ten sample means associated with these samples.
(iii) Use your answers to (ii) to derive the distribution of the random variable $\overline{X}$.
(iv) Calculate the mean of $\overline{X}$ (a) using your answers from part (i), (b) directly from the probability distribution of $\overline{X}$ from part (iii).
(v) Explain why it is not appropriate to use the formula $Var(\overline{X}) = \sigma^2/n$ in this particular example, and then calculate the $Var(\overline{X})$ from the probability distribution found in part (iii).

(i) Probability distribution - just add up the number of occurrences of each number of A's and divide by the total (5).

# A's	Probability
0	$^1/_5$
1	$^1/_5$
2	$^1/_5$
3	$^2/_5$

$$E(X) = \sum_x xf(x) = 0(^1/_5) + 1(^1/_5) + 2(^1/_5) + 3(^2/_5) = {}^1/_5 + {}^2/_5 + {}^6/_5 \Rightarrow \boxed{\mathbf{E(X) = {}^9/_5 = 1.8}}$$

$$Var(X) = \sum_x (x-\mu)^2 f(x) =$$

$$(0 - 1.8)^2(^1/_5) + (1 - 1.8)^2(^1/_5) + (2 - 1.8)^2(^1/_5) + (3 - 1.8)^2(^2/_5) \Rightarrow \boxed{\mathbf{Var(X) = {}^{34}/_{25} = 1.36}}$$

(ii) Just using the first letter for the students.

Samples	# A's	Sample Mean
ABC	4	$^4/_3$ = 1.333
ABD	6	$^6/_3$ =2
ABE	5	$^5/_3$ = 1.667
ACD	7	$^7/_3$ = 2.333
ACE	6	$^6/_3$ =2
ADE	8	$^8/_3$ = 2.667
BCD	4	$^4/_3$ = 1.333
BCE	3	$^3/_3$ =1
BDE	5	$^5/_3$ = 1.667
CDE	6	$^6/_3$ =2

(iii) Probability distribution - just add up the number of occurrences of each sample mean and divide by the total (10).

Sample Mean	Probability
1	0.1
$^4/_3$	0.2
$^5/_3$	0.2
2	0.3
$^7/_3$	0.1
$^8/_3$	0.1

(iv) (a) $E(\overline{X}) = E(X) =$ $\boxed{\mathbf{1.8}}$

(b) $E(\overline{X}) = \sum_{\overline{x}} \overline{x}_i f_{\overline{x}}(\overline{x}_i)$ =

1(0.1) + ($^4/_3$)(0.2) + ($^5/_3$)(0.2) + 2(0.3) + ($^7/_3$)(0.1) + ($^8/_3$)(0.1) = $\boxed{\mathbf{1.8}}$

(v) The distribution in part (iii) doesn't come from a random sample because it was generated by looking at every combination of the five students rather than random draws.

$Var(\overline{X}) = \sum_{\overline{x}} \left(\overline{x}_i - E(\overline{X})\right)^2 f_{\overline{x}}(\overline{x}_i)$ = (1 - 1.8)^2(0.1) + ($^4/_3$ - 1.8)^2(0.2) + ($^5/_3$ - 1.8)^2(0.2) + (2 - 1.8)^2(0.3) + ($^7/_3$ - 1.8)^2(0.1) + ($^8/_3$ - 1.8)^2(0.1) = $\boxed{\mathbf{^{17}/_{75} = 0.22667}}$

2. (a) Generate 100 samples of 10 observations drawn from a standard normal distribution.
(i) For each sample calculate sample mean and sample variance.
(ii) Plot a histogram of the sample means and sample variances, indicating on your plot the true mean or variance of the random variables. Comment on the relationship between the sample mean and the true mean, the sample variance and the true variance.
(b) Repeat the above exercise for exponentially distributed random variables with parameter $\theta = 2$.
(c) Repeat the above exercise for the Weibull distribution with distribution function

$$F(x) = 1 - \exp\left[-(x/2)^{1.5}\right]$$

(i) Standard Normal: don't need $f(x)$ of $F(x)$ because we can use **invnorm** command in Stata, $\boldsymbol{E(X) = 0}$, $\boldsymbol{Var(X) = 1}$

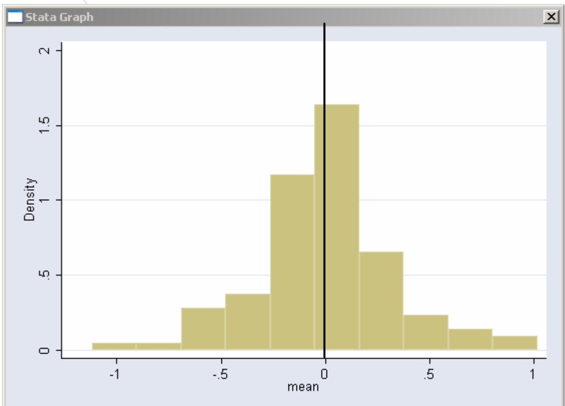
Sample program for Stata (from Scott). I added the red line to get the program to work.

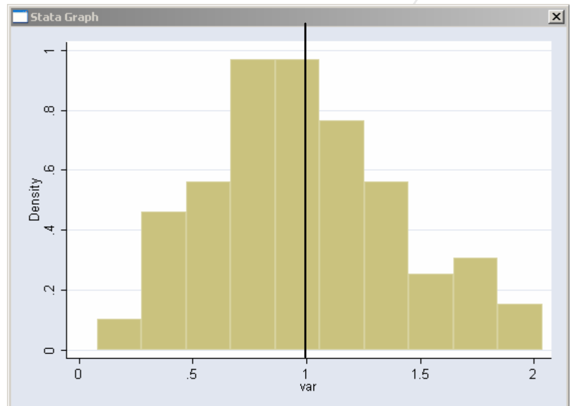
```
program define nsim /*arguments are obs iterations*/
version 6
args n it
tempname sim
postfile `sim' mean var using results, replace
quietly {
local i=1
while `i' <= `it' {
drop _all
set obs `n'
gen z = invnorm(uniform())
summarize z
post `sim' r(mean) r(Var)
local i = `i' + 1
}
}
postclose `sim'
use results,clear

end
```

Results:

Variable	Obs	Mean	Std. Dev.	Min	Max
mean	100	.0005718	.3383416	-1.118136	1.017259
var	100	.9885811	.4143226	.0791966	2.038543





(ii) Exponential Distribution: $f(x) = \theta e^{-\theta x}$, $\frac{1}{\theta}$ = $\boldsymbol{E(X)}$ **= 0.5**, $\frac{1}{\theta^2}$ = $\boldsymbol{Var(X)}$ **= 0.25**

CDF: $F(x) = \int_0^x \theta e^{-\theta x} dx = -e^{-\theta x}\Big]_0^x = -e^{-\theta x} - (-e^0) = 1 - e^{-\theta x}$

Solve CDF for x:

$$e^{-\theta x} = 1 - F(x) \;\Rightarrow\; -\theta x = \ln(1 - F(x)) \;\Rightarrow\; \boxed{x = -\frac{\ln(1 - F(x))}{\theta}}$$

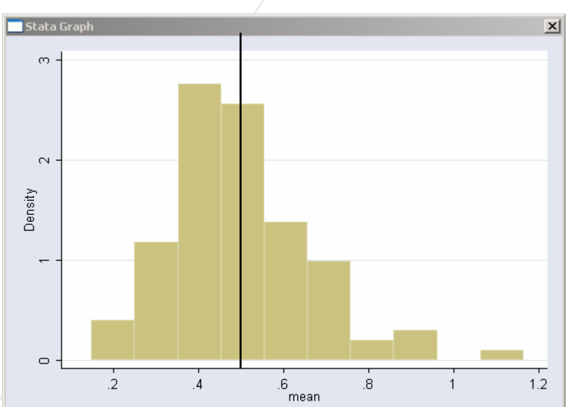
Generate Uniform(0,1) and plug it in for $F(x)$ to generate values for x

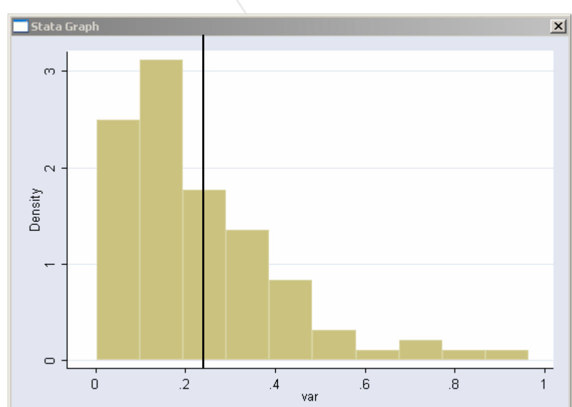
Just change green line in program above to this:

```
gen z = -ln(1 - uniform())/2
```

Results:

Variable	Obs	Mean	Std. Dev.	Min	Max
mean	100	.4997093	.1652599	.149499	1.164545
var	100	.2295738	.1771686	.002916	.963869





(iii) Weibull Distribution: $f(x) = \dfrac{dF(x)}{dx} = -\exp\left[-(x/2)^{1.5}\right] \cdot \left(-1.5(x/2)^{0.5}\right) \cdot 1/2 =$

$$1.5\left(\frac{x}{2}\right)^{0.5} \cdot \frac{1}{2} \cdot \exp\left[-(x/2)^{1.5}\right]$$

$$E(X) = \int_0^\infty x f(x)dx = \int_0^\infty x\left[1.5\left(\frac{x}{2}\right)^{0.5} \cdot \frac{1}{2} \cdot \exp\left[-(x/2)^{1.5}\right]\right]dx =$$

$$\int_0^\infty 1.5\left(\frac{x}{2}\right)^{0.5} \cdot \frac{1}{2} \cdot x\exp\left[-(x/2)^{1.5}\right]dx$$

Let $y = \left(\dfrac{x}{2}\right)^{1.5} \Rightarrow \dfrac{x}{2} = y^{2/3} \Rightarrow x = 2y^{2/3}$

Note: $\dfrac{dy}{dx} = 1.5\left(\dfrac{x}{2}\right)^{0.5} \cdot \dfrac{1}{2}$ so multiply $E(X)$ by 1 = $\dfrac{dy/dx}{dy/dx} = \dfrac{dy/dx}{1.5(x/2)^{0.5} \cdot 1/2}$

$$E(X) = \int_0^\infty 1.5\left(\frac{x}{2}\right)^{0.5} \cdot \frac{1}{2} \cdot x\exp\left[-(x/2)^{1.5}\right]dx \cdot \frac{dy/dx}{1.5(x/2)^{0.5} \cdot 1/2} = \int_0^\infty x\exp\left[-(x/2)^{1.5}\right]dy$$

Now substitute $x = 2y^{2/3}$ for the first term and $\left(\dfrac{x}{2}\right)^{1.5} = y$ for the second

$E(X) = \displaystyle\int_0^\infty 2y^{2/3}e^{-y}dy = 2\int_0^\infty y^{5/3-1}e^{-y}dy = 2\Gamma(5/3)$ = 2(0.9027) = ***E*(*X*) = 1.8054**

$$E(X^2) = \int_0^\infty x^2 f(x)dx = \int_0^\infty x^2\left[1.5\left(\frac{x}{2}\right)^{0.5} \cdot \frac{1}{2} \cdot \exp\left[-(x/2)^{1.5}\right]\right]dx =$$

$$\int_0^\infty 1.5\left(\frac{x}{2}\right)^{0.5} \cdot \frac{1}{2} \cdot x^2\exp\left[-(x/2)^{1.5}\right]dx$$

Use the same y from before $\Rightarrow x^2 = 4y^{4/3}$

$$E(X^2) = \int_0^\infty 1.5\left(\frac{x}{2}\right)^{0.5} \cdot \frac{1}{2} \cdot x^2\exp\left[-(x/2)^{1.5}\right]dx \cdot \frac{dy/dx}{1.5(x/2)^{0.5} \cdot 1/2} =$$

$$\int_0^\infty x^2\exp\left[-(x/2)^{1.5}\right]dy$$

Now substitute $x^2 = 4y^{4/3}$ for the first term and $\left(\dfrac{x}{2}\right)^{1.5} = y$ for the second

$E(X^2) = \displaystyle\int_0^\infty 4y^{4/3}e^{-y}dy = 4\int_0^\infty y^{7/3-1}e^{-y}dy = 4\Gamma(7/3)$ = 4(1.1906) = 4.7624

$Var(X) = E(X^2) - [E(X)]^2$ = 4.7624 - (1.8054)^2 = ***Var*(*X*) = 1.5029**

Solve CDF for *x*:

$$\exp\left[-(x/2)^{1.5}\right] = 1 - F(x) \;\Rightarrow\; -(x/2)^{1.5} = \ln(1 - F(x)) \;\Rightarrow$$
$$x/2 = \left[-\ln(1 - F(x))\right]^{1/1.5} \;\Rightarrow\; \boxed{x = 2\left[-\ln(1 - F(x))\right]^{2/3}}$$

