Questions Done

by Michael Seo

Jan 2005 - 1, 6, 7

Jan 2007 - 2, 4, 5

Jan 2008 - 1, 2 (same as June 2015 #6), 3, 4, 5

Aug 2008 - 1 (same as June 2014 #1), 2 (same as June 2015 #5), 3(a,b), 4

Jan 2009 - 1, 3, 4, 5

Jan 2010 - 1, 3(a,b,c), 4, 5

June 2010 - 1, 3, 4(a,b,c), 5

Jan 2011 - 1, 2, 3, 4

June 2011 - 1, 3, 4

Jan 2012 - 1, 2, 5

Jan 2013 - 1, 3(a,b), 4, 5

June 2013 - 3(a,b), 4, 5

June 2014 - 1, 3(a,b), 4, 5

June 2015 - 1, 3, 4, 5, 6

Q1 (Jan 2005)

Suppose Y is a $Binomial(n, \pi)$ random variable.

a) Write down the density function as a one-parameter exponential family and identify what is the canonical link

Solution:

$$f(y \mid \pi) = \pi^y (1-\pi)^{n-y} = exp(ylog\frac{\pi}{1-\pi} + nlog(1-\pi))$$

b) Assume now that N independent observations $y_1, ..., y_N$ are available, such that $Y_i \sim Bin(n_i, \pi_i)$, with corresponding p-dimensional covariate vectors $x_1, ..., x_N$ where $x_i = (x_{i1}, ..., x_{ip})$. If the canonical link is used, derive the likelihood equations to estimate the vector of regression coefficients, β.

Solution:

$$L(\beta) = \prod_{i=1}^{N} \binom{n_i}{y_i} \pi_i(\beta)^{y_i} (1-\pi_i(\beta))^{m_i - y_i}$$

$$logL(\beta) = const + \sum \{y_i log\pi_i(\beta) + (m_i - y_i) log(1-\pi_i(\beta))\}$$

We use canonical link and so,

$$log\frac{\pi_i}{1-\pi_i} = X_i^T \beta \Rightarrow \pi_i = \frac{1}{1+exp(-x_i^T\beta)} \text{ and } \pi_i' = \frac{exp(-x_i^T\beta)}{\{1+exp(-x_i^T\beta)\}^2} = \pi_i(1-\pi_i)$$

Then,

$$\begin{aligned}
\frac{\partial logL(\beta)}{\partial \beta} &= \sum (\frac{y_i \pi_i'(\beta)}{\pi_i(\beta)} + \frac{(m_i - y_i)(-\pi_i'(\beta))}{1-\pi_i(\beta)}) x_i \\
&= \sum \frac{y_i \pi_i'(\beta) - m_i \pi_i'(\beta)\pi_i(\beta)}{\pi_i(\beta)(1-\pi_i(\beta))} x_i = \sum \{y_i - m_i p_i(\beta)\} x_i
\end{aligned}$$

c) What is the sufficient statistic for estimating β? Why?

Solution:

Theorem 6.2.10 Let $X_1, ..., X_n$ be iid observations from a pdf or pmf $f(x \mid \theta)$ that belongs to an exponential family given by

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

where $\theta = (\theta_1, \theta_2, ..., \theta_d)$, $d \leq k$. Then

$$T(X) = (\sum_{j=1}^{n} t_1(X_j), ..., \sum_{j=1}^{n} t_k(X_j))$$

Thus, $\sum y_i$ is the sufficient statistic.

d) Find the likelihood equation for the one-parameter exponential family with canonical link in general.

Solution:

$$logL(\beta, \phi) = \frac{y_i^T x_i^T \beta - b(x_i^T\beta)}{a_i(\phi)} + c(y_i, \phi)$$

$$S(\beta, \phi) = \sum_{i=1}^{n} \frac{(y_i - b'(x_i^T\beta)) x_i}{a_i(\phi)}$$

$$\frac{\partial S(\beta, \phi)}{\partial \beta^T} = \sum_{i=1}^{n} \frac{-b''(x_i^T\beta)}{a_i(\phi)} x_i x_i^T$$

Q6 (Jan 2005)

Let (X_i, Y_i), $i = 1, ..., n$ be a random sample from a bivariate normal distribution, with mean (μ_X, μ_Y), variances (σ_X^2, σ_Y^2) and correlation ρ. The objective is to construct an approximate confidence interval for the ratio of the means $\theta = \mu_Y/\mu_X$.

a) First show that if X, Y are random variables with means μ_X, μ_Y, then

$$var(X/Y) \approx (\frac{\mu_X}{\mu_Y})^2(\frac{VarX}{\mu_X^2} + \frac{VarY}{\mu_Y^2} - 2\frac{Cov(X,Y)}{\mu_X\mu_Y})$$

Solution:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim BN(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix})$$

Then $g(\mu_X, \mu_Y) = \frac{\mu_X}{\mu_Y}$, then $g'(\mu_X, \mu_Y) = \begin{bmatrix} \frac{1}{\mu_Y} & -\frac{\mu_X}{\mu_Y^2} \end{bmatrix}$

$$\begin{aligned} Var(X/Y) &\approx \begin{bmatrix} \frac{1}{\mu_Y} & -\frac{\mu_X}{\mu_Y^2} \end{bmatrix} \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} \begin{bmatrix} \frac{1}{\mu_Y} \\ -\frac{\mu_X}{\mu_Y^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\sigma_X^2}{\mu_Y} - \frac{\rho\mu_X\sigma_X\sigma_Y}{\mu_Y^2} & \frac{\rho\sigma_X\sigma_Y}{\mu_Y} - \frac{\mu_X\sigma_Y^2}{\mu_Y^2} \end{bmatrix} \begin{bmatrix} \frac{1}{\mu_Y} \\ -\frac{\mu_X}{\mu_Y^2} \end{bmatrix} \\ &= \frac{\sigma_X^2}{\mu_Y^2} - 2\frac{\rho\mu_X\sigma_X\sigma_Y}{\mu_Y^3} + \frac{\mu_X^2\sigma_Y^2}{\mu_Y^4} \\ &= \frac{\mu_X^2}{\mu_Y^2}(\frac{VarX}{\mu_X^2} + \frac{VarY}{\mu_Y^2} - 2\frac{Cov(X,Y)}{\mu_X\mu_Y}) \end{aligned}$$

b) Now use this approximation and the Central Limit Theorem to derive a confidence interval for θ.

Solution:

$$\sqrt{n}(\frac{\bar{X}}{\bar{Y}} - \frac{\mu_X}{\mu_Y}) \sim N(0, \frac{\mu_X^2}{\mu_Y^2}(\frac{VarX}{\mu_X^2} + \frac{VarY}{\mu_Y^2} - 2\frac{Cov(X,Y)}{\mu_X\mu_Y}))$$

$\theta = \mu_Y/\mu_X$. Then $g(\mu) = \mu^{-1}$ and $g'(\mu) = \frac{-1}{\mu^2}$

$$\begin{aligned} \sqrt{n}(\frac{\bar{Y}}{\bar{X}} - \frac{\mu_Y}{\mu_X}) &\sim N(0, \frac{\mu_X^2}{\mu_Y^2}(\frac{VarX}{\mu_X^2} + \frac{VarY}{\mu_Y^2} - 2\frac{Cov(X,Y)}{\mu_X\mu_Y})(\frac{\mu_Y}{\mu_X})^4) \\ &= N(0, (\frac{\mu_Y}{\mu_X})^2(\frac{VarX}{\mu_X^2} + \frac{VarY}{\mu_Y^2} - 2\frac{Cov(X,Y)}{\mu_X\mu_Y})) \end{aligned}$$

Then 95% confidence interval is

$$\frac{\bar{Y}}{\bar{X}} \pm 1.96\sqrt{(\frac{\mu_Y}{\mu_X})^2(\frac{VarX}{\mu_X^2} + \frac{VarY}{\mu_Y^2} - 2\frac{Cov(X,Y)}{\mu_X\mu_Y})/n}$$

Q7 (Jan 2005)

Assume that $X_1, X_2, ..., X_n$ are independent random variables with a Poisson distribution with mean θ_1 up to an unknown "change point" $m\,(m \leq n)$ and θ_2 after m.

i) Derive the likelihood function for the unknown parameter $\phi = (\theta_1, \theta_2, m)$. Assume that the three components of ϕ are independently distributed a-priori. What is a conjugate prior for ϕ?

Solution:

$$\begin{aligned}
L(\theta_1, \theta_2, m) &= \prod_{i=1}^{m} \frac{\theta_1^{X_i}}{X_i!} e^{-\theta_1} \prod_{i=m+1}^{n} \frac{\theta_2^{X_i}}{X_i} e^{-\theta_2} \\
&\propto e^{-m\theta_1} \theta_1^{\sum_{i=1}^{m} x_i} e^{-(n-m)\theta_2} \theta_2^{\sum_{i=m+1}^{n} x_i}
\end{aligned}$$

For prior distributions, we assume $\theta_1 \sim Gamma(a, b)$, $\theta_2 \sim Gamma(c, d)$, and that m has a discrete uniform distribution on the values $1, ..., n$. Then the posterior is,

$$\begin{aligned}
P(\theta_1, \theta_2, m \mid x) &\propto L(\theta_1, \theta_2, m) P(\theta_1) P(\theta_2) P(m) \\
&\propto e^{-m\theta_1} \theta_1^{\sum_{i=1}^{m} x_i} e^{-(n-m)\theta_2} \theta_2^{\sum_{i=m+1}^{n} x_i} \theta_1^{a-1} e^{-b\theta_1} \theta_2^{c-1} e^{-d\theta_2} \frac{1}{n} \\
&\propto \theta_1^{\sum_{i=1}^{m} x_i + a - 1} e^{-(m+b)\theta_1} \theta_2^{\sum_{i=m+1}^{n} x_i + b - 1} e^{-(n-m+d)\theta_2}
\end{aligned}$$

ii) Describe a Gibbs sampler to simulate values from the posterior density of ϕ and give the forms of the relevant conditional densities.

Solution:

$$\theta_1 \mid \theta_2, m, x \sim Gamma(\sum_{i=1}^{m} x_i + a,\, m + b)$$

$$\theta_2 \mid \theta_1, m, x \sim Gamma(\sum_{i=m+1}^{n} x_i + b,\, n - m + d)$$

To get the full conditional for m, note the joint density of the data is

$$\begin{aligned}
p(x \mid \theta_1, \theta_2, m) &= \prod_{i=1}^{m} \frac{\theta_1^{X_i}}{X_i!} e^{-\theta_1} \prod_{i=m+1}^{n} \frac{\theta_2^{X_i}}{X_i} e^{-\theta_2} \\
&= [\prod_{i=1}^{n} \frac{1}{X_i!}] e^{m(\theta_2 - \theta_1)} e^{-n\theta_2} \theta_1^{\sum_{i=1}^{m} X_i} [\prod_{i=m+1}^{n} \theta_2^{X_i}] [\frac{\prod_{i=1}^{m} \theta_2^{X_i}}{\prod_{i=1}^{m} \theta_2^{X_i}}] \\
&= [\prod_{i=1}^{n} \frac{\theta_2^{X_i}}{X_i!} e^{-n\theta_2}] \times (e^{m(\theta_2 - \theta_1)} (\frac{\theta_1}{\theta_2})^{\sum_{i=1}^{m} X_i}) \\
&= f(x, \theta_2) g(x \mid m)
\end{aligned}$$

By Bayes' Law, for any particular value m^* of m,

$$p(m^* \mid x) = \frac{f(x, \theta_2) g(x \mid m^*) p(m^*)}{\sum_{m=1}^{n} f(x, \theta_2) g(x \mid m^*) p(m^*)}$$

Since $p(m) = 1/n$ (constant), we have

$$p(m^* \mid x) = p(m^* \mid x, \theta_1, \theta_2) \propto \frac{g(x \mid m^*)}{\sum_{m=1}^{n} g(x \mid m)}$$

This ratio defines a probability vector for m that we use at each iteration to sample a value of m from $\{1, 2, ..., n\}$.

Q2 (Jan 2007)

Let $X_1, ..., X_n$ $(n > 2)$ be a random sample from the two-parameter exponential distribution with density

$$f(x \mid \mu, \sigma) = \sigma^{-1} exp\{-(x-\mu)/\sigma\} I_{(\mu,\infty)}(x),$$

where $\mu > 0$, $\sigma > 0$ and $I_A(x) = 1$ if $x \in A$ and 0 otherwise. Let $\theta = (\mu, \sigma)$ and let $X_{(i)}$ denote the ith order statistic.

a) Show that $\{X_{(1)}, \sum_{i=1}^n (X_i - X_{(1)})\}$ is a sufficient statistic for θ.

Solution:

$$\begin{aligned} L(\mu, \sigma \mid X_1, ...X_n) =& \sigma^{-n} exp(-\sum (x_i - \mu)/\sigma) I_{(\mu,\infty)}(x_{(1)}) \\ =& \sigma^{-n} exp(-\frac{\sum x_i}{\sigma} + \frac{n\mu}{\sigma}) I_{(\mu,\infty)}(x_{(1)}) \end{aligned}$$

Theorem 6.2.6 (Factorization Theorem) Let $f(x \mid \theta)$ denote the joint pdf or pmf of a sample X. A statistic $T(X)$ is a sufficient statistic for θ if and only if there exist functions $g(t \mid \theta)$ and $h(x)$ such that, for all sample points x and all parameter points θ,

$$f(x \mid \theta) = g(T(x) \mid \theta) h(x)$$

Using factorization theorem, $g(t_1, t_2 \mid \mu, \sigma) = \sigma^{-n} exp(-\frac{t_1}{\sigma} + \frac{n\mu}{\sigma}) I_{(\mu,\infty)}(t_2)$ where $t_1(x) = \sum x_i$ and $t_2(x) = x_{(1)}$. $(x_{(1)}, \sum x_i - x_{(1)})$ is a one-to-one transformation of t_1 and t_2, and thus is also sufficient statistic.

b) Define $Z_i = X_{(i)} - X_{(i-1)}$ for $i = 1, ..., n$, where $X_0 = 0$. Show that $Z_1, ..., Z_n$ are independent and that $2(n-i+1)Z_i/\sigma$ has the χ^2 distribution with two degrees of freedom. Note: the density function of a χ^2 distribution with ν degrees of freedom is

$$f(x \mid \nu) = \frac{1}{\Gamma(\nu/2) 2^{\nu/2}} x^{\nu/2-1} e^{-x/2}$$

c) Show that $X_{(1)}$ and $\sum_{i=1}^n (X_i - X_{(1)})$ are independent.

Solution:

$$\sum_{i=1}^n (X_i - X_{(1)}) = \sum_{i=1}^n (X_{(i)} - X_{(1)}) = \sum_{i=2}^n (X_{(i)} - X_{(1)}) = \sum_{i=2}^n \sum_{k=2}^i z_{(k)}$$

Note $X_{(1)} - X_{(1)} = 0$, $X_{(2)} - X_{(1)} = Z_{(2)}$, $X_{(3)} - X_{(1)} = X_{(3)} - X_{(2)} + X_{(2)} - X_{(1)} = Z_{(3)} + Z_{(2)}$

Since $X_{(1)} = Z_{(1)}$ and $\sum (X_i - X_{(1)})$ doesn't have $Z_{(1)}$ and above question we proved Z are independent, so $X_{(1)}$ and $\sum (X_i - X_{(1)})$ are independent

f) Find a maximum likelihood estimator (MLE) for σ and find its asymptotic distribution.

Solution:

$$\begin{aligned} L &= \sigma^{-n} exp(-\frac{\sum (x_i - \mu)}{\sigma}) I_{(\mu,\infty)}(x_{(1)}) \\ logL &= -nlog\sigma - \frac{\sum x_i - n\mu}{\sigma} \\ \frac{d}{d\mu} logL &= \frac{n}{\sigma} \end{aligned}$$

At a fixed σ, likelihood is an increasing function of μ. $\hat{\mu} = x_{(1)}$

$$\frac{d}{d\sigma} logL = -\frac{n}{\sigma} + \frac{\sum x_i - n\mu}{\sigma^2} = 0$$

and $\hat{\sigma} = \frac{\sum x_i - n x_{(1)}}{n} = \bar{x} - x_{(1)}$

Q4 (Jan 2007)

Assume that, for $i = 1, ..., n$, $X_i \sim Bernoulli(\theta)$ are iid random variables, with $0 < \theta < 1$.
a) Consider a $Beta(\alpha, \beta)$ prior on θ and compute the Bayes estimator of $\lambda = \theta(1-\theta)$ under squared error loss

Solution:

$P(\theta \mid X_1, ...X_n) \sim beta(\alpha + \sum x_i,\ n + \beta - \sum x_i)$

$$\begin{aligned} E(a - \theta(1-\theta))^2 &= \int (a - \theta(1-\theta))^2 p(\theta \mid X_1, ..., X_n) d\theta \\ \frac{dE(L(\theta, a))}{da} &= 2a - 2\int \theta(1-\theta) p(\theta \mid X_1, ..., X_n) d\theta \\ \delta^{bayes} &= \frac{(\alpha + \sum x_i)(n + \beta - \sum x_i)}{(\alpha + \beta + n)(\alpha + \beta + n + 1)} \end{aligned}$$

b) Discuss the bias and consistency properties of the Bayes estimator of λ and compare the Bayes estimator to the UMVUE of λ.

Solution:

$\hat{\theta} = \bar{x}$. Then, $\hat{\lambda} = \hat{\theta}(1 - \hat{\theta}) = \bar{x}(1 - \bar{x})$

$$E(\bar{x}(1-\bar{x})) = E(\bar{x}) - E(\bar{x}^2) = E(\bar{x}) - (Var(\bar{x}) + E(\bar{x})^2) = \theta - \frac{\theta(1-\theta)}{n} - \theta^2 = \frac{(n-1)\theta(1-\theta)}{n}$$

Thus, UMVUE is $\phi(\sum x_i) = \bar{x}(1 - \bar{x})\frac{n}{n-1}$

Q5 (Jan 2007)

Let $X_1, X_2, ..., X_n$ be the survival times of n patients with advanced lung cancer. Suppose X_i's are independent, having a common exponential distribution with mean $\lambda^{-1} > 0$. The object of the study is to estimate λ^{-1}, and an unbiased estimator is desired. Consider the following two estimators:

$$T_1(X) = \bar{X}$$

$$T_2(X) = n\{min(X_1, X_2, ..., X_n)\}$$

a) Verify that both estimators are unbiased.

Solution:

$X_1, ..., X_n \sim exp(\lambda) = \lambda exp(-\lambda x_i)$ with mean of $\frac{1}{\lambda}$

$$E(\bar{X}) = \frac{1}{n}E(\sum x_i) = \frac{1}{\lambda}$$

Let $Y = min(X_1, ..., X_n)$. Then,

$$P(Y < y) = 1 - P(min(X_1, .., X_n) \geq y) = 1 - P(X_1 \geq y)...P(X_n \geq y) = 1 - exp(-\lambda ny)$$

$f(y) = \lambda n \times exp(-\lambda ny) \sim exp(\lambda n)$ with mean of $\frac{1}{\lambda n}$

$$E(ny) = nE(y) = n(\frac{1}{\lambda n}) = \frac{1}{\lambda}$$

b) Compute the variances of the estimates

Solution:

$$Var(\bar{X}) = \frac{1}{n^2}Var(\sum x_i) = \frac{1}{n^2}\frac{n}{\lambda^2} = \frac{1}{n\lambda^2}$$

$$Var(ny) = n^2 Var(y) = n^2\frac{1}{\lambda^2 n^2} = \frac{1}{\lambda^2}$$

c) Explain which estimator should be chosen, if the criterion is to minimize MSE

Solution:

$MSE = Var + Bias^2$. We should choose sample mean estimator.

e) Suppose it turns out that exact survival times had been observed for only r $(0 < r < n)$ of the n patients, while the observed survival times for the remaining $n - r$ patients are right censored. Assuming that the censoring is non-informative, derive the MLE of λ^{-1}. Show that the asymptotic variance of the MLE lies between the variances computed in part (b).

Solution:

$$L(\lambda \mid x) = \prod_{i=1}^{r} \lambda exp(-\lambda x_i) \prod_{j=1}^{n-r} P(X > x_j) = \lambda^r exp(-\lambda(\sum_{i=1}^{n} x_i)$$

$$\begin{aligned} log(L(\lambda \mid x)) =& rlog(\lambda) - \lambda(\sum_{i=1}^{n} x_i) \\ \frac{dlogL(\lambda \mid x)}{d\lambda} =& \frac{r}{\lambda} - (\sum_{i=1}^{n} x_i) = 0 \Rightarrow \hat{\lambda}^{-1} = \frac{\sum_{i=1}^{n} x_i}{r} \end{aligned}$$

Then,

$$I(\lambda) = E(-\frac{d^2L(\lambda \mid x)}{d^2\lambda}) = E(\frac{r}{\lambda^2}) = \frac{r}{\lambda^2}$$

$$I(\lambda) = \frac{dg}{d\lambda}I(1/\lambda)\frac{dg}{d\lambda} = \frac{-1}{\lambda^2}I(1/\lambda)\frac{-1}{\lambda^2} \Rightarrow I(1/\lambda) = r\lambda^2$$

$Var(1/\hat{\lambda}) = \frac{1}{r\lambda^2}$ and

$$\frac{1}{n\lambda^2} < \frac{1}{r\lambda^2} < \frac{1}{\lambda^2}$$

Q1 (Jan 2008)

Let $Y_1, ..., Y_n$ be independent Bernoulli variables, and let $x_1, ..., x_n$ be associated explanatory variables. For $i = 1, ..., n$, define $p_i = P(Y_i = 1 \mid x_i)$, where

$$log(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1 x_i$$

and (β_0, β_1) are unknown parameters.
a) Find a sufficient statistic for (β_0, β_1).
Solution:

$$\begin{aligned} f(y_1, ..., y_n \mid x_1, ...x_n) &= \prod_{i=1}^n f(y_i \mid x_1, ..., x_n) \\ &= \prod_{i=1}^n p_i^{y_i}(1-p_i)^{1-y_i} \\ &= \prod_{i=1}^n exp\{y_i log\frac{p_i}{1-p_i} + log(1-p_i)\} \\ &= \prod_{i=1}^n exp\{y_i(\beta_0 + \beta_1 x_i) + log(\frac{1}{1+exp(\beta_0+\beta_1 x_i)})\} \\ &= exp\{\beta_0 \sum y_i + \beta_1 \sum y_i x_i + \sum log\frac{1}{1+exp(\beta_0+\beta_1 x_i)}\} \end{aligned}$$

Sufficient statistic for (β_0, β_1) is $(\sum y_i x_i, \sum y_i)$.
b) Show whether the sufficient statistic in part (a) is minimal or not.
Solution:
Theorem 6.2.13 Let $f(x \mid \theta)$ be the pmf or pdf of a sample X. Suppose there exists a function $T(x)$ such that, for every two sample points x and y, the ratio $f(x \mid \theta)/f(y \mid \theta)$ is constant as a function of θ if and only if $T(x) = T(y)$. Then $T(X)$ is a minimal sufficient statistic for θ.

$$\frac{f(y_1, ..., y_n \mid x_1, .., x_n)}{f(y_1^*, ...y_n^* \mid x_1, ...x_n)} = exp(\frac{\sum y_i\beta_0 + \sum y_i x_i \beta_1}{\sum y_i^*\beta_0 + \sum y_i^* x_i \beta_1})$$

is constant iff

$$\sum y_i = \sum y_i^* \text{ and } \sum y_i x_i = \sum y_i^* x_i$$

c) Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the MLE of β_0 and β_1. Derive the likelihood equations for estimating $\hat{\beta}_0$ and $\hat{\beta}_1$.
Solution:

$$logL(\beta_0, \beta_1) = (\sum y_i)\beta_0 + (\sum y_i x_i)\beta_1 - \sum log(1 + exp(\beta_0 + \beta_1 x_i))$$

$$\frac{\partial logL(\beta_0, \beta_1)}{\partial \beta_0} = \sum y_i - \sum \frac{exp(\beta_0+\beta_1 x_i)}{1+exp(\beta_0+\beta_1 x_i)} = \sum (y_i - p_i) = 0$$

$$\frac{\partial logL(\beta_0, \beta_1)}{\partial \beta_1} = \sum y_i x_i - \sum \frac{exp(\beta_0+\beta_1 x_i)}{1+exp(\beta_0+\beta_1 x_i)} x_i = \sum (y_i - p_i) x_i = 0$$

d) Derive the estimated variance-covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1)$.
Solution:

$$I(\beta_0, \beta_1) = -E(\begin{bmatrix} \frac{\partial^2 l}{\partial \beta_0^2} & \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 l}{\partial \beta_1^2} \end{bmatrix}) = E(\begin{bmatrix} \sum p_i(1-p_i) & \sum x_i p_i(1-p_i) \\ \sum x_i p_i (1-p_i) & \sum x_i^2 p_i(1-p_i) \end{bmatrix})$$

Invert the information matrix to find variance covariance matrix.

Q3 (Jan 2008)

When drawing inference about a population mean, it is important to take into account correlation between observations from the sample. This question concerns drawing inference about a sample mean when the sample is drawn over time and the observations are serially correlated.