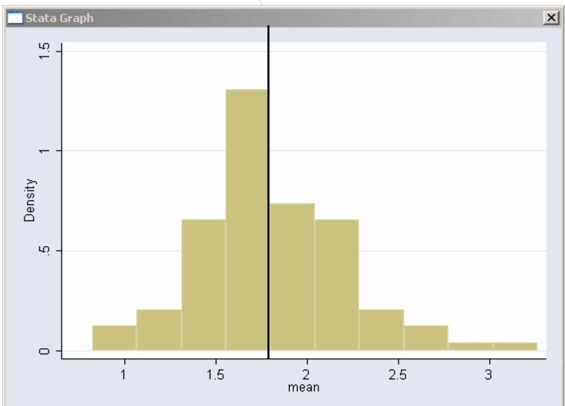
Generate Uniform(0,1) and plug it in for $F(x)$ to generate values for x

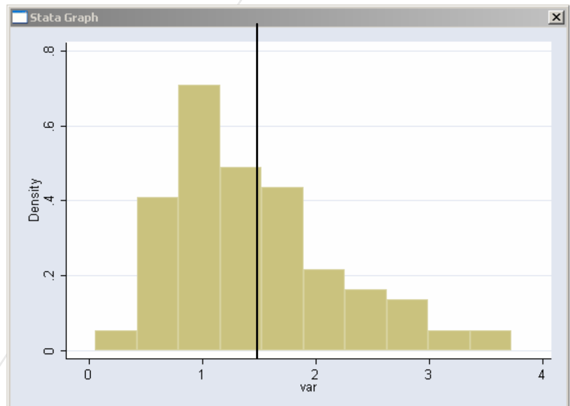
Just change green line in program above to this:

```
gen z = 2*(-ln(1 - uniform()))^(2/3)
```

Results:

Variable	Obs	Mean	Std. Dev.	Min	Max
mean	100	1.807677	.4113603	.8243586	3.266533
var	100	1.44065	.7496428	.0505821	3.727319





3. Suppose that the random variable X has an unknown distribution with a mean of μ and a variance of σ^2. Given a <u>random</u> sample of n observations, calculate $\bar{x}_{n-1}$ (the sample mean of the first n - 1 observations), and then form $\hat{\mu}$ defined by

$$\hat{\mu} = 0.5(\bar{x}_{n-1} + x_n)$$

(i) Show that $\hat{\mu}$ is an unbiased estimator of μ and then explain why $\hat{\mu}$ is not consistent for μ.
(ii) Find $Var(\hat{\mu})$, compare it with $Var(\bar{x}_{n-1})$, and then discuss the efficiency of $\hat{\mu}$. Does $\hat{\mu}$ attain the Cramer-Rao lower bound? Explain.

(i) $E(\hat{\mu}) = E[0.5(\bar{x}_{n-1} + x_n)] = 0.5[E(\bar{x}_{n-1}) + E(x_n)] = 0.5[\mu + \mu] = \mu$ $\therefore$ unbiased

A consistent estimator is one that converges in probability to the actual parameter.

To have convergence in probability we need $\lim_{n\to\infty} \Pr[|\hat{\mu} - \mu| < \varepsilon] = 1$.

That will not happen in this case because $\bar{x}_{n-1}$ converges to μ, but x_n doesn't.

$\therefore$ $\hat{\mu}$ is not a consistent estimator.

(ii) $$Var(\bar{x}_{n-1}) = Var\left(\frac{1}{n-1}\sum_{i=1}^{n-1} x_i\right) = \frac{1}{(n-1)^2}\sum_{i=1}^{n-1} Var(x_i) = \frac{1}{(n-1)^2}\sum_{i=1}^{n-1}\sigma^2 = \frac{1}{(n-1)^2}(n-1)\sigma^2$$

$$= \boxed{\frac{\sigma^2}{n-1}}$$

$$Var(\hat{\mu}) = Var[0.5(\bar{x}_{n-1} + x_n)] = (0.5)^2[Var(\bar{x}_{n-1}) + Var(x_n)] = \boxed{0.25\left[\frac{\sigma^2}{n-1} + \sigma^2\right]} =$$

$$0.25\left[\frac{\sigma^2}{n-1} + \frac{n\sigma^2 - \sigma^2}{n-1}\right] = \frac{0.25n\sigma^2}{n-1} > \frac{\sigma^2}{n-1} \text{ when } n > 4$$

Since $Var(\hat{\mu}) > Var(\bar{x}_{n-1})$ (when $n > 4$), $\hat{\mu}$ **does not** attain the Cramer-Rao lower bound.

4. If X is a random variable such that $E(X) = 3$ and $E(X^2) = 13$
(a) Use Chebyshev's inequality to determine a lower bound for the probability $\Pr(-2 < X < 8)$.
(b) Find the above probability if X has a normal distribution.

(a) Chebyshev's Inequality: for constant $k > 0$, $\Pr[|X - E(X)| \geq k] \leq Var(X)/k^2$

Write for lower bound by flipping the inequalities: $\Pr[|X - E(X)| < k] > 1 - Var(X)/k^2$

$Var(X) = E(X^2) - (E(X))^2 = 13 - 9 = 4$

For this problem: $\Pr[3 - k < X < k + 3] > 1 - 4/k^2$

Since $3 - k = -2$ and $k + 3 = 8$, $k = 5$

$\therefore$ $\Pr[-2 < X < 8] > 1 - 4/25$ $\boxed{\mathbf{21/25 = 0.84}}$

(b) $\Pr[-2 < X < 8] = \Pr\left[\frac{-2-3}{2} < \frac{X-3}{2} < \frac{8-3}{2}\right] = \Pr\left[\frac{-5}{2} < z < \frac{5}{2}\right] =$

$\Pr\left[z < \frac{5}{2}\right] - \Pr\left[z < \frac{-5}{2}\right]$ = 0.9938 - 0.0062 = $\boxed{\mathbf{0.9876}}$

5. The density function of a random variable is $f(x;\theta) = (\ln\theta)\theta^{-x}$, $x > 0$ and $\theta > 1$. A random sample of size n is drawn.
(i) Derive the score function and the MLE of θ.
(ii) Show that the mean and variance of X are $E(X) = 1/(\ln\theta)$ and $Var(X) = 1/(\ln\theta)^2$. Use this to derive the information $I(\theta)$ and the Cramer-Rao lower bound for the variance of $\bar{x}$. Is the lower bound attained?
(iii) Construct a random variable Z_n, which is a function of $\bar{x}$ and θ, such that its asymptotic distribution is $N(0,1)$.

(i) $L(x;\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n} (\ln\theta)\theta^{-x_i} = (\ln\theta)^n \theta^{-\sum_{i=1}^{n} x_i}$

$\ln L(x;\theta) = n\ln(\ln\theta) - \left(\sum_{i=1}^{n} x_i\right)\ln\theta$

Score = $S(\hat{\theta}) = \frac{\partial \ln L(x;\theta)}{\partial\theta} \Rightarrow \boxed{S(\hat{\theta}) = \frac{n}{\theta\ln\theta} - \frac{1}{\theta}\sum_{i=1}^{n} x_i}$

MLE: Solve $S(\hat{\theta}) = 0$

$\frac{n}{\theta\ln\theta} - \frac{1}{\theta}\sum_{i=1}^{n} x_i = 0 \Rightarrow \frac{n}{\theta\ln\theta} = \frac{1}{\theta}\sum_{i=1}^{n} x_i \Rightarrow \frac{n}{\ln\theta} = \sum_{i=1}^{n} x_i \Rightarrow \ln\theta = \frac{n}{\sum_{i=1}^{n} x_i} \Rightarrow$

$\hat{\theta} = \exp\left[n \Big/ \sum_{i=1}^{n} x_i\right] \Rightarrow \boxed{\hat{\theta} = \exp\left[\frac{1}{\bar{x}}\right]}$

(ii) Substitute $1/\gamma = \ln(\theta)$, now $f(x;\theta) = (\ln\theta)e^{-x\ln(\theta)} = \frac{1}{\gamma}e^{\frac{-x}{\gamma}}$ ... exponential distribution,

therefore, $E(X) = \gamma = \frac{1}{\ln\theta}$ and $Var(X) = \gamma^2 = \frac{1}{(\ln\theta)^2}$

$I(\theta) = -E\left[\frac{\partial^2 \ln L(\theta;\mathbf{x})}{\partial\theta^2}\right] = -E\left[\frac{-n}{(\theta\ln\theta)^2}\cdot\left(\ln\theta + \frac{\theta}{\theta}\right) + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i\right] =$

$\frac{n}{(\theta\ln\theta)^2}\cdot(\ln\theta + 1) - \frac{1}{\theta^2}\sum_{i=1}^{n} E(x_i) = \frac{n}{\theta^2\ln\theta} + \frac{n}{\theta^2(\ln\theta)^2} - \frac{1}{\theta^2}\sum_{i=1}^{n}\frac{1}{\ln\theta} =$

$\frac{n}{\theta^2\ln\theta} + \frac{n}{\theta^2(\ln\theta)^2} - \frac{n}{\theta^2\ln\theta} = \boxed{I(\theta) = \frac{n}{\theta^2(\ln\theta)^2}}$

Cramer-Rao Inequality - let $x_i \sim f(x;\theta)$ and $T = T(x_1,...,x_n)$ be a statistic such that $E(T) = u(\theta)$ (some function of θ). Assume regularity conditions. Then

$$Var(T) \geq [u'(\theta)]^2/I(\theta)$$

Let $T = \bar{x}$

From $\hat{\theta} = \exp\left[\frac{1}{\bar{x}}\right]$ (in part (i), we get $\bar{x} = \frac{1}{\ln\theta}$

Therefore, $E(T) = u(\theta) = \frac{1}{\ln\theta}$

$u'(\theta) = \frac{-1}{(\ln\theta)^2}\cdot\frac{1}{\theta}$

From Cramer-Rao inequality, $Var(\bar{x}) \geq \frac{[u'(\theta)]^2}{I(\theta)} = \frac{\left[\frac{-1}{(\ln\theta)^2}\cdot\frac{1}{\theta}\right]^2}{\frac{n}{\theta^2(\ln\theta)^2}} = \frac{1}{\theta^2(\ln\theta)^4}\cdot\frac{\theta^2(\ln\theta)^2}{n}$

$$\boxed{Var(\bar{x}) \geq \frac{1}{n(\ln\theta)^2}}$$

(iii) **Asymptotic Normality** - $\lim_{n\to\infty}\mathbf{I}(\boldsymbol{\theta})/n = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ under the regularity conditions of the CRLB

$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}))$; i.e., MLEs have normal distribution at the limit

In this case, we don't need matrices because there is only 1 parameter, therefore,

$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, n/I(\theta))$

To make this $N(0,1)$, we multiply both sides by $\sqrt{I/n}$:

$\sqrt{n}(\hat{\theta} - \theta)\cdot\sqrt{I/n} \xrightarrow{d} N(0, n/I(\theta))\cdot\sqrt{I/n} \;\Rightarrow\; \sqrt{I}(\hat{\theta} - \theta) \xrightarrow{d} N(0,1)$

Therefore, set random variable $Z_n = \sqrt{I}(\hat{\theta} - \theta) = \sqrt{\frac{n}{\theta^2(\ln\theta)^2}}\left(e^{1/\bar{x}} - \theta\right) =$

$$\boxed{Z_n = \frac{\sqrt{n}}{\theta\ln\theta}\left(e^{1/\bar{x}} - \theta\right)}$$

6. The Pareto distribution is frequently used as a model to study income distribution. It has the following CDF (not PDF): $F(x) = 1 - (\theta_1/x)^{\theta_2}$ for $x \geq \theta_1$, and zero elsewhere. θ_1 and θ_2 are both positive.

(i) Derive the density function $f(x)$.

(ii) Given a random sample of n observations, find the MLE of θ_1 and θ_2.

(iii) Derive the information matrix for θ_1, θ_2.

(iv) Derive the Cramer-Rao lower bounds for the variances of the estimators of θ_1 and θ_2. (You do not need to verify whether these lower bounds are attained.)

(i) $\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$

$$f(x,\boldsymbol{\theta}) = \frac{dF(x)}{dx} = -\theta_2\left(\frac{\theta_1}{x}\right)^{\theta_2-1}\left(-\frac{\theta_1}{x^2}\right) = \boxed{\theta_2\theta_1^{\theta_2}\left(\frac{1}{x}\right)^{\theta_2+1}}$$

(ii) $L(x;\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i,\boldsymbol{\theta}) = \prod_{i=1}^{n}\theta_2\theta_1^{\theta_2}\left(\frac{1}{x}\right)^{\theta_2+1} = \theta_2^n\theta_1^{n\theta_2}\prod_{i=1}^{n}\frac{1}{x^{\theta_2+1}}$

$$\ln L(x;\boldsymbol{\theta}) = n\ln(\theta_2) + n\theta_2\ln(\theta_1) - (\theta_2+1)\sum_{i=1}^{n}\ln(x_i)$$

$$\mathbf{S}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \dfrac{\partial \ln L(x;\boldsymbol{\theta})}{\partial\theta_1} \\ \dfrac{\partial \ln L(x;\boldsymbol{\theta})}{\partial\theta_2} \end{bmatrix} = \begin{bmatrix} \dfrac{n\theta_2}{\theta_1} \\ \dfrac{n}{\theta_2} + n\ln(\theta_1) - \sum_{i=1}^{n}\ln(x_i) \end{bmatrix}$$

MLE: Solve $\mathbf{S}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

From Eqn 1, can't set either θ_i equal to zero so try to get θ_1 as large as possible.

Given requirement that $x \ge \theta_1$, the largest θ_1 will be $\boxed{\hat{\theta}_1 = \min(x_i)}$

$$\frac{n}{\theta_2} + n\ln(\min(x_i)) - \sum_{i=1}^{n}\ln(x_i) = 0 \Rightarrow \boxed{\hat{\theta}_2 = \frac{n}{\sum_{i=1}^{n}\ln(x_i) - n\ln(\min(x_i))}}$$

(iii) $\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$, where $I_{ij} = -E\left[\dfrac{\partial^2 \ln L(x;\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right]$

$$I_{11} = -E\left[\frac{\partial^2 \ln L(x;\boldsymbol{\theta})}{\partial\theta_1^2}\right] = -E\left[\frac{\partial}{\partial\theta_1}\left(\frac{n\theta_2}{\theta_1}\right)\right] = -E\left[\frac{-n\theta_2}{\theta_1^2}\right] = \frac{n\theta_2}{\theta_1^2}$$

$$I_{12} = I_{21} = -E\left[\frac{\partial^2 \ln L(x;\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2}\right] = -E\left[\frac{\partial}{\partial\theta_2}\left(\frac{n\theta_2}{\theta_1}\right)\right] = -E\left[\frac{n}{\theta_1}\right] = \frac{-n}{\theta_1}$$

$$I_{22} = -E\left[\frac{\partial^2 \ln L(x;\boldsymbol{\theta})}{\partial\theta_2^2}\right] = -E\left[\frac{\partial}{\partial\theta_2}\left(\frac{n}{\theta_2} + n\ln(\theta_1) - \sum_{i=1}^{n}\ln(x_i)\right)\right] = -E\left[\frac{-n}{\theta_2^2}\right] = \frac{n}{\theta_2^2}$$

$$\boxed{\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{n\theta_2}{\theta_1^2} & \dfrac{-n}{\theta_1} \\ \dfrac{-n}{\theta_1} & \dfrac{n}{\theta_2^2} \end{bmatrix}}$$

(iv) CRLB is $[\mathbf{I}(\boldsymbol{\theta})]^{-1}$

Using $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-cb}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

$$[\mathbf{I}(\boldsymbol{\theta})]^{-1} = \frac{1}{\frac{n\theta_2}{\theta_1^2}\left(\frac{n}{\theta_2^2}\right) - \frac{-n}{\theta_1}\left(\frac{-n}{\theta_1}\right)}\begin{bmatrix} \frac{n}{\theta_2^2} & \frac{n}{\theta_1} \\ \frac{n}{\theta_1} & \frac{n\theta_2}{\theta_1^2} \end{bmatrix} = \frac{1}{\frac{n^2}{\theta_1^2\theta_2} - \frac{n^2}{\theta_1^2}}\begin{bmatrix} \frac{n}{\theta_2^2} & \frac{n}{\theta_1} \\ \frac{n}{\theta_1} & \frac{n\theta_2}{\theta_1^2} \end{bmatrix} =$$

$$\frac{\theta_1^2\theta_2}{n^2 - n^2\theta_2}\begin{bmatrix} \frac{n}{\theta_2^2} & \frac{n}{\theta_1} \\ \frac{n}{\theta_1} & \frac{n\theta_2}{\theta_1^2} \end{bmatrix} = \boxed{[\mathbf{I}(\boldsymbol{\theta})]^{-1} = \begin{bmatrix} \frac{\theta_1^2}{n\theta_2(1-\theta_2)} & \frac{\theta_1\theta_2}{n(1-\theta_2)} \\ \frac{\theta_1\theta_2}{n(1-\theta_2)} & \frac{\theta_2^2}{n(1-\theta_2)} \end{bmatrix}}$$

Documentation

Prof Werner practically did the entire homework assignment for me... and thankfully didn't hit me after my 50th or so question. I sat down with her on 16 Oct and went over some general points on the homework. For problem 1, Prof Werner told me part (iv)(a) just uses $E(X)$ from part (i). She also told me the answer to part (v) (i.e., not from a random sample). On problem 2, Prof Werner verified my CDF for the exponential distribution. For problem 3, Prof Werner said to assume it was a random sample (i.e., x_i's independent). She pretty much gave me the answer to the second part of part (i). She also showed me how to do part (ii) without computing the CRLB. For problem 4, Prof Werner confirmed that $k = 5$. For problem 5, Prof Werner reviewed the methodology for finding $L(x;\theta)$, $\ln L(x;\theta)$, the score, and the MLE. For problem 6, Prof Werner verified my answers for $L(x;\theta)$, $\ln L(x;\theta)$.

I met with Prof Werner again in and out of class on 21 Oct. She confirmed that I had to solve the CDFs for x in parts (b) and (c) of problem 2. She also figured out the mean and variance of the Weibull distribution for me. On problem 4, Prof Werner showed us how to reverse Chebyshev's inequality to make it a lower bound. Josh Kneifel found a math error because I was dividing by 13 instead of by 2. On problem 5, Prof Werner told us to make the substitution to see we had an exponential distribution in order to find $E(X)$ and $Var(X)$ in part (ii). She also pointed out that $\overline{x}$ should be in the MLE for $\widehat{\theta}$ and told me to use the second derivative for computing the information matrix. She gave the answer to the CRLB and part (iii) in class. For problem 6 part (ii), Prof Werner told us in class to use $\hat{\theta}_1 = \min(x_i)$ because we can't set either θ equal to zero (similar to an example she did in class). She reminded me after class that the information matrix would be 2 x 2 and that I had to take the inverse to get the CRLB.

Scott (TA) save us a template for generating the samples in Stata. He also gave us some basic commands to get the overall mean and variance of the samples and to plot the histograms.

1. In order to be able to work at home, Professor Slutsky needs to provide his cat Flash with cat toys that that he will leave him alone. There are two types of cat toys that Professor Slutsky is trying to choose between: Hartz *Stuffin' Mice* which cost $2 per toy, and Pet Essentials *Furry Mice* which cost $3 per toy. Depending on the production quality and the friskiness of Flash, Professor Slutsky has ascertained that the lifetime, x, in weeks of cat toys has an exponential distribution with probability density function

$$f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}},\ x > 0 \qquad (1)$$

Professor Slutsky knows that cat toys come in either one of two qualities: high quality with $\theta = 2$; and low quality with $\theta = 1$. unfortunately, due to poor eye sight, Professor Slutsky cannot tell by inspection alone whether or not a cat toy is of high or low quality. Having purchased *Stuffin' Mice* for years, he knows that they are of low quality, but wants to know whether or not it is worth spending the extra dollar for the *Furry Mice*.

(a) What is the hypothesis that Professor Slutsky wishes to test?
(b) Professor Slutsky, vaguely remembering his statistics course, decides to buy three *Furry Mice* and accept the null hypothesis, H_0, if the three mice each last for more than 1.5 weeks. Assuming that the lifetime of each toy is independent of the lifetime of the other toys, calculate the following probabilities:
  (i) Pr[Accepting H_0 | H_0 true]
  (ii) Pr[Rejecting H_0 | H_0 true]
  (iii) Pr[Accepting H_0 | H_1 true]
  (iv) Pr[Rejecting H_0 | H_1 true]
(c) Which of (i), (ii), (iii), and (iv) is the probability of making a type I error? Which is the probability of making a type II error?
(d) What is the power of Professor Slutsky's test?
(e) Having studied statistics more recently than Professor Slutsky, you suggest that he test the null hypothesis by noting whether or not the average lifetime of the *Furry Mice* is greater than or less than 1.5 weeks. having ascertained that the probability distribution for the sum of the lifetimes ($\tilde{x} = x_1 + x_2 + x_3$) of the three mice is given by

$$f(x;\theta) = \frac{\tilde{x}^2}{2\theta^3}e^{-\frac{\tilde{x}}{\theta}},\ \tilde{x} > 0 \qquad (2)$$

For your stated H_0, calculate
  (i) Pr[Accepting H_0 | H_0 true]
  (ii) Pr[Rejecting H_0 | H_0 true]
  (iii) Pr[Accepting H_0 | H_1 true]
  (iv) Pr[Rejecting H_0 | H_1 true]
(f) Which, if either, is the better test? Explain you conclusion.
(g) Suppose that the three *Furry Mice* last for 1.6, 0.8 and 2.4 weeks. How confident is Professor Slutsky that it is worth spending the extra dollar on these mice? What would you suggest Professor Slutsky do in order to resolve his doubts?

(a) Wants to know if the new toy (*Furry Mice*) will be worth the money; he'd be spending 50 percent more (\$3 vs. \$2), so the *Furry Mice* has to last at least 50 percent longer than the *Stuffin' Mice*. That's θ = 1.5, but there are only two possibilities: 1 or 2. Therefore,

H_0: θ = 2

H_1: θ = 1

(b) (i) Pr[Accepting H_0 | H_0 true]... Pr[Exp(2) ≥ 1.5]^3 = (1 - Pr[Exp(2) < 1.5])^3 = (1 - 0.4724)^3 = **0.1054**

(ii) Pr[Rejecting H_0 | H_0 true]... 1 - Pr[Accepting H_0 | H_0 true] = 1 - 0.1054 = **0.8946**

(iii) Pr[Accepting H_0 | H_1 true]... Pr[Exp(1) ≥ 1.5]^3 = (1 - Pr[Exp(1) < 1.5])^3 = (1 - 0.2231)^3 = **0.0111**

(iv) Pr[Rejecting H_0 | H_1 true]... 1 - Pr[Accepting H_0 | H_0 true] = 1 - 0.0111 = **0.9889**

(c) Type I error = Pr[Rejecting H_0 | H_0 true] **(ii)**

Type II error = Pr[Accepting H_0 | H_1 true] **(iii)**

(d) Power = 1 - β (Type II error) = 1 - 0.0111 = **0.9889**

(e) $$F(\tilde{x};\theta) = \int_0^{\tilde{x}} \frac{x^2}{2\theta^3} e^{-\frac{x}{\theta}} dx = \left. -\frac{x^2}{2\theta^2} e^{-\frac{x}{\theta}} - \frac{2x}{2\theta} e^{-\frac{x}{\theta}} - e^{-\frac{x}{\theta}} \right]_0^{\tilde{x}} = -\frac{\tilde{x}^2}{2\theta^2} e^{-\frac{\tilde{x}}{\theta}} - \frac{\tilde{x}}{\theta} e^{-\frac{\tilde{x}}{\theta}} - e^{-\frac{\tilde{x}}{\theta}} + 1$$

(i) Pr[Accepting H_0 | H_0 true]... Pr[$f(\tilde{x};2) \geq 4.5$] = 1 - Pr[$f(\tilde{x};2) < 4.5$] = 1 - $F(4.5;2)$ =

$$1 - \left(-\frac{(4.5)^2}{2(2)^2} e^{-\frac{4.5}{2}} - \frac{4.5}{2} e^{-\frac{4.5}{2}} - e^{-\frac{4.5}{2}} + 1 \right) = \mathbf{0.6093}$$

(ii) Pr[Rejecting H_0 | H_0 true]... 1 - Pr[Accepting H_0 | H_0 true] = 1 - 0.6093 = **0.3907**

(iii) Pr[Accepting H_0 | H_1 true]... Pr[$f(\tilde{x};1) \geq 4.5$] = 1 - Pr[$f(\tilde{x};1) < 4.5$] = 1 - $F(4.5;1)$ =

$$1 - \left(-\frac{(4.5)^2}{2(1)^2} e^{-\frac{4.5}{1}} - \frac{4.5}{1} e^{-\frac{4.5}{1}} - e^{-\frac{4.5}{1}} + 1 \right) = \mathbf{0.1736}$$

(iv) Pr[Rejecting H_0 | H_1 true]... 1 - Pr[Accepting H_0 | H_0 true] = 1 - 0.1736 = **0.8264**

(f) Slutsky Test - α = 0.8946; power = 0.9889; New Test - α = 0.3907; power = 0.8264; significance is nice because we're sure to not reject that the toy is of high quality when it actually is, but in this case power is probably more important to Prof. Slutsky. Power gives the probability of rejecting the hypothesis that the more expensive toy is of high quality when it really isn't. Since Prof. Slutsky is worried about spending too much money, power is more important to him and he should use his test. Could increase rejection region of second test to 8.3 (vs. 4.5; this was found by trial and error in Excel). This gives same power (0.9889), and α = 0.783 which is better.