Suppose that $X_1, ..., X_n$ $(n > 1)$ are correlated and follow a first order moving average model, defined as follows. Let $Z_0, ..., Z_n$ be iid normal with $E(Z_i) = 0$ and known variance $var(Z_i) = \sigma^2$. Suppose that

$$X_i = \mu + Z_i + \theta Z_{i-1}.$$

a) Show that $E(X_i) = \mu$ for any $i = 1, 2, ..., n$.
Solution:

$$E(X_i) = E(\mu + Z_i + \theta Z_{i-1}) = \mu$$

b) Let h be an integer-valued time lag. The covariance between X_{i+h} and X_i is denoted by

$$\gamma(h) = cov(X_{i+h}, X_i).$$

Derive the covariance function.

$$\begin{aligned}\gamma(0) = cov(X_i, X_i) &= cov(\mu + Z_i + \theta Z_{i-1}, \mu + Z_i + \theta Z_{i-1}) \\ &= cov(Z_i, Z_i) + 2\theta cov(Z_i, Z_{i-1}) + \theta^2 cov(Z_{i-1}, Z_{i-1}) \\ &= \sigma^2 + \theta^2\sigma^2 = (1 + \theta^2)\sigma^2\end{aligned}$$

$$\begin{aligned}\gamma(1) = cov(X_i, X_{i+1}) &= cov(\mu + Z_i + \theta Z_{i-1}, \mu + Z_{i+1} + \theta Z_i) \\ &= cov(Z_i, \theta Z_i) = \theta\sigma^2\end{aligned}$$

$$\gamma(2) = cov(X_i, X_{i+2}) = cov(\mu + Z_i + \theta Z_{i-1}, \mu + Z_{i+2} + \theta Z_{i+1}) = 0$$

for any $| h | \geq 2$, covariance is zero.
c) Derive the distribution of $\bar{X}_n$. Make sure that you explicitly specify $E(\bar{X}_n)$ and $var(\bar{X}_n)$

$$Var(\bar{X}_n) = var(\frac{1}{n}\sum X_i) = \frac{1}{n^2}\{\sum_{i=1}^{n} var(X_i) + \sum_{i \neq j} cov(X_i, X_j)\} = \frac{1}{n^2}\{n(1 + \theta^2)\sigma^2 + 2(n-1)\theta\sigma^2\}$$

$$\sqrt{n}(\bar{X}_n - \mu) \to N(0, Var(\bar{X}_n))$$

d) Derive a 95% confidence interval for μ.

$$(\bar{X}_n - 1.96\frac{sd(\bar{X}_n)}{\sqrt{n}}, \bar{X}_n + 1.96\frac{sd(\bar{X}_n)}{\sqrt{n}})$$

e) In conclusion, by ignoring correlation in the data (when in fact it exists) results in error on inference about μ. Let us illustrate this case. Suppose that $\theta = 1$. Note that the 95% confidence interval

$$(\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}})$$

is not correct if correlation exists because it does not have 95% coverage probability. Show that

$$P(| \bar{X}_n - \mu | < 1.96\frac{\sigma}{\sqrt{n}}) < 0.95$$

Solution:

$$Var(\bar{X}) = \frac{1}{n^2}\{n(1 + \theta^2)\sigma^2 + 2(n-1)\theta\sigma^2\} = \frac{1}{n^2}\{2n\sigma^2 + 2(n-1)\sigma^2\} = \frac{4n\sigma^2 - 2\sigma^2}{n^2}$$

$$\frac{4n\sigma^2 - 2\sigma^2}{n^2} = \frac{(4n-2)}{n}\frac{\sigma^2}{n} > \frac{\sigma^2}{n}$$

since $\frac{4n-2}{n} > 1$ and

$$P(| \bar{X}_n - \mu | < 1.96\frac{\sigma}{\sqrt{n}}) < 0.95$$

Q4 (Jan 2008)

Consider the linear model $Y = X\beta + \epsilon$, where Y is a $N \times 1$ random vector, X is a $N \times P$ matrix, β is a $P \times 1$ parameter vector and ϵ is a random vector that satisfies

$$E(\epsilon) = 0 \,\text{and}\, cov(\epsilon) = \sigma^2 I,$$

When the columns of X are linearly independent, the least squares estimate of the parameter vector β is

$$\hat{\beta}_{LSE} = (X^T X)^{-1} X^T Y.$$

Sometimes, especially when P (the dimension of β) is large, the columns of X may exhibit strong multi-collinearity. Consequently, it becomes difficult (or even impossible) to invert the matrix $X^T X$. Strong multi-collinearity suggests removing some covariates, which is equivalent to setting to 0 some components of the parameter estimates of the vector β.

One approach to this problem is ridge regression, which uses a penalized least squares criterion that penalizes model complexity (e.g. redundant covariates). The optimizing criterion in ridge regression is

$$\begin{aligned} S_\lambda(b) &= \| Y - Xb \|^2 + \lambda \| b \|^2 \\ &= \sum_{i=1}^{N} (Y_i - X_i b)^2 + \lambda \sum_{p=1}^{P} b_p^2 \end{aligned}$$

where X_i is the ith row of X, Y_i is the ith element of Y, $b = (b_1, ..., b_p)^T$, and λ is a non-negative tuning parameter. For a fixed λ, we shall denote the ridge regression estimator of β to be

$$\tilde{\beta}_\lambda = \arg \min_{b \in R^P} S(b)$$

that is, the value of b that minimizes the penalized least squares criterion. Throughout, squared norm of a vector a is the sum of its squared components; that is, $\| a \|^2 = \sum_i a_i^2$.

a) What is the limiting value of the ridge regression estimator as $\lambda \to 0$? As $\lambda \to \infty$?

Solution:

$\lambda \to 0$ this is least square regression.

$\lambda \to \infty$ all β are 0.

Note: in ridge regression, we do not put penalization on intercept. However, the intercept can be dropped if we do centering before fitting the model.

b) In this question and the next, we shall demonstrate how to obtain the minimizer of $S(b)$. First, show that the penalty part can be written as

$$\lambda \| b \|^2 = \| 0_P - \sqrt{\lambda} I_P b \|^2$$

where 0_p is a P-vector of zeros and I_P is the $P \times P$ identity matrix.

Solution:

$$\| 0_P - \sqrt{\lambda} I_P b \|^2 = \| \sqrt{\lambda} I_P b \|^2 = \lambda \| I_P b \|^2 = \lambda \| b \|^2$$

c) Next, define $Y_\lambda = \begin{bmatrix} Y \\ 0_P \end{bmatrix}$ and $X_\lambda = \begin{bmatrix} X \\ \sqrt{\lambda} I_P \end{bmatrix}$. Show that the ridge regression estimator for β, for a fixed λ, is

$$\tilde{\beta}_\lambda = (X^T X + \lambda I_P)^{-1} X^T Y.$$

Solution:

$$S_\lambda(b) = (Y - Xb)^T (Y - Xb) + \lambda b^T b = Y^T Y - 2Y^T X b + b^T X^T X b + \lambda b^T b$$

NoteL $\frac{\partial a^T \beta}{\partial \beta} = a$, $\frac{\partial \beta^T A \beta}{\partial \beta} = (A + A^T)\beta$

$$\frac{\partial S_\lambda(b)}{\partial b} = -2X^T Y + 2X^T X b + 2\lambda b = 0$$

then,

$$(X^T X + \lambda I)b = X^T Y$$
$$\hat{b} = (X^T X + \lambda I_P)^{-1} X^T Y$$

And,

$$\frac{\partial^2 S_\lambda(b)}{\partial b^2} = 2X^T X + 2\lambda I_P > 0$$

Hence $\hat{b}$ gets the minimum for $S_\lambda(b)$.

d) When $(X^T X)^{-1}$ exists, show that $E(\hat{\beta}_{LSE}) = \beta$.

Solution:

$$E(\hat{\beta}_{LSE}) = E[(X^T X)^{-1} X^T Y] = (X^T X)^{-1} X^T X \beta = \beta$$

e) Even though $\hat{\beta}_{LSE}$ is an unbiased estimator for β, its squared norm overestimates the squared norm of β. That is, $E\{\| \hat{\beta}_{LSE} \|^2\} > \| \beta \|^2$. To demonstrate this, show in particular that

$$E\{\| \hat{\beta}_{LSE} \|^2\} = \| \beta \|^2 + \sigma^2 trace\{(X^T X)^{-1}\}.$$

You may use the following fact: for an $N \times 1$ random vector Y having mean μ and covariance matrix Σ, and for any compatible matrix A,

$$E(Y^T A Y) = \mu^T A \mu + trace(A\Sigma)$$

Solution:

$$\begin{aligned}
E\{\| \hat{\beta}_{LSE} \|^2\} &= E[Y^T X (X^T X)^{-1} (X^T X)^{-1} X^T Y] \\
&= \beta^T X^T X (X^T X)^{-1} (X^T X)^{-1} X^T X \beta + trace(X (X^T X)^{-1} (X^T X)^{-1} X^T \sigma^2 I) \\
&= \beta^T \beta + \sigma^2 trace(X^T X (X^T X)^{-1} (X^T X)^{-1}) \\
&= \| \beta \|^2 + \sigma^2 trace((X^T X)^{-1})
\end{aligned}$$

Q5 (Jan 2008)

Let $X_1, X_2, ..., X_n$ be a random sample from a population distribution function F. Assume there are no ties in the data. Define the empirical distribution function $\hat{F}_n$ as

$$\begin{aligned}\hat{F}_n(x) &= (1/n)(\text{number of } X_i \le x)\\ &= (1/n)\sum_{i=1}^{n} I(X_i \le x)\end{aligned}$$

a) Show that for any x, $\hat{F}_n(x)$ is an unbiased and consistent estimator of $F(x)$.

Solution:

For fixed x, $\sum_{i=1}^{n} I(X_i \le x)$ is the sum of n independent Bernoulli random variables with success probability $F(x)$.

Therefore, $n\hat{F}_n(x) \sim Binomial(n, F(x))$

$$E(\hat{F}_n(x)) = \frac{1}{n}E(n\hat{F}_n(x)) = \frac{1}{n}nF(x) = F(x)$$

Thus, $\hat{F}_n$ is unbiased.

$$Var(\hat{F}_n(x)) = \frac{1}{n^2}var(n\hat{F}_n(x)) = \frac{1}{n^2}nF(x)(1-F(x)) = \frac{F(x)(1-F(x))}{n}$$

For any $\epsilon > 0$,

$$P(\mid \hat{F}_n(x) - F(x) \mid \ge \epsilon) \le \frac{var(F_n(x))}{\epsilon^2} = \frac{F(x)(1-F(x))}{n\epsilon^2}$$

As $n \to \infty$, $P(\mid \hat{F}_n(x) - F(x) \mid \ge \epsilon) \to 0$ and is therefore consistent.

b) How would you generalize the above procedure if you wish to estimate $p = P(X \in J)$ for some fixed interval or set J

Solution:

$$\hat{P}(X \in J) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in J)$$

c) Show that $\hat{F}_n$ can be written as a function of a suitable sufficient statistic.

Solution:

Order statistic is a sufficient statistic. For any x, $\hat{F}_n(x)$ can be written as $\frac{1}{n}\sum_{i=1}^{n} I(X_{(i)} \le x)$

d) Suppose F is the family of distribution functions putting probability masses $p_1, p_2, ..., p_n$ on $X_1, X_2, ..., X_n$ and no where else. Show that the corresponding likelihood function defined as $L = \prod_{i=1}^{n} p_i$ is maximized if $F = \hat{F}_n$.

Solution:

Want to maximize $L = \prod p_i$ with $\sum p_i = 1$.

Using Lagrange multiplier, $H = \prod_{i=1}^{n} p_i + \lambda(\sum_{i=1}^{n} p_i - 1)$

$$\frac{\partial H}{\partial \lambda} = \sum p_i - 1 = 0$$

$$\frac{\partial H}{\partial p_j} = p_j^{-1}\prod_{i=1}^{n} p_i + \lambda = 0$$

This implies $p_i = \frac{1}{n}$, $i = 1, ..., n$ and $\lambda = -(\frac{1}{n})^{n-1}$.

Therefore, $L = \prod p_i$ is maximized if $F = \hat{F}_n = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)$.

Q3 (Aug 2008)

Let (X,Y) be a random vector with joint probability density function $f_{X,Y}(x,y)$. Denote the marginal densities of X and Y to be, respectively, $f_X(x)$ and $f_Y(y)$. Mutual information between X and Y, if it exists, is defined to be

$$MI_{X,Y} = E_{X,Y} log \frac{f_{X,Y}(X,Y)}{f_X(X)f_Y(Y)} = \int\int f_{X,Y}(x,y) log \frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)} dxdy$$

a) On the connection between mutual information and correlation. Suppose that (X,Y) is bivariate normal with mean vector μ and variance-covariance structure with variances denoted σ_{XX} and σ_{YY} and correlation $\rho_{X,Y}$. Derive an expression for mutual information for (X,Y) and show that (for fixed σ_{XX}, σ_{YY} and μ) it is an increasing function of $\mid \rho_{X,Y} \mid$.

Solution:

$$MI_{X,Y} = \int\int f_{X,Y}(x,y) log f(x,y) dxdy - \int\int f_{X,Y}(x,y) log f(x) dxdy - \int\int f_{X,Y}(x,y) log f(y) dxdy$$

And,

$$\begin{aligned}
\int\int f(x,y) log f(x) dxdy &= \int log f(x) f(x) dx \\
&= \int (2\pi\sigma_X^2)^{-\frac{1}{2}} exp(-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}) log((2\pi\sigma_X^2)^{-\frac{1}{2}} exp(-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}))dx \\
&= -\frac{1}{2} log(2\pi\sigma_X^2) + (-\frac{1}{2\sigma_X^2}) \int (x-\mu_X)^2 (2\pi\sigma_X^2)^{-\frac{1}{2}} exp(-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2})dx \\
&= -\frac{1}{2} log(2\pi\sigma_X^2) - \frac{1}{2} = -\frac{1}{2} log(2\pi\sigma_X^2 e)
\end{aligned}$$

For the bivariate term, let $X = (x,y)$, then $f(x) = (2\pi)^{-\frac{n}{2}} \mid \Sigma \mid^{-\frac{1}{2}} exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$

$$\begin{aligned}
\int f(x) log f(x) dx &= E(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) - \frac{1}{2} log((2\pi)^n \mid \Sigma \mid) \\
&= -\frac{1}{2} E\{(x-\mu)^T\Sigma^{-1}(x-\mu)\} - \frac{1}{2} log((2\pi)^n \mid \Sigma \mid) \\
&= -\frac{1}{2} tr(E((x-\mu)^T\Sigma^{-1}(x-\mu)) - \frac{1}{2} ln((2\pi)^n \mid \Sigma \mid) \\
&= -\frac{1}{2} tr(\Sigma\Sigma^{-1}) - \frac{1}{2} ln((2\pi)^n \mid \Sigma \mid) \\
&= -\frac{1}{2} ln((2\pi e)^n \mid \Sigma \mid) = -\frac{1}{2} ln((2\pi e)^n (\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2))
\end{aligned}$$

Then,

$$\begin{aligned}
MI_{XY} &= \frac{1}{2} ln(\frac{(2\pi\sigma_X^2 e)(2\pi\sigma_Y^2 e)}{(2\pi e)^2(\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2)}) \\
&= \frac{1}{2} ln(\frac{\sigma_X^2\sigma_Y^2}{\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2}) = -\frac{1}{2} ln(\frac{\sigma_Y^2 + \sigma_{XY}^2/\sigma_X^2}{\sigma_Y^2}) \\
&= -\frac{1}{2} ln(1-\rho^2)
\end{aligned}$$

b) Show that $MI_{X,Y} \geq 0$ with equality holding if and only if X and Y are independent. (Hint: use Jensen's ineqaulity).

Theorem 4.7.7 (Jensen's Inequality) For any random variable X, if $g(x)$ is a convext function, then

$$Eg(X) \geq g(EX).$$

One immediate application of Jensen's Inequality shows that $EX^2 \geq (EX)^2$, since $g(x) = x^2$ is convex.
Solution:

$$\begin{aligned}
-\int\int f(x,y)log\frac{f(x,y)}{f(x)f(y)}dxdy &= \int\int f(x,y)log\frac{f(x)f(y)}{f(x,y)}dxdy \\
&\leq log\int\int f(x,y)\frac{f(x)f(y)}{f(x,y)}dxdy \\
&= log\int\int f(x)f(y)dxdy \\
&= log1 = 0
\end{aligned}$$

Then,

$$\int\int f(x,y)log\frac{f(x,y)}{f(x)f(y)}dxdy \geq 0$$

equality iff $f(x,y) = f(x)f(y)$ iff X and Y are independent.

Q4 (Aug 2008)

Consider the time series model $X_{t+1} = \phi X_t + \epsilon_{t+1}$ where ϵ_t is iid $N(0,1)$ and $\phi \in (-1,1)$. It is well known that time series X_t can be represented by

$$X_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$$

a) First, show that $EX_t < \infty$ and that $EX_t = 0$ for any t.
Solution:

$$E(X_t) = E(\phi X_{t-1} + \epsilon_t) = \phi E(X_{t-1}) + E(\epsilon_t) = \phi E(X_{t-1}) + 0$$

With stationary assumption $E(X_t) = E(X_{t-1})$. Let μ denote this common mean.

$$\mu = \frac{0}{1-\phi} = 0$$

b) Derive $Var(X_t)$
Solution:

$$\begin{aligned} Var(X_t) &= Var(\phi X_{t-1}) + Var(\epsilon_t) \\ &= \phi^2 Var(X_{t-1}) + 1 \end{aligned}$$

By the stationary assumption, $Var(x_t) = Var(x_{t-1})$. Substitute $Var(x_t)$ for $Var(x_{t-1})$ and then solve for $Var(x_t)$. Because $Var(x_t) > 0$, it follows that $(1-\phi^2) > 0$ and therefore $\mid \phi \mid < 1$.
c) Show that for any lag h,

$$corr(X_{t+h}, X_t) = \phi^{|h|}$$

Solution:
Covariance and correlation between observations one time period apart

$$\gamma_1 = E(x_t x_{t+1}) = E(x_t(\phi x_t + \epsilon_{t+1})) = E(\phi x_t^2 + x_t \epsilon_{t+1}) = \phi Var(x_t)$$

$$\rho_1 = \frac{Cov(x_t, x_{t+1})}{Var(x_t)} = \frac{\phi Var(x_t)}{Var(x_t)} = \phi$$

Covariance and correlation between observations h time periods apart

$$x_t = \phi x_{t-1} + \epsilon_t$$

$$x_{t-h} x_t = \phi x_{t-h} x_{t-1} + x_{t-h} \epsilon_t$$

$$E(x_{t-h} x_t) = E(\phi x_{t-h} x_{t-1}) + E(x_{t-h} \epsilon_t)$$

$$\gamma_h = \phi \gamma_{h-1}$$

Then, $\gamma_h = \phi^h \gamma_0 = \phi^h Var(x_t)$ and

$$\rho_h = \frac{\gamma_h}{Var(x_t)} = \frac{\phi^h Var(x_t)}{Var(x_t)} = \phi^h$$

Q1 (Jan 2009)

Let $X_1, ..., X_n \sim Bernoulli(\theta)$ and consider the following questions involving tests of hypotheses about θ.

a) Please comment on the form of the Bayes Factor when both the null and the alternative hypotheses consist of single points. Can the Bayes Factor be computed for all types of prior distributions (both proper and improper)? (Recall that the Bayes Factor is defined as the ratio of the posterior odds of the null divided by the prior odds of the null).

Solution:

$$\text{Bayes Factor} = \frac{\frac{P(H_0|y)}{P(H_1|y)}}{\frac{P(H_0)}{P(H_1)}} = \frac{P(y \mid H_0)}{P(y \mid H_1)} = \frac{\int P(\theta_0 \mid H_0)P(y \mid \theta_0, H_0)d\theta_0}{\int P(\theta_1 \mid H_1)P(y \mid \theta_1, H_1)d\theta_1}$$

The Bayes factor is only defined when the marginal density of y under H is proper. Bayes factor can't be computed when prior is improper.

b) Consider first the question of testing the null hypothesis $H_0 : \theta \leq \theta_0$ versus the alternative $H_A : \theta > \theta_0$. The prior distribution for θ is $Beta(a, b)$. Derive the form of the Bayes test of H_0

Solution:

$$\frac{P(H_0 \mid y)}{P(H_1 \mid y)} = \frac{P(H_0)}{P(H_1)}\frac{P(y \mid H_0)}{P(y \mid H_1)} = \frac{P(H_0)}{P(H_1)} \times \frac{\int P(\theta_0 \mid H_0)P(y \mid \theta_0, H_0)d\theta_0}{\int P(\theta_1 \mid H_1)P(y \mid \theta_1, H_1)d\theta_1}$$

Then,

$$\text{Bayes Factor} = \frac{\int_0^{\theta_0} \frac{\Gamma(a+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha+\sum y_i-1}(1-\theta)^{n-\sum y_i+\beta-1}d\theta_0}{\int_{\theta_0}^1 \frac{\Gamma(a+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha+\sum y_i-1}(1-\theta)^{n-\sum y_i+\beta-1}d\theta_0}$$

c) Consider now the question of testing the null hypothesis $H_0 : \theta = \theta_0$ versus the alternative $H_A : \theta \neq \theta_0$. The prior distribution is a mixed distribution, that puts mass π_0 on θ_0 and is distributed as $Beta(a, b)$ for $\theta \neq \theta_0$. Derive the form of the Bayes test of H_0 vs H_A and compute the Bayes Factor.

Solution:

$$\begin{aligned}
m(x) &= \int f(y \mid \theta)f(\theta)d\theta \\
&= \pi_0 f(y \mid \theta_0) + (1-\pi_0)\int f(y \mid \theta)f(\theta)d\theta \\
&= \pi_0\theta_0^{\sum y_i}(1-\theta_0)^{n-\sum y_i} + (1-\pi_0)\int_0^1 \frac{\Gamma(a+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha+\sum y_i-1}(1-\theta)^{n-\sum y_i+\beta-1}d\theta \\
&= \pi_0\theta_0^{\sum y_i}(1-\theta_0)^{n-\sum y_i} + (1-\pi_0)\frac{\Gamma(a+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+\sum y_i)\Gamma(n-\sum y_i+\beta)}{\Gamma(n+\alpha+\beta)}
\end{aligned}$$

And

$$P(\theta_0 = 0 \mid y) = \frac{\pi_0\theta_0^{\sum y_i}(1-\theta_0)^{n-\sum y_i}}{\pi_0\theta_0^{\sum y_i}(1-\theta_0)^{n-\sum y_i} + (1-\pi_0)\frac{\Gamma(a+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+\sum y_i)\Gamma(n-\sum y_i+\beta)}{\Gamma(n+\alpha+\beta)}}$$

Then,

$$\frac{P(\theta_0 = 0 \mid y)}{1-P(\theta_0 = 0 \mid y)} = \frac{\pi_0\theta_0^{\sum y_i}(1-\theta_0)^{n-\sum y_i}}{(1-\pi_0)\frac{\Gamma(a+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+\sum y_i)\Gamma(n-\sum y_i+\beta)}{\Gamma(n+\alpha+\beta)}}$$

$$\text{Bayes Factor}=\frac{\theta_0^{\sum y_i}(1-\theta_0)^{n-\sum y_i}}{\frac{\Gamma(a+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+\sum y_i)\Gamma(n-\sum y_i+\beta)}{\Gamma(n+\alpha+\beta)}}$$

Q2 (Jan 2009)

Suppose $X_1, ..., X_n$ is a random sample from the uniform distribution on the interval $[0, \theta]$.
a) Find the method of moments estimator for θ, and its mean and variance.
Solution:
$E(X) = \frac{\theta}{2}$ and $Var(X) = \frac{\theta^2}{12}$ then by central limit theorem, $\sqrt{n}(\bar{X} - \frac{\theta}{2}) \sim N(0, \frac{\theta^2}{12})$
method of moment
$m_1 = \frac{1}{n}\sum x_i$ and $\mu_1 = E(X) = \frac{\theta}{2}$. Then, $\frac{1}{n}\sum x_i = \frac{\theta}{2} \Rightarrow \tilde{\theta} = 2\bar{x}$
Then, by delta theorem, $g(x) = 2x$

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \sim N(0, \sigma^2 g'(x)^2)$$
$$\sqrt{n}(2\bar{x} - \theta) \sim N(0, \frac{\theta^2}{3})$$

Thus, $\tilde{\theta} \sim N(\theta, \frac{\theta^2}{3n})$
b) Find the MLE of θ.
Solution:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(x_i < \theta) = \frac{1}{\theta^n} I(max(x_1, ..., x_n) < \theta)$$
$$\frac{\partial L(\theta)}{\partial \theta} = -n\theta^{-n-1} < 0$$

Likelihood is decreasing respect to θ. Thus likelihood is maximized at $\hat{\theta}_{MLE} = max(x_1, ..., x_n) = X_{(n)}$
c) Find the probability density of the MLE and calculate its mean and variance.
Solution:

$$P(max(x_1, ..., x_n)) < x) = P(x_1 < x)^n = (\frac{x}{\theta})^n$$

Then,

$$f(x) = \frac{n}{\theta}(\frac{x}{\theta})^{n-1} = \frac{n}{\theta^n} x^{n-1}$$

and

$$E(x) = \int_0^\theta \frac{n}{\theta^n} x^n dx = \frac{n}{n+1} \frac{\theta^{n+1}}{\theta^n} = \frac{n}{n+1}\theta$$
$$E(x^2) = \int_0^\theta \frac{n}{\theta^n} x^{n+1} dx = \frac{n}{n+2}\theta^2$$
$$Var(x) = \frac{n}{n+2}\theta^2 - (\frac{n}{n+1})^2\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2$$

d) Compare the variance and bias of the method of moments estimator and the MLE.
Solution:
Method of moments estimator $\tilde{\theta} \sim N(\theta, \frac{\theta^2}{3n})$. Estimator is unbiased and variance is $\frac{\theta^2}{3n}$
MLE $\hat{\theta} \sim N(\frac{n}{n+1}\theta, \frac{n}{(n+2)(n+1)^2}\theta^2)$. Estimator is biased. Bias $= -\frac{1}{n+1}\theta$ and variance is $\frac{n}{(n+2)(n+1)^2}\theta^2$
e) Find a modification of the MLE that renders it unbiased
Solution:
$E(\frac{n+1}{n} x_{(n)}) = \theta$ and $Var(\frac{n+1}{n} x_{(n)}) = \frac{(n+1)^2}{n^2} \frac{n}{(n+2)(n+1)^2}\theta^2 = \frac{1}{n(n+2)}\theta^2$
f) Recommend an estimator of θ, with justification.
Solution:
$MSE_{mm} = \frac{\theta^2}{3n}$
$MSE_{MLE} = \frac{\theta^2}{(n+1)^2} + \frac{n}{(n+2)(n+1)^2}\theta^2$
$MSE_{\frac{n+1}{n}x_{(n)}} = \frac{1}{n(n+2)}\theta^2$

Q3 (Jan 2009)

Let X be a Bernoulli random with probability mass function

$$P(X = x) = \theta^x (1-\theta)^{1-x} I_{(0,1)}(x)$$

a) Let $X_1, ..., X_n$ be iid Bernoulli random variables with the same pmf as X. Derive the maximum likelihood estimator (MLE) for θ.

Solution:

Log likelihood is

$$l(\theta) = \sum x_i log(\theta) + (1 - x_i) log(1-\theta)$$

$$\frac{\partial l}{\partial \theta} = \sum \frac{x_i}{\theta} - \frac{1-x_i}{1-\theta} = 0 \Rightarrow \sum \frac{x_i - \theta}{\theta(1-\theta)} = 0 \Rightarrow \hat{\theta}_{MLE} = \bar{x}$$

b) Find asymptotic distribution of MLE

Solution:

$$I(\theta) = E\left(\frac{(\sum x_i - n\theta)^2}{\theta^2(1-\theta)^2}\right) = \frac{n}{\theta(1-\theta)}$$

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N(0, \theta(1-\theta))$$

c) Suppose that you are interested in estimating the function $g(\theta)$ defined to be

$$g(\theta) = \theta log(\theta) + (1-\theta) log(1-\theta)$$

Using the delta method, derive the asymptotic distribution of

$$\sqrt{n}\{g(\hat{\theta}_n) - g(\theta)\}$$

and obtain an approximate 95% confidence interval for $g(\theta)$.

Solution:

$$\begin{aligned}\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) &\to N(0, g'(\theta)^2\theta(1-\theta)) \\ &= N(0, (log\frac{\theta}{1-\theta})^2\theta(1-\theta))\end{aligned}$$

d) Derive an expression for the asymptotic bias of the MLE $g(\hat{\theta}_n)$, i.e.

$$Bias\{g(\hat{\theta}_n), g(\theta)\} = E\{g(\hat{\theta}_n)\} - g(\theta)$$

and show that $Bias\{g(\hat{\theta}_n), g(\theta)\}$ is of order $O(1/n)$. For this test, you only need to derive the bias using the second order Taylor series expansion of $g(\hat{\theta}_n)$.