(g) x_1 = 1.6, x_2 = 0.8, x_3 = 2.4 ⇒ $\tilde{x}$ = (1.6 + 0.8 + 2.4) = 4.8

1 - Pr[$f(\tilde{x};1) \geq 4.8$] = 1 - (1 - Pr[$f(\tilde{x};1) < 4.8$]) = $F(4.8;1)$ =

$$\left(-\frac{(4.8)^2}{2(1)^2} e^{-\frac{4.8}{1}} - \frac{2(4.8)}{2(1)} e^{-\frac{4.8}{1}} - e^{-\frac{4.8}{1}} + 1 \right) = \mathbf{0.8574}$$

Buy more toys to improve the confidence.

2. Mendelian theory indicates that the shape and color of a certain variety of pea ought to be divided into four groups: round and yellow, round and green, angular and yellow, and angular and green, according to the ratios 9/3/3/1. For n = 556 peas, the following were observed.

Round and yellow	315
Round and green	108
Angular and yellow	101
Angular and green	32

Test whether the data agrees with Mendelian theory.

H_0: $p_1 = {}^9/_{16}$, $p_2 = {}^3/_{16}$, $p_3 = {}^3/_{16}$ (implies $p_4 = 1 - p_1 - p_2 - p_3 = {}^1/_{16}$)
H_a: $p_1 \neq {}^9/_{16}$, $p_2 \neq {}^3/_{16}$, $p_3 \neq {}^3/_{16}$

Multinomial Distribution

$$f(\mathbf{x};\boldsymbol{\theta}) = p_1^{x_1} p_2^{x_2} p_3^{x_3} (1 - p_1 - p_2 - p_3)^{x_4}, \quad x_1 + x_2 + x_3 + x_4,$$

where $\mathbf{x} = [x_1 \quad x_2 \quad x_3 \quad x_4]^T$ and $\boldsymbol{\theta} = [p_1 \quad p_2 \quad p_3]^T$

$$L(\mathbf{x};\boldsymbol{\theta}) = \prod_{i=1}^{N} f(\mathbf{x}_i;\boldsymbol{\theta}) = \prod_{i=1}^{N} p_1^{x_{1i}} p_2^{x_{2i}} p_3^{x_{3i}} (1 - p_1 - p_2 - p_3)^{x_{4i}} =$$

$$p_1^{n_1} p_2^{n_2} p_3^{n_3} (1 - p_1 - p_2 - p_3)^{N - n_1 - n_2 - n_3}, \text{ where } n_j = \sum_{i=1}^{N} x_{ji}, \quad j = 1,2,3$$

Note: $E(n_j) = \sum_{i=1}^{N} \frac{n_j}{N} = \sum_{i=1}^{N} p_j = Np_j$

Note: for simplicity substitute $n_4 = N - n_1 - n_2 - n_3$

$$L(\mathbf{x};\boldsymbol{\theta}) = p_1^{n_1} p_2^{n_2} p_3^{n_3} (1 - p_1 - p_2 - p_3)^{n_4}$$

$$\ln L(\mathbf{x};\boldsymbol{\theta}) = n_1 \ln(p_1) + n_2 \ln(p_2) + n_3 \ln(p_3) + n_4 \ln(1 - p_1 - p_2 - p_3)$$

$$\text{Score } S(\mathbf{x},\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial \ln L(\mathbf{x};\boldsymbol{\theta})}{\partial p_1} \\ \frac{\partial \ln L(\mathbf{x};\boldsymbol{\theta})}{\partial p_2} \\ \frac{\partial \ln L(\mathbf{x};\boldsymbol{\theta})}{\partial p_3} \end{bmatrix} = \begin{bmatrix} n_1 \frac{1}{p_1} - n_4 \frac{1}{1 - p_1 - p_2 - p_3} \\ n_2 \frac{1}{p_2} - n_4 \frac{1}{1 - p_1 - p_2 - p_3} \\ n_3 \frac{1}{p_3} - n_4 \frac{1}{1 - p_1 - p_2 - p_3} \end{bmatrix}$$

To find MLE, solve $S(\mathbf{x},\boldsymbol{\theta}) = \mathbf{0}$ for $\boldsymbol{\theta}$; lots of complicated algebra will show that

$$\hat{p}_j = \frac{n_j}{N}, \quad j = 1,2,3 \tag{4}$$

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}, \text{ where } I_{i,j} = -E\left[\frac{\partial^2 \ln L(x;\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right]$$

(I'm only doing this because of very liberal use of cut and paste from the midterm.)

$$I_{11} = -E\left[\frac{\partial}{\partial p_1}\left(n_1\frac{1}{p_1} - n_4\frac{1}{1-p_1-p_2-p_3}\right)\right] = -E\left[-n_1\frac{1}{p_1^2} - n_4\frac{1}{(1-p_1-p_2-p_3)^2}\right] =$$

$$\frac{1}{p_1^2}E(n_1) + \frac{1}{(1-p_1-p_2-p_3)^2}E(n_4) =$$

$$\frac{1}{p_1^2}Np_1 + \frac{1}{(1-p_1-p_2-p_3)^2}N(1-p_1-p_2-p_3) = \frac{N}{p_1} + \frac{N}{(1-p_1-p_2-p_3)}$$

$$I_{12} = -E\left[\frac{\partial}{\partial p_2}\left(n_1\frac{1}{p_1} - n_4\frac{1}{1-p_1-p_2-p_3}\right)\right] = -E\left[-n_4\frac{1}{(1-p_1-p_2-p_3)^2}\right] =$$

$$\frac{1}{(1-p_1-p_2-p_3)^2}E(n_4) = \frac{1}{(1-p_1-p_2-p_3)^2}N(1-p_1-p_2-p_3) =$$

$$\frac{N}{(1-p_1-p_2-p_3)}$$

$$I_{13} = -E\left[\frac{\partial}{\partial p_3}\left(n_1\frac{1}{p_1} - n_4\frac{1}{1-p_1-p_2-p_3}\right)\right] = \frac{N}{(1-p_1-p_2-p_3)}$$ (same as I_{12})

$$I_{22} = -E\left[\frac{\partial}{\partial p_2}\left(n_2\frac{1}{p_2} - n_4\frac{1}{1-p_1-p_2-p_3}\right)\right] = -E\left[-n_2\frac{1}{p_2^2} - n_4\frac{1}{(1-p_1-p_2-p_3)^2}\right] =$$

$$\frac{1}{p_2^2}E(n_2) + \frac{1}{(1-p_1-p_2-p_3)^2}E(n_4) =$$

$$\frac{1}{p_2^2}Np_2 + \frac{1}{(1-p_1-p_2-p_3)^2}N(1-p_1-p_2-p_3) = \frac{N}{p_2} + \frac{N}{(1-p_1-p_2-p_3)}$$

$$I_{23} = -E\left[\frac{\partial}{\partial p_3}\left(n_2\frac{1}{p_2} - n_4\frac{1}{1-p_1-p_2-p_3}\right)\right] = -E\left[-n_4\frac{1}{(1-p_1-p_2-p_3)^2}\right] =$$

$$\frac{1}{(1-p_1-p_2-p_3)^2}E(n_4) = \frac{1}{(1-p_1-p_2-p_3)^2}N(1-p_1-p_2-p_3) =$$

$$\frac{N}{(1-p_1-p_2-p_3)}$$

$$I_{33} = -E\left[\frac{\partial}{\partial p_3}\left(n_3\frac{1}{p_3} - n_4\frac{1}{1-p_1-p_2-p_3}\right)\right] = -E\left[-n_3\frac{1}{p_3^2} - n_4\frac{1}{(1-p_1-p_2-p_3)^2}\right] =$$

$$\frac{1}{p_3^2}E(n_3) + \frac{1}{(1-p_1-p_2-p_3)^2}E(n_4) =$$

$$\frac{1}{p_3^2}Np_3 + \frac{1}{(1-p_1-p_2-p_3)^2}N(1-p_1-p_2-p_3) = \frac{N}{p_3} + \frac{N}{(1-p_1-p_2-p_3)}$$

(Wasn't that fun?)

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{p_1} + \frac{N}{(1-p_1-p_2-p_3)} & \frac{N}{(1-p_1-p_2-p_3)} & \frac{N}{(1-p_1-p_2-p_3)} \\ \frac{N}{(1-p_1-p_2-p_3)} & \frac{N}{p_2} + \frac{N}{(1-p_1-p_2-p_3)} & \frac{N}{(1-p_1-p_2-p_3)} \\ \frac{N}{(1-p_1-p_2-p_3)} & \frac{N}{(1-p_1-p_2-p_3)} & \frac{N}{p_3} + \frac{N}{(1-p_1-p_2-p_3)} \end{bmatrix}$$

Critical Value - estimating 3 parameters (p_1, p_2, p_3); 95% χ^2_3 = 7.814 ∴ reject H_0 if test statistic exceeds 7.814.

Summary of Data - just to make things easier to find

$$\boldsymbol{\theta}_0 = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} 9/16 \\ 3/16 \\ 3/16 \end{bmatrix} = \begin{bmatrix} 0.5625 \\ 0.1875 \\ 0.1875 \end{bmatrix} \qquad \hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_3 \end{bmatrix} = \begin{bmatrix} 315/556 \\ 108/556 \\ 101/556 \end{bmatrix} = \begin{bmatrix} 0.5665 \\ 0.1942 \\ 0.1817 \end{bmatrix}$$

$$S(\mathbf{x};\boldsymbol{\theta}_0) = \begin{bmatrix} 315\frac{1}{0.5625} - 32\frac{1}{0.0625} \\ 108\frac{1}{0.1875} - 32\frac{1}{0.0625} \\ 101\frac{1}{0.1875} - 32\frac{1}{0.0625} \end{bmatrix} = \begin{bmatrix} 48 \\ 64 \\ 26.6667 \end{bmatrix}$$

$$\mathbf{I}(\boldsymbol{\theta}_0) = \begin{bmatrix} \frac{556}{0.5625} + \frac{556}{0.0625} & \frac{556}{0.0625} & \frac{556}{0.0625} \\ \frac{556}{0.0625} & \frac{556}{0.1875} + \frac{556}{0.0625} & \frac{556}{0.0625} \\ \frac{556}{0.0625} & \frac{556}{0.0625} & \frac{556}{0.1875} + \frac{556}{0.0625} \end{bmatrix} =$$

$$\begin{bmatrix} 9884.44 & 8896 & 8896 \\ 8896 & 11861.33 & 8896 \\ 8896 & 8896 & 11861.33 \end{bmatrix}$$

$$\mathbf{I}^{-1}(\boldsymbol{\theta}_0) = \begin{bmatrix} 0.000443 & -0.00019 & -0.00019 \\ -0.00019 & 0.000274 & -0.000063 \\ -0.00019 & -0.000063 & 0.000274 \end{bmatrix}$$

$$\mathbf{I}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \frac{556}{0.5665}+\frac{556}{0.0576} & \frac{556}{0.0576} & \frac{556}{0.0576} \\ \frac{556}{0.0576} & \frac{556}{0.1942}+\frac{556}{0.0576} & \frac{556}{0.0576} \\ \frac{556}{0.0576} & \frac{556}{0.0576} & \frac{556}{0.1817}+\frac{556}{0.0576} \end{bmatrix} =$$

$$\begin{bmatrix} 10641.88 & 9660.5 & 9660.5 \\ 9660.5 & 12522.87 & 9660.5 \\ 9660.5 & 9660.5 & 12721.25 \end{bmatrix}$$

Likelihood Ratio Test -

$\xi_{LR} = -2\left(\ln L(\mathbf{x};\boldsymbol{\theta}_0) - \ln(\mathbf{x};\hat{\boldsymbol{\theta}})\right) =$

$-2(315\ln(0.5625)+108\ln(0.1875)+101\ln(0.1875)+32\ln(0.0625)-$
$315\ln(0.5665)-108\ln(0.1942)-101\ln(0.1817)-32\ln(0.0576))$ = **0.475** < 7.814 $\therefore$ don't reject H_0. The data seems to agree with Mendelian theory

Lagrange Multiplier Test -

$\xi_{LM} = S^T(\mathbf{x};\boldsymbol{\theta}_0)I^{-1}(\boldsymbol{\theta}_0)S(\mathbf{x};\boldsymbol{\theta}_0) =$

$$\begin{bmatrix} 48 & 64 & 26.6667 \end{bmatrix} \begin{bmatrix} 0.000443 & -0.00019 & -0.00019 \\ -0.00019 & 0.000274 & -0.000063 \\ -0.00019 & -0.000063 & 0.000274 \end{bmatrix} \begin{bmatrix} 48 \\ 64 \\ 26.6667 \end{bmatrix}$$ = **0.470** < 7.814 $\therefore$

don't reject H_0. The data seems to agree with Mendelian theory

Wald Test -

$$(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = \begin{bmatrix} 0.5665 \\ 0.1942 \\ 0.1817 \end{bmatrix} - \begin{bmatrix} 0.5625 \\ 0.1875 \\ 0.1875 \end{bmatrix} = \begin{bmatrix} 0.004 \\ 0.0067 \\ -0.0058 \end{bmatrix}$$

$$\xi_W = (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T \mathbf{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = \begin{bmatrix} 0.004 \\ 0.0067 \\ -0.0058 \end{bmatrix}^T \begin{bmatrix} 0.000443 & -0.00019 & -0.00019 \\ -0.00019 & 0.000274 & -0.000063 \\ -0.00019 & -0.000063 & 0.000274 \end{bmatrix} \begin{bmatrix} 0.004 \\ 0.0067 \\ -0.0058 \end{bmatrix}$$

= **0.487** < 7.815 $\therefore$ don't reject H_0. The data seems to agree with Mendelian theory

Documentation.

Prof Werner gave me the formula for $F(\tilde{x};\theta)$ in 1e. I worked all of problem 1 with Josh Kneifel and Prof Werner... lovely arguments on the wording of the question and violent agreement about our calculations.

1. Assume that the following data come from the linear model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad i = 1,2,\ldots,n$$

y	-6.1	-0.5	7.2	6.9	-0.2	-2.1	-3.9	3.8
x	-2.0	0.6	1.4	1.3	0.0	-1.6	-1.7	0.7

Find the maximum likelihood estimates of β_0, β_1, and σ^2

$$\boxed{\hat{\boldsymbol{\beta}} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1.1630 \\ 3.2338 \end{bmatrix}}$$

$$\hat{\sigma}^2 = \frac{1}{n}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 = \frac{24.0838}{8} = \mathbf{3.01}$$

$$s^2 = \frac{1}{n-k}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} = \frac{1}{n-k}\sum_{i=1}^{n}\varepsilon_i^2 = \frac{24.0838}{6} = \mathbf{4.01}$$

2. The model:

$$Y = \alpha_1 + \alpha_2 E_2 + \alpha_3 E_3 + u$$

is estimated by OLS, where E_2 and E_3 are dummy variables indicating membership of the second and third educational classes, respectively. Show that the OLS estimates are:

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \overline{Y}_1 \\ \overline{Y}_2 - \overline{Y}_1 \\ \overline{Y}_3 - \overline{Y}_1 \end{bmatrix}$$

where $\overline{Y}_i$ denotes the mean value of Y in the i^{th} educational class.

n_i = number in each educational class; $\sum_{i=1}^{3} n_i = n$

$$
\begin{array}{c} \\ \mathbf{x} = \end{array}
\begin{array}{c}
\begin{array}{ccc} \text{Const} & E_2 & E_3 \end{array} \\
\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix}
\end{array}
\Rightarrow \mathbf{x}^T\mathbf{x} =
\begin{bmatrix} \overbrace{1 \quad 1 \quad \cdots}^{n_1} & \overbrace{1 \quad 1 \quad \cdots}^{n_2} & \overbrace{1 \quad 1 \quad \cdots}^{n_3} \\ 0 \quad 0 \quad \cdots & 1 \quad 1 \quad \cdots & 0 \quad 0 \quad \cdots \\ 0 \quad 0 \quad \cdots & 0 \quad 0 \quad \cdots & 1 \quad 1 \quad \cdots \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix}
= \begin{bmatrix} n & n_2 & n_3 \\ n_2 & n_2 & 0 \\ n_3 & 0 & n_3 \end{bmatrix}
$$

Find $(\mathbf{x}^T\mathbf{x})^{-1}$:

$$
\left[\begin{array}{ccc|ccc} n & n_2 & n_3 & 1 & 0 & 0 \\ n_2 & n_2 & 0 & 0 & 1 & 0 \\ n_3 & 0 & n_3 & 0 & 0 & 1 \end{array}\right]
$$

Subtract rows 2 and 3 from row 1; realize $n - n_2 - n_3 = n_1$

$$
\left[\begin{array}{ccc|ccc} n-n_2-n_3 & 0 & 0 & 1 & -1 & -1 \\ n_2 & n_2 & 0 & 0 & 1 & 0 \\ n_3 & 0 & n_3 & 0 & 0 & 1 \end{array}\right]
= \left[\begin{array}{ccc|ccc} n_1 & 0 & 0 & 1 & -1 & -1 \\ n_2 & n_2 & 0 & 0 & 1 & 0 \\ n_3 & 0 & n_3 & 0 & 0 & 1 \end{array}\right]
$$

Divide row 1 by n_1; row 2 by n_2; row 3 by n_3

$$
\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/n_1 & -1/n_1 & -1/n_1 \\ 1 & 1 & 0 & 0 & 1/n_2 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1/n_3 \end{array}\right]
$$

Subtract row 1 from row 2 and row 3

$$
\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/n_1 & -1/n_1 & -1/n_1 \\ 0 & 1 & 0 & -1/n_1 & 1/n_2 + 1/n_1 & 1/n_1 \\ 0 & 0 & 1 & -1/n_1 & 1/n_1 & 1/n_3 + 1/n_1 \end{array}\right]
$$

$$
\therefore (\mathbf{x}^T\mathbf{x})^{-1} = \begin{bmatrix} 1/n_1 & -1/n_1 & -1/n_1 \\ -1/n_1 & 1/n_2 + 1/n_1 & 1/n_1 \\ -1/n_1 & 1/n_1 & 1/n_3 + 1/n_1 \end{bmatrix}
$$

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 & \cdots & 1 & 1 & \cdots \\ 0 & 0 & \cdots & 1 & 1 & \cdots & 0 & 0 & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 1 & \cdots \end{bmatrix} \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n_2} \\ Y_{31} \\ Y_{32} \\ \vdots \\ Y_{3n_3} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{3}\sum_{j=1}^{n_i} Y_{ij} \\ \sum_{j=1}^{n_2} Y_{2j} \\ \sum_{j=1}^{n_3} Y_{3j} \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y} = \begin{bmatrix} 1/n_1 & -1/n_1 & -1/n_1 \\ -1/n_1 & 1/n_2 + 1/n_1 & 1/n_1 \\ -1/n_1 & 1/n_1 & 1/n_3 + 1/n_1 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{3}\sum_{j=1}^{n_i} Y_{ij} \\ \sum_{j=1}^{n_2} Y_{2j} \\ \sum_{j=1}^{n_3} Y_{3j} \end{bmatrix} =$$

$$\begin{bmatrix} \frac{1}{n_1}\sum_{i=1}^{3}\sum_{j=1}^{n_i} Y_{ij} - \frac{1}{n_1}\sum_{j=1}^{n_2} Y_{2j} - \frac{1}{n_1}\sum_{j=1}^{n_3} Y_{3j} \\ -\frac{1}{n_1}\sum_{i=1}^{3}\sum_{j=1}^{n_i} Y_{ij} + \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{2j} + \frac{1}{n_1}\sum_{j=1}^{n_2} Y_{2j} + \frac{1}{n_1}\sum_{j=1}^{n_3} Y_{3j} \\ -\frac{1}{n_1}\sum_{i=1}^{3}\sum_{j=1}^{n_i} Y_{ij} + \frac{1}{n_1}\sum_{j=1}^{n_2} Y_{2j} + \frac{1}{n_3}\sum_{j=1}^{n_3} Y_{3j} + \frac{1}{n_1}\sum_{j=1}^{n_3} Y_{3j} \end{bmatrix} = \begin{bmatrix} \frac{1}{n_1}\sum_{j=1}^{n_1} Y_{1j} \\ \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{2j} - \frac{1}{n_1}\sum_{j=1}^{n_1} Y_{1j} \\ \frac{1}{n_3}\sum_{j=1}^{n_3} Y_{3j} - \frac{1}{n_1}\sum_{j=1}^{n_1} Y_{1j} \end{bmatrix} =$$

$$\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 - \bar{Y}_1 \\ \bar{Y}_3 - \bar{Y}_1 \end{bmatrix}$$ (which is what we wanted to show)

3. Suppose that C and Y represent per capita consumption and per capita disposable income respectively, and that J.M. Keynes thinks that they are related by the equation of the form:

$$C = bY + u \quad u \sim N(0, \sigma^2)$$

Keynes wants to estimate the parameter b (which measures the marginal propensity to consume). The random variable u is unobservable, but Keynes can obtain n pairs of observations (C_1,Y_1), (C_2,Y_2), ..., (C_n,Y_n), which relate observations of (C_i,Y_i) over n different years.

(i) Find an expression for the ordinary least squares estimator, $\hat{b}_{OLS}$.

(ii) Use ordinary least squares and the following data to estimate the marginal propensity to consume.

Year	1930	1931	1932	1933	1934	1935
C_i	1059	1016	919	897	934	985
Y_i	1128	1077	921	893	952	1035

(i) $\mathbf{x} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix}$

$$\mathbf{x}^T\mathbf{x} = \begin{bmatrix} Y_1 & \cdots & Y_n \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{i=1}^{n} Y_i^2$$

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} Y_1 & \cdots & Y_n \end{bmatrix} \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix} = \sum_{i=1}^{n} C_i Y_i$$

$$\hat{b}_{OLS} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y} = \boxed{\frac{\sum_{i=1}^{n} C_i Y_i}{\sum_{i=1}^{n} Y_i^2}}$$

(ii) $\hat{b}_{OLS}$ = **0.9652**

4. The hardness, y, of the shells of eggs laid by a certain breed of chickens was assumed to be roughly linearly related to the amount, x, of a certain food supplement put into the diet of chickens. The model assumed is the classical linear regression model. Data were collected and are given below:

y_i	0.70	0.98	1.16	1.75	0.76	0.82	0.95	1.24	1.75	1.95
x_i	0.12	0.21	0.34	0.61	0.13	0.17	0.21	0.34	0.62	0.71

(i) Test the hypothesis that β_1 = 1.00 versus the hypothesis that $\beta_1 \neq$ 1.00. Use a Type I error probability of 5 percent.
(ii) Test the hypothesis that β_1 > 1 versus the hypothesis that $\beta_1 \leq$ 1.

(i) H_0: β_1 = 1.00
H_a: $\beta_1 \neq$ 1.00

$$\mathbf{R} = \begin{bmatrix} 0 & 1 \end{bmatrix}$$

$$\mathbf{q} = \begin{bmatrix} 1.00 \end{bmatrix}$$

$$F = \frac{\left\{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})^T \left[\mathbf{R}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^T\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})\right\}/m}{\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}/(n-k)} \sim F_{m,n-k}$$

Rejection region: F < FINV(0.975,1,8) = 0.0010 or F > FINV(0.025,1,8) = 7.5709
F = 415.7 > 7.5709 $\therefore$ reject H_0 and conclude $\beta_1 \neq$ 1.00

(ii) H_0: β_1 > 1.00
H_a: $\beta_1 \leq$ 1.00
Same set up, but now rejection region for β_1 = 1.00: F > FINV(0.05,1,8) = 5.318
F = 415.7 > 5.318 $\therefore$ don't reject H_0 (not sufficient evidence to say $\beta_1 \leq$ 1.00)

5. A production function model is specified as:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

where Y_i = log output, X_{2i} = log labor input, and X_{3i} = log capital input. The data refer to a sample of 23 firms, and the observations are measured as from the sample means:

$$\begin{array}{ll} \sum x_{2i}^2 = 12 & \sum x_{2i}x_{3i} = 8 \\ \sum x_{3i}^2 = 12 & \sum y_i x_{2i} = 10 \\ \sum y_i^2 = 10 & \sum y_i x_{3i} = 8 \end{array}$$

(i) Estimate β_2, β_3, their standard errors, and R^2.
(ii) Test the hypothesis that $\beta_2 + \beta_3 = 1$.
(iii) Suppose now that you want to impose the restriction that $\beta_2 + \beta_3 = 1$. What is the least squares estimate of β_2 and its standard error? What is the value of R^2 in this case? Compare the results with those obtained in part (i) and comment.