Solution:

$$\begin{aligned}g(\hat{\theta}_n) &= g(\theta) + g'(\theta)(\hat{\theta}_n - \theta) + g''(\theta)\frac{(\hat{\theta}_n - \theta)^2}{2} \\ g(\hat{\theta}_n) - g(\theta) &= log\frac{\theta}{1-\theta}(\hat{\theta}_n - \theta) + \frac{1}{\theta(1-\theta)}\frac{(\hat{\theta}_n - \theta)^2}{2} \\ E\{g(\hat{\theta}_n)\} - g(\theta) &= \frac{1}{2\theta(1-\theta)}\frac{\theta(1-\theta)}{n} = \frac{1}{2n}\end{aligned}$$

e) Based on your results above, apply the bias-correction to the approximate 95% confidence interval for $g(\theta)$.

Solution:

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \to N(\frac{1}{2n}, (log\frac{\theta}{1-\theta})^2\theta(1-\theta))$$

and

$$g(\theta) + \frac{1}{2n} \pm 1.96\sqrt{\frac{(log\frac{\theta}{1-\theta})^2\theta(1-\theta)}{n}}$$

Q4 (Jan 2009)

Let X be an exponential random variable having a probability density function

$$f(x \mid \lambda) = \frac{1}{\lambda} exp(-x/\lambda) I_{(0,\infty)}(x)$$

where $\lambda > 0$.

a) Show that for any $u > 0$, $P(X > u) = exp(-u/\lambda)$.

Solution:

$$P(X > u) = 1 - P(X < u) = 1 - (1 - exp(-\frac{u}{\lambda})) = exp(-u/\lambda)$$

b) Suppose that you are asked to test the hypothesis

$$H_1 : \lambda = \lambda_1 \,\text{vs.}\, H_2 : \lambda = \lambda_2$$

where without any loss of generality, $0 < \lambda_1 < \lambda_2$. The likelihood ratio statistic is of the form

$$R(X, \lambda_1, \lambda_2) = \frac{f(x \mid \lambda_2)}{f(x \mid \lambda_1)}$$

Set the probability of type I error to be α. Find the rejection/critical region

$$\{x \mid R(x, \lambda_1, \lambda_2) > \tau\}$$

Solution:

Choose τ s.t. $P(R(x, \lambda_1, \lambda_2) > \tau \mid \lambda_1) = \alpha$

$$\frac{\frac{1}{\lambda_2} exp(-\frac{x}{\lambda_2})}{\frac{1}{\lambda_1} exp(-\frac{x}{\lambda_1})} = \frac{\lambda_1}{\lambda_2} exp((\frac{1}{\lambda_1} - \frac{1}{\lambda_2})x) > \tau$$

$$x > \frac{log(\frac{\lambda_2}{\lambda_1}\tau)}{\frac{1}{\lambda_1} - \frac{1}{\lambda_2}}$$

Choose τ s.t. $P(X > \frac{log(\frac{\lambda_2}{\lambda_1}\tau)}{\frac{1}{\lambda_1} - \frac{1}{\lambda_2}} \mid \lambda_1) = \alpha$. Then,

$$exp(-\frac{log(\frac{\lambda_2}{\lambda_1}\tau)}{\frac{1}{\lambda_1} - \frac{1}{\lambda_2}}/\lambda_1) = \alpha \Rightarrow \frac{log(\frac{\lambda_2}{\lambda_1}\tau)}{\frac{1}{\lambda_1} - \frac{1}{\lambda_2}} = -\lambda_1 log\alpha$$

$$\tau = exp[(\frac{1}{\lambda_1} - \frac{1}{\lambda_2})(-\lambda_1 log\alpha)] \times \frac{\lambda_1}{\lambda_2}$$

c) Under the above approach to hypothesis testing, which is the classical paradigm, the threshold τ is obtained once we have decided on the probability of type I error α. For this problem, we shall derive the threshold based on a different paradigm. It is obtained by minimizing both the probability of type I and type II errors. In order to simplify this problem, consider testing the hypothesis

$$H_1 : \lambda = 1 \,\text{vs.}\, H_2 : \lambda = 2$$

Denote the probability of type I and type II errors (as functions of the threshold k) to be, respectively,

$$\alpha(k) = P\{R(X, \lambda_1, \lambda_2) > k \mid \lambda = 1\}$$
$$\beta(k) = P\{R(X, \lambda_1, \lambda_2) < k \mid \lambda = 2\}$$

Find k that minimizes the sum of the two errors

$$g(k) = \alpha(k) + \beta(k)$$

Solution:

$$R(X,\lambda_1,\lambda_2) > k \text{ and } \frac{\frac{1}{\lambda_2}exp(-\frac{x}{\lambda_2})}{\frac{1}{\lambda_1}exp(-\frac{x}{\lambda_1})} > k$$

$$x\frac{\lambda_1}{\lambda_2}exp((\frac{1}{\lambda_1}-\frac{1}{\lambda_2})x) > k \Rightarrow x > \frac{log\frac{\lambda_2}{\lambda_1}k}{\frac{1}{\lambda_1}-\frac{1}{\lambda_2}} = 2log2k$$

Then,

$$\alpha(k) = P(X > 2log2k \mid \lambda = 1) = exp(-\frac{2log2k}{1}) = \frac{1}{4k^2}$$
$$\beta(k) = P(X < 2log2k \mid \lambda = 2) = 1 - exp(-\frac{2log2k}{2}) = 1 - \frac{1}{2k}$$

So,

$$\alpha(k) + \beta(k) = 1 - \frac{1}{2k} + \frac{1}{4k^2}$$

Take derivative respect to k

$$\frac{1}{2k^2} - 2\frac{1}{4k^3} = 0 \Rightarrow \frac{1}{2k^2} = \frac{1}{2k^3} \Rightarrow k^3 = k^2 \Rightarrow k^2(k-1) = 0$$

The $k = 1$ minimizes the sum of the two errors. Double derivative is $-2\frac{1}{2k^3} + 6\frac{1}{4k^4} = \frac{-k+1.5}{k^4}$. At $k = 1$, double derivative is positive and thus is minimum.

Q5 (Jan 2009)

This question concerns drawing inference about a mixture of two Poisson distributions having rate parameters λ_0 and λ_1, respectively. We will first consider the situation where the distribution is a labeled mixture; that is, for each individual we observe the pair (Z, Y), where

$$Y \sim Poisson(\lambda_0) \text{ if } Z = 0$$
$$Y \sim Poisson(\lambda_1) \text{ if } Z = 1.$$

Assume that Z has a Bernoulli distribution with probability θ; i.e., $P(Z = 1) = \theta$.
a) Given a sample $(Z_1, Y_1), ..., (Z_n, Y_n)$, show that the loglikelihood $l(\lambda_0, \lambda_1, \theta \mid Y, Z)$ satisfies

$$l(\lambda_0, \lambda_1, \theta \mid Y, Z) \propto \sum Z_i(Y_i log\lambda_1 - \lambda_1) + (1 - Z_i)(Y_i log\lambda_0 - \lambda_0)$$

and derive the MLE for each parameter.
Solution:

$$f(Y, Z; \lambda_0, \lambda_1, \theta) = \prod_{i=1}^{n} (\frac{\lambda_1^{y_i} e^{-\lambda_1}}{y_i!})^{z_i} (\frac{\lambda_0^{y_i} e^{-\lambda_0}}{y_i!})^{1-z_i}$$

$$L(\lambda_0, \lambda_1, \theta) \propto \sum_{i=1}^{n} z_i(y_i log\lambda_1 - \lambda_1) + (1 - z_i)(y_i log\lambda_0 - \lambda_0)$$

Take derivative,

$$\frac{\partial L}{\partial \lambda_1} = \sum_{i=1}^{n} z_i(\frac{y_i}{\lambda_1} - 1) = 0 \Rightarrow \hat{\lambda}_1 = \frac{\sum_{i=1}^{n} z_i y_i}{\sum_{i=1}^{n} z_i} \text{ and } \frac{\partial^2 L}{\partial \lambda^2} = -\frac{\sum z_i y_i}{\lambda_1^2} < 0$$

$$\frac{\partial L}{\partial \lambda_0} = \sum_{i=1}^{n} (1 - z_i)(\frac{y_i}{\lambda_0} - 1) = 0 \Rightarrow \hat{\lambda}_0 = \frac{\sum_{i=1}^{n} (1 - z_i) y_i}{\sum_{i=1}^{n} (1 - z_i)} \text{ and } \frac{\partial^2 L}{\partial \lambda_0^2} = -\frac{\sum (1 - z_i) y_i}{\lambda_0^2} < 0$$

Now consider the case where the mixture is unlabeled; that is, each individual has only Y observed, and the sample is $Y_1, ..., Y_n$. A common approach to inference is to treat this as a missing data problem and to use the EM algorithm. The complete data is (Y, Z) and the observed data is Y.
The EM algorithm starts with initial values $(\theta^*, \lambda_0^*, \lambda_1^*)$ for the parameters $(\theta, \lambda_0, \lambda_1)$, and then iterates between two steps. In the E-step, one computes the expected values of the complete-data loglikelihood, conditional on observed data. In the M-step, the parameter values are updated by maximizing the expected loglikelihood. The algorithm iterates between the E and M steps until convergence is reached. The questions below assume you are starting the first iteration of the EM algorithm, with the objective of updating the initial values $(\theta^*, \lambda_0^*, \lambda_1^*)$.
b) For the loglikelihood in part (a), the E-step requires finding the expected value of the missing Z for each individual; denote this by $\hat{Z}_i = E(Z_i \mid Y_i)$. Find an expression for $\hat{Z}_i$ in terms of Y_i and the initial values $(\theta^*, \lambda_0^*, \lambda_1^*)$.
Solution:

$$\hat{Z}_i = E(Z_i \mid Y_i) = 0 \times f(Z_i = 0 \mid Y_i) + 1 \times f(Z_i = 1 \mid Y_i)$$

Here,

$$f(Z_i \mid Y_i) = \frac{f(Z_i, Y_i)}{f(Y_i)} = \frac{(\frac{\lambda_1^{y_i} e^{-\lambda_1}}{y_i!})^{z_i} (\frac{\lambda_0^{y_i} e^{-\lambda_0}}{y_i!})^{1-z_i} \theta^{z_i} (1 - \theta)^{1-z_i}}{\theta \frac{\lambda_1^{y_i} e^{-\lambda_1}}{y_i!} + (1 - \theta) \frac{\lambda_0^{y_i} e^{-\lambda_0}}{y_i!}}$$

Then,

$$\hat{Z}_i = \frac{\theta \frac{\lambda_1^{y_i} e^{-\lambda_1}}{y_i!}}{\theta \frac{\lambda_1^{y_i} e^{-\lambda_1}}{y_i!} + (1 - \theta) \frac{\lambda_0^{y_i} e^{-\lambda_0}}{y_i!}}$$

Now, we put the initial values $(\theta^*, \lambda_0^*, \lambda_1^*)$ into $\hat{Z}_i$ above.

$$\hat{Z}_i = \frac{\theta^* \frac{\lambda_1^{*y_i} e^{-\lambda_1^*}}{y_i!}}{\theta^* \frac{\lambda_1^{*y_i} e^{-\lambda_1^*}}{y_i!} + (1-\theta^*) \frac{\lambda_0^{*y_i} e^{-\lambda_0^*}}{y_i!}}$$

c) The M step involves maximizing the loglikelihood in part (a), where Z_i is replaced with $\hat{Z}_i$. Find expressions for the updated MLE's of λ_0 and λ_1 in terms of the $\hat{Z}_i$ and Y_i. Note $\hat{Z}_i$ is a function of the initial parameter values, not the parameters themselves.

Solution:

$$\hat{\lambda}_1 = \frac{\sum_{i=1}^n \hat{Z}_i Y_i}{\sum_{i=1}^n \hat{Z}_i}$$

$$\hat{\lambda}_0 = \frac{\sum_{i=1}^n (1-\hat{Z}_i) Y_i}{\sum_{i=1}^n (1-\hat{Z}_i)}$$

d) How would you compute an updated estimate of θ?

Solution:

We could use sample mean of updated $\hat{Z}_i$ to estimate updated $\hat{\theta} = \frac{\sum_{i=1}^n \hat{Z}_i}{n}$

e) Algorithms for drawing inference about unlabeled mixtures can be numerically unstable. Thinking conceptually about this problem, describe situations where you believe estimation routines would be stable, and where they would be unstable.

Solution:

If rate parameter λ_0 and λ_1 have similar values, in the step of calculating updated $\hat{Z}_i$, replacing λ_0^* by λ_1^* or λ_1^* by λ_0^* does not make much difference for the value of $\hat{Z}_i$ which might cause the estimation unstable.

Q1 (Jan 2010)

First, consider the model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, ..., n$ where

$$\begin{bmatrix} x_i \\ \epsilon_i \end{bmatrix} \sim N_2(\begin{bmatrix} \mu_x \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_\epsilon^2 \end{bmatrix})$$

We have

$$\begin{bmatrix} Y_i \\ x_i \end{bmatrix} \sim N_2(\begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, \begin{bmatrix} \sigma_y^2 & \sigma_{xY} \\ \sigma_{xY} & \sigma_x^2 \end{bmatrix})$$

where $\sigma_y^2 = \beta_1^2\sigma_x^2 + \sigma_\epsilon^2$, $\mu_y = \beta_0 + \beta_1\mu_x$, and $\sigma_{xY} = \beta_1\sigma_x^2$.

a) Derive $E(Y_i \mid x_i)$ and $Var(Y_i \mid x_i)$

Solution:

$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim N(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix})$

$$y_1 \mid y_2 = a \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Then,

$$\begin{aligned} E(Y_i \mid x_i) &= \mu_y + \sigma_{xY}\frac{1}{\sigma_x^2}(x_i - \mu_x) \\ &= \beta_0 + \beta_1\mu_x + \beta_1\sigma_x^2\frac{1}{\sigma_x^2}(x_i - \mu_x) \\ &= \beta_0 + \beta_1\mu_x + \beta_1 x_i - \beta_1\mu_x \\ &= \beta_0 + \beta_1 x_i \end{aligned}$$

and

$$\begin{aligned} Var(Y_i \mid x_i) &= \sigma_y^2 - \sigma_{xY}\frac{1}{\sigma_x^2}\sigma_{xY} \\ &= \beta_1^2\sigma_x^2 + \sigma_\epsilon^2 - \beta_1\sigma_x^2\frac{1}{\sigma_x^2}\beta_1\sigma_x^2 \\ &= \sigma_\epsilon^2 \end{aligned}$$

Second, suppose now that one does not observe x_i, $i = 1, 2, ..., n$, but observes $w_i = x_i + u_i$, where

$$\begin{bmatrix} x_i \\ \epsilon_i \\ u_i \end{bmatrix} \sim N_3(\begin{bmatrix} \mu_x \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_\epsilon^2 & 0 \\ 0 & 0 & \sigma_\epsilon^2 \end{bmatrix})$$

Assume that Y is conditionally independent of w: $E(Y_i \mid x_i, w_i) = E(Y_i \mid x_i)$. Suppose the true model is $E(Y_i \mid x_i) = \beta_0 + \beta_1 x_i$, but the observed data are (Y_i, w_i), $i = 1, 2, ..., n$.

b) Relate $E(Y_i \mid w_i)$ to $E(x_i \mid w_i)$

Solution:

$$\begin{aligned} E(Y_i \mid w_i) &= E(E(Y_i \mid x_i, w_i) \mid w_i) \\ &= E(E(Y_i \mid x_i) \mid w_i) \\ &= E(\beta_0 + \beta_1 x_i \mid w_i) \\ &= \beta_0 + \beta_1 E(x_i \mid w_i) \end{aligned}$$

c) What is the joint distribution of x_i and w_i and what is $E(x_i \mid w_i)$?

Solution:

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ \epsilon_i \\ u_i \end{bmatrix} = \begin{bmatrix} x_i \\ x_i + u_i \end{bmatrix} = \begin{bmatrix} x_i \\ w_i \end{bmatrix}$$

Then,

$$E(\begin{bmatrix} x_i \\ w_i \end{bmatrix}) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_x \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_x \end{bmatrix}$$

$$Var(\begin{bmatrix} x_i \\ w_i \end{bmatrix}) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_\epsilon^2 & 0 \\ 0 & 0 & \sigma_\epsilon^2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ \sigma_x^2 & 0 & \sigma_\epsilon^2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 + \sigma_\epsilon^2 \end{bmatrix}$$

Hence,

$$\begin{bmatrix} x_i \\ w_i \end{bmatrix} \sim N_2(\begin{bmatrix} \mu_x \\ \mu_x \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 + \sigma_\epsilon^2 \end{bmatrix})$$

and

$$E(x_i \mid w_i) = \mu_x + \sigma_x^2 \frac{1}{\sigma_x^2 + \sigma_\epsilon^2}(w_i - \mu_x)$$

d) Show that $E(Y_i \mid w_i) = \beta_0^* + \beta_1^* w_i$ (Use your answers to (b) and (c)). What is the relationship between β_0^*, β_1^* and β_0, β_1?

Solution:

$$\begin{aligned} E(Y_i \mid w_i) &= \beta_0 + \beta_1[\mu_x + \sigma_x^2 \frac{1}{\sigma_x^2 + \sigma_\epsilon^2}(w_i - \mu_x)] \\ &= \beta_0 + \beta_1 \mu_x - \beta_1 \frac{\mu_x \sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2} + \beta_1 \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2} w_i \end{aligned}$$

We see that

$$\beta_0^* = \beta_0 + \beta_1 \mu_x - \beta_1 \frac{\mu_x \sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2} \text{ and } \beta_1^* = \beta_1 \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2}$$

Q3 (Jan 2010)

The causal effect of a treatment (e.g. a drug) on an outcome can be defined using potential outcomes. The idea is to define two outcomes for each individual: an individual would have outcome Y^1 if taking the treatment and would have outcome Y^0 if not. The causal effect of the treatment on the individual is defined as $Y^1 - Y^0$. Of general interest is the average causal effect which is defined as $ACE = E(Y^1 - Y^0)$.

This problem illustrates how to estimate ACE in randomized trials. Suppose that we have a sample of n individuals. We randomize each individual with probability π to the treatment. Let Z_i be the randomization indicator. $Z_i = 1$ if the ith individual receives the treatment; $Z_i = 0$ if randomized to the control. The observed outcome is $Y_i = Y_i^1 Z_i + Y_i^0(1 - Z_i)$.

a) Show that ACE is identified by the observed data by proving

$$ACE = E(Y \mid Z = 1) - E(Y \mid Z = 0)$$

Hint: It is sufficient to prove $E(Y \mid Z = 1) = E(Y^1)$. You need to use the fact that Z is randomized; i.e. Z is independent of Y^0 and Y^1.

Solution:

$$ACE = E(Y^1 - Y^0) = E(Y^1) - E(Y^0) = E(Y \mid Z = 1) - E(Y \mid Z = 0)$$

b) Show that the ACE can be unbiasedly estimated by

$$\frac{1}{n}\{\frac{\sum_{i=1}^n Y_i Z_i}{\pi} - \frac{\sum_{i=1}^n Y_i(1 - Z_i)}{1 - \pi}\}$$

Note that Y_i and Z_i are not independent.

Solution:

$$\begin{aligned}
E(\frac{1}{n}(\frac{\sum Y_i Z_i}{\pi} - \frac{\sum Y_i(1 - Z_i)}{1 - \pi})) &= \frac{1}{n\pi}E(\sum Y_i Z_i) - \frac{1}{n(1-\pi)}E(\sum Y_i(1 - Z_i)) \\
&= E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)
\end{aligned}$$

since

$$\begin{aligned}
E(\sum Y_i Z_i) &= \sum E(Y_i Z_i) \\
E(Y_i Z_i) &= E(E(Y_i Z_i \mid Z_i)) \\
&= \sum_{a=0}^{1} P(Z_i = a)E(Y_i Z_i \mid Z_i = a) \\
&= \pi E(Y_i \mid Z_i = 1) + (1 - \pi)E(Y_i \times 0 \mid Z_i = 0) \\
&= \pi E(Y^1)
\end{aligned}$$

And,

$$\begin{aligned}
E(Y_i(1 - Z_i)) &= E(E(Y_i(1 - Z_i) \mid Z_i)) \\
&= \sum_{a=0}^{1} P(Z_i = a)E(Y_i(1 - Z_i) \mid Z_i = a) \\
&= \pi E(Y_i(1 - 1) \mid Z_i = 1) + (1 - \pi)E(Y_i \mid Z_i = 0) \\
&= (1 - \pi)E(Y^0)
\end{aligned}$$

c) Estimate π by $\hat{\pi} = \frac{1}{n}\sum_{i=1}^n Z_i$. Show that the ACE can be consistently estimated by

$$\frac{1}{n}\{\frac{\sum_{i=1}^n Y_i Z_i}{\hat{\pi}} - \frac{\sum_{i=1}^n Y_i(1 - Z_i)}{1 - \hat{\pi}}\}$$

Solution:
We have that $\frac{1}{n}\sum_{i=1}^{n} Y_i Z_i \xrightarrow{p} \pi E(Y^1)$ and $\frac{1}{n}\sum_{i=1}^{n} Z_i \xrightarrow{p} \pi$, then

$$\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i Z_i}{\hat{\pi}} \xrightarrow{p} \frac{\pi E(Y^1)}{\pi} = E(Y^1)$$

Similarly,

$$\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i (1 - Z_i)}{1 - \hat{\pi}} \xrightarrow{p} \frac{(1-\pi) E(Y^0)}{1-\pi} = E(Y^0)$$

Q4 (Jan 2010)

After a major snowstorm, many people in Providence were left without power. Two students, Chris and Kate, were discussing their dilemmas. Chris was relatively lucky and had power restored after 48 hours; however, kate was still without power after 54 hours. If we are willing to assume that the time without power is exponentially distributed with mean λ (i.e. the pdf is $f(x \mid \lambda) = \frac{1}{\lambda} exp(-x/\lambda) I_{(0,\infty)}(x)$) and that Chris and Kate's outcomes are independent of each other, answer the following questions.

a) Write down the likelihood function of λ.

Solution:

$F(x;\lambda) = 1 - e^{-x/\lambda}$ and define tail function

$$G(x;\lambda) = 1 - F(x;\lambda) = e^{-x/\lambda}$$

Assume that the first r observations are fully observed, while for $x_{r+1}, ..., x_n$ we only know that $x_j > t_j$ for some known positive constant. Likelihood is the "probability of the observed data" and for censored data, it is given by $P(x_j > t_j) = G(t_j;\lambda)$. Full likelihood is

$$L(\lambda) = \prod_{i=1}^{r} f(x_i;\lambda) \prod_{i=r+1}^{n} G(t_j;\lambda) = \frac{1}{\lambda^r} exp(-\frac{\sum_{i=1}^{r} x_i}{\lambda}) \times exp(-\frac{\sum_{i=r+1}^{n} t_i}{\lambda})$$

Then

$$l(\lambda) = -r log(\lambda) - \frac{\sum_{i=1}^{r} x_i}{\lambda} - \frac{\sum_{i=r+1}^{n} t_i}{\lambda}$$

b) Find the maximum likelihood estimate of λ.

Solution:

$$\frac{\partial l(\lambda)}{\partial \lambda} = -\frac{r}{\lambda} + \frac{\sum_{i=1}^{r} x_i}{\lambda^2} + \frac{\sum_{i=r+1}^{n} t_i}{\lambda^2} = 0$$

Then,

$$\hat{\lambda} = \frac{\sum_{i=1}^{r} x_i + \sum_{i=r+1}^{n} t_i}{r} = \frac{48+54}{1} = 102$$

c) Find the maximum likelihood estimate for the probability that for another individual the power will be out for more than 120 hours.

Solution:

$$P(X > 120) = e^{-120/\hat{\lambda}} = e^{-120/102} = 0.308$$

d) Using a non-informative prior distribution for λ, $p(\lambda) \propto 1$ for $\lambda > 0$, find the posterior distribution of λ given Chris and Kate's data.

Solution:

$$P(\lambda \mid x) \propto P(\lambda) L(\lambda) = (\frac{1}{\lambda})^r exp(-\frac{1}{\lambda}(\sum_{i=1}^{r} x_i + \sum_{i=r+1}^{n} t_i)) \sim Gamma(r+1, \sum_{i=1}^{r} x_i + \sum_{i=r+1}^{n} t_i)$$

e) Kate is interested in the probability that her power will be restored by 120 hours given that it hasn't yet been restored after 54 hours. Express this as a function of λ, say $\phi(\lambda)$.

Solution:

$$P(x < 120 \mid x > 54) = P(x < 66) = 1 - exp(-66/\lambda)$$

f) Given Chris and Kate's data, what is the posterior mean for $\phi(\lambda)$?

Solution:

$$E(\lambda \mid y) = \alpha\beta = (r+1)(\sum_{i=1}^{r} x_i + \sum_{i=r+1}^{n} t_i) = 204$$

$$
\begin{aligned}
E(\phi(\lambda) \mid y) &= \int (1 - exp(-66/\lambda)) \frac{102^2}{\Gamma(2)} (\frac{1}{\lambda})^{2-1} exp(-102/\lambda) dx \\
&= 1 - \int \frac{102^2}{\Gamma(2)} (\frac{1}{\lambda})^{2-1} exp(-168/\lambda) dx \\
&= 1 - \frac{102^2}{168^2} \int \frac{168^2}{\Gamma(2)} (\frac{1}{\lambda})^{2-1} exp(-168/\lambda) dx \\
&= 1 - \frac{102^2}{168^2} = 1 - 0.3686 = 0.6313
\end{aligned}
$$

Q5 (Jan 2010)

Let $X_1, ..., X_n, X_{n+1}, ..., X_{2n}$ be iid exponential random variable with pdf

$$f(x \mid \lambda) = \frac{1}{\lambda} exp(-x/\lambda) I_{(0,\infty)}(x)$$

Recall that $E(X) = \lambda$ and $Var(X) = \lambda^2$

Suppose that $X_1, ..., X_n$ are observed directly while $X_{n+1}, ..., X_{2n}$ are not directly observed. Here, we observe Y_i, $i = 1, ..., n$ where

$$Y_i = \begin{cases} 1, & \text{if } X_{n+i} > 5 \\ 0, & \text{otherwise.} \end{cases}$$

a) Using $X_1, ..., X_n$, the MLE for λ is $U_{1n} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. State the approximate (asymptotic) distribution of U_{1n}.

Solution:

By CLT, we have $\frac{\sqrt{n}(U_{1n}-\lambda)}{\lambda} \overset{D}{\sim} N(0,1)$ when $n \to \infty$

b) Using $Y_1, ..., Y_n$, derive the MLE for λ. Denote the MLE to be $U_{2n} = U_2(Y_1, ..., Y_n)$. Here, you may directly use: $B_1, ..., B_n$ iid Bernoulli with $E(B_i) = \beta$. The MLE for β is $\hat{\beta}_n = \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

Solution:

Let $\beta = P(X_{n+i} > 5)$. Then, we have

$$\beta = \int_5^\infty \frac{1}{\lambda} exp(-\frac{x}{\lambda})dx = [-exp(-\frac{x}{\lambda})]_5^\infty = exp(-\frac{5}{\lambda})$$

Then $\lambda = -\frac{5}{log(\beta)}$. We know that $\hat{\beta}_n = \bar{Y}_n$. By the invariance of MLE, we have

$$\hat{\lambda}_{MLE} = -\frac{5}{log(\hat{\beta})} = -\frac{5}{log(\bar{Y}_n)} = U_{2n}$$

c) Derive the asymptotic distribution of U_{2n}.