(i) "as from the sample means" $\Rightarrow \sum x_{2i} = \sum x_{3i} = \sum y_i = 0$

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y}$$

$$\mathbf{x}^T\mathbf{x} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{21} & x_{22} & \cdots & x_{2n} \\ x_{31} & x_{32} & \cdots & x_{3n} \end{bmatrix} \begin{bmatrix} 1 & x_{21} & x_{31} \\ 1 & x_{22} & x_{32} \\ \vdots & \vdots & \vdots \\ 1 & x_{2n} & x_{3n} \end{bmatrix} = \begin{bmatrix} n & \sum x_{2i} & \sum x_{3i} \\ \sum x_{2i} & \sum x_{2i}^2 & \sum x_{2i}x_{3i} \\ \sum x_{3i} & \sum x_{2i}x_{3i} & \sum x_{3i}^2 \end{bmatrix} =$$

$$\begin{bmatrix} 23 & 0 & 0 \\ 0 & 12 & 8 \\ 0 & 8 & 12 \end{bmatrix}$$

$$(\mathbf{x}^T\mathbf{x})^{-1} = \begin{bmatrix} 0.0435 & 0 & 0 \\ 0 & 0.15 & -0.1 \\ 0 & -0.1 & 0.15 \end{bmatrix}$$

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} \sum y_i \\ \sum y_i x_{2i} \\ \sum y_i x_{3i} \end{bmatrix} = \begin{bmatrix} 0 \\ 10 \\ 8 \end{bmatrix}$$

$$(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y} = \begin{bmatrix} 0.0435 & 0 & 0 \\ 0 & 0.15 & -0.1 \\ 0 & -0.1 & 0.15 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.7 \\ 0.2 \end{bmatrix} \therefore \boxed{\hat{\boldsymbol{\beta}}_{OLS} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_3 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.7 \\ 0.2 \end{bmatrix}}$$

$\hat{\boldsymbol{\beta}}_{OLS} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{x}^T\mathbf{x})^{-1})$

$\hat{\sigma}^2 = \frac{1}{n}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}$ = 1.4/23 = 0.06087

Note 1: $\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} = \sum_{i=1}^{n} e_i^2$ = RSS (see below)

Note 2: This $\hat{\sigma}^2$ is based on MLE for σ^2, but could use $s^2 = \frac{1}{n-k}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}$ for unbiased estimator in which case s^2 = 1.4/20 = 0.07

$$\hat{\sigma}^2(\mathbf{x}^T\mathbf{x})^{-1} = \frac{1.4}{23}\begin{bmatrix} 0.0435 & 0 & 0 \\ 0 & 0.15 & -0.1 \\ 0 & -0.1 & 0.15 \end{bmatrix} = \begin{bmatrix} 0.002647 & 0 & 0 \\ 0 & 0.00913 & -0.00609 \\ 0 & -0.00609 & 0.00913 \end{bmatrix}$$

Using $s^2(\mathbf{x}^T\mathbf{x})^{-1}$, $\frac{1.4}{20}\begin{bmatrix} 0.0435 & 0 & 0 \\ 0 & 0.15 & -0.1 \\ 0 & -0.1 & 0.15 \end{bmatrix} = \begin{bmatrix} 0.003043 & 0 & 0 \\ 0 & 0.0105 & -0.007 \\ 0 & -0.007 & 0.0105 \end{bmatrix}$

$\therefore$ $\hat{\sigma}_{\hat{\beta}_2} = \hat{\sigma}_{\hat{\beta}_3}$ = sqrt(0.00913) = 0.0956

$s_{\hat{\beta}_2} = s_{\hat{\beta}_3}$ = sqrt(0.0105) = 0.1025

RSS = $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i})^2 = \sum_{i=1}^{n}(y_i - 0.7x_{2i} - 0.2x_{3i})^2$ =

$\sum_{i=1}^{n}(y_i^2 - 0.7y_i x_{2i} - 0.2y_i x_{3i} - 0.7y_i x_{2i} + 0.49x_{2i}^2 + 0.14x_{2i}x_{3i} - 0.2y_i x_{3i} + 0.14x_{2i}x_{3i} + 0.04x_{3i}^2)$

$= \sum_{i=1}^{n}(y_i^2 - 1.4y_i x_{2i} - 0.4y_i x_{3i} + 0.49x_{2i}^2 + 0.28x_{2i}x_{3i} + 0.04x_{3i}^2)$ =

$\sum_{i=1}^{n} y_i^2 - 1.4\sum_{i=1}^{n} y_i x_{2i} - 0.4\sum_{i=1}^{n} y_i x_{3i} + 0.49\sum_{i=1}^{n} x_{2i}^2 + 0.28\sum_{i=1}^{n} x_{2i}x_{3i} + 0.04\sum_{i=1}^{n} x_{3i}^2$ =

10 - 1.4(10) - 0.4(8) + 0.49(12) + 0.28(8) + 0.04(12) = 1.4

TSS = $\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2$ = 10

$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$ = 1 - 1.4/10 = $\boxed{R^2 = 0.86}$

(ii) H_0: $\beta_2 + \beta_3 = 1$

H_a: $\beta_2 + \beta_3 \neq 1$

$\mathbf{R} = [0 \quad 1 \quad 1]$

$\mathbf{q} = [1]$

$$F = \frac{\left\{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})^T\left[\mathbf{R}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^T\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})\right\}/m}{\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}/(n-k)} \sim F_{m,n-k}$$

Rejection region: F < FINV(0.975,1,20) = 0.0010 or F > FINV(0.025,1,20) = 5.87

$$F = \frac{-0.1(1/0.1)(-0.1)}{0.07} = 1.429 \quad \therefore \ \underline{\text{do not}} \text{ reject } H_0$$

(iii) $y_i = \beta_1 + \beta_2 x_{2i} + (1-\beta_2)x_{3i} \Rightarrow y_i - x_{3i} = \beta_1 + \beta_2(x_{2i} - x_{3i})$

$$\tilde{\mathbf{y}} = \begin{bmatrix} y_1 - x_{31} \\ \vdots \\ y_n - x_{3n} \end{bmatrix}, \ \tilde{\mathbf{x}} = \begin{bmatrix} 1 & x_{21} - x_{31} \\ \vdots & \vdots \\ 1 & x_{2n} - x_{3n} \end{bmatrix}$$

$$\tilde{\mathbf{x}}^T \tilde{\mathbf{x}} = \begin{bmatrix} 1 & \cdots & 1 \\ x_{21} - x_{31} & \cdots & x_{2n} - x_{3n} \end{bmatrix} \begin{bmatrix} 1 & x_{21} - x_{31} \\ \vdots & \vdots \\ 1 & x_{2n} - x_{3n} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_{2i} - x_{3i} \\ \sum_{i=1}^n x_{2i} - x_{3i} & \sum_{i=1}^n (x_{2i} - x_{3i})^2 \end{bmatrix} =$$

$$\begin{bmatrix} n & \sum_{i=1}^n x_{2i} - \sum_{i=1}^n x_{3i} \\ \sum_{i=1}^n x_{2i} - \sum_{i=1}^n x_{3i} & \sum_{i=1}^n (x_{2i}^2 - 2x_{2i}x_{3i} + x_{3i}^2) \end{bmatrix} =$$

$$\begin{bmatrix} n & \sum_{i=1}^n x_{2i} - \sum_{i=1}^n x_{3i} \\ \sum_{i=1}^n x_{2i} - \sum_{i=1}^n x_{3i} & \sum_{i=1}^n x_{2i}^2 - 2\sum_{i=1}^n x_{2i}x_{3i} + \sum_{i=1}^n x_{3i}^2 \end{bmatrix} = \begin{bmatrix} 23 & 0 \\ 0 & 12 - 2(8) + 12 \end{bmatrix} = \begin{bmatrix} 23 & 0 \\ 0 & 8 \end{bmatrix}$$

$$(\tilde{\mathbf{x}}^T \tilde{\mathbf{x}})^{-1} = \begin{bmatrix} 0.0435 & 0 \\ 0 & 0.125 \end{bmatrix}$$

$$\tilde{\mathbf{x}}^T \tilde{\mathbf{y}} = \begin{bmatrix} 1 & \cdots & 1 \\ x_{21} - x_{31} & \cdots & x_{2n} - x_{3n} \end{bmatrix} \begin{bmatrix} y_1 - x_{31} \\ \vdots \\ y_n - x_{3n} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n (y_i - x_{3i}) \\ \sum_{i=1}^n (y_i - x_{3i})^2 \end{bmatrix} =$$

$$\begin{bmatrix} \sum_{i=1}^n (y_i - x_{3i}) \\ \sum_{i=1}^n (y_i^2 - 2y_i x_{3i} + x_{3i}^2) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i - \sum_{i=1}^n x_{3i} \\ \sum_{i=1}^n y_i^2 - 2\sum_{i=1}^n y_i x_{3i} + \sum_{i=1}^n x_{3i}^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 10 - 2(8) + 12 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

$$(\tilde{\mathbf{x}}^T \tilde{\mathbf{x}})^{-1} \tilde{\mathbf{x}}^T \tilde{\mathbf{y}} = \begin{bmatrix} 0.0435 & 0 \\ 0 & 0.125 \end{bmatrix} \begin{bmatrix} 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.75 \end{bmatrix} \ \therefore \ \boxed{\hat{\boldsymbol{\beta}}_{OLS} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.75 \end{bmatrix}}$$

$$\hat{\boldsymbol{\beta}}_{OLS} \sim N(\boldsymbol{\beta}, \sigma^2 (\mathbf{x}^T \mathbf{x})^{-1})$$

$$\hat{\sigma}^2 = \frac{1}{n} \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = 1.5/23$$

Note 1: $\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = \sum_{i=1}^n e_i^2$ = RSS (see below)

Note 2: This $\hat{\sigma}^2$ is based on MLE for σ^2, but could use $s^2 = \frac{1}{n-k}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}$ for unbiased estimator in which case s^2 = 1.5/21

$$\hat{\sigma}^2(\tilde{\mathbf{x}}^T\tilde{\mathbf{x}})^{-1} = \frac{1.5}{23}\begin{bmatrix} 0.0435 & 0 \\ 0 & 0.125 \end{bmatrix} = \begin{bmatrix} 0.002836 & 0 \\ 0 & 0.008152 \end{bmatrix}$$

Using $s^2(\tilde{\mathbf{x}}^T\tilde{\mathbf{x}})^{-1}$, $\frac{1.5}{21}\begin{bmatrix} 0.0435 & 0 \\ 0 & 0.125 \end{bmatrix} = \begin{bmatrix} 0.003106 & 0 \\ 0 & 0.008929 \end{bmatrix}$

$\therefore$ $\boxed{\begin{array}{l} \hat{\sigma}_{\hat{\beta}_2} = \text{sqrt}(0.008152) = 0.0903 \\ s_{\hat{\beta}_2} = \text{sqrt}(0.008929) = 0.0945 \end{array}}$

RSS = $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left[(y_i - x_{3i}) - \hat{\beta}_1 - \hat{\beta}_2(x_{2i} - x_{3i})\right]^2 = \sum_{i=1}^{n}[y_i - 0.75x_{2i} - 0.25x_{3i}]^2 =$

$\sum_{i=1}^{n}(y_i^2 - 1.5y_ix_{2i} - 0.5y_ix_{3i} + 0.5625x_{2i}^2 + 0.375x_{2i}x_{3i} + 0.0625x_{3i}^2) =$

$\sum_{i=1}^{n} y_i^2 - 1.5\sum_{i=1}^{n} y_ix_{2i} - 0.5\sum_{i=1}^{n} y_ix_{3i} + 0.5625\sum_{i=1}^{n} x_{2i}^2 + 0.375\sum_{i=1}^{n} x_{2i}x_{3i} + 0.0625\sum_{i=1}^{n} x_{3i}^2 =$

10 - 1.5(10) - 0.5(8) + 0.5625(12) + 0.375(8) + 0.0625(12) = 1.5

TSS = $\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2$ = 10

$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$ = 1 - 1.5/10 = $\boxed{R^2 = 0.85}$

In order to enforce $\beta_2 + \beta_3 = 1$, we had to make both parameters bigger. As the numbers turn out, the difference is shared equally between them (0.05 larger than in part i). Why that's the case, I have now idea. It does makes sense that forcing this condition results in a lower R^2 value (0.86 vs. 0.85) because we are no longer using all the data to minimize the sum of the squared residuals (we're still minimizing them, but the altered residuals aren't the same). One good thing about the second version is that the parameters are not correlated as they were in part i. I know the independent variables (x_{2i} and x_{3i}) are not supposed to be correlated, but we didn't talk about the parameters. The strange thing is that the standard errors for β_2 is actually smaller in the second case (although for β_1 it's higher).

6. You will be e-mailed a dataset on heights, sec, mother's height, and father's height.
(i) Using this dataset, estimate the unconstrained regression of height on sex, mother's height, and father's height. Interpret your results.
(ii) You wish to test the hypothesis that the coefficients on mother's height and father's height sum to one, and that the coefficient on sex is equal to -8. What is the $\mathbf{R}$ matrix and the $\mathbf{r}$ vector that correspond to these restrictions?
(iii) Conduct a test of this hypothesis that has an F-statistic. What are your conclusions?