Solution:

We know that

$$\sqrt{n}(\bar{Y}_n - \beta) \overset{D}{\to} N(0, \beta(1-\beta))$$

Define $g(\beta) = -\frac{5}{log\beta}$, then we have

$$\sqrt{n}(g(\hat{\beta}) - g(\beta)) \overset{D}{\to} N(0, \beta(1-\beta)\, g'(\beta)^2)$$

and

$$\beta(1-\beta)\, g'(\beta)^2 = \beta(1-\beta)(\frac{5}{(log\beta)^2\beta})^2 = \frac{25(1-\beta)}{(log\beta)^4\beta} = \frac{25(1-e^{-5/\lambda})}{(-\frac{5}{\lambda})^4 e^{-5/\lambda}} = \frac{\lambda^4(e^{5/\lambda}-1)}{25}$$

Hence

$$\sqrt{n}(U_{2n} - \lambda) \overset{D}{\to} N(0, \frac{\lambda^4(e^{5/\lambda}-1)}{25})$$

d) Denote an estimator $U_n(\alpha)$ to be

$$U_n(\alpha) = \alpha U_{1n} + (1-\alpha)U_{2n}$$

where $\alpha \in [0,1]$. Derive an approximate 95% confidence interval estimator for λ based on the estimator $U_n(\alpha)$.

Solution:

We have

$$\sqrt{n}(U_{1n} - \lambda) \overset{D}{\to} N(0, \lambda^2)$$

$$\sqrt{n}(U_{2n} - \lambda) \xrightarrow{D} N(0, \frac{\lambda^4(e^{5/\lambda} - 1)}{25})$$

Then, we have

$$\sqrt{n}(U_n(\alpha) - \lambda) \to N(0, \alpha^2\lambda^2 + (1-\alpha)^2\frac{\lambda^4(e^{5/\lambda} - 1)}{25})$$

A 95% CI is given by

$$\hat{\lambda} \pm 1.96 \times \sqrt{\alpha^2\lambda^2 + (1-\alpha)^2\frac{\lambda^4(e^{5/\lambda} - 1)}{25}}$$

Q1 (June 2010)

The goal is to estimate the parameter μ from a random sample $X = [X_1, ..., X_n]'$. We consider two estimators, namely

$$T_1 = T_1(X) \,\text{and}\, T_2 = T_2(X), \text{ where}$$

$\mathbb{E}T_k = \mu + \delta_k$ and $\mathbb{V}arT_k = \sigma_k^2$, $k = 1, 2$. Using these two estimators, we form a class of weighted estimator

$$\{T(\alpha) = \alpha T_1 + (1-\alpha)T_2,\, \alpha \in (0,1)\}.$$

Here, we would like to find the weight α that minimizes the mean-squared error

$$\xi(\alpha) = E(T(\alpha) - \mu)^2$$

a) For the first part of this problem, assume that $Cov(T_1, T_2) = 0$

Solution:

i. For any α, derive the bias of the estimator $T(\alpha)$

$$\begin{aligned} E(T(\alpha)) &= E(\alpha T_1 + (1-\alpha)T_2) \\ &= \alpha E(T_1) + (1-\alpha)E(T_2) \\ &= \alpha(\mu + \delta_1) + (1-\alpha)(\mu + \delta_2) \\ &= \mu + \alpha\delta_1 + (1-\alpha)\delta_2 \end{aligned}$$

Bias is $\alpha\delta_1 + (1-\alpha)\delta_2$

ii) For any α, derive the variance of the estimator $T(\alpha)$

$$\begin{aligned} Var(T(\alpha)) &= Var(\alpha T_1) + Var((1-\alpha)T_2) \\ &= \alpha^2 Var(T_1) + (1-\alpha)^2 Var(T_2) \\ &= \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2 \end{aligned}$$

iii) Derive the weight α^* that minimizes $\xi(\alpha)$

$$2\alpha\sigma_1^2 - 2(1-\alpha)\sigma_2^2 = 0$$

$$\alpha^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

b) For the second part of this problem, we shall assume that T_1 and T_2 are correlated; more precisely, $Cov(T_1, T_2) = \tau$ and that $\tau < \min_k \delta_k^2 + \sigma_k^2$. Determine the weight α^* that minimizes $\xi(\alpha)$.

Solution:

$$\begin{aligned} Var(T(\alpha)) &= Var(\alpha T_1) + Var((1-\alpha)T_2) + 2Cov(\alpha T_1, (1-\alpha)T_2) \\ &= \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2 + 2\alpha(1-\alpha)\tau \end{aligned}$$

Then,

$$2\alpha\sigma_1^2 - 2(1-\alpha)\sigma_2^2 + 2\tau - 2\alpha\tau = 0 \Rightarrow \alpha^* = \frac{\sigma_2^2 - \tau}{\sigma_1^2 + \sigma_2^2 - \tau}$$

c) We assume further that the estimators T_1 and T_2 are jointly asymptotically Gaussian with mean $[\mu + \delta_1,\, \mu + \delta_2]$ and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \tau \\ \tau & \sigma_2^2 \end{bmatrix}$$

Denote $T^* = T(\alpha^*)$ to be the optimal estimator. Give an approximate distribution of T^*.

Solution:

$$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix} \sim N(\begin{bmatrix} \mu + \delta_1 \\ \mu + \delta_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \tau \\ \tau & \sigma_2^2 \end{bmatrix})$$

$T^* = \alpha^* T_1 + (1-\alpha^*)T_2$
Result 5.3 If $X \sim N_p(\mu, V)$ and $Y = a + BX$ where a is $q \times 1$, and B is $q \times p$, then $Y \sim N_q(a + B\mu, BVB^T)$.
Then,

$$T^* = \begin{bmatrix} \alpha^* & (1-\alpha^*) \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}$$

and

$$\begin{aligned} T^* &\sim N(\alpha^*(\mu+\delta_1) + (1-\alpha^*)(\mu+\delta_2),\ \alpha^{*2}\sigma_1^2 + 2\alpha^*(1-\alpha^*)\tau + (1-\alpha^*)^2\sigma_2^2) \\ &= N(\mu + \alpha^*\delta_1 + (1-\alpha^*)\delta_2,\ \alpha^{*2}\sigma_1^2 + 2\alpha^*(1-\alpha^*)\tau + (1-\alpha^*)^2\sigma_2^2) \end{aligned}$$

Q3 (June 2010)

Suppose that we have independent samples of a response y_{ij} corresponding to drug dose $i\,(i = 1, 2, 3)$ and drug formulation $j\,(j = 1, ..., 4)$. Consider a two-way ANOVA model with missing cells:

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

where e_{ij}'s are iid $N(0, \sigma^2)$ and $y_{6\times 1} = \begin{bmatrix} y_{12} & y_{13} & y_{22} & y_{23} & y_{31} & y_{34} \end{bmatrix}$

a) Is $\beta_1 - \beta_3$ estimable? That is, does there exist a linear function of the y_{ij}'s that is unbiased for $\beta_1 - \beta_3$?

Solution:

Result 3.1 Under the linear mean model, $\lambda^T b$ is (linearly) estimable iff there exists a vector a such that $E(a^T y) = \lambda^T b$ for all b, or $\lambda^T = a^T X$ or $\lambda = X^T a$

The design matrix looks like

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

There doesn't exist a linear function of y_{ij} that is unbiased for $\beta_1 - \beta_3$

b) Consider testing whether there is any drug-formulation effect, that is, testing the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$. Is H_0 a testable hypothesis? Breifly explain.

Solution:

Can test $\beta_2 = \beta_3$ and $\beta_1 = \beta_4$, but can't link β_2, β_3 with β_1, β_4

c) The investigator decides to construct a test of the hypothesis H_0 using his own instincts. He takes the differences of the pairs of observations with the same drug dose:

$$z_1 = y_{12} - y_{13}$$
$$z_2 = y_{22} - y_{23}$$
$$z_3 = y_{31} - y_{34}$$

and construct the statistic $G = \frac{(z_1+z_2)^2/2+z_3^2}{(z_1-z_2)^2}$, intending to reject H_0 if G is too large.

Derive the distribution of z, where $z = (z_1, z_2, z_3)^T$

Solution:

$$\begin{aligned} E(z_1) &= E(y_{12}) - E(y_{13}) = \beta_2 - \beta_3 \\ E(z_2) &= E(y_{22}) - E(y_{23}) = \beta_2 - \beta_3 \\ E(z_3) &= E(y_{31}) - E(y_{34}) = \beta_1 - \beta_4 \\ Var(z_1) &= Var(y_{12}) + Var(y_{13}) = 2\sigma^2 \\ Var(z_2) &= Var(z_3) = 2\sigma^2 \end{aligned}$$

Then,

$$Cov(z_1, z_2) = Cov(y_{12} - y_{13}, y_{22} - y_{23}) = 0$$

since e_{ij} is iid. Similarly,

$$Cov(z_2, z_3) = Cov(z_1, z_3)$$

$$Z \sim MVN(\begin{bmatrix} \beta_2 - \beta_3 \\ \beta_2 - \beta_3 \\ \beta_1 - \beta_4 \end{bmatrix}, 2\sigma^2 I)$$

d) Prove that $c_1[(z_1+z_2)^2/2+z_3^2]$ has a χ^2 distribution with degree of freedom d_1 under H_0. Find appropriate values of c_1 and d_1.
Solution:

$$z_1 + z_2 \sim N(2\beta_2 - 2\beta_3,\, 4\sigma^2)$$
$$\frac{z_1 + z_2}{\sqrt{2}} \sim N(\sqrt{2}\beta_2 - \sqrt{2}\beta_3,\, 2\sigma^2)$$
$$z_3 \sim N(\beta_1 - \beta_4, 2\sigma^2)$$

Result 5.9 If $X \sim N_p(\mu, I_p)$, then $W = X^T X = \sum_{i=1}^p X_i^2 \sim \chi_p^2(\frac{1}{2}\mu^T \mu)$

Result 5.10 If $X \sim N_p(\mu, V)$, where V is nonsingular, then $W = X^T V^{-1} X \sim \chi_p^2(\frac{1}{2}\mu^T V^{-1}\mu)$.

Pf: Since V is nonsingular, we can construct a nonsingular matrix A such that $AA^T = V$. So define $Z = A^{-1}X$, then $Z \sim N_p(A^{-1}\mu,\, A^{-1}AA^T A^{-T} = I_p)$, and using Result 5.9 we have

$$Z^T Z = X^T A^{-T} A^{-1} X = X^T V^{-1} X \sim \chi_p^2(\frac{1}{2}\mu^T A^{-T} A^{-1}\mu = \frac{1}{2}\mu^T V^{-1}\mu)$$

Using Result 5.9. If $c_1 = \frac{1}{2\sigma^2}$

$$\sqrt{c_1}\frac{z_1 + z_2}{\sqrt{2}} \sim N(\frac{\beta_2 - \beta_3}{\sigma},\, 1)$$

and

$$\sqrt{c_1} z_3 \sim N(\frac{(\beta_1 - \beta_4)}{\sqrt{2}\sigma},\, 1)$$

Then,

$$\frac{1}{2\sigma^2}[(z_1 + z_2)^2/2 + z_3^2] \sim \chi_2^2(\frac{1}{2}\frac{2(\beta_2 - \beta_3)^2 + (\beta_1 - \beta_4)^2}{2\sigma^2})$$

e) Derive the distribution of G under H_0
Solution:
$z_1 - z_2 \sim N(0,\, 4\sigma^2)$

$$Cov(z_1 + z_2,\, z_1 - z_2) = Var(z_1) - Cov(z_1, z_2) + Cov(z_2, z_1) - Var(z_2) = 0$$

$$Cov(z_1 - z_2, z_3) = Cov(z_1, z_3) - Cov(z_2, z_3) = 0$$

This implies that $(z_1 - z_2)^2$ and $(z_1 + z_2)^2/2 + z_3^2$ are independent.

$$\frac{(z_1 - z_2)^2}{2\sigma^2} \sim \chi_1^2$$

Then,

$$G = \frac{[(z_1 + z_2)^2/2 + z_3^2]/2}{(z_1 - z_2)^2} = F_{2,1}(\frac{1}{2}\frac{2(\beta_2 - \beta_3)^2 + (\beta_1 - \beta_4)^2}{2\sigma^2}) \text{ under } H_a$$

$$G \sim F_{2,1} \text{ under } H_0$$

Q4 (June 2010)

Consider the problem of classification in two classes, based on the vector $X = (X_1, ..., X_p)$ which follows a p-dimensional normal distribution. The two classes have different mean μ_1, μ_2 and common covariance matrix Σ. Assume equal misclassification costs and equal prior probabilities for the two classes.

i) As a first step, derive the form of the optimal discriminant function for this problem. Define $\mu = \frac{1}{2}(\mu_1+\mu_2)$ and show that the optimal discrimination function has the form $d(X) = \beta^T(X - \mu)$. What is the vector β?

Solution:

Linear discriminant analysis arises in the special case when we assume common covariance. Multivariate normal distribution:

$$f(x) = \frac{1}{(2\pi)^{p/2} \mid \Sigma \mid^{1/2}} exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$$

In comparing two classes 1 and 2,

$$\begin{aligned}
log\frac{P(G=1 \mid X=x)}{P(G=2 \mid X=x)} &= log\frac{f_1(x)}{f_2(x)} + log\frac{\pi_k}{\pi_l} \\
&= -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) \\
&= -\frac{1}{2}(\mu_1+\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2) + x^T\Sigma^{-1}(\mu_1-\mu_2) \\
&= (x - \frac{1}{2}(\mu_1+\mu_2))^T\Sigma^{-1}(\mu_1-\mu_2) \\
&= (\Sigma^{-1}(\mu_1-\mu_2))^T(x-\mu)
\end{aligned}$$

Thus, $\beta = \Sigma^{-1}(\mu_1 - \mu_2)$

ii) Show that if X is from class 1, the distribution of $d(X)$ is $N(\frac{1}{2}D^2,\, D^2)$, where D^2 is the Mahalanobis distance between the two classes. Similarly show that $d(X) \sim N(-\frac{1}{2}D^2, D^2)$ when X is from class 2.

Solution:

If $X \sim N(\mu_1, \Sigma)$,

$$\begin{aligned}
d(X) = \beta^T(X-\mu) &\sim N(\beta^T(\mu_1-\mu),\, \beta^T\Sigma\beta) \\
&= N(\frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2),\, (\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)) \\
&= N(\frac{1}{2}D^2, D^2)
\end{aligned}$$

Similarly, if $X \sim N(\mu_2, \Sigma)$,

$$d(X) \sim N(-\frac{1}{2}D^2, D^2)$$

iii) Show that the probability of misclassification is equal to $\Phi(-\frac{1}{2}D)$ for either class.

Solution:

The LDA rule classifies to class 1 if $d(x) = \beta^T(x-\mu) > 0$

Misclassification rate for class 1:

$$\begin{aligned}
P(\text{classify } 1 \mid \text{true class is } 2) &= P(d(x) > 0 \mid \text{true class is } 2) \\
&= P(\frac{d(x)+\frac{1}{2}D^2}{D} > \frac{\frac{1}{2}D^2}{D}) \\
&= P(Z > \frac{1}{2}D) = \Phi(-\frac{1}{2}D)
\end{aligned}$$

Similarly,

$$P(d(x) < 0 \mid \text{true class } 1) = P(\frac{d(x)-\frac{1}{2}D^2}{D} < -\frac{1}{2}D) = \Phi(-\frac{1}{2}D)$$

Q5 (June 2010)

A scientist using an apparatus of known standard deviation 0.12 takes nine independent measurements of some quantity. The measurements are assumed to be normally distributed, with the stated standard deviation and unknown mean θ, where the scientist is willing to place a vague prior distribution on θ, i.e. $\pi(\theta) \propto 1$. If the sample mean obtained is 17.653, obtain limits between which a tenth measurement will lie with 95% probability.

Solution:

$$p(\theta \mid y) \propto exp(-\frac{1}{2}\frac{n(\theta - \bar{y})^2}{\sigma^2}) \sim N(\bar{y}, \frac{\sigma^2}{9})$$

Then posterior predictive distribution is:

$$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \theta) p(\theta \mid y)\, d\theta$$

And,

$$E(\tilde{y} \mid y) = E(E(\tilde{y} \mid \theta, y) \mid y) = E(\theta \mid y) = \bar{y}$$

$$Var(\tilde{y} \mid y) = E(Var(\tilde{y} \mid \theta, y) \mid y) + Var(E(\tilde{y} \mid \theta, y) \mid y) = E(\sigma^2 \mid y) + Var(\theta \mid y) = \sigma^2 + \frac{\sigma^2}{9}$$

Then,

$$\tilde{y} \mid y \sim N(\bar{y}, \frac{10}{9}\sigma^2)$$

$$\frac{\tilde{y} \mid y - \bar{y}}{\sqrt{\frac{10}{9}\sigma^2}} \sim N(0,1) \Rightarrow P(-1.96 < \frac{\tilde{y} \mid y - \bar{y}}{\sqrt{\frac{10}{9}\sigma^2}} < 1.96) = 0.95$$

We know that $\bar{y} = 17.653$ and $\sigma = 0.12$,

$$17.4 < \tilde{y} \mid y < 17.9$$

Q1 (Jan 2011)

Assume that $X_1, ..., X_n$ is a random sample from a $N(\mu, \sigma^2)$ distribution. Both the mean and the variance of the normal distribution are unknown. Consider now the problem of testing the null hypothesis $H_0 : a \leq \mu \leq b$ versus the alternative $H_1 : \mu < a$ or $\mu > b$.

i) Derive the maximum likelihood estimates of μ, σ^2 under H_0. (Hint: Consider seperately the following situations: (1) $a \leq \bar{X} \leq b$, (2) $\bar{X} < a$ and (3) $\bar{X} > b$)

Solution:

$$
\begin{aligned}
L(\mu, \sigma^2 \mid x) &= (2\pi\sigma^2)^{-\frac{n}{2}} exp(-\frac{1}{2} \sum (x_i - \mu)^2/\sigma^2) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} exp(-\frac{1}{2\sigma^2} \sum (x_i - \bar{x})^2 + -\frac{1}{2\sigma^2} n(\bar{x} - \mu)^2) \\
logL(\mu, \sigma^2 \mid x) &= -\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2}\frac{1}{\sigma^2} \sum (x_i - \bar{x})^2 + -\frac{1}{2}\frac{1}{\sigma^2} n(\bar{x} - \mu)^2 \\
\frac{d}{d\mu} l(\mu, \sigma^2 \mid x) &= \frac{n}{\sigma^2}(\bar{x} - \mu)
\end{aligned}
$$

Case 1: $a \leq \bar{x} \leq b$

$\bar{x}$ is the maximum. $\hat{\mu} = \bar{x}$ and $\hat{\sigma} = \frac{1}{n} \sum (x_i - \bar{x})^2$

Case 2: $\bar{x} < a$

$\mu > \bar{x}$ under H_0

$$
\frac{\partial}{\partial\mu} l(\mu, \sigma^2 \mid x) = \frac{n}{\sigma^2}(\bar{x} - \mu) < 0
$$

Likelihood is decreasing in μ and is maximized at a. $\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - a)^2$

Case 3: $\bar{x} > b$

$\bar{x} > \mu$ under H_0

$$
\frac{\partial}{\partial\mu} l(\mu, \sigma^2 \mid x) = \frac{n}{\sigma^2}(\bar{x} - \mu) > 0
$$

Likelihood is increasing in μ and is maximized at b. $\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - b)^2$

ii) Derive the form of the likelihood ratio test statistic for this hypothesis.

Solution:

See questions 8.37 and 8.38 in casella berger for more reference.

Case 2: $\hat{\mu} = a$ and $\hat{\sigma}_a^2 = \frac{1}{n} \sum (x_i - a)^2$

LRT is equivalent to the t-test because $\lambda < c$ when

$$
\begin{aligned}
\lambda(x) &= \frac{(2\pi\hat{\sigma}_a^2)^{-n/2} exp(-\frac{1}{2\hat{\sigma}_a^2} \sum (x_i - a)^2)}{(2\pi\hat{\sigma}^2)^{-n/2} exp(-\frac{1}{2\hat{\sigma}^2} \sum (x_i - \bar{x})^2)} = (\frac{\hat{\sigma}^2}{\hat{\sigma}_a^2})^{\frac{n}{2}} = (\frac{\frac{1}{n} \sum (x_i - \bar{x})^2}{\frac{1}{n} \sum (x_i - a)^2})^{\frac{n}{2}} \\
&= (\frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2 + n(\bar{x} - a)^2})^{\frac{n}{2}} = (\frac{\frac{n-1}{n} s^2}{\frac{n-1}{n} s^2 + (\bar{x} - a)^2})^{\frac{n}{2}} = (\frac{\frac{n-1}{n}}{\frac{n-1}{n} + \frac{(\bar{x}-a)^2}{s^2}})^{\frac{n}{2}} < c'
\end{aligned}
$$

which is same as rejecting when $(\bar{x} - a)/(s/\sqrt{n})$ is large

Q2 (Jan 2011)

PART I Let $X_1, X_2, ..., X_n$ be independently and identically distributed random variables with finite mean μ and finite variance σ^2. Which of the following statement(s) about the sample mean ($\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$) are true by the central limit theorem (CLT)? Answer TRUE or FALSE. If false, explain why.

i) When n is large, $\bar{X} \sim N(\mu, \sigma^2/n)$

Solution: T

ii) $\bar{X} \xrightarrow{d} N(\mu, \sigma^2/n)$

Solution: F (Convergence in distribution is only define on set distribution not on moving distribution)

iii) $\sqrt{n}(\bar{X}) \xrightarrow{d} N(\mu, \sigma^2)$

Solution: F ($E(\sqrt{n}\bar{x}) = \sqrt{n}E(\bar{x})$ depends on n (is not a set value)

iii) $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$

Solution: T

Part II A statistician conducts a simulation to verify the CLT. He considers generating random variables $X_1, ..., X_n$ from the following mixture distribution:

$$f(\mu, \sigma^2) = 0.5N(-\mu, \sigma^2) + 0.5N(\mu, \sigma^2)$$

i) Show that $E(X_i) = 0$ and $Var(X_i) = \mu^2 + \sigma^2$.

Solution:

$X \mid Y \sim N(Y, \sigma^2)$

$$Y = \begin{cases} \mu & w.p. \ \frac{1}{2} \\ -\mu & w.p. \ \frac{1}{2} \end{cases}$$

$E(X) = E(E(X \mid Y)) = E(Y) = 0$ and

$$\begin{aligned} Var(X) &= E(Var(X \mid Y)) + Var(E(X \mid Y)) \\ &= E\,\sigma^2 + Var\,Y \\ &= \sigma^2 + \mu^2\frac{1}{2} + (-\mu)^2\frac{1}{2} = \sigma^2 + \mu^2 \end{aligned}$$

Now he chooses $\mu = 10$, $\sigma^2 = 10$, and $n = 100$, and conducts the simulation through the following steps.

Step 1. He generates 50 independent random numbers from $N(-10, 10)$.

Step 2. He generates 50 independent random numbers from $N(10, 10)$.

Step 3. He combines the generated random numbers from Steps 1&2 to form a sample and calculates the sample mean.

Step 4. He repeats Steps 1-3 1000 times.

By the CLT, the sample means should approximately have a normal distribution with mean 0 and variance $\frac{100+10}{100} = 1.1$ However, the statistician finds his simulated sample means actually have a variance of ≈ 0.1.

ii) Why do his sample means have a variance of ≈ 0.1?

Solution:

$Y_i \sim N(-10, 10)$ and $Z_i \sim N(10, 10)$

$$\begin{aligned} \bar{X} &= \frac{\sum_{i=1}^{50} Y_i + \sum_{i=1}^{50} Z_i}{100} \\ &= \frac{1}{100}Y_1 + ... + \frac{1}{100}Z_1 \end{aligned}$$

$Var(\frac{1}{100}Y_i) = Var(\frac{1}{100}Z_i) = \frac{10}{100^2}$ and

$$Var(\bar{x}) = \frac{100 \times 10}{100^2} = 0.1$$

Q3 (Jan 2011)

i) Show that if $y \mid \theta$ is exponentially distributed with rate θ, then the gamma prior distribution is conjugate for inferences about θ given an iid sample of y values.

Solution:

$P(y \mid \theta) = \theta e^{-\theta y}$ and $P(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}$

$$P(\theta \mid y) \propto \theta^{a+n-1}e^{-(\sum y_i + b)\theta} \sim gamma(a+n,\, b + \sum y_i)$$

ii) Show that the equivalent prior specification for the mean, $\phi = 1/\theta$, is inverse gamma. That is, you are going to derive the inverse-gamma prior by finding the equivalent prior for the mean. Show that the inverse-gamma prior is conjugate for inference about ϕ.

Solution:

$p(\phi) = \frac{b^a}{\Gamma(a)}(\frac{1}{\phi})^{a-1}e^{-b\frac{1}{\phi}} \mid \frac{-1}{\phi^2} \mid = \frac{b^a}{\Gamma(a)}(\frac{1}{\phi})^{a+1}e^{-b\frac{1}{\phi}}$

$$P(\phi \mid y) = (\frac{1}{\phi})^{n+a+1}exp(-\frac{1}{\phi}(\sum x_i + b)) \sim Inv\,Gamma(\alpha + n,\, \sum x_i + b)$$

iii) The length of life of a light bulb manufactured by a certain process has an exponential distribution with unknown rate θ. Suppose the prior distribution for θ is a gamma distribution with coefficient of variation 0.5. (The coefficient of variation is defined as the standard deviation divided by the mean). A random sample of light bulbs is to be tested and the lifetime of each obtained. If the coefficient of variation of the distribution of θ is to be reduced to 0.1, how many light bulbs need to be tested?

Solution:

Coefficient of variation

$$\frac{sd}{mean} = \frac{\sqrt{\frac{a}{b^2}}}{\frac{a}{b}} = \frac{1}{\sqrt{a}} = 0.5 \Rightarrow a = 4$$

Similarly,

$$\frac{\sqrt{\frac{a+n}{(\sum y_i + b)^2}}}{\frac{a+n}{\sum y_i + b}} = \frac{1}{\sqrt{a+n}} = 0.1 \Rightarrow n = 96$$

iv) If the coefficient of variation refers to the distribution of ϕ instead of θ, how would your answer about the number of light bulbs be changed?

$$\begin{aligned} E(\phi) &= \int_0^\infty \phi \frac{b^a}{\Gamma(a)}(\frac{1}{\phi})^{a+1}e^{-b\frac{1}{\phi}}\,d\phi \\ &= \frac{b^a}{\Gamma(a)}\frac{\Gamma(a-1)}{b^{a-1}}\int \frac{b^{a-1}}{\Gamma(a-1)}(\frac{1}{\phi})^{(a-1)+1}exp(-b\frac{1}{\phi})\,d\phi \\ &= \frac{b}{a-1} \end{aligned}$$

Also, $E(\phi^2) = \frac{b^2}{(a-1)(a-2)}$ and $Var(\phi) = \frac{b^2(a-1)}{(a-1)^2(a-2)} - \frac{b^2(a-2)}{(a-1)^2(a-2)} = \frac{b^2}{(a-1)^2(a-2)}$

$$\frac{\sqrt{\frac{b^2}{(a-1)^2(a-2)}}}{\frac{b}{a-1}} = \frac{1}{\sqrt{a-2}} = 0.5 \Rightarrow a = 6$$

And,

$$\frac{\sqrt{\frac{(b+\sum y_i)^2}{(a+n-1)^2(a+n-2)}}}{\frac{b+\sum y_i}{a+n-1}} = \frac{1}{\sqrt{a+n-2}} = 0.1 \Rightarrow n = 96$$

There is no change in the number of light bulbs.

Q4 (Jan 2011)

Consider the linear model $y_{ij} = \beta + \epsilon_{ij}$, $i = 1, 2$; $j = 1, 2$ where

$$\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \end{bmatrix} \sim N(0, \sigma_1^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}) \text{ is independent of } \begin{bmatrix} \epsilon_{21} \\ \epsilon_{22} \end{bmatrix} \sim N(0, \sigma_2^2 I)$$

$\sigma_2^2 = 2\sigma_1^2$, and $\rho = .5$.

i) Derive a simple, non-matrix formula for the best linear unbiased estimator (BLUE) of β in this model. This estimator is denoted $\hat{\beta}_{BLUE}$.

Solution:

$y = X\beta + \epsilon$, $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma_1^2 V$ where

$$V = \begin{bmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \text{ and } V^{-1} = \begin{bmatrix} \frac{1}{1-\rho^2} & -\frac{\rho}{1-\rho^2} & 0 & 0 \\ -\frac{\rho}{1-\rho^2} & \frac{1}{1-\rho^2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

$V = GG^T$ and $R = G^{-1}$. Then, $Ry = RX\beta + R\epsilon$

$$\begin{aligned} Cov(R\epsilon) &= Rcov(\epsilon)R^T = R\sigma^2 VR \\ &= \sigma^2 RGG^T R^T = \sigma^2 I \end{aligned}$$

Then,

$$\hat{\beta} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}\tilde{y} = (X^T R^T RX)^{-1} X^T R^T Ry = (X^T V^{-1} X)^{-1} X^T V^{-1} y$$

where

$$\begin{aligned} X^T V^{-1} X &= \frac{1-\rho}{1-\rho^2} + \frac{1-\rho}{1-\rho^2} + \frac{1}{2} + \frac{1}{2} = \frac{2(1-\rho)}{1-\rho^2} + 1 = \frac{2}{1+\rho} + 1 \\ (X^T V^{-1} X)^{-1} &= (\frac{2}{1+\rho} + 1)^{-1} = (\frac{3+\rho}{1+\rho})^{-1} = \frac{1+\rho}{3+\rho} = \frac{3}{7} \end{aligned}$$

Also,

$$\begin{aligned} X^T V^{-1} y &= \begin{bmatrix} \frac{1-\rho}{1-\rho^2} & \frac{1-\rho}{1-\rho^2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} \\ &= \frac{1}{1+\rho}(y_{11} + y_{12}) + \frac{1}{2}(y_{21} + y_{22}) \\ &= \frac{2}{3}(y_{11} + y_{12}) + \frac{1}{2}(y_{21} + y_{22}) \end{aligned}$$

$\hat{\beta}_{GLS} = \frac{3}{7} \times (\frac{2}{3}(y_{11} + y_{12}) + \frac{1}{2}(y_{21} + y_{22})) = \frac{2}{7}(y_{11} + y_{12}) + \frac{3}{14}(y_{21} + y_{22})$

ii) Assuming the model is correct, compute the ratio $\frac{var(\bar{y})}{var(\hat{\beta}_{BLUE})}$ where

$$\bar{y} = \frac{1}{4} \sum_{i=1}^{2} \sum_{j=1}^{2} y_{ij}$$

is the sample mean.

Solution:

$$\begin{aligned} Var(\hat{\beta}) &= Var((X^T V^{-1} X)^{-1} X^T V^{-1} y) \\ &= (X^T V^{-1} X)^{-1} X^T V^{-1} Var(y) V^{-T} X (X^T V^{-1} X)^{-T} \\ &= \sigma_1^2 (X^T V^{-1} X)^{-1} = \sigma_1^2 \frac{1+\rho}{3+\rho} = \frac{3}{7}\sigma_1^2 \end{aligned}$$

And,

$$\begin{aligned}
Var(\bar{y}) &= Var(\frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2} y_{ij}) \\
&= \frac{1}{16}(Var(y_{11}+y_{12}) + Var(y_{21}+y_{22})) \\
&= \frac{1}{16}(Var(y_{11}) + Var(y_{12}) + 2Cov(y_{11}, y_{12}) + Var(y_{21}) + Var(y_{22})) \\
&= \frac{1}{16}(2\sigma_1^2 + 2\sigma_1^2\rho + 2\sigma_1^2 + 2\sigma_1^2) = \frac{1}{16}(6\sigma_1^2 + 2\sigma_1^2\rho) = \frac{7}{16}\sigma_1^2
\end{aligned}$$

Then,

$$\frac{Var(\bar{y})}{Var(\hat{\beta}_{BLUE})} = \frac{\frac{7}{16}\sigma_1^2}{\frac{3}{7}\sigma_1^2} = \frac{49}{48}$$

iii) Now consider the linear model

$$y = X\beta + \epsilon,\ \epsilon \sim N_n(0, \sigma^2 I_n)$$

where X is $n \times (k+1)$ with $rank(X) = k+1 < n$. Suppose this model is fit to data y yielding the ordinary least squares regression parameter estimate $\hat{\beta}$ and MSE denoted by s^2. We wish to form a prediction interval for an additional observation y_0 conforming to this model. That is, $y_0 = x_0^T\beta + \epsilon_0$ where x_0 is a vector of explanatory variables corresponding to y_0 (x_0^T can be considered as an additional row of X) and $\epsilon_0 \sim N(0, \sigma^2)$ independent of ϵ. For $\hat{y}_0 = x_0^T\hat{\beta}$, show that

$$\frac{y_0 - \hat{y}_0}{s\sqrt{1 + x_0^T(X^TX)^{-1}x_0}} \sim t_{n-k-1}$$

Solution:

$$\hat{\beta} \sim N(\beta, \sigma^2(X^TX)^{-1})$$

$$x_0^T\hat{\beta} \sim N(x_0^T\beta,\ x_0^T\sigma^2(X^TX)^{-1}x_0)$$

Since $y_0 \sim N(x_0^T\beta, \sigma^2)$

$$y_0 - x_0^T\hat{\beta} \sim N(0,\ [1 + x_0^T(X^TX)^{-1}x_0]\sigma^2)$$

Then,

$$Z = \frac{y_0 - x_0^T\hat{\beta}}{\sigma\sqrt{1 + x_0^T(X^TX)^{-1}x_0}} \sim N(0,1)$$

$$V = \frac{Y^T(I-P)Y}{\sigma^2} \sim \chi^2_{n-k-1}$$

Since $X\hat{\beta} = P_XY$ and $P_X(I - P_X) = 0$, P_XY and $Y^T(I-P)Y$ are independent.

$$t = \frac{\frac{y_0 - x_0^T\hat{\beta}}{\sigma\sqrt{1 + x_0^T(X^TX)^{-1}x_0}}}{\sqrt{\frac{Y^T(I-P)Y}{\sigma^2(n-k-1)}}} = \frac{y_0 - x_0^T\hat{\beta}}{\sqrt{s^2(1 + x_0^T(X^TX)^{-1}x_0)}} = \frac{y_0 - \hat{y}_0}{s\sqrt{1 + x_0^T(X^TX)^{-1}x_0}} \sim t_{n-k-1}$$

Q1 (June 2011)

Let θ_1 denote the sensitivity and θ_2 the specificity of a diagnostic test designed to detect the presence of tuberculosis. The test was administered to n subjects drawn from a population of interest in a particular study and resulted in positive findings for m of them. Assume that only the test results are available in the study, without reference standard information on the true disease status of any subject.

Review:

		True Condition	
		Positive	Negative
Predicted Condition	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Sensitivity = TP/P = TP/(TP+FN)

Specificity = TN/N = TN/(FP+TN)

Prevalence = condition positive/ total population = (TP + FN)/(TP+FN+FP+TN)

Prevalence x Sensitivity = TP/(TP+FN+FP+TN)

Let π be prevalence, θ_1 sensitivity, θ_2 specificity

Let n be # of subjects tested and m be the observed number of positive.

Let Y_1 be the number of true positive, Y_2 be the number of false negative (This is the missing data)

$$\begin{aligned} L(n,m,Y_1,Y_2 \mid \pi,\theta_1,\theta_2) &= (\pi\theta_1)^{y_1}(\pi(1-\theta_1))^{y_2}[(1-\pi)(1-\theta_2)]^{m-y_1}[(1-\pi)\theta_2]^{n-m-y_2} \\ &= \pi^{y_1+y_2}(1-\pi)^{n-y_1-y_2}\theta_1^{y_1}(1-\theta_1)^{y_2}\theta_2^{n-m-y_2}(1-\theta_2)^{m-y_1} \end{aligned}$$

i Assume that the values of θ_1 and θ_2 are known.

a) Represent the data in the traditional table of test results cross classified by true disease status. Construct a likelihood function for the prevalence π, which includes both observed and missing data. Is a maximum likelihood estimate for π possible to construct?

Solution:

$$L(n,m,Y_1,Y_2 \mid \pi) = \pi^{y_1+y_2}(1-\pi)^{n-y_1-y_2}$$

$$l(y \mid \pi) = (y_1+y_2)log\pi + (n-y_1-y_2)log(1-\pi)$$

$$\frac{dl(y \mid \pi)}{d\pi} = \frac{y_1+y_2}{\pi} - \frac{n-y_1-y_2}{1-\pi} = 0 \Rightarrow \hat{\pi} = \frac{y_1+y_2}{n}$$

since we do not know information about y_1 and y_2, we cannot construct MLE.

b) Construct a Bayesian estimate of π. Please state carefully the assumptions you are making.

Solution:

Assuming prior for prevalence is beta

$$p(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}$$

$$p(\pi \mid y) \propto \pi^{\alpha+y_1+y_2-1}(1-\pi)^{\beta+n-y_1-y_2-1} \sim Beta(\alpha+y_1+y_2, \beta+n-y_1-y_2)$$

ii) Assume that the actual values of sensitivity and specificity are not known but beta priors with known means c_1 and c_2 respectively can be assumed on these test characteristics. Assume also a suitable prior for prevalence.

Solution:

Beta has mean $\frac{\alpha}{\alpha+\beta} = c_1 \Rightarrow \alpha = \frac{c_1}{1-c_1}\beta$

$$\theta_1 \sim beta(\alpha_1, \beta_1)$$

$$\theta_2 \sim beta(\alpha_2, \beta_2)$$

where $\alpha_1 = \frac{c_1}{1-c_1}\beta_1$

a) Derive the form of the joint posterior distribution for the parameters in the model (prevalence, sensitivity, specificity).

Solution:

$$p(\pi, \theta_1, \theta_2 \mid y) \propto \pi^{\alpha+y_1+y_2-1}(1-\pi)^{\beta+n-y_1-y_2-1}\theta_1^{\alpha_1+y_1-1}(1-\theta_1)^{\beta_1+y_2-1}\theta_2^{\alpha_2+n-m-y_2-1}(1-\theta_2)^{\beta_2+m-y_1-1}$$

b) Does the marginal distribution of each parameter have closed form?

Solution:

$$\pi \mid y \sim Beta(\alpha + y_1 + y_2,\ \beta + n - y_1 - y_2)$$
$$\theta_1 \mid y \sim Beta(\alpha_1 + y_1, \beta_1 + y_2)$$
$$\theta_2 \mid y \sim Beta(\alpha_2 + n - m - y_2, \beta_2 + m - y_1)$$

c) Describe how Bayesian estimates and posterior credible regions can be obtained for each parameter.

Solution:

Latent data y_1 and y_2 are not observed. Direct use of posterior in calculating π, θ_1, θ_2 not possible. We can use gibbs sampler.

$$y_1 \mid \theta_1, \theta_2, \pi, m \propto (\theta_1\pi)^{y_1}((1-\pi)(1-\theta_2))^{m-y_1} \sim Binomial(m,\ \frac{\pi\theta_1}{\pi\theta_1 + (1-\pi)(1-\theta_2)})$$

$$y_2 \mid \theta_1, \theta_2, \pi, n, m \sim Binomial(m-n, \frac{\pi(1-\theta_1)}{\pi(1-\theta_1) + (1-\pi)\theta_2})$$

d) If flat priors were used for the sensitivity and specificity of the test, how would this affect estimates and credible regions?

Solution:

Much wider interval estimates for posterior

iii) How would the Bayesian model of part (ii) be used to estimate the probability that two out of the next ten subjects selected from the study population will test positive for tuberculosis?

Solution:

Each iteration calculate:

$$\begin{aligned} P(screen\, positive) &= P(screen\, positive \mid disease)P(disease) + P(screen\, positive \mid no\, disease)P(no\, disease) \\ &= \theta_1\pi + (1-\theta_2)(1-\pi) \end{aligned}$$

Use binomial to calculate

$$P(x=2) = \binom{10}{2}p^2(1-p)^8$$

where $p = P(screen\, positive)$

Note: Most of this come from a paper "Bayesian Estimation of Disease Prevention and the paremeters of diagnostics tets in the absence of a gold standard" by Lawrence Joseph and et al.

Q3 (June 2011)

The Poisson regression model is a common generalized linear model for count data. Let Y represent the count, and X denote one independent variable. We are interested in how X is associated with the mean of Y. You may assume that n is large.

i) Write down the Poisson regression model with one covariate using the canonical link. Derive likelihood equations for the MLEs of your parameters.

$$exp(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi))$$

$P(Y = y) = e^{-\mu}\frac{\mu^y}{y!} = exp(ylog\mu - \mu - log\, y!)$
with canonical link

$$\theta = \eta = log(\mu) = X\beta \Rightarrow \mu = exp(X\beta)$$

$$L(\beta; y) = \prod_{i=1}^{n} exp(y_i log\mu_i - \mu_i - log\, y_i!)$$

$$l = \sum y_i x_i \beta - \sum exp(x_i\beta) - \sum log\, y_i!$$

Then

$$\frac{\partial l}{\partial \beta} = \sum y_i x_i - \sum exp(x_i\beta)x_i \stackrel{set}{=} 0$$

ii) One limitation of the Poisson model is that the variance of $Y \mid X$ must equal the mean. Count data often show over-dispersion, with the variance exceeding the mean. Can you discover from the data if there is over-dispersion? If so, describe how.

Solution:

Calculate deviance

$$\begin{aligned} D(y, \hat{\mu}) &= 2\{l(y, y) - l(y, \hat{\mu})\} \\ &= 2\sum(y_i log y_i - y_i - y_i log\hat{\mu}_i + \hat{\mu}_i) \\ &= 2\sum(y_i log\frac{y_i}{\hat{\mu}_i} + \hat{\mu}_i - y_i) \end{aligned}$$

The residual deviance has an approximate χ^2 distribution with $(n - p)$ degrees of freedom. Since the expected value of a χ^2 distribution is equal to its degree of freedom, it follows that the residual deviance for a well-fitting model should be approximately equal to its degrees of freedom. Equivalently, we may say that the mean deviance (deviance/df) should be closer to one. If the variance of the data is greater than that under poisson sampling, the residual mean deviance is likely to be greater than 1.

iii) One way of dealing with the over-dispersion is a mixture model. Suppose that $Y \mid \lambda \sim Poisson(\lambda)$. If λ has a gamma distribution, then marginally Y has a gamma-Poisson mixture distribution. The gamma distribution has density

$$f(\lambda) = \frac{\lambda^{\alpha-1}\beta^{\alpha}}{\Gamma(\alpha)} exp(-\beta\lambda)$$

where α is the shape parameter and β is the rate parameter. The mean and variance of λ are α/β and α/β^2. If the shape of the gamma distribution is fixed and does not depend on X, one can reparametrize the gamma distribution with two parameters μ and α, where $\mu = \alpha/\beta$ is the mean. What is the mean and variance of Y? For what limiting value of α does this reduce to the Poisson regression problem?

Solution:

$$E(Y) = E(E(Y \mid \lambda)) = E(\lambda) = \mu = \frac{\alpha}{\beta}$$

and

$$\begin{aligned} Var(Y) &= Var(E(Y \mid \lambda)) + E(Var(Y \mid \lambda)) \\ &= Var(\lambda) + E(\lambda) \\ &= \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2} \end{aligned}$$

We need $E(Y) = Var(Y)$ for this to be a poisson. Hence $\frac{\alpha}{\beta} \to \mu$ and $\frac{\alpha}{\beta^2} \to 0$

iv) When you replace the Poisson distribution with the gamma-Poisson mixture distribution, if other components of the original model are the same, do you still have a generalized linear model? Why or why not?

Solution:

$$\begin{aligned} P(Y = y) &= \int_0^\infty e^{-\lambda} \frac{\lambda^y}{y!} \frac{\lambda^{\alpha-1}\beta^\alpha}{\Gamma(\alpha)} exp(-\beta\lambda)\, d\lambda \\ &= \frac{\Gamma(\alpha+y)}{y!\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta+1)^{\alpha+y}} \\ &= \frac{\Gamma(\alpha+y)}{y!\Gamma(\alpha)} (\frac{\beta}{\beta+1})^\alpha (1 - \frac{\beta}{\beta+1})^y \end{aligned}$$

when α is an integer

$$P(Y = y) = \frac{(\alpha+y-1)!}{y!(\alpha-1)!} (\frac{\beta}{\beta+1})^\alpha (1 - \frac{\beta}{\beta+1})^y$$

which is negative binomial.

$$P(Y = y) = exp(ylog\frac{1}{\beta+1} + \alpha log\frac{\beta}{\beta+1} + log\frac{\Gamma(\alpha+y)}{y!\Gamma(\alpha)})$$

This is still a GLM.

v) Explain how to estimate the regression parameters.

Solution:

Calculate $S(\beta) = \frac{\partial l}{\partial \beta}$ and $S'(\beta) = \frac{\partial^2 l}{\partial \beta^2}$

$$0 = S(\beta) = S(\beta^0) + (\beta - \beta^0)S'(\beta^0) \Rightarrow \beta = \beta^0 - S'(\beta^0)^{-1}S(\beta^0)$$

Update $\beta^{(t+1)} = \beta^{(t)} - [S'(\beta^{(t)})]^{-1}S(\beta^{(t)})$ until convergence.

vi) Suppose the rate, instead of the shape, of the gamma distribution is fixed and does not depend on X. What is the mean and variance of Y? For what limiting value of α does this reduce to the Poisson regression problem? Does this change your model from (iii)? Does this change your maximum likelihood model fitting?

Solution:

In this case α is unknown and β is known. EY, VarY, limiting α are the same as (iii). This will change the model fitting since it can't be written in the GLM form.

Q4 (June 2011)

Let $Z \sim N(0,1)$. In the sequence of steps below, you will show that for any $x > 0$,

$$\frac{1}{\sqrt{2\pi}} exp(-x^2/2)(\frac{1}{x} - \frac{1}{x^3}) < P(Z > x) < \frac{1}{\sqrt{2\pi}} exp(-x^2/2)(\frac{1}{x})$$

i. Show that for any $x > 0$,

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty exp(-y^2/2)(1 - \frac{3}{y^4})\, dy = \frac{1}{\sqrt{2\pi}} exp(-x^2/2)(\frac{1}{x} - \frac{1}{x^3})$$

One way to show this is by differentiation and an application of the second fundamental law of calculus.
Solution:
First fundamental theorem of calc. $\int_a^x f(x)\, dx = F(x)$
Second fundamental theorem of calc. $\int_a^b f(x)\, dx = F(b) - F(a)$
Let $F(x) = \frac{1}{\sqrt{2\pi}} exp(-x^2/2)(\frac{1}{x} - \frac{1}{x^3})$.

$$\begin{aligned} F'(x) &= \frac{1}{\sqrt{2\pi}} exp(-x^2/2)[-\frac{1}{x^2} + \frac{3}{x^4} + (-x)(\frac{1}{x} - \frac{1}{x^3})] \\ &= \frac{1}{\sqrt{2\pi}} exp(-x^2/2)(-1 + \frac{3}{x^4}) = f(x) \end{aligned}$$

Then,

$$\int_x^\infty f(x) = 0 - \frac{1}{\sqrt{2\pi}} exp(-x^2/2)(\frac{1}{x} - \frac{1}{x^3})$$

ii) Show that for any $x > 0$,

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty exp(-y^2/2)(1 + \frac{1}{y^2})\, dy = \frac{1}{\sqrt{2\pi}} exp(-x^2/2)(\frac{1}{x})$$

Solution:
Let $F(x) = \frac{1}{\sqrt{2\pi}} exp(-x^2/2)(\frac{1}{x})$.

$$F'(x) = \frac{1}{\sqrt{2\pi}} exp(-x^2/2)((-x)\frac{1}{x} + \frac{-1}{x^2}) = \frac{1}{\sqrt{2\pi}} exp(-x^2/2)(-1 - \frac{1}{x^2}) = f(x)$$

Then,

$$\int_x^\infty f(x) = F(\infty) - F(x) = -\frac{1}{\sqrt{2\pi}} exp(-x^2/2)(\frac{1}{x})$$

iii) Use i and ii to show what we wanted to show.
Solution:

$$P(Z > x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2})\, dx > \frac{1}{\sqrt{2\pi}} \int_x^\infty exp(-\frac{y^2}{2})(1 - \frac{3}{y^4})\, dy = \frac{1}{\sqrt{2\pi}} exp(-x^2/2)(\frac{1}{x} - \frac{1}{x^3})$$

Similarly,

$$P(Z > x) < \frac{1}{\sqrt{2\pi}} \int_x^\infty exp(-y^2/2)(1 + \frac{1}{y^2})\, dy = \frac{1}{\sqrt{2\pi}} exp(-x^2/2)(\frac{1}{x})$$

iv) Finally, argue that, for some large positive number x,

$$P(Z > x) \approx \frac{1}{\sqrt{2\pi x}} exp(-x^2/2)$$

Let X_1, $X_2, ...$ be iid random variables with $E(X_i) = 0$ and $var(X_i) = \sigma^2 < \infty$. Let $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. For any positive number ϵ, define $P_{n,\epsilon} = P(\bar{X}_n > \epsilon)$. Show that, as $n \to \infty$,

$$P_{n,\epsilon} \approx \frac{\sigma}{\epsilon\sqrt{n}}\frac{1}{\sqrt{2\pi}}exp(-\frac{n\epsilon^2}{2\sigma^2})$$

Solution:
Let $g(x) = \frac{1}{\sqrt{2\pi}x}exp(-\frac{x^2}{2})$. We have shown that $g(x)(1-\frac{1}{x^2}) < P(Z > x) < g(x)$. When x is large, $1-\frac{1}{x^2}$ is close to 1, hence $P(Z > x) \approx \frac{1}{\sqrt{2\pi}}\frac{1}{x}exp(-\frac{x^2}{2})$
By CLT $\frac{\sqrt{n}(\bar{x}-0)}{\sigma} \xrightarrow{d} N(0,1)$

$$P_{n,\epsilon} = P(\bar{X}_n > \epsilon) = P(\frac{\sqrt{n}\bar{x}}{\sigma} > \frac{\sqrt{n}\epsilon}{\sigma})$$

For large n,

$$P(\bar{X}_n > \epsilon) \approx \frac{\sigma}{\epsilon\sqrt{n}}\frac{1}{\sqrt{2\pi}}exp(-\frac{n\epsilon^2}{2\sigma^2})$$

Q1 (Jan 2012)

For a random variable U_n having an asymptotic normal distribution, we write $U_n \sim AN(\mu_n, \sigma_n^2)$, which means

$$\frac{U_n - \mu_n}{\sigma_n} \xrightarrow{d} Z \sim N(0,1)$$

a) Suppose that $X_n \sim AN(\mu_n, \sigma_n^2)$ where $\mu_n \to \mu$ and $\sigma_n^2 \to 0$. Show that $X_n \xrightarrow{p} \mu$.

Solution:

$$\begin{aligned}
\lim_{n\to\infty} P(\mid X_n - \mu \mid \geq \epsilon) &\leq \frac{E(X_n - \mu)^2}{\epsilon^2} \\
&= \frac{E(X_n^2) - 2E(X_n)\mu + \mu^2}{\epsilon^2} \\
&= \frac{\sigma_n^2 + \mu_n^2 - 2\mu_n\mu + \mu^2}{\epsilon^2} \\
&\to 0 \, \mathrm{as} \, n \to \infty
\end{aligned}$$

b) Suppose that $X_n \sim AN(\mu, \sigma_n^2)$ and $S_n = a_n + b_n$ where $b_n/a_n \xrightarrow{p} 0$. If $\sigma_n/a_n \xrightarrow{p} c$ where $0 < c < \infty$, show that

$$\frac{X_n - \mu}{S_n} \sim AN(0, c^2)$$

Solution:

$$\frac{S_n}{\sigma_n} = \frac{a_n + b_n}{\sigma_n} = \frac{a_n}{\sigma_n} + \frac{b_n}{\sigma_n} = \frac{a_n}{\sigma_n} + \frac{b_n}{a_n}\frac{a_n}{\sigma_n} \xrightarrow{p} \frac{1}{c} + 0 \times \frac{1}{c} = \frac{1}{c}$$

Then,

$$\frac{X_n - \mu}{\sigma_n} = \frac{\frac{X_n - \mu}{\sigma_n}}{\frac{S_n}{\sigma_n}} \to N(0,\, c^2)$$

Q2 (Jan 2012)

a) A distribution is said to be a member of the "scaled exponential family" if its pdf takes the form

$$exp[\{y\theta - b(\theta)\}/\sigma^2 + c(y, \sigma)]$$

Assume that Y follows a lognormal distribution with $E(Y) = \mu$ and $var(Y) = \sigma^2\mu^2$. Determine whether the distribution of Y is a member of the scaled exponential family.

Solution:

log normal distribution:

$$\begin{aligned} f(y) &= \frac{1}{y\sqrt{2\pi}\sigma} exp\{-\frac{(logy - \mu)^2}{2\sigma^2}\} \\ &= \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{\mu^2}{2\sigma^2})\frac{1}{y} exp(\frac{-(logy)^2 + 2\mu log\, y}{2\sigma^2}) \end{aligned}$$

This is exponential distribution with natural parameter $(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2})$ and natural statistics $(ln^2(X), ln(X))$. Not sure if this is scaled exponential family however.

Just as a note: if Y has the normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in (0, \infty)$, then Y has moment generating function given by

$$E(e^{tY}) = exp(\mu t + \frac{1}{2}\sigma^2 t^2)$$

Hence $E(X^t) = E(e^{tY})$ and $E(X) = exp(\mu + \frac{1}{2}\sigma^2)$ and $Var(X) = exp[2(\mu + \sigma^2)] - exp(2\mu + \sigma^2)$

b) Let $Y_1, ..., Y_n$ denote a sample of n independent observations with mean μ_i and variance $var(Y_i) = \sigma^2\mu_i^2$, where σ is unknown and $log(\mu_i) = \beta_0 + \beta_1(x_i - \bar{x})$ (note that Y is not assumed to be lognormal here). In this model, x_i is a covariate of interest and $\bar{x}$ is the sample mean of the x_i's.