(i) Height = 55.28 - 12.92 Gender + 0.317 Mom + 0.406 Dad + ε

On initial inspection it would appear gender is the biggest determinant of height, but it's not. What gender does imply is that women (gender = 1) are almost 13 cm shorter than men (assuming their parent's are the same height). This 13 cm isn't much when we factor in the average heights of moms and dads which have a greater impact on height than gender does (in absolute terms): + 51.8 for average mom and + 72.4 for average dad. Father's height is more significant because father's are taller in general, but also more significant on the margin (for each cm of height a father adds 0.406 cm to his offspring versus only 0.317 cm a mom passes on).

(ii) H_0: Mom + Dad = 1 and Gender = -8
H_a: one (or both) of these don't hold

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{q} = \begin{bmatrix} 1 \\ -8 \end{bmatrix}$$

(iii) Using Stata

```
. test mom + dad = 1
 ( 1)  mom + dad = 1
       F(  1,    29) =    2.95
            Prob > F =    0.0965
. test gender = -8, accumulate
 ( 1)  mom + dad = 1
 ( 2)  gender = -8
       F(  2,    29) =    7.44
            Prob > F =    0.0025
```

Up to a 99.75% confidence level, we can reject that both Mom + Dad = 1 and Gender = -8. Based on the results of the first test, it would appear Gender = -8 is the part that causes the joint test to fail.

Documentation

Prof Werner showed how to find $\mathbf{x}^T\mathbf{x}$ for problem 2 in class.

Prof Werner told me "as from the sample means" $\Rightarrow \sum x_{2i} = \sum x_{3i} = \sum y_i = 0$ in problem 5. She also told me how to set it up to do part iii.

Scott showed us how to use the `insheet`, `regress` and `test` commands in Stata to do problem 6.

1. A multiple choice exam has N questions, each of which has k possible answers. A student knows the correct answer to n of these questions. For the remaining $N - n$ questions, he checks the answers completely at random. To take into consideration the chance of students randomly selecting the correct response, the professor has decided that for each correct answer the student gets one point, and for each wrong answer he gets $-1/k$ points.
(a) Derive an expression for the density function for the students score.
(b) What is the expected value of the total score for the test? Do you have any recommendations that you could make?

(a) Probability of getting uncertain question correct is $1/k$
Number of uncertain questions student gets correct is binomial distribution with parameters $N - n$ (# of uncertain questions) and $1/k$ (probability of success)
Let X = number of uncertain questions the student gets correct ~ Bin($N - n$, $1/k$)

$$f(x) = \binom{N-n}{x}\frac{1}{k^x}\left(1-\frac{1}{k}\right)^{N-n-x}, \quad E(X) = (N-n)\frac{1}{k}, \quad Var(X) = (N-n)\frac{1}{k}\left(1-\frac{1}{k}\right)$$

Number of uncertain questions student gets incorrect is $N - n - X$
Score equals number of certain questions (n) plus number of uncertain questions student gets correct (X), minus $1/k$ times number of uncertain questions student gets incorrect ($N - n - X$)
Therefore, score is a function of X: $S(X) = n + X - 1/k(N - n - X) = n + X - (N - n)/k + X/k =$

$$S(x) = n - \frac{1}{k}(N-n) + x\left(1+\frac{1}{k}\right), \text{ where } 0 \le x \le N - n$$

That is, the student gets n points and we automatically subtract $1/k$ points for all uncertain questions, then add $1 + 1/k$ points for the uncertain questions he gets right.
Double check min and max scores:
$S(0) = n - 1/k(N - n)$ (gets none of the uncertain questions right)
$S(N - n) = n - 1/k(N - n) + (N - n) + 1/k(N - n) = N$ (gets all uncertain questions right)
$S(X)$ is a strictly increasing function of X, so there is a unique x that gives a particular score s.

The density of the score is $$\boxed{g(s) = \binom{N-n}{x}\frac{1}{k^x}\left(1-\frac{1}{k}\right)^{N-n-x}, \text{ where } x = \frac{s - n + \frac{1}{k}(N-n)}{1+\frac{1}{k}}}$$

This is valid for $n - 1/k(N - n) \le s \le N$

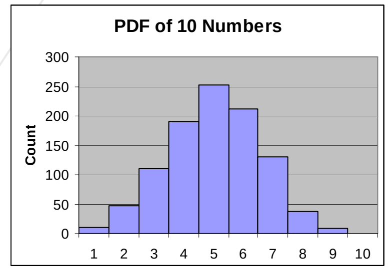
(b) The score is a linear function of X, therefore, the expected value of the function is equal to the function of the expected value, i.e., $E(S(X)) = S(E(X))$
Substitute the value for $E(X)$ given above into the formula for $S(X)$:

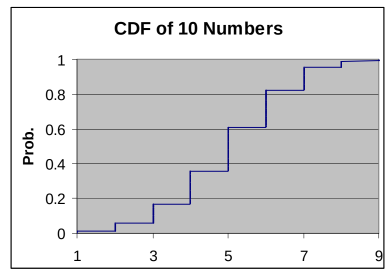
$$S(E(X)) = n - \frac{1}{k}(N-n) + \left[(N-n)\frac{1}{k}\right]\left(1+\frac{1}{k}\right) = \boxed{n + (N-n)\frac{1}{k^2}}$$

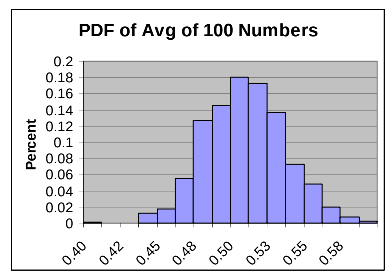
Ideally, the expected value of the score would be n in order to measure exactly what the student knew. Unfortunately, even by subtracting $1/k$ points, the expected value will be greater than n (although less than it would be without subtracting the points). Also, $(N - n)1/k^2$ will never be zero. In order to get it closer to zero, the professor should make k as large as possible (i.e., **more choices for each question**).

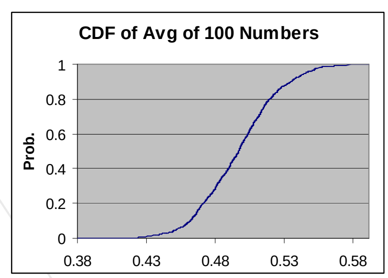
If penalty of $1/(k - 1)$ for incorrect answers, then $E(S) = n$.

2. Numbers are selected at random from the interval (0,1).
(a) If 10 numbers are selected, what is the probability that exactly 5 of them are less than 1/2?
(b) If 10 numbers are selected, on average how many of them are less than 1/2?
(c) If 100 numbers are selected, what is the probability that the average of the numbers is less than 1/2?









(a) Uniform(0,1)

$\Pr[\text{Uni}(0,1) < 1/2] = \int_0^{1/2} 1dx = x\Big]_0^{1/2} = 1/2$

X = # Uni(0,1) samples < 1/2 ~ Binomial with n = 10 and p = 1/2

$$f(5) = \binom{10}{5}\frac{1}{2^5}\left(1-\frac{1}{2}\right)^{10-5} = 252\frac{1}{2^{10}} = \boxed{{}^{63}/_{256} = \mathbf{0.246}}$$

(b) $E(X) = 10(1/2) = \boxed{5}$

Double checked with simulation in Excel:
- Trials: 1000
- Average: <u>5.03</u>
- StDev: 1.56
- $F(5)$ = 0.611
- $F(4)$ = 0.358
- Pr[5 samples < 1/2] = $F(5) - F(4)$ = <u>0.253</u>

(c) Use Theorem: If random sample is from a population with mean μ and variance σ^2, then the sample mean is a random variable with mean μ and variance σ^2/n.

In this case, population mean and variance are from Uni(0,1):
$E(X) = 1/2$, $Var(X) = 1/12$

$\bar{x}$ = average of sample of 100 Uni(0,1); $\bar{x} \sim N(1/2, 1/1200)$

$$\Pr[\bar{x} < 1/2] = \Pr\left[\frac{\bar{x}-1/2}{\sqrt{1/1200}} < \frac{1/2-1/2}{\sqrt{1/1200}}\right] = \Pr[z < 0] = \boxed{\mathbf{0.5}}$$

Double checked with simulation in Excel:
- Trials: 1000
- Average: <u>0.50</u>
- StDev: 0.03

3. Let $x_1, x_2, ..., x_m$ be a random sample from a normal distribution with mean μ_x and variance σ^2, and $y_1, y_2, ..., y_n$ be a sample from a normal distribution with mean μ_y and the same variance σ^2. Also x_i and y_j are independent for all i, j. Let $\bar{x}$ and $\bar{y}$ be the sample means, and s_x^2 and s_y^2 be the sample variances.

(a) Write down the distribution of $z = \bar{x} - \bar{y}$, it's mean and variance.

(b) What is the distribution of

$$u = \frac{(m-1)s_x^2 + (n-1)s_y^2}{\sigma^2}$$

(c) Using z and u, construct a random variable that has a t-distribution. What are its degrees of freedom?

(a) Theorem: If random sample is from a population with mean μ and variance σ^2, then the sample mean is a random variable with mean μ and variance σ^2/n.

$\therefore\ \bar{x} \sim N(\mu_x, \sigma^2/m)$ and $\bar{y} \sim N(\mu_y, \sigma^2/n)$

$$z = \bar{x} - \bar{y} \sim \boxed{N\left(\mu_x - \mu_y, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right)}$$

(p287 of *Statistics for Engineering and the Sciences*. 3rd ed. Mendenhall & Sincich)

$$f(z) = \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right)}} \exp\left[-\frac{1}{2\left(\frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right)}\left(z - (\mu_x - \mu_y)\right)^2\right]$$

$$E(z) = \mu_x - \mu_y \qquad Var(z) = \frac{\sigma^2}{m} + \frac{\sigma^2}{n}$$

(b) $\frac{(m-1)s_x^2}{\sigma^2} \sim \chi^2_{m-1}$ and $\frac{(n-1)s_y^2}{\sigma^2} \sim \chi^2_{n-1}$

(p305 of *Statistics for Engineering and the Sciences*. 3rd ed. Mendenhall & Sincich)

From class notes: If $X_1 \sim \chi^2_m$ and $X_2 \sim \chi^2_m$ and they're independent, then $X_1 + X_2 \sim \chi^2_{m+n}$

$$u = \frac{(m-1)s_x^2}{\sigma^2} + \frac{(n-1)s_y^2}{\sigma^2} \sim \boxed{\chi^2_{m+n-2}}$$

$$f(u) = \frac{(u)^{\frac{m+n-2}{2}-1} e^{-\frac{u}{2}}}{2^{\frac{m+n-2}{2}}\Gamma\left(\frac{m+n-2}{2}\right)} \qquad E(u) = m+n-2 \qquad Var(u) = 2(m+n-2)$$

(c) From class notes: $z \sim N(0,1)$ and $X \sim \chi^2_n \Rightarrow t = \dfrac{z}{\sqrt{X/n}} \sim t_n$

$$z = \bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right) \Rightarrow z' = \frac{z - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{m} + \frac{\sigma^2}{n}}} \sim N(0,1)$$

$$\therefore \boxed{\frac{z'}{\sqrt{u/(m+n-2)}} \sim t_{m+n-2}}$$

4. Suppose that the random variable X has a Pareto density

$$f(x;\theta_1,\theta_2) = (\theta_2/\theta_1)(\theta_1/x)^{\theta_2+1}$$

for $\theta_1, \theta_2 > 0$ and $x \geq \theta_1$

(a) Derive the density of the random variable $Y = \ln(X)$.

(b) After extensive testing (to be done in the second half of the course) you have determined that $\theta_1 = 1$. Given this, find the $E(Y)$ and $Var(Y)$.

(c) Given your answer to (b), what is $E(Y)$ and $Var(Y)$ for the general case where θ_1 is not equal to 1?

(a) $Y = \ln(X) \Rightarrow X = e^Y = s(y)$

$$\frac{ds(y)}{dy} = e^y$$

$$g(y) = f(s(y))\left|\frac{ds(y)}{dy}\right| = \frac{\theta_2}{\theta_1}(\theta_1/e^y)^{\theta_2+1}e^y = \boxed{\theta_2\theta_1^{\theta_2}e^{-\theta_2 y},\ y \geq \ln(\theta_1)}$$

(b) $g(y) = \theta_2 e^{-\theta_2 y},\ y \geq 0$ ~ Exponential

$$\boxed{\begin{aligned} E(Y) &= \frac{1}{\theta_2} \\ Var(Y) &= \frac{1}{\theta_2^2} \end{aligned}}$$

(c) Since $\theta_1^{\theta_2}$ is a constant, $\boxed{E(Y) = \dfrac{\theta_1^{\theta_2}}{\theta_2} \text{ and } Var(Y) = \dfrac{\theta_1^{2\theta_2}}{\theta_2^2}}$

5. Suppose that $\mathbf{X}$ is random variables with multinomial density:

$$f(x_1, x_2, x_3; p_1, p_2) = p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{x_3}, \; x_i = 0,1, \; \sum x_i = 1$$

(a) Given a random sample of N observations, write down the likelihood and log-likelihood functions.