(i) Set up the quasi-likelihood score equation for (β_0, β_1). Show your derivation. If your equation involves first/second derivatives or integrals, please evaluate them.

Solution:

$\eta_i = log(\mu_i) = \beta_0 + \beta_1(x_i - \bar{x})$

Regular GLM:

$b'(\theta) = E(y) = \mu_i$ and $Var(y) = b''(\theta)\phi$

$$l(y_i) = \frac{y_i\theta - b(\theta)}{\phi} + c(y, \phi)$$

Then,

$$\begin{aligned} S(\beta_0) = \frac{\partial}{\partial\beta_0} l(y_i) &= \frac{\partial}{\partial\theta} l(y_i) \times \frac{1}{\frac{\partial\mu_i}{\partial\theta}} \frac{\partial\mu_i}{\partial\beta_0} \\ &= \frac{y_i - b'(\theta)}{\phi} \frac{\phi}{Var(y_i)} \mu_i \\ &= \frac{y_i - \mu}{Var(y_i)} \mu_i \end{aligned}$$

Similarly,

$$S(\beta_1) = \frac{y_i - \mu}{Var(y_i)} \mu_i (x_i - \bar{x})$$

For n observations $l(y; \beta) = \sum_{i=1}^n l_i(y; \beta)$

$$\begin{cases} \frac{\partial l(y;\beta)}{\partial\beta_0} = \sum \frac{y_i - \mu_i}{\sigma^2\mu_i^2} \mu_i = 0 \\ \frac{\partial l(y;\beta)}{\partial\beta_1} = \sum \frac{y_i - \mu_i}{\sigma^2\mu_i^2} \mu_i (x_i - \bar{x}) = 0 \end{cases}$$

b) Show that the quasi-likelihood estimates of β_0 and β_1 are uncorrelated.
Solution:

$$
\begin{aligned}
I(\beta) &= E(\frac{\partial}{\partial\beta_i} logL(\beta) \frac{\partial}{\partial\beta_j} logL(\beta)) \\
&= E \begin{bmatrix} \sum(\frac{y_i-\mu_i}{\sigma^2\mu_i^2})^2\mu_i^2 & \sum(\frac{y_i-\mu_i}{\sigma^2\mu_i^2})^2\mu_i^2(x_i-\bar{x}) \\ \sum(\frac{y_i-\mu_i}{\sigma^2\mu_i^2})^2\mu_i^2(x_i-\bar{x}) & \sum(\frac{y_i-\mu_i}{\sigma^2\mu_i^2})^2\mu_i^2(x_i-\bar{x})^2 \end{bmatrix}
\end{aligned}
$$

Since $E(\sum(\frac{y_i-\mu_i}{\sigma^2\mu_i^2})^2\mu_i^2(x_i-\bar{x})) = \frac{1}{\sigma^2\mu_i^2}\sum(x_i-\bar{x}) = 0$

c) Derive the asymptotic variances of β_0 and β_1
Solution:

$$
Var(\beta) = I(\beta)^{-1} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & (\frac{\sum(x_i-\bar{x})^2}{\sigma^2})^{-1} \end{bmatrix}
$$

Q5 (Jan 2012)

Suppose X, V_1, and V_2 are independent random variables with $X \sim N(0,\theta)$, $V_1 \sim N(0,1)$, and $V_2 \sim N(0,1)$, where $\theta \in [0,\infty)$ is an unknown parameter. Suppose we observe $Y_1 = X + V_1$ and $Y_2 = X + V_2$. Let $\hat{\theta}$ denote the maximum likelihood estimator of θ based on Y_1 and Y_2. Find $\hat{\theta}$ and describe its distribution.

Solution:

$Y_1 \sim N(0,\theta+1)$ and $Y_2 \sim N(0,\theta+1)$

$Cov(Y_1,Y_2) = Cov(X+V_1, X+V_2) = Var(X) = \theta$

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N_2(0, \begin{bmatrix} \theta+1 & \theta \\ \theta & \theta+1 \end{bmatrix})$$

Then,

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{2\pi \mid \Sigma \mid^{1/2}} exp(-\frac{1}{2}[y_1\, y_2]\Sigma^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix})$$

where $\mid \Sigma \mid = 2\theta+1$ and $\Sigma^{-1} = \frac{1}{2\theta+1}\begin{bmatrix} \theta+1 & -\theta \\ -\theta & \theta+1 \end{bmatrix}$

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{2\pi\sqrt{2\theta+1}} exp(-\frac{1}{2}\frac{1}{2\theta+1}[(y_1^2+y_2^2-2y_1y_2)\theta + y_1^2 + y_2^2])$$

$$log\, f = -log 2\pi - \frac{1}{2}log(2\theta+1) - \frac{1}{2}(y_1-y_2)^2\frac{\theta}{2\theta+1} - \frac{1}{2}\frac{y_1^2+y_2^2}{2\theta+1}$$

$$\begin{aligned} \frac{dlogf}{d\theta} &= -\frac{1}{2}\frac{2}{2\theta+1} - \frac{1}{2}(y_1-y_2)^2\frac{1}{(2\theta+1)^2} - \frac{1}{2}(y_1^2+y_2^2)\frac{-2}{(2\theta+1)^2} \\ &= -\frac{1}{2\theta+1} - \frac{1}{2}\frac{y_1^2 - 2y_1y_2 + y_2^2 - 2y_1^2 - 2y_2^2}{(2\theta+1)^2} \\ &= -\frac{1}{2\theta+1} + \frac{1}{2}\frac{(y_1+y_2)^2}{(2\theta+1)^2} = 0 \end{aligned}$$

Then,

$$\hat{\theta} = \begin{cases} \frac{(y_1+y_2)^2-2}{4} & \text{when } (y_1+y_2)^2 > 2 \\ 0 & \text{otherwise} \end{cases}$$

To find variance,

$$\frac{\partial^2 logf}{\partial^2\theta} = \frac{2}{(2\theta+1)^2} - 2\frac{(y_1+y_2)^2}{(2\theta+1)^3}$$

and

$$I(\theta) = E(-\frac{\partial^2 logf}{\partial^2\theta}) = \frac{-2}{(2\theta+1)^2} + \frac{2E(y_1+y_2)^2}{(2\theta+1)^3} = \frac{-2(2\theta+1)+2(4\theta+2)}{(2\theta+1)^3} = \frac{2(2\theta+1)}{(2\theta+1)^3}$$

since $y_1+y_2 \sim N(0, \theta+1+\theta+1+2\theta) = N(0,4\theta+2)$. Then,

$$\sqrt{n}(\hat{\theta}-\theta) \to N(0, (\frac{2}{(2\theta+1)^2})^{-1})$$

Q1 (Jan 2013)

Suppose $2m+n$ independent observations are collected from three different populations, where m observations have mean θ, another m observations have mean $\theta+\phi$, and n observations have mean $\theta-2\phi$. The observations come from a normal distribution having common variance σ^2.

a) Find the least squares estimators $\hat{\theta}$ and $\hat{\phi}$

Solution:

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & -2 \\ 1 & -2 \end{bmatrix} \text{ and compute } (X^TX)^{-1}X^Ty$$

$$X^TX = \begin{bmatrix} 2m+n & m-2n \\ m-2n & m+4n \end{bmatrix}$$

$$(X^TX)^{-1} = \frac{1}{m^2+13mn} \begin{bmatrix} m+4n & -m+2n \\ -m+2n & 2m+n \end{bmatrix}$$

b) Show that these estimates are uncorrelated if $m = 2n$

Solution:

If $m = 2n$, estimates are uncorrelated

c) Derive the LR test of the hypothesis $\phi = 0$ and state its distribution when the hypothesis is true.

Solution:

$$F = \frac{[Q(\hat{b}_H) - Q(\hat{b})]/s}{Q(\hat{b})/(N-r)}$$

which can be shown to have the $F_{s,N-r}(\lambda)$.

$$Q(\hat{b}_H) = y^T(I - P_{X_0})y = RSS(reduced)$$

$$Q(\hat{b}) = y^T(I - P_X)y = RSS(full)$$

$$Q(\hat{b}_H) - Q(\hat{b}) = y^T(P_X - P_{X_0})y$$

$P_X = X(X^TX)^{-1}X^T$ too hard to calculate

$$H_0 :\ K^Tb = [1,\ 0] \begin{bmatrix} \theta \\ \phi \end{bmatrix} = 0$$

$$K^T\hat{b} \sim N(K^T(X^TX)^{-1}X^Ty,\ \sigma^2K^T(X^TX)^{-1}K) \text{ and define } H = K^T(X^TX)^{-1}K$$

$$F = \frac{(K^T\hat{b} - m)^T(\sigma^2H)^{-1}(K^T\hat{b} - m)}{RSS/(N-r)}$$

Q3 (Jan 2013)

Let a set of possible models be $\{M_\tau : \tau \in \{1, 2, ...\}\}$, and each model M_τ has a parameter θ_τ. For example, in linear regression with p predictors, M_τ is one of 2^p possible models and θ_τ is the regression coefficients under model M_τ. This question, however, addresses a more general setting.

Suppose $X_1, ..., X_n$ are iid from a distribution with the density function $f^*(x) = f(x \mid \theta^*_{\tau^*})$. That is, these are from the true model M_{τ^*} with the true parameter $\theta^*_{\tau^*}$.

Let $\hat{\theta}_\tau(\chi_n)$ be the maximum likelihood estimator (MLE) under model M_τ using n iid samples $\chi_n = \{X_1, ..., X_n\}$. One metric of measuring the risk of $\hat{\theta}_\tau(\chi_n)$ is

$$R(\tau, n) = -\mathbb{E}_{\chi_n} \int f^*(x) log\, f(x \mid \hat{\theta}_\tau(\chi_n)) dx$$

where the expectation is over the random quantity $\hat{\theta}_\tau(\chi_n)$ which depends on the random sample χ_n. It is impossible to calculate $R(\tau, n)$ in practice because $f^*(x)$ is unknown. A quantity related to $R(\tau, n)$ which can be estimated using random subsamples is defined as follows. Define the risk based on a subsample χ_k of size k $(k \leq n-1)$ by

$$R(\tau, k) = -\mathbb{E}_{X_{k+1}} \mathbb{E}_{\chi_k} log\, f(X_{k+1} \mid \hat{\theta}_\tau(\chi_k))$$

where X_{k+1} is a sample from $f^*(x)$ not within χ_k. This question seeks an asymptotically unbiased estimator of $R(\tau, k)$.

For this question, assume the suitable regularity conditions for MLE asymptotics are satisfied. For example, it is fine to switch the order of expectations and derivatives, and the terms associated with the third order (or higher) are of samller order and thus negligible. For simplicity, we will also assume the model is true for problem 1-3 below, so $\tau = \tau^*$.

a) Show that $R(\tau, k) = F(\tau) + C(\tau, k)$ where the model lack of fit $F(\tau)$ is

$$F(\tau) = -\int f^*(x)\, log f^*(x)\, dx$$

and the estimation cost $C(\tau, k)$ is

$$C(\tau, k) = \mathbb{E}_{\chi_k} \Big[\int f^*(x) \{log\, f^*(x) - log\, f(x \mid \hat{\theta}_\tau(\chi_k))\} dx \Big]$$

Solution:

$$\begin{aligned} R(\tau, k) &= -\mathbb{E}_{X_{k+1}} \mathbb{E}_{\chi_k} log\, f(X_{k+1} \mid \hat{\theta}_\tau(\chi_k)) \\ &= -\int\int f^*(x)\, log\, f(x \mid \hat{\theta}_\tau(\chi_k))\, dx d\theta \\ &= -\int\int f^*(x)\, log\, f^*(x)\, dx d\theta + \int\int f^*(x)\, \{log\, f^*(x) - log\, f(x \mid \hat{\theta}_\tau(\chi_k))\} dx d\theta \end{aligned}$$

b) Show that the following estimator of $F(\tau)$

$$-\hat{l}(\tau)/n = -n^{-1} \sum_{i=1}^{n} log\, f(X_i \mid \hat{\theta}_\tau(\chi_n))$$

is asymptotically biased. An asymptotically unbiased estimate of $F(\tau)$ is

$$-\hat{l}(\tau)/n + p_\tau/(2n)$$

where p_τ is the number of parameters under model M_τ.

Solution:

Here is the review of related stuff.

Asymptotic efficiency of MLEs ($\hat{\theta}_{MLE}$ expanded at θ_0)

$$0 = S(\hat{\theta}) = S(\theta_0) + (\hat{\theta} - \theta_0)S'(\theta_0) + Remainder$$

Then,

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{-\sqrt{n}S(\theta_0)}{S'(\theta_0)} = \frac{-\frac{1}{\sqrt{n}}S(\theta_0)}{\frac{1}{n}S'(\theta_0)} \sim N(0, \frac{1}{I(\theta_0)})$$

Pf:

$$E(S(\theta_0)) = \int \frac{f'(x_i \mid \theta)}{f(x_i \mid \theta)} f(x_i \mid \theta)dx = \frac{\partial}{\partial \theta} \int f(x_i \mid \theta)dx = 0$$
$$Var(S(\theta_0)) = E(S^2(\theta_0)) = I(\theta_0)$$

Then,

$$\frac{1}{\sqrt{n}}S(\theta_0) = \sqrt{n}\frac{1}{n}\sum \frac{\partial log\, f(x_i \mid \theta)}{\partial \theta} \sim N(0,\, I(\theta_0))$$

and also,

$$-\frac{1}{n}S'(\theta_0) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} log\, f(x_i \mid \theta) \xrightarrow{p} I(\theta_0)$$

Asymptotic distribution of the LRT (expand $l(\theta \mid x)$ around $\hat{\theta}$)

$$l(\theta \mid x) = l(\hat{\theta} \mid x) + l'(\hat{\theta} \mid x)(\theta - \hat{\theta}) + l''(\hat{\theta} \mid x)\frac{(\theta - \hat{\theta})^2}{2!}$$

$$\begin{aligned} -2log\lambda(x) &= -l''(\hat{\theta} \mid x)(\theta - \hat{\theta})^2 \\ &\to nI(\theta_0)(\theta - \hat{\theta})^2 \to \chi_1^2 \end{aligned}$$

since $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \frac{1}{I(\theta)})$ and $l'(\hat{\theta} \mid x) = 0$

Back to question:

$$l(\theta) = l(\hat{\theta}) + (\theta - \hat{\theta})l'(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^T l''(\hat{\theta})(\theta - \hat{\theta})$$

equivalently,

$$\frac{1}{n}\sum log\, f(x_i \mid \theta) = \frac{1}{n}\sum log\, f(x_i \mid \hat{\theta}) + 0 + \frac{1}{2}(\theta - \hat{\theta})^T l''(\hat{\theta})(\theta - \hat{\theta})$$
$$F(\tau) \to -\frac{1}{n}\sum log\, f(x_i \mid \hat{\theta}) + \frac{E(\chi_{p_\tau})}{2n} = -\hat{l}(\tau)/n + p_\tau/(2n)$$

since $-E(l''(\theta)) = nI(\theta)$, $\sqrt{n}(\theta - \hat{\theta}) \to N(0,\, \frac{1}{I(\theta)})$, and $(\sqrt{nI(\theta_\tau)}(\theta - \hat{\theta}))^2 \to \chi^2_{p_\tau}$

c) Show that the following estimate, as $k \to \infty$, $p_\tau/(2k)$ is asymptotically unbiased for $C(\tau, k)$

d) Define the estimate of $R(\tau, k)$ by

$$s(\tau, k) = 2n\{-\hat{l}(\tau)/n + p_\tau/(2n) + p_\tau/(2k)\}$$

Suppose we compare two nested models, model M_i and M_j, which differ in parameter size by 1 (i.e. $p_j = p_i + 1$). When considering testing the hypothesis: $H_0 : M_i$ is true vs $H_1 : M_j$ is true, a test bsaed on $s(\tau, k)$ rejects H_0 if $s(\tau = j, k) < s(\tau = i, k)$. Find the asymptotic level of significance for this test. Express the result as a function of n and k.

Q4 (Jan 2013)

An experiment was conducted on a sample of healthy volunteers. Each subject was given one of three doses (1,2, or 3 milligrams) of a drug. Each dose group (level) consists of the same number of volunteers. The response was the change in diastolic blood pressure one hour after taking the drug.

Variables L and Q were defined as $L = -1, 0, 1$ and $Q = 1, -2, 1$ for doses $1, 2,$ and 3 mg, respectively. The following output was generated by fitting a model with constant variance and identity link. The dispersion parameter was estimated by dividing the residual deviance by its degree of freedom. Note that $D(y, \hat{\mu})/\phi \sim \chi^2(df)$, where $D(y, \hat{\mu})$ is the residual deviance, ϕ is the dispersion parameter, and df is the deviance.

Output: Deviance degree of freedom $df = 3$ and the estimated covariance matrix was diagonal.

Parameter	Estimate	Variance
Intercept	1	2
L	2	3
Q	3	1

a) How many observations (units) are there?

Solution:

$df = n - p = n - 3 = 3$. Thus, $n = 6$

b) Set up a generalized linear model for this data. Assume that y_i's $(i = 1, ...n)$ are the change in diastolic blood pressure one hour after taking the drug and $\hat{\mu}_i$'s are their fitted values based on the MLE, where n denotes the number of units in part (a). Express the scaled deviance $D(y, \hat{\mu})/\phi$ and prove that it is chi-square distribution.

Solution:

Log Likelihood

$$l(\theta, \phi) = \sum_{i=1}^{n} \{(y_i\theta_i - b(\theta_i))/a(\phi) + c(y_i, \phi)\} = \sum_{i=1}^{n} \{(y_i\theta_i - \frac{1}{2}\theta^2)/\phi + c(y_i, \phi)\}$$

with identity link $\theta_i = x_i\beta$ **This is normal

$$\theta = \mu = b'(\theta) \to b(\theta) = \frac{1}{2}\theta^2$$

With constant variance assumption, $a(\phi) = \phi$. In normal case, $a(\phi) = \sigma^2$

Null Deviance = -2(LL(Null Model) - LL(Saturated Model)) on $df = df_{Sat} - df_{Null}$

Residual Deviance = -2(LL(Proposed Model) - LL(Saturated Model)) on $df = df_{Sat} - df_{res}$

$$\begin{aligned}
D(y, \hat{\mu})/\phi =& -2(l(\hat{\mu}) - l(\tilde{\mu})) \\
=& -2(\sum_{i=1}^{n}(y_i\hat{\mu} - \frac{1}{2}\hat{\mu}^2)/\phi + c(y_i, \phi) - \sum_{i=1}^{n}(y_i^2 - \frac{1}{2}y_i^2)/\phi + c(y_i, \phi)) \\
=& \sum_{i=1}^{n}(-2y_i\hat{\mu} + \hat{\mu}^2 + y_i^2)/\phi = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu})^2}{\phi} \\
\frac{D(y, \hat{\mu})}{\phi} =& \frac{1}{\sigma^2}\sum(y_i - \hat{\mu})^2 = \frac{1}{\sigma^2}y^T(I - P)y \sim \chi_3^2
\end{aligned}$$

c) Calculate the residual deviance for this fitted model based on the above output. Your final answer should be a number.

$(X^TX)^{-1} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 12 \end{bmatrix}$ and since $Var(\hat{\beta}) = \hat{\sigma}^2(X^TX)^{-1} \Rightarrow \hat{\sigma}^2 = 12 = \frac{1}{n-p}\sum(y_i - X\beta)^2$

Then $D(y, \hat{\mu}) = 36$

Q5 (Jan 2013)

Part 1

Suppose that Y_i, $i \in \{1, ..., n\}$, are iid normal random variables with mean μ and variance σ^2, where Y_i for $i = 1, ...r$ are observed and Y_i for $i = r+1, ..., n$ are missing. Also assume that the missing Y_is are missing at random.

1a) We are interested in finding the maximum likelihood for μ and σ^2 using the EM algorithm. Derive the two steps of the EM algorithm.

Solution:

$$L_C(\mu, \sigma \mid Y) = (2\pi\sigma^2)^{-\frac{n}{2}} exp(-\frac{1}{2\sigma^2}(\sum y_i^2 - 2\mu \sum y_i + n\mu^2))$$

$$logL_C(\mu, \sigma \mid Y) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{\sum y_i^2}{2\sigma^2} + \frac{\mu \sum y_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}$$

We have $E_{\mu^{(t)},\sigma^{2(t)}}(Y_j \mid Y_i) = \mu^{(t)}$ and $E_{\mu^{(t)},\sigma^{2(t)}}(Y_j^2 \mid Y_i) = \mu^{(t)2} + \sigma^{2(t)}$ where $i = 1, ..., r$ and $j = r+1, ..., n$

$$E_{\mu^{(t)},\sigma^{2(t)}}(logL_c \mid Y_{obs}) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{\sum y_i^2}{2\sigma^2} - \frac{(n-r)(\sigma^{2(t)} + \mu^{(t)2})}{2\sigma^2} + \frac{\mu \sum y_i}{\sigma^2} + \frac{\mu(n-r)\mu^{(t)}}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}$$

$$\frac{dE_{\mu^{(t)},\sigma^{2(t)}}(logL_c \mid Y_{obs})}{d\mu} = \frac{\sum y_i}{\sigma^2} + \frac{(n-r)\mu^{(t)}}{\sigma^2} - \frac{2n\mu}{2\sigma^2} = \frac{\sum y_i + (n-r)\mu^{(t)} - n\mu}{\sigma^2} = 0$$

Then,

$$\mu^{(t+1)} = \frac{\sum_{i=1}^r y_i + (n-r)\mu^{(t)}}{n}$$

Similarly,

$$\frac{dE_{\mu^{(t)},\sigma^{2(t)}}(logL_c \mid Y_{obs})}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum y_i^2 + (n-r)(\sigma^{2(t)} + \mu^{(t)2})}{2\sigma^4} - \frac{\mu \sum y_i + \mu(n-r)\mu^{(t)}}{\sigma^4} + \frac{n\mu^2}{2\sigma^4} = 0$$

$$-n\sigma^2 + \sum y_i^2 + (n-r)(\sigma^{2(t)} + \mu^{(t)2}) - 2n\mu^{(t+1)2} + n\mu^{(t+1)2} = 0$$

$$\sigma^{2(t+1)} = \frac{1}{n}(\sum y_i^2 + (n-r)(\sigma^{2(t)} + \mu^{(t)2})) - \mu^{(t+1)2}$$

1b) The problem is common in classical statistical analysis, and can be thought of as a sample of size r from a normal distribution. Derive the MLEs of μ and σ^2 in this case directly from the likelihood function. Compare these estimators to the ones obtained from the EM algorithm in part a). Does the EM algorithm converge to the same point estimates?

Solution:

The limit of $\mu^{(t)}$ will satisfy

$$\mu = \frac{\sum_{i=1}^r y_i + (n-r)\mu}{n} \Rightarrow \hat{\mu}_{MLE} = \frac{\sum_{i=1}^r y_i}{r}$$

For $\sigma^{2(t)}$

$$n\sigma^2 = \sum\{y_i^2 + (n-r)(\sigma^2 + \hat{\mu}^2)\} - n\hat{\mu}^2$$

$$r\sigma^2 = \sum y_i^2 - r\hat{\mu} = \sum_{i=1}^r (y_i - \bar{y})^2 \Rightarrow \hat{\sigma}^2_{MLE} = \frac{1}{r}\sum_{i=1}^r (y_i - \bar{y})^2$$

Part 2

Assume $Y \sim Binomial(n, \theta)$ and that $\theta \sim Beta(\alpha, \beta)$.
a) Find the Bayes Rule for the Squared Error Loss. For $\frac{\alpha}{\alpha+\beta} = 0.5$ identify the values for which the risk of the Bayes rule is lower than the risk of MLE.
Solution:
Posterior for θ is

$$P(\theta \mid Y) \propto Beta(Y + \alpha,\, n - Y + \beta)$$

Under squared error loss, bayes rule is

$$\delta^{\pi}(X) = E(\theta \mid Y) = \frac{Y + \alpha}{n + \alpha + \beta}$$

Then

$$\begin{aligned}
E(\delta^{\pi}(X) - \theta)^2 &= Var_{\theta}\delta(X) + (E_{\theta}\delta(X) - \theta)^2 \\
&= Var(\frac{Y+\alpha}{n+\alpha+\beta}) + (E(\frac{Y+\alpha}{n+\alpha+\beta}) - \theta)^2 \\
&= \frac{Var(Y)}{(n+\alpha+\beta)^2} + (\frac{n\theta+\alpha}{n+\alpha+\beta} - \frac{n\theta+\alpha\theta+\beta\theta}{n+\alpha+\beta})^2 \\
&= \frac{n\theta(1-\theta)}{(n+\alpha+\beta)^2} + \frac{(\alpha(1-\theta)-\beta\theta)^2}{(n+\alpha+\beta)^2} \\
&= \frac{Var(Y)}{(n+\alpha+\beta)^2} + \frac{(2\beta-3\beta\theta)^2}{(n+\alpha+\beta)^2} \text{ since } \alpha = 2\beta
\end{aligned}$$

Risk of MLE is

$$R(\delta_{MLE}, \theta) = \frac{1}{n^2}Var(Y) = \frac{n\theta(1-\theta)}{n^2}$$

2b) Find the Bayes Rule for the loss function $L(\theta, a) = \frac{(a-\theta)^2}{\theta(1-\theta)}$
Solution:
Jun Shao 2nd Ex 4.9

$$\begin{aligned}
E(W(\theta)(\theta-a)^2) &= \int W(\theta)\theta^2\pi(\theta \mid x)d\theta - \int 2aW(\theta)\theta\pi(\theta \mid x)d\theta + \int a^2W(\theta)\pi(\theta \mid x)d\theta \\
\frac{dE(W(\theta)(\theta-a)^2)}{da} &= -\int 2W(\theta)\theta\pi(\theta \mid x)d\theta + \int 2aW(\theta)\pi(\theta \mid x)d\theta = 0 \\
\delta(x) &= \frac{\int \theta W(\theta)\pi(\theta \mid x)d\theta}{\int W(\theta)\pi(\theta \mid x)d\theta}
\end{aligned}$$

Let $W(\theta) = \frac{1}{\theta(1-\theta)}$. Then,

$$\begin{aligned}
\delta(x) &= \frac{\int \frac{1}{\theta(1-\theta)}\theta\frac{\Gamma(n+\alpha+\beta)}{\Gamma(Y+\alpha)\Gamma(n-Y+\beta)}\theta^{Y+\alpha-1}(1-\theta)^{n-Y+\beta-1}d\theta}{\int \frac{1}{\theta(1-\theta)}\frac{\Gamma(n+\alpha+\beta)}{\Gamma(Y+\alpha)\Gamma(n-Y+\beta)}\theta^{Y+\alpha-1}(1-\theta)^{n-Y+\beta-1}d\theta} \\
&= \frac{\int \theta^{Y+\alpha-1}(1-\theta)^{n-Y+\beta-1-1}d\theta}{\int \theta^{Y+\alpha-1-1}(1-\theta)^{n-Y+\beta-1-1}d\theta} \\
&= \frac{\frac{\Gamma(Y+\alpha)\Gamma(n-Y+\beta-1)}{\Gamma(n+\alpha+\beta-1)}}{\frac{\Gamma(Y+\alpha-1)\Gamma(n-Y+\beta-1)}{\Gamma(n+\alpha+\beta-2)}} = \frac{Y+\alpha-1}{n+\alpha+\beta-2}
\end{aligned}$$

Q3 (June 2013)

We wish to estimate the parameter, θ, of an exponential distribution. We consider that $X_1,...,X_n$ are exchangeable with $X_i \mid \theta \sim Exp(\theta)$ so that $E(X_i \mid \theta) = 1/\theta$

a) Show that $Gamma(\alpha, \beta)$ prior distribution is conjugate.

$$p(\theta \mid x) \sim Gamma(\alpha + n,\, \beta + \sum x_i)$$

b) We wish to produce an estimate, d, for θ, with loss function

$$L(\theta, d) = \frac{(\theta - d)^2}{\theta^3}$$

Find the Bayes rule and Bayes risk assuming no data was observed, and using the conjugate prior distribution. (assuming that $\alpha > 3$)

Solution:

Minimize posterior expected loss

$$\int L(\theta, \delta)\pi(\theta \mid x)d\theta = \int_0^\infty \frac{(\theta - a)^2}{\theta^3}\pi(\theta \mid x)d\theta$$

$$0 = \int \frac{-2\theta + 2a}{\theta^3}\pi(\theta \mid x)d\theta$$

$$\int \frac{\theta}{\theta^3}\pi(\theta \mid x)d\theta = \int \frac{a}{\theta^3}\pi(\theta \mid x)d\theta$$

Then,

$$a = \frac{\int \frac{1}{\theta^2}\pi(\theta \mid x)d\theta}{\int \frac{1}{\theta^3}\pi(\theta \mid x)d\theta}$$

$$a = \frac{\frac{(\beta+\sum x_i)^{\alpha+n}}{\Gamma(\alpha+n)} \int \theta^{\alpha+n-2-1} e^{-(\beta+\sum x_i)\theta} d\theta}{\frac{(\beta+\sum x_i)^{\alpha+n}}{\Gamma(\alpha+n)} \int \theta^{\alpha+n-3-1} e^{-(\beta+\sum x_i)\theta} d\theta} = \frac{\frac{\Gamma(a+n-2)}{(\beta+\sum x_i)^{\alpha+n-2}}}{\frac{\Gamma(a+n-3)}{(\beta+\sum x_i)^{\alpha+n-3}}} = \frac{\alpha + n - 3}{\beta + \sum x_i}$$

$$\delta_{Bayes}(x) = \frac{\alpha + n - 3}{\beta + \sum x_i} \text{ and } \delta_{Bayes}(x) = \frac{\alpha - 3}{\beta} \text{ assuming no data}$$

Q4 (June 2013)

The coefficient of variation (CV) is defined as the ratio of the standard deviation to the mean of a distribution. This problem addressed the asymptotic properties of an estimator of CV.

a) If $\mu \neq 0$, show that

$$\frac{s_n}{\bar{X}_n} \to \frac{\sigma}{\mu}$$

and $\frac{s_n}{\bar{X}_n}$ is asymptotically normal with mean $\frac{\sigma}{\mu}$ and variance $\frac{1}{n}[\frac{\sigma^2\mu_2}{\mu_4} - \frac{\mu_3}{\mu^3} + \frac{\mu_4 - \mu_2^2}{4\mu^2\sigma^2}]$

Solution:

Example 5.33 BS. Consider the "1/n" version of the sample variance $s_n^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Then,

$$s_n^2 - \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\{(X_i - \mu)^2 - \sigma^2\} - (\bar{X} - \mu)^2$$

and $h(X_i) = [(X_i - \mu)^2 - \sigma^2]$,

$$\sqrt{n}(s_n^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4)$$

Example 5.36 BS. Suppose that $X_1, ..., X_n$ are iid (μ, σ^2) and $\mu_4 = E(X_1 - \mu)^4$ exists,

$$\sqrt{n}(\bar{X} - \mu, s_n^2 - \sigma^2) \xrightarrow{d} BN(0, \Sigma) \, \text{as} \, n \to \infty$$

where $\Sigma_{11} = \sigma^2$, $\Sigma_{12} = \mu_3$, and $\Sigma_{22} = \mu_4 - \sigma^4$

$g(\theta) = g(\mu, \sigma^2) = \frac{\sqrt{\sigma^2}}{\mu}$ and $g'(\theta) = \begin{bmatrix} -\frac{\sigma}{\mu^2} & \frac{1}{2\mu\sigma} \end{bmatrix}$

$$\sqrt{n}(\frac{s_n}{\bar{X}_n} - \frac{\sigma}{\mu}) \to AN(0, \begin{bmatrix} -\frac{\sigma}{\mu^2} & \frac{1}{2\mu\sigma} \end{bmatrix} \begin{bmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{bmatrix} \begin{bmatrix} -\frac{\sigma}{\mu^2} \\ \frac{1}{2\mu\sigma} \end{bmatrix})$$

$$\begin{bmatrix} -\frac{\sigma}{\mu^2} & \frac{1}{2\mu\sigma} \end{bmatrix} \begin{bmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{bmatrix} = \begin{bmatrix} -\frac{\sigma^3}{\mu^2} + \frac{\mu_3}{2\mu\sigma} & -\frac{\sigma\mu_3}{\mu^2} + \frac{\mu_4 - \sigma^4}{2\mu\sigma} \end{bmatrix}$$

$$\begin{aligned} \begin{bmatrix} -\frac{\sigma^3}{\mu^2} + \frac{\mu_3}{2\mu\sigma} & -\frac{\sigma\mu_3}{\mu^2} + \frac{\mu_4 - \sigma^4}{2\mu\sigma} \end{bmatrix} \begin{bmatrix} -\frac{\sigma}{\mu^2} \\ \frac{1}{2\mu\sigma} \end{bmatrix} &= -\frac{\sigma}{\mu^2}(-\frac{\sigma^3}{\mu^2} + \frac{\mu_3}{2\mu\sigma}) + \frac{1}{2\mu\sigma}(-\frac{\sigma\mu_3}{\mu^2} + \frac{\mu_4 - \sigma^4}{2\mu\sigma}) \\ &= \frac{\sigma^4}{\mu^4} - \frac{\mu_3}{\mu^3} + \frac{\mu_4 - \sigma^4}{4\mu^2\sigma^2} \end{aligned}$$

b) If $\mu = 0$ then show that

$$\frac{1}{\sqrt{n}}\frac{s_n}{\bar{X}_n} \to \frac{1}{N(0,1)}$$

Solution:

Delta method cannot be used for $\frac{s_n}{\bar{X}_n}$ at $\mu = 0$ since the derivative does not exist. We need to use continuity theorem.

$$\frac{\sqrt{n}\bar{X}_n}{s_n} \to N(0,1)$$

Q5 (June 2013)

Let Y be an ordinal response variable with J categories. Let $\pi_j(x) = P(Y = j \mid x)$ at a fixed setting x. Consider the adjacent categories logic model

$$L_j = log(\pi_j(x)/\pi_{j+1}(x)) = \alpha_j + \beta^T x$$

where β is a $p \times 1$ vector.

a) L_j is equal to logit of what probability?

Solution:

$$log\frac{x}{1-x} = log\frac{\pi_j}{\pi_{j+1}} \Rightarrow x = \frac{\pi_j}{\pi_j + \pi_{j+1}} = \frac{P(Y = j \mid x)}{P(Y = j \,\text{or}\, Y = j+1 \mid x)}$$

b) How is this model related to the baseline-category logit model? Express the parameters in Model (α_j and β) using the parameters in the equivalent baseline-category model.

Solution:

$$log\frac{\pi_j}{\pi_1} = \alpha_j^* + \beta_j^{*T} x \,\text{and}\, log\frac{\pi_{j+1}}{\pi_1} = \alpha_{j+1}^* + \beta_{j+1}^{*T} x \,\text{with}\, \alpha_J^* = \beta_J^* = 0$$

$$log\frac{\pi_j}{\pi_{j+1}} = (\beta_j^* - \beta_{j+1}^*)^T x + (\alpha_j^* - \alpha_{j+1}^*)$$

Adjacent Model

$$log\frac{\pi_j}{\pi_{j+1}} = \alpha_j + \beta^T x$$

Compare two models

$$(\beta_j^* - \beta_{j+1}^*)^T x + (\alpha_j^* - \alpha_{j+1}^*) = \alpha_j + \beta^T x$$

Then we find $\alpha_j = \alpha_j^* - \alpha_{j+1}^*$ and $\beta = \beta_j^* - \beta_{j+1}^*$

c) Based on your answer, can we use software to fit the baseline category logit model to fit the adjacent category logit model?

Solution:

$$\begin{cases} \alpha_J^* = 0 \\ \alpha_{J-1}^* - \alpha_J^* = \alpha_{J-1} \\ \alpha_1^* - \alpha_2^* = \alpha_1 \end{cases} \Rightarrow \begin{cases} \alpha_1^* = \alpha_1 + ... + \alpha_{J-1} \\ \alpha_2^* = \alpha_2 + ... + \alpha_{J-1} \\ \alpha_J^* = 0 \end{cases}$$

By $\beta = \beta_j^* - \beta_{j+1}^* \forall j = 1, ..., J$

$$\begin{cases} \beta_J^* = 0 \\ \beta_{J-1}^* = \beta \\ \beta_{J-2}^* = 2\beta \\ \beta_1^* = (J-1)\beta \end{cases}$$

$\therefore log(\frac{\pi_j}{\pi_J}) = \alpha_j + ... + \alpha_{J-1} + \beta^T (J-j)x$

Q1 (June 2014)

The truncated binomial distribution arises naturally in genetic studies and other settings. Formally, a random variable X has the truncated binomial distribution if $P(X = x) = P(Y = x \mid Y \geq 1)$ for $Y \sim Binom(n, \theta)$. Assume that $n > 1$.

a) Derive the density of X

Solution:

$$P(Y \leq y \mid Y \geq 1) = \frac{P(0 < Y \leq y)}{P(Y \geq 1)} = \frac{F(y; \lambda) - F(0; \lambda)}{1 - P(Y = 0)}$$

Taking derivative

$$f(y \mid y > 0) = \frac{P(Y = y)}{1 - P(Y = 0)} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{1-(1-\theta)^n}$$

$$E(x) = \frac{n\theta}{1-(1-\theta)^n}$$

b) Show that X is a complete and sufficient statistic for θ and derive the expectation of X.

Solution:

We can write in exponential family and x is sufficient and complete.

$$f(x \mid \theta) = \binom{n}{x}\frac{(1-\theta)^n}{1-(1-\theta)^n} exp(xlog\frac{\theta}{1-\theta})$$

where $h(x) = \binom{n}{x}$, $c(\theta) = \frac{(1-\theta)^n}{1-(1-\theta)^n}$, $w_1(\theta) = log\frac{\theta}{1-\theta}$ and $t_1(x) = x$

c) Find a uniformly minimum variance unbiased estimator (UMVUE) of $q(\theta) = \theta/(1-(1-\theta)^n)$. Show that it is the UMVUE.

Solution:

Theorem 7.3.23 Casella Berger: Let T be a complete sufficient statistic for a parameter θ, and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value.

$E(x) = \frac{n\theta}{1-(1-\theta)^n}$ and $\phi(T) = \frac{x}{n}$ is the UMVUE of $\frac{\theta}{1-(1-\theta)^n}$

Q3 (June 2014)

The geometric probability distribution can be interpreted as the distribution of the number of iid Bernoulli trials observed until a first success. Its probability mass function is

$$f(y \mid \pi) = \pi^y(1-\pi),\ y = 0, 1, 2, ...$$

The distribution is a member of the exponential family which takes the general form

$$f(y \mid \theta, \phi) = exp\{(y\theta - b(\theta))/a(\phi) + c(y, \phi)\}$$

a) Show that the geometric distribution is a member of the exponential family of distributions. Identify θ, ϕ, $a(\phi)$, $b(\theta)$, and $c(y, \phi)$

Solution:

$$f(y \mid \pi) = exp\{ylog\pi + log(1-\pi)\}$$

where $\theta = log\pi$, $a(\phi) = 1$, $c(y, \phi) = 0$, $b(\theta) = -log(1-\pi) = -log(1-e^\theta)$

b) Express the common expression of mean and variance of exponential family distribution and then derive $E(Y)$ and $V(Y)$. What's the canonical link?

Solution:

$E(Y) = b'(\theta) = \frac{e^\theta}{1-e^\theta}$ and $Var(Y) = b''(\theta)a(\phi) = \frac{(1-e^\theta)e^\theta + e^\theta e^\theta}{(1-e^\theta)^2} = \frac{e^\theta}{(1-e^\theta)^2}$

Canonical link: $\theta = g(\mu) = log\frac{\mu}{1+\mu}$

Q4 (June 2014)

Gender	Cholesterol	Yes Heart Disease	No Heart disease	Total
Male	High	16	256	272
	Low	28	2897	2925
Female	High	13	319	332
	Low	23	2565	2588
	Total	80	6037	6117

a) Define the odds ratio of having a heart disease for male individuals with high and low Cholesterol.

Solution:

$$\frac{16/256}{28/2897} = 6.47$$

b) Estimate the odds ratio above as well as the asscoiated 95% confidence interval. Use $Z_{\alpha/2} = 1.96$. Interpret the result.

Solution:

Consider $log\frac{\hat{p}_1}{1-\hat{p}_1} - log\frac{\hat{p}_2}{1-\hat{p}_2} = 1.867$

By delta method, its asymptotic variance is $\frac{1}{n_1\hat{p}_1(1-\hat{p}_1)} + \frac{1}{n_2\hat{p}_2(1-\hat{p}_2)} = \frac{1}{272\times\frac{16}{272}\times\frac{256}{272}} + \frac{1}{2925\times\frac{28}{2925}\times\frac{2897}{2925}} = 0.102$

95% CI for log OR is $(1.867 - 1.96 \times \sqrt{0.102},\ 1.867 + 1.96\sqrt{0.102}) = (1.241,\ 2.493)$

95% CI for OR is (3.459, 12.097)

c) State the log-linear model that only expresses the main effects of the three characteristics on the expected counts. Interpret the assumption of the model, and compute the fitted values in the top left count of the table (male, high cholesterol, with the disease) according to the model.

Solution:

Let π_{ijk} be the probability of one of individual to be in ijk-th cell. With n total sample size, we have $\mu_{ijk} = n\pi_{ijk}$. If the three variables with 2 categories respectively are independent, $\pi_{ijk} = \pi_{i++}\times\pi_{+j+}\times\pi_{++k}$. We have

$$log\mu_{ijk} = logn + log\pi_{i++} + log\pi_{+j+} + log\pi_{++k} = logn + \lambda_i^G + \lambda_j^C + \lambda_k^D$$

Then,

$$\hat{\mu}_{111} = n\pi_{111} = n \times \frac{n_{1++}}{n} \times \frac{n_{+1+}}{n} \times \frac{n_{++1}}{n} = \frac{(272+2925)\times(272+332)\times 80}{6117^2} = 4.129$$

d) We would like to conduct the deviance goodness-of-fit test for the model in (c). Derive the deviance test statistic. State the null and the alternative hypothesis, the formula for the test statistic and the decision rule at the confidence level of 95%.

Solution:

$$D(y,\hat{\mu}) = 2\sum_i\sum_j\sum_k\{n_{ijk}log(n_{ijk}/\hat{\mu}_{ijk}) - (n_{ijk} - \hat{\mu}_{ijk})\} = 2\sum_i\sum_j\sum_k\{n_{ijk}log(n_{ijk}/\hat{\mu}_{ijk})\}$$

Our hypothesis is

$$H_0 : \lambda_i^G = \lambda_j^C = \lambda_k^D = 0 \text{ and } H_a : \text{At least one } \lambda \neq 0$$

$D(y,\hat{\mu}) \sim \chi^2_{2\times2\times2-(1+(2-1)+(2-1)+(2-1))} = \chi^2_4$

e) Compute the odds ratio for association between cholesterol and heart disease sepeartely for men and women. Compute the marginal odds ratio for association between cholesterol and heart disease. Does the magnitude of the marginal odds ratio lie between the two conditional odds ratio?

Solution:

Men: $\frac{16/256}{28/2897} = 6.47$, Women: $\frac{13/319}{23/2565} = 4.545$, Marginal: $\frac{(16+13)/(256+319)}{(28+23)/(2897+2565)} = 5.401$

Q5 (June 2014)

Consider $J(\geq 2)$ independent primary care doctor, for each primary care doctor j we estimate the average BMI for his patients θ_j with the data $\{y_{ij},\ i = 1, ..., n\}$, which are modeled to be an i.i.d. sample from $N(\theta_j, \sigma^2)$. The average BMI $\theta = \{\theta_1, ..., \theta_J\}$ themselves are modeled as i.i.d. sample from $N(\mu, \tau^2)$. Our goal here is to make inference about θ based on the data $Y = \{y_{ij},\ i = 1, ..., n,\ j = 1, ..., J\}$. For simplicity, here we assume that σ^2 is known. However, $\{\mu, \tau^2\}$ are unknown and thus need to be estimated from Y as well.

a) Find posterior distribution for θ_j

Solution:

We have $p(\theta_j) \propto exp(-\frac{(\theta_j-\mu)^2}{2\tau^2})$ and $p(y_{ij} \mid \theta_j) \propto exp(-\frac{(y_{ij}-\theta_j)^2}{2\sigma^2})$

$$\begin{aligned}
p(\theta_j \mid y) &\propto exp(-\frac{(\theta_j-\mu)^2}{2\tau^2}) \prod_{i=1}^{n} exp(-\frac{(y_{ij}-\theta_j)^2}{2\sigma^2}) \\
&\propto exp(-\frac{1}{2}(\frac{\theta_j^2 - 2\mu\theta_j}{\tau^2} + \frac{n\theta_j^2 - 2\sum_{i=1}^{n} y_{ij}\theta_j}{\sigma^2})) \\
&= exp(-\frac{1}{2}\frac{(\sigma^2 + n\tau^2)\theta_j^2 - 2(\mu\sigma^2 + n\bar{y}_{\cdot j}\tau^2)\theta_j}{\sigma^2\tau^2}) \\
&\propto exp(-\frac{1}{2}\frac{(\theta_j - \frac{\mu\sigma^2+n\tau^2\bar{y}_{\cdot j}}{\sigma^2+n\tau^2})^2}{\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}})
\end{aligned}$$

Hence $\theta_j \sim N(\frac{\mu\sigma^2+n\tau^2\bar{y}_{\cdot j}}{\sigma^2+n\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+n\tau^2})$

Intuitively, $E(\theta_j \mid Y, \mu, \tau^2)$ is a weighted sum of $\bar{y}_{\cdot j}$ and μ, with weights $\frac{n}{\sigma^2}$ and $\frac{1}{\tau^2}$. $Var(\theta_j \mid Y, \mu, \tau^2)$ is the inverse of the combined precision.

b) Since $\{\mu, \tau^2\}$ are unknown, we cannot use the results in (a) to make inference about θ. One method that can be used in practice is to estimate $\{\mu, \tau^2\}$ by

$$\hat{\mu} = \bar{y}_{\cdot\cdot} \text{ and } \hat{\tau}^2 = \frac{MS_B - MS_W}{n}$$

where $\bar{y}_{\cdot\cdot} = \frac{\sum_i \sum_j y_{ij}}{nJ}$ and

$$MS_B = \frac{n\sum_j(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2}{J-1} \text{ and } MS_W = \frac{\sum_i \sum_j (y_{ij} - \bar{y}_{\cdot j})^2}{J(n-1)}$$

are the between-group and within-group mean squares. Show that $\hat{\mu}$ and $\hat{\tau}^2$ are unbiased.

Solution:

$E(Y_{ij}) = E(E(Y_{ij} \mid \theta)) = E(\theta_j) = \mu$

$$\hat{\mu} = E(\bar{y}_{\cdot\cdot}) = \frac{\sum_i \sum_j E(y_{ij})}{nJ} = \frac{nJ\mu}{nJ} = \mu$$

Also, $Var(\bar{y}_{\cdot j}) = E(Var(\bar{y}_{\cdot j} \mid \theta)) + Var(E(\bar{y}_{\cdot j} \mid \theta)) = \frac{\sigma^2}{n} + \tau^2$ and $Var(\bar{y}_{\cdot\cdot}) = \frac{\sum_j Var(\bar{y}_{\cdot j})}{J} = \frac{1}{J}(\frac{\sigma^2}{n} + \tau^2)$

$$\begin{aligned}
E\,MS_B &= \frac{n}{J-1} E \sum_j (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2 \\
&= \frac{n}{J-1} E \{\sum_j (\bar{y}_{\cdot j} - \mu)^2 - J(\bar{y}_{\cdot\cdot} - \mu)^2\} \\
&= \frac{n}{J-1} \sum_j E(\bar{y}_{\cdot j} - \mu)^2 - \frac{nJ}{J-1} E(\bar{y}_{\cdot\cdot} - \mu)^2 \\
&= \frac{n}{J-1} \sum_j Var(\bar{y}_{\cdot j}) - \frac{nJ}{J-1} Var(\bar{y}_{\cdot\cdot}) \\
&= \frac{nJ}{J-1}(\frac{\sigma^2}{n} + \tau^2) - \frac{nJ}{J-1}(\frac{1}{J}(\frac{\sigma^2}{n} + \tau^2)) \\
&= \frac{nJ}{J-1} \frac{J-1}{J} (\frac{\sigma^2}{n} + \tau^2) \\
&= \sigma^2 + n\tau^2
\end{aligned}$$

Also, $Var(y_{ij}) = Var(E(y_{ij} \mid \theta)) + E(Var(y_{ij} \mid \theta)) = \tau^2 + \sigma^2$

$$\begin{aligned}
E\,MS_W &= \frac{1}{J(n-1)} E \sum_i \sum_j (y_{ij} - \bar{y}_{\cdot j})^2 \\
&= \frac{1}{J(n-1)} E\{\sum_i \sum_j (y_{ij} - \mu)^2 - (\bar{y}_{\cdot j} - \mu)^2\} \\
&= \frac{1}{J(n-1)} \{\sum_i \sum_j Var(y_{ij}) - Var(\bar{y}_{\cdot j})\} \\
&= \frac{nJ}{J(n-1)} \{(\tau^2 + \sigma^2) - \frac{\sigma^2}{n} + \tau^2\} \\
&= \frac{n}{n-1}(\frac{n-1}{n}\sigma^2) = \sigma^2
\end{aligned}$$

Therefore, $E(\hat{\tau}^2) = \frac{\sigma^2 + n\tau^2 - \sigma^2}{n} = \tau^2$

c) Although $\hat{\tau}^2$ is an unbiased and consistent estimator for τ^2, it has a rather undesirable property: for any finite sample size, there is a positive probability that $\hat{\tau}^2 < 0$. Prove that

$$Pr(\hat{\tau}^2 < 0 \mid \mu, \tau^2) = Pr(F_{J-1,\, J(n-1)} < \frac{\sigma^2}{\sigma^2 + n\tau^2}) > 0$$

You may use the fact that MS_B and MS_W are independent based on Fisher-Cochran Theorem. What would happen when $n \to \infty$?

Solution:

$\frac{(J-1)MS_B}{\sigma^2 + n\tau^2} \sim \chi^2_{J-1}$ and $\frac{J(n-1)MS_W}{\sigma^2} \sim \chi^2_{J(n-1)}$

Definition 5.8 Let U_1 and U_2 be independent random variables with, $U_1 \sim \chi^2_{p_1}$ and $U_2 \sim \chi^2_{p_2}$; then $F = \frac{U_1/p_1}{U_2/p_2}$ has the F-distribution with p_1 and p_2 degrees of freedom.

$$\begin{aligned}
P(\hat{\tau}^2 < 0) &= P(MS_B - MS_W < 0) \\
&= P(\frac{MS_B}{MS_W} < 1) \\
&= P(\frac{MS_B/(\sigma^2 + n\tau^2)}{MS_W/\sigma^2} < \frac{\sigma^2}{\sigma^2 + n\tau^2}) > 0
\end{aligned}$$

When $n \to \infty$, $\frac{\sigma^2}{\sigma^2+n\tau^2} \to 0,\ J(n-1) \to \infty$

$$lim_{n\to\infty} P(F_{J-1,\,J(n-1)} < \frac{\sigma^2}{\sigma^2+n\tau^2}) = P(\chi^2_{J-1} < 0) = 0$$

d) A full Bayesian method specify hyper-prior on $\{\mu, \tau^2\}$ and obtains the posterior distribution of $\{\mu, \tau^2\}$ given Y. Show that regardless of the choice of the hyper-prior, a Bayesian estimator for τ^2 cannot be negative. Also, show that assuming that $p(\mu \mid \tau^2) = 1$, the conditional distribution of $\mu \mid \tau^2, Y$ is $N(\bar{y}_{\cdot\cdot}, \frac{\sigma^2+n\tau^2}{nJ})$.

Solution (not sure):

We have $\bar{y}_{\cdot j} \mid \mu, \tau \sim N(\mu, \frac{\sigma^2}{n} + \tau^2)$

$$\begin{aligned}
p(\mu, \tau \mid y) &\propto p(\mu, \tau) \prod_{j=1}^{J} N(\bar{y}_{\cdot j} \mid \mu, \frac{\sigma^2}{n} + \tau^2) \\
&\propto p(\mu \mid \tau) p(\tau) \prod_{j=1}^{J} N(\bar{y}_{\cdot j} \mid \mu, \frac{\sigma^2}{n} + \tau^2) \\
&\propto p(\tau)\, exp(-\frac{\sum_{j=1}^{J}(\mu - \bar{y}_{\cdot j})^2}{2(\frac{\sigma^2}{n} + \tau^2)})
\end{aligned}$$

Then,

$$\begin{aligned}
P(\mu \mid \tau, y) = \frac{P(\mu, \tau \mid y)}{P(\tau \mid y)} &\propto \frac{P(\tau)}{P(\tau \mid y)} exp(-\frac{\sum_{j=1}^{J}(\mu - \bar{y}_{\cdot j})^2}{2(\frac{\sigma^2}{n} + \tau^2)}) \\
&\propto exp(-\frac{J\mu^2 - 2\mu \sum_j \bar{y}_{\cdot j}}{\frac{\sigma^2/n+\tau^2}{J}}) \\
&\propto exp(-\frac{1}{2}\frac{(\mu - \bar{y}_{\cdot\cdot})^2}{\frac{\sigma^2/n+\tau^2}{J}})
\end{aligned}$$

e) Suppose in addition to $p(\mu \mid \tau^2) = 1$ we also assume $p(\tau^2) = 1$. Find the corresponding posterior density $p(\tau \mid Y)$ and determine if and when it is a proper density (i.e. if it can be normalized to be proper density).

$$\begin{aligned}
P(\tau \mid y) &= \frac{P(\mu, \tau \mid y)}{P(\mu \mid \tau,\, y)} \\
&\propto \frac{P(\mu, \tau) P(y \mid \mu, \tau)}{P(\mu \mid \tau, y)} \\
&\propto \frac{P(\mu, \tau) \prod_{j=1}^{J} N(\bar{y}_{\cdot j} \mid \mu, \sigma^2/n + \tau^2)}{N(\mu \mid \bar{y}_{\cdot\cdot},\, \frac{\sigma^2/n+\tau^2}{J})} \\
&\propto (\sigma^2/n + \tau^2)^{\frac{1-J}{2}} exp(-\frac{\sum_j(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2}{2(\sigma^2/n + \tau^2)})
\end{aligned}$$

If $J > 3$, then the series converges since in a p-series given by

$$\sum_{n=1}^{\infty} 1/n^p = 1/1^p + 1/2^p + ...$$

If $p > 1$, then the series converges and if $0 < p \leq 1$ then the series diverges.