(b) What are the maximum likelihood estimators of p_1 and p_2?

(c) Show that the maximum likelihood estimators, $\hat{p}_1$ and $\hat{p}_2$, are unbiased estimators of p_1 and p_2.

(d) Derive the information matrix for p_1 and p_2.

(e) What are the Cramer-Rao lower bounds for $Var(p_1)$ and $Var(p_2)$?

(a) $L(\mathbf{x};\mathbf{p}) = \prod_{i=1}^{N} f(\mathbf{x}_i, \mathbf{p})$, where $\mathbf{x}_i = (x_{1i}, x_{2i}, x_{3i})$ and $\mathbf{p} = (p_1, p_2)$

$$L(\mathbf{x};\mathbf{p}) = \prod_{i=1}^{N} p_1^{x_{1i}} p_2^{x_{2i}} (1 - p_1 - p_2)^{x_{3i}} = p_1^{\sum_{i=1}^{N} x_{1i}} p_2^{\sum_{i=1}^{N} x_{2i}} (1 - p_1 - p_2)^{\sum_{i=1}^{N} x_{3i}}$$

$$\ln L(\mathbf{x};\mathbf{p}) = \left(\sum_{i=1}^{N} x_{1i}\right)\ln(p_1) + \left(\sum_{i=1}^{N} x_{2i}\right)\ln(p_2) + \left(\sum_{i=1}^{N} x_{3i}\right)\ln(1 - p_1 - p_2)$$

For simplicity let $\theta_1 = \left(\sum_{i=1}^{N} x_{1i}\right)$, $\theta_2 = \left(\sum_{i=1}^{N} x_{2i}\right)$, $\theta_3 = \left(\sum_{i=1}^{N} x_{3i}\right) = N - \theta_1 - \theta_2$

(b) To find MLE, take partials of $\ln L(\mathbf{x};\mathbf{p})$ wrt $\mathbf{p}$, set them equal to zero and solve for $\mathbf{p}$

$$\frac{\partial \ln L(\mathbf{x};\mathbf{p})}{\partial p_1} = \theta_1 \frac{1}{p_1} - \theta_3 \frac{1}{1 - p_1 - p_2} = 0 \qquad (1)$$

$$\frac{\partial \ln L(\mathbf{x};\mathbf{p})}{\partial p_2} = \theta_2 \frac{1}{p_2} - \theta_3 \frac{1}{1 - p_1 - p_2} = 0 \qquad (2)$$

Solve Eqn (1) for p_2 by first getting common denominators

$$\theta_1 \frac{1 - p_1 - p_2}{p_1(1 - p_1 - p_2)} - \theta_3 \frac{p_1}{p_1(1 - p_1 - p_2)} = 0$$

Get rid of the denominator by multiplying both sides by the denominator

$$\theta_1 - \theta_1 p_1 - \theta_1 p_2 - \theta_3 p_1 = 0$$

Combine terms for p_1

$$\theta_1 - (\theta_1 + \theta_3) p_1 - \theta_1 p_2 = 0$$

Solve for p_2 by taking it to the other side and dividing by θ_1

$$p_2 = \frac{\theta_1 - (\theta_1 + \theta_3) p_1}{\theta_1} \qquad (3)$$

Now work on Eqn (2) to solve for p_1; first get common denominators

$$\theta_2 \frac{1 - p_1 - p_2}{p_2(1 - p_1 - p_2)} - \theta_3 \frac{p_2}{p_2(1 - p_1 - p_2)} = 0$$

As before, get rid of the denominator by multiplying both sides by the denominator

$$\theta_2 - \theta_2 p_1 - \theta_2 p_2 - \theta_3 p_2 = 0$$

Combine terms for p_2

$$\theta_2 - \theta_2 p_1 - (\theta_2 + \theta_3) p_2 = 0$$

Now substitute Eqn (3) for p_2

$$\theta_2 - \theta_2 p_1 - (\theta_2 + \theta_3) \frac{\theta_1 - (\theta_1 + \theta_3) p_1}{\theta_1} = 0$$

Break up fraction so θ_1's cancel

$$\theta_2 - \theta_2 p_1 - (\theta_2 + \theta_3) + \frac{(\theta_2 + \theta_3)(\theta_1 + \theta_3) p_1}{\theta_1} = 0$$

Combine terms with p_1 on one side and everything else on the other. Get common denominators on the left term for p_1 and multiply out the term on the right.

$$\frac{-\theta_2 \theta_1 p_1}{\theta_1} + \frac{(\theta_2 \theta_1 + \theta_3^2 + \theta_1 \theta_3 + \theta_2 \theta_3) p_1}{\theta_1} = \theta_3$$

Multiply both sides by θ_1 and notice the $\theta_1 \theta_2$ terms cancel

$$(\theta_3^2 + \theta_1 \theta_3 + \theta_2 \theta_3) p_1 = \theta_1 \theta_3$$

Now solve for p_1. We can cancel the θ_3 and substitute $N = \theta_1 + \theta_2 + \theta_3$

$$\hat{p}_1 = \frac{\theta_1 \theta_3}{\theta_3(\theta_3 + \theta_1 + \theta_2)} = \frac{\theta_1}{N} \tag{4}$$

Now substitute Eqn (4) into Eqn (3) to finish solving for p_2

$$\hat{p}_2 = \frac{\theta_1 - (\theta_1 + \theta_3)\theta_1 / N}{\theta_1} = \frac{N\theta_1 - (\theta_1 + \theta_3)\theta_1}{\theta_1 N}$$

Substitute $\theta_1 + \theta_3 = N - \theta_2$

$$\hat{p}_2 = \frac{N - (N - \theta_2)}{N} = \frac{\theta_2}{N}$$

Wipe sweat off brow...

$$\boxed{\hat{p}_1 = \frac{\left(\sum_{i=1}^{N} x_{1i}\right)}{N} \quad \hat{p}_2 = \frac{\left(\sum_{i=1}^{N} x_{2i}\right)}{N}}$$

(c) $$E(\hat{p}_1) = E\left[\frac{\left(\sum_{i=1}^{N} x_{1i}\right)}{N}\right] = \frac{1}{N}\sum_{i=1}^{N} E(x_{1i}) = \frac{1}{N}\sum_{i=1}^{N} p_1 = \frac{1}{N} N p_1 = p_1 \therefore \text{ unbiased}$$

$$E(\hat{p}_2) = E\left[\frac{\left(\sum_{i=1}^{N} x_{2i}\right)}{N}\right] = \frac{1}{N}\sum_{i=1}^{N} E(x_{2i}) = \frac{1}{N}\sum_{i=1}^{N} p_2 = \frac{1}{N} N p_2 = p_2 \therefore \text{ unbiased}$$

(d) $\mathbf{I}(\mathbf{p}) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$, where $I_{i,j} = -E\left[\dfrac{\partial^2 \ln L(x;\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right]$

To make things quicker and easier to read:

$$E(\theta_1) = E\left(\sum_{i=1}^{N} x_{1i}\right) = \sum_{i=1}^{N} E(x_{1i}) = \sum_{i=1}^{N} p_1 = Np_1$$

$$E(\theta_2) = E\left(\sum_{i=1}^{N} x_{2i}\right) = \sum_{i=1}^{N} E(x_{2i}) = \sum_{i=1}^{N} p_2 = Np_2$$

$$E(\theta_3) = E\left(\sum_{i=1}^{N} x_{3i}\right) = \sum_{i=1}^{N} E(x_{3i}) = \sum_{i=1}^{N} (1-p_1-p_2) = N(1-p_1-p_2)$$

$$I_{11} = -E\left[\frac{\partial}{\partial p_1}\left(\theta_1 \frac{1}{p_1} - \theta_3 \frac{1}{1-p_1-p_2}\right)\right] = -E\left[-\theta_1 \frac{1}{p_1^2} - \theta_3 \frac{1}{(1-p_1-p_2)^2}\right] =$$

$$\frac{1}{p_1^2} E(\theta_1) + \frac{1}{(1-p_1-p_2)^2} E(\theta_3) = \frac{1}{p_1^2} Np_1 + \frac{1}{(1-p_1-p_2)^2} N(1-p_1-p_2) =$$

$$\frac{N}{p_1} + \frac{N}{(1-p_1-p_2)}$$

$$I_{12} = -E\left[\frac{\partial}{\partial p_2}\left(\theta_1 \frac{1}{p_1} - \theta_3 \frac{1}{1-p_1-p_2}\right)\right] = -E\left[\theta_3 \frac{1}{(1-p_1-p_2)^2}\right] =$$

$$-\frac{1}{(1-p_1-p_2)^2} E(\theta_3) = -\frac{1}{(1-p_1-p_2)^2} N(1-p_1-p_2) = -\frac{N}{(1-p_1-p_2)}$$

$$I_{22} = -E\left[\frac{\partial}{\partial p_2}\left(\theta_2 \frac{1}{p_2} - \theta_3 \frac{1}{1-p_1-p_2}\right)\right] = -E\left[-\theta_2 \frac{1}{p_2^2} - \theta_3 \frac{1}{(1-p_1-p_2)^2}\right] =$$

$$\frac{1}{p_2^2} E(\theta_2) + \frac{1}{(1-p_1-p_2)^2} E(\theta_3) = \frac{1}{p_2^2} Np_2 + \frac{1}{(1-p_1-p_2)^2} N(1-p_1-p_2) =$$

$$\frac{N}{p_2} + \frac{N}{(1-p_1-p_2)}$$

$$\boxed{\mathbf{I}(\mathbf{p}) = \begin{bmatrix} \dfrac{N}{p_1} + \dfrac{N}{(1-p_1-p_2)} & \dfrac{N}{(1-p_1-p_2)} \\ \dfrac{N}{(1-p_1-p_2)} & \dfrac{N}{p_2} + \dfrac{N}{(1-p_1-p_2)} \end{bmatrix}}$$

(e) CRLB is $[\mathbf{I}(\boldsymbol{\theta})]^{-1}$

Using $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \dfrac{1}{ad-cb}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

$$\frac{1}{ad-cb} = \frac{1}{\left(\frac{N}{p_1}+\frac{N}{(1-p_1-p_2)}\right)\left(\frac{N}{p_2}+\frac{N}{(1-p_1-p_2)}\right)-\left(\frac{N}{(1-p_1-p_2)}\right)^2} =$$

$$\frac{1}{\left(\frac{N^2}{p_1 p_2}+\frac{N^2}{p_1(1-p_1-p_2)}+\frac{N^2}{p_2(1-p_1-p_2)}+\frac{N^2}{(1-p_1-p_2)^2}\right)-\left(\frac{N}{(1-p_1-p_2)}\right)^2} =$$

$$\frac{1}{\left(\frac{N^2(1-p_1-p_2)}{p_1 p_2(1-p_1-p_2)}+\frac{N^2 p_2}{p_1 p_2(1-p_1-p_2)}+\frac{N^2 p_1}{p_1 p_2(1-p_1-p_2)}\right)} =$$

$$\frac{p_1 p_2(1-p_1-p_2)}{N^2-N^2 p_1-N^2 p_2+N^2 p_2+N^2 p_1} = \frac{p_1 p_2(1-p_1-p_2)}{N^2}$$

$$[\mathbf{I}(\boldsymbol{\theta})]^{-1} = \frac{p_1 p_2(1-p_1-p_2)}{N^2}\begin{bmatrix} \frac{N}{p_2}+\frac{N}{(1-p_1-p_2)} & -\frac{N}{(1-p_1-p_2)} \\ -\frac{N}{(1-p_1-p_2)} & \frac{N}{p_1}+\frac{N}{(1-p_1-p_2)} \end{bmatrix} =$$

$$\frac{p_1 p_2(1-p_1-p_2)}{N^2}\begin{bmatrix} \frac{N(1-p_1)}{p_2(1-p_1-p_2)} & \frac{N}{(1-p_1-p_2)} \\ \frac{N}{(1-p_1-p_2)} & \frac{N(1-p_2)}{p_1(1-p_1-p_2)} \end{bmatrix} = \begin{bmatrix} \frac{p_1(1-p_1)}{N} & -\frac{p_1 p_2}{N} \\ -\frac{p_1 p_2}{N} & \frac{p_2(1-p_2)}{N} \end{bmatrix} =$$

$$\boxed{\begin{array}{l} Var(p_1) \geq \frac{p_1(1-p_1)}{N} \\ Var(p_2) \geq \frac{p_2(1-p_2)}{N} \end{array}}$$