Q1 (June 2015)

The probit model assumes that $P(Y_i = 1 \mid X_i) = \Phi(X_i^T \beta)$. For this question assume that the prior $p(\beta) \propto 1$.
a) As a starting point for any MCMC algorithm it is generally useful to start around the maximum a posteriori probability (MAP) estimate. One way to estimate the MAP is by using the EM algorithm, where we introduce a latent variable Z_i, for each unit i. Specificially:

$$Y_i \mid Z_i, \beta, X_i = 1_{Z_i > 0}$$

$$Z_i \mid \beta, X_i \sim N(X_i^T \beta, 1)$$

Show that the new formulation is equivalent to $P(Y_i = 1 \mid X_i)$.
Solution:

$$\begin{aligned} P(Y_i = 1 \mid X_i) &= P(Y_i = 1,\, Z_i \geq 0 \mid X_i) + P(Y_i = 1,\, Z_i < 0 \mid X_i) \\ &= P(Y_i = 1 \mid X_i,\, Z_i \geq 0)\, P(Z_i \geq 0 \mid X_i) + P(Y_i = 1 \mid X_i,\, Z_i < 0) P(Z_i < 0 \mid X_i) \\ &= 1 \times P(Z_i - X_i^T \beta \geq -X_i^T \beta) = 1 - \Phi(-X_i^T \beta) = \Phi(X_i^T \beta) \end{aligned}$$

b) When the complete likelihood belongs to the exponential family

$$p(Y,\, Z \mid \theta) = h(Y, Z)\, exp(\theta^T T(Y, Z) - S(\theta))$$

the EM algorithm boils down to the following two steps:
- E-step: compute the expected value of the complete data sufficient statistic(s) $T(Y, Z)$, given the observed data and the current parameter estimates: $\nu^{(t)} := E(T(Y, Z) \mid \theta^{(t)}, Y)$
- M-step: set $\theta^{(t+1)}$ to the value which makes the unconditional expectation of the complete data sufficient statistic(s) equal to $\nu^{(t)} : E(T(Y, Z) \mid \theta^{(t+1)}) = \nu^{(t)}$
Prove this derivation of the EM algorithm.
Solution:

$$\begin{aligned} logL &= log\, h(Y, Z) + \theta^T T(Y, Z) - S(\theta) \\ \frac{\partial E_{\theta^{(t)}} logL}{\partial \theta} &= \frac{\partial E_{\theta^{(t)}} log\, h(Y, Z)}{\partial \theta} + \frac{\partial E_{\theta^{(t)}} \theta^T T(Y, Z)}{\partial \theta} - \frac{\partial S(\theta)}{\partial \theta} \\ &= 0 + E_{\theta^{(t)}}\, T(Y, Z) - E\, T(Y, Z) = 0 \\ &\Rightarrow E(T(Y, Z) \mid \theta^{(t+1)}) = E(T(Y, Z) \mid \theta^{(t)}, Y) \end{aligned}$$

c) Show that the complete data (log)-likelihood belongs to the exponential family, and identify the sufficient statistics.
Solution:

$$\begin{aligned} f(y_i, z_i) &= f(y_i \mid z_i) f(z_i) = 1_{z_i \geq 0} (2\pi)^{-\frac{1}{2}} exp(-\frac{(z_i - x_i^T \beta)^2}{2}) \\ f(y, z) &= 1_{z_1 \geq 0} ... 1_{z_n \geq 0} exp(-\frac{(Z - X\beta)^T (Z - X\beta)}{2}) \\ &= 1_{z_1 \geq 0} ... 1_{z_n \geq 0} exp(-\frac{Z^T Z - 2\beta^T X^T Z + \beta^T X^T X \beta}{2}) \end{aligned}$$

Therefore the sufficient statistics is $T(Z) = X^T Z$ with $\theta = \beta$
d) Derive the EM algorithm for probit regression.
Hint: The steps of the EM algorithm are:
E-step for MAP:

$$Q(\theta \mid \theta^{(t)}) = E_{\theta^{(t)}} (log\, p(\theta, z \mid y)) = \int log\, p(\theta, z \mid y)\, p(z \mid \theta^{(t)}, y)\, dz$$

The M-step

$$\theta^{(t+1)} := \ arg\, max_\theta Q(\theta \mid \theta^{(t)})$$

It is also useful to know that if $U \sim N(\mu, \sigma^2)$ and $V = [U \mid U \geq a]$ is a truncated Normal variable then

$$E(V) = \mu + \sigma \times \frac{\phi((a-\mu)/\sigma)}{1 - \Phi((a-\mu)/\sigma)}$$

and that if $U \sim N(\mu, \sigma^2)$ and $W = [U \mid U < a]$ then

$$E(W) = \mu - \sigma \times \frac{\phi((a-\mu)/\sigma)}{\Phi((a-\mu)/\sigma)}$$

Solution:

$$\begin{aligned}
E(X^T Z \mid \beta^{(t)}, Y = 1) &= X^T E(Z \mid \beta^{(t)}, Z \geq 0) = X^T (X\beta^{(t)}) + X^T \frac{\phi(-X\beta^{(t)})}{1 - \Phi(-X\beta^{(t)})} \\
&= X^T X\beta^{(t)} + X^T \frac{\phi(X\beta^{(t)})}{\Phi(X\beta^{(t)})} \\
E(X^T Z \mid \beta^{(t)}, Y = 0) &= X^T E(Z \mid \beta^{(t)}, Z < 0) = X^T (X\beta^{(t)}) - X^T \frac{\phi(-X\beta^{(t)})}{\Phi(-X\beta^{(t)})} \\
&= X^T X\beta^{(t)} - X^T \frac{\phi(X\beta^{(t)})}{1 - \Phi(X\beta^{(t)})} \\
E(X^T Z \mid \beta^{(t)}, Y) &= X^T X\beta^{(t)} + \frac{X^T \phi(X\beta^{(t)})}{\Phi(X\beta^{(t)})(1 - \Phi(X\beta^{(t)}))}(Y - \Phi(X\beta^{(t)})) \\
&= X^T Y^{(t)}
\end{aligned}$$

where we let $Y^{(t)} = X\beta^{(t)} + \frac{\phi(X\beta^{(t)})(Y - \Phi(X\beta^{(t)}))}{(1 - \Phi(X\beta^{(t)}))\Phi(X\beta^{(t)})}$

$$E(X^T Z \mid \beta^{(t)}) = X^T X\beta^{(t+1)} = X^T Y^{(t)} = E(X^T Z \mid \beta^{(t)}, Y)$$

$\beta^{(t+1)}$ is found by regressing $Y^{(t)} \sim X$

Q3 (June 2015)

The data are in the form of n pairs of points (x_i, y_i) where x_i denotes the sampling time at which the ith blood is taken and y_i denotes the ith measured concentration, $i = 1, ..., n$. The table below shows an example of pharmacokinetics data. The data are collected following the administration of a single 30mg dose of the drug cadralazine to a cardiac failure patient. The response y_i represents the concentration at time x_i, $i = 1, ..., 8$

Observation number i	Time (hours) x_i	Concentration (mg/liter) y_i
1	2	1.63
2	4	1.01
3	6	0.73
4	8	0.55
5	10	0.41
6	24	0.01
7	28	0.06
8	32	0.02

a) The most straight forward model for these data is to assume

$$log\, y_i = \mu(\beta) + \epsilon_i = log\Big[\frac{D}{V} exp(-k_e x_i)\Big] + \epsilon_i,$$

where $\epsilon_i \mid \sigma^2 \overset{iid}{\sim} N(0, \sigma^2)$, $\beta = [V, k_e]$ and the dose is $D = 30$. The parameters are the volume of distribution $V > 0$ and the elimination rate k_e.

1. For this model, obtain expressions for the log-likelihood function $L(\beta, \sigma^2)$, the score function $S(\beta, \sigma^2)$ and the expected information matrix $I(\beta, \sigma^2)$.

Solution:

$logy_i \sim N(\mu(\beta), \sigma^2)$

$$L(\beta, \sigma) = (\sqrt{2\pi}\sigma)^{-n} exp(-\frac{\sum_i (log\, y_i - \mu(\beta))^2}{2\sigma^2})$$
$$logL(\beta, \sigma) = -\frac{\sum_i \{\mu^2 - 2\mu log\, y_i + (log\, y_i)^2\}}{2\sigma^2} - nlog(\sqrt{2\pi\sigma^2})$$

Then,

$$S(\beta) = \frac{\partial logL(\beta, \sigma)}{\partial \beta} = \frac{\sum_{i=1}^n \{logy_i - \mu\}}{\sigma^2} \times \frac{\partial \mu}{\partial \beta}, \text{ where } \beta = (V, k_e)$$

where

$$\frac{\partial \mu(\beta)}{\partial V} = \frac{-\frac{D}{V^2} exp(-k_e x_i)}{\frac{D}{V} exp(-k_e x_i)} = -\frac{1}{V}$$
$$\frac{\partial \mu(\beta)}{\partial k_e} = \frac{\frac{D}{V} exp(-k_e x_i) \times (-x_i)}{\frac{D}{V} exp(-k_e x_i)} = -x_i$$

and

$$S(\sigma^2) = \frac{\sum_i (logy_i - \mu)^2}{2\sigma^4} - \frac{n}{2}\frac{1}{\sigma^2}$$

We have score function

$$S(\beta, \sigma^2) = \begin{bmatrix} -\frac{\sum_{i=1}^n logy_i - \mu}{\sigma^2}\frac{1}{V} \\ -\frac{\sum_{i=1}^n logy_i - \mu}{\sigma^2} x_i \\ \frac{\sum_i (logy_i - \mu)^2}{2\sigma^4} - \frac{n}{2}\frac{1}{\sigma^2} \end{bmatrix}$$

We find each component of $I(\beta, \sigma^2) = E\{S(\beta,\sigma^2)S(\beta,\sigma^2)^T\}$

$$I_{11} = \frac{1}{V^2}\frac{1}{\sigma^4}\sum_{i=1}^{n} E(logy_i - \mu(\beta))^2 = \frac{1}{V^2}\frac{1}{\sigma^4} n\sigma^2 = \frac{n}{V^2\sigma^2}$$

$$I_{22} = \frac{1}{\sigma^4}\sum_{i=1}^{n} E\{(logy_i - \mu(\beta))^2 x_i^2\} = \frac{1}{\sigma^4}\sum_{i=1}^{n} x_i^2\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{\sigma^2}$$

$$I_{12} = \frac{1}{\sigma^4 V} E(\sum_{i=1}^{n}(logy_i - \mu(\beta))^2 x_i) = \frac{1}{\sigma^4 V}\sum_{i=1}^{n}\sigma^2 x_i = \frac{\sum x_i}{\sigma^2 V}$$

$$\begin{aligned} I_{13} &= -E[\frac{\sum_{i=1}^{n}(logy_i - \mu(\beta))}{\sigma^2}\frac{1}{V}][\frac{\sum_i (logy_i - \mu(\beta))x_i}{2\sigma^4} - \frac{n}{2}\frac{1}{\sigma^2}] \\ &= -E(\frac{\sum(logy_i - \mu(\beta)) \times \sum(logy_i - \mu(\beta))^2}{2V\sigma^6} - \frac{n}{2\sigma^2}\frac{\sum(logy_i - \mu(\beta))}{\sigma^2 V}) = 0 \end{aligned}$$

$$I_{23} = -E(\frac{\sum(logy_i - \mu(\beta))}{\sigma^2}x_i)(\frac{\sum_i (logy_i - \mu(\beta))^2}{2\sigma^4} - \frac{n}{2}\frac{1}{\sigma^2}) = 0$$

$$I_{33} = E[-(\frac{\sum_i (logy_i - \mu)^2}{\sigma^6} + \frac{n}{2}\frac{1}{\sigma^4})] = \frac{n\sigma^2}{\sigma^6} - \frac{n}{2}\frac{1}{\sigma^4} = \frac{n}{2\sigma^4}$$

Putting all this together,

$$I = \begin{bmatrix} \frac{n}{V^2\sigma^2} & \frac{\sum x_i^2}{\sigma^2 V} & 0 \\ \frac{\sum x_i^2}{\sigma^2 V} & \frac{\sum x_i^2}{\sigma^2} & 0 \\ 0 & 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

2. Obtain the MLE and provide an asymptotic 95% confidence interval for each element of β
Solution:
$S(\beta,\sigma^2) = 0$ and $\sum_{i=1}^{n} logy_i = \sum_{i=1}^{n} log\frac{D}{V}exp(-k_e x_i) = log\frac{D^n}{V^n}exp(-k_e\sum x_i)$
Use three equations below to find $\hat{V}$, $\hat{k}_e$, and $\hat{\sigma}^2$

$$-\frac{\sum_{i=1}^{n} logy_i - \mu(\beta)}{\sigma^2}\frac{1}{V} = 0 \Rightarrow log\frac{D^n}{V^n}exp(-k_e\sum x_i) = \sum logy_i$$

$$\sum(logy_i)x_i = \sum x_i\mu(\beta) \Rightarrow \sum(logy_i)x_i = \sum x_i log\frac{D}{V}exp(-k_e x_i)$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i (log(y_i) - \mu)^2$$

95% C.I. for V: $\hat{V} \pm 1.96 \times \frac{\hat{V}^2\hat{\sigma}^2}{n}$ and 95% C.I. k_e: $\hat{k}_e \pm 1.96 \times \frac{\hat{\sigma}^2}{\sum x_i^2}$

3. The clearance $Cl = V \times k_e$ and elimination half-life $x_{1/2} = log2/k_e$ are parameters of interest in this experiment. Find the MLEs of these parameters along with asymptotic 95% confidence intervals.
Solution:
We know that $\frac{\partial l(\beta)}{\partial\beta} = \frac{\partial log f(y;g(\beta))}{\partial g(\beta)}\frac{\partial g(\beta)}{\partial\beta}$ and $I(\beta) = (\frac{\partial g(\beta)}{\partial\beta})^T I(g(\beta))\frac{\partial g(\beta)}{\partial\beta}$
$h(x,y) = xy$, $\frac{\partial h(x,y)}{\partial x} = y$ and $\frac{\partial h(x,y)}{\partial y} = x$

$$I(g(\beta)) = \begin{bmatrix} k_e & V \end{bmatrix} \begin{bmatrix} \frac{n}{V^2\sigma^2} & \frac{\sum x_i^2}{\sigma^2 V} \\ \frac{\sum x_i^2}{\sigma^2 V} & \frac{\sum x_i^2}{\sigma^2} \end{bmatrix}^{-1} \begin{bmatrix} k_e \\ V \end{bmatrix}$$

Then use this to find $Cl \sim N(V \times k_e, Var(above))$

$x_{1/2} : g(x) = \frac{log2}{x}$ and $\frac{\partial g(x)}{\partial x} = -\frac{log2}{x^2}$

$$Var(x_{1/2}) = (\frac{\partial g(x)}{\partial x})^2_{x=k_e} \times (\frac{\sum x_i^2}{\sigma^2})^{-1} = \frac{(log2)^2}{k_e^4}(\frac{\sum x_i^2}{\sigma^2})^{-1}$$

b) An alternative model is to consider the fractional polynomial model for the mean function

$$\mu(\beta) = exp(\beta_0 + \beta_1 x + \beta_2/x)$$

where $\beta = [\beta_0,\ \beta_1,\ \beta_2]$. We may choose a gamma distribution for the response y_i

$$y \mid \beta, \alpha \overset{iid}{\sim} Gamma(1/\alpha, \mu(\beta)\alpha)$$

to give $E(Y) = \mu$ and $Var(Y) = \alpha\mu^2$

1. For this model, obtain expressions for the log-likelihood function $L(\beta, \alpha)$, the score function $S(\beta, \alpha)$, and the expected information matrix $I(\beta, \alpha)$. (Recall that if a random variable $X \sim Gamma(a, b)$, then the density of X is $p(x) = \frac{1}{\Gamma(a)b^a}x^{a-1}exp(-x/b)$, with $E(X) = ab$ and $Var(X) = ab^2$

Solution:

$$logL(a, b) = -\frac{x}{b} + (a-1)logx - log(\Gamma(a)b^a)$$

$$logL(\alpha, \beta) = \sum_{i=1}^{n}[-\frac{y_i}{\alpha\mu_i} + (\frac{1}{\alpha} - 1)logy_i - log\Gamma(\frac{1}{\alpha})(\alpha\mu)^{\frac{1}{\alpha}}]$$

Then find score function

$$S(\beta) = \sum_{i=1}^{n}\frac{Y_i - \mu}{\alpha\mu^2}\frac{\partial\mu}{\partial\beta}$$

where $\frac{\partial\mu(\beta)}{\partial\beta_0} = \mu,\ \frac{\partial\mu(\beta)}{\partial\beta_1} = \mu x,\ \frac{\partial\mu(\beta)}{\partial\beta_2} = \frac{\mu}{x}$

$$S(\alpha) = \sum_{i=1}^{n} -\frac{y_i}{\alpha^2\mu_i} - \frac{logy_i}{\alpha^2} + \frac{d}{d\alpha}log\Gamma(\frac{1}{\alpha}) + \frac{1}{\alpha^2}log(\alpha\mu) - \frac{1}{\alpha^2} = g(\alpha, y_i)$$

then,

$$S(\beta, \alpha) = \begin{bmatrix} \sum_{i=1}^{n}\frac{Y_i-\mu}{\alpha\mu} \\ \sum_{i=1}^{n}\frac{Y_i-\mu}{\alpha\mu}x_i \\ \sum_{i=1}^{n}\frac{Y_i-\mu}{\alpha\mu}\frac{1}{x_i} \\ g(\alpha) \end{bmatrix}$$

$I = E(S(\beta, \alpha)S(\beta, \alpha)^T)$

$$I_{11} = \frac{E[\sum(Y_i - \mu)]^2}{\alpha^2\mu^2} = \frac{n\alpha\mu^2}{\alpha^2\mu^2} = \frac{n}{\alpha}$$

$$I_{22} = \frac{1}{\alpha^2\mu^2}E(\sum_{i=1}^{n}(Y_i - \mu)x_i)^2 = \frac{1}{\alpha^2\mu^2}\sum\alpha\mu^2x_i^2 = \frac{\sum_{i=1}^{n}x_i^2}{\alpha}$$

Similarly,

$$I = \begin{bmatrix} \frac{n}{\alpha} & \frac{1}{\alpha}\sum x_i & \frac{1}{\alpha}\sum\frac{1}{x_i} \\ \frac{1}{\alpha}\sum x_i & \frac{\sum x_i^2}{\alpha} & \frac{n}{\alpha} \\ \frac{1}{\alpha}\sum\frac{1}{x_i} & \frac{n}{\alpha} & \frac{1}{\alpha}\sum\frac{1}{x_i^2} \end{bmatrix}$$

Q4 (June 2015)

Part A

Consider the linear model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \text{ with } \epsilon_{ij} \sim N(0, \sigma^2),\ i = 1, ..5,\ j = 1, ..., 20$$

Let $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)^T$. Consider the solution to the normal equations, $\hat{\mu} = 0$, $\hat{\alpha} = [\bar{Y}_{1\cdot}, ..., \bar{Y}_{5\cdot}] = [10, 12, 5, 10, 8]$, and $\sum_{i=1}^{5} \sum_{j=1}^{20} Y_{ij}^2 = 12460$

a) Calculate a test statistic for testing $H_0 : \mu + \alpha_2 = 10$ vs. $H_1 : \mu + \alpha_2 > 10$. Write the rejection rule for H_0 and its associated degree of freedom(s).

Solution:

$$X = \begin{bmatrix} \mathbf{1}_{20} & \mathbf{1}_{20} & 0 & 0 & 0 & 0 \\ \mathbf{1}_{20} & 0 & \mathbf{1}_{20} & 0 & 0 & 0 \\ \mathbf{1}_{20} & 0 & 0 & \mathbf{1}_{20} & 0 & 0 \\ \mathbf{1}_{20} & 0 & 0 & 0 & \mathbf{1}_{20} & 0 \\ \mathbf{1}_{20} & 0 & 0 & 0 & 0 & \mathbf{1}_{20} \end{bmatrix} \text{ and } b = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix}$$

Then,

$$(X^TX)^g = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/20 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/20 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/20 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/20 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/20 \end{bmatrix}$$

$$k^T\hat{b} - m = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \\ 12 \\ 5 \\ 10 \\ 8 \end{bmatrix} - 10 = 2 \text{ and } k^T(X^TX)^g k = \frac{1}{20}$$

$$k^T\hat{b} = k^T(X^TX)^g X^T y \sim N(k^Tb, \sigma^2 k^T(X^TX)^g k) = N(\mu + \alpha_2, \frac{1}{20}\sigma^2)$$

$$(k^T\hat{b} - m)/\sqrt{\sigma^2 k^T(X^TX)^g k} \sim N(\,(k^Tb - m)/\sqrt{\sigma^2 k^T(X^TX)^g k}, 1)$$

and we know

$$\hat{\sigma}^2 = \frac{RSS}{n-r} = \frac{1}{100-5}\sum_{i=1}^{5}\sum_{j=1}^{20}(y_{ij} - \bar{y}_{i\cdot})^2 = \frac{1}{95}\sum_{i=1}^{5}(\sum_{j=1}^{20} y_{ij}^2 - 20\bar{y}_{i\cdot}^2) = \frac{1}{95}\sum_{i=1}^{5}\sum_{j=1}^{20} y_{ij}^2 - 20(100+144+25+100+64) = 40$$

Result 5.16 Let $X \sim N_p(\mu, V)$ and A be symmetric with rank s; if $BVA = 0$, then BX and X^TAX are independent.

Since $X\hat{b} = P_x y$ and $(I - P_X)y$ are independent, $k\hat{\beta}$ and RSS are independent.

Definition: Let $U \sim N(\mu, 1)$ and $V \sim \chi_k^2$. If U and V are independent, then $T = U/\sqrt{V/k}$ has the noncentral Student's t-distribution with k degrees of freedom and noncentrality μ, denoted $T \sim t_k(\mu)$

$$t = \frac{k^T\hat{b} - m}{\sqrt{\sigma^2 k^T(X^TX)^g k}} / \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} = \frac{k^T\hat{b} - m}{\sqrt{\hat{\sigma}^2 k^T(X^TX)^g k}} = \frac{2}{\sqrt{2}} = \sqrt{2} \sim t_{95}((k^Tb - m)/\sqrt{\sigma^2 k^T(X^TX)^g k})$$

Reject if $| t | > c_{\alpha/2}$ provides a level α test for $\alpha = Pr(| T | > c_{\alpha/2} \mid T \sim t_{95})$

b) Are $\alpha_1 - \alpha_2$ and $\alpha_2 - \alpha_3$ orthogonal?

Solution:
Since $\begin{bmatrix}0 & 1 & -1 & 0 & 0 & 0\end{bmatrix}^T \begin{bmatrix}0 & 0 & 1 & -1 & 0 & 0\end{bmatrix} \neq 0$, $\alpha_1 - \alpha_2$ and $\alpha_2 - \alpha_3$ are not orthogonal.
c) Calculate the F-statistics for testing the null hypothesis: $H_0 : \alpha_1 = \alpha_2 = \alpha_3$. Write the rejection rule for H_0 and its degree of freedom(s).
Solution:

$$K = \begin{bmatrix}0 & 1 & -1 & 0 & 0 & 0\\ 0 & 0 & 1 & -1 & 0 & 0\end{bmatrix}$$
$$K^T\hat{b} - m \sim N_s(K^T b - m, \sigma^2 H)$$

where $s = 2$, $K^T b - m = \begin{bmatrix}\bar{y}_1 - \bar{y}_2\\ \bar{y}_2 - \bar{y}_3\end{bmatrix} = \begin{bmatrix}-2\\ 7\end{bmatrix}$, $H = K^T(X^TX)^gK = \begin{bmatrix}2/20 & -1/20\\ -1/20 & 2/20\end{bmatrix}$, and $H^{-1} = \frac{400}{3}\begin{bmatrix}2/20 & 1/20\\ 1/20 & 2/20\end{bmatrix} = \frac{1}{3}\begin{bmatrix}40 & 20\\ 20 & 40\end{bmatrix}$

$$F = \frac{(K^T\hat{b} - m)^T H^{-1}(K^T\hat{b} - m)/s}{RSS/(N-r)} = \frac{260}{40} = 6.5 \sim F_{2,95}(\phi)$$

where $\phi = \frac{1}{2}(K^T b - m)^T(\sigma^2 H)^{-1}(K^T b - m)$
When the null hypothesis is true $\phi = 0$. Under the alternative hypothesis, F has noncentral F distribution.

Part B

(a) Suppose that $X_{n\times p}$ has $rank(X) = r$. For $a_{n\times 1}$ show that $P_X a = 0$ if and only if $X^T a = 0$.
Solution:
For any vector $a \in \mathbb{R}$, $P_X a = 0$ iff $X^T a = 0$
If $X^T a = 0$, then $P_X a = X(X^TX)^g X^T a = 0$. If $P_X a = 0$, then $X^T P_X a = X^T a = 0$
(b) Assume the Gauss-Markov model $Y = X\beta + \epsilon$, with $rank(X_{n\times p}) = r$, $Var(\epsilon) = \sigma^2 I_n$ where $\sigma^2 > 0$. Suppose that $\lambda^T\beta$ is estimable and $rank(\lambda_{p\times 1}) = 1$. Prove that if $\hat{\beta}$ solves the normal equations, then $0 < Var(\lambda^T\hat{\beta}) < \infty$ (note that both inequalities are strict).
Solution:
The solutions to the normal equations are given by $\hat{\beta} = (X^TX)^g X^T Y - (I - (X^TX)^g X^T X)w$, for all w.
Since $L^T\beta$ is estimable, by definition there exists a vector a such that $L^T = a^T X$

$$\begin{aligned}
L^T\hat{\beta} &= L^T(X^TX)^g X^T Y + L^T(I - (X^TX)^g X^T X)w\\
&= a^T X(X^TX)^g X^T Y + a^T X(I - (X^TX)^g X^T X)w\\
&= a^T P_X Y + a^T(X - P_X X)w\\
&= a^T P_X Y + a^T(X - X)w = a^T P_X Y
\end{aligned}$$

$a^T P_X Y$ does not depend on the arbitrary w and the choice of the generalized inverse $(X^TX)^g$ as P_X does not depend on the choice of the generalized inverse. This shows that $L^T\hat{\beta}$ is the same for all solutions $\hat{\beta}$.

$$\begin{aligned}
Var(L^T\hat{\beta}) &= \sigma^2 L^T(X^TX)^g X^T X(X^TX)^g L\\
&= \sigma^2 a^T X(X^TX)^g X^T X(X^TX)^g X^T a\\
&= \sigma^2 a^T X(X^TX)^g X^T a = \sigma^2 a^T P_X a
\end{aligned}$$

P_X is a positive semi-definite matrix. Hence, $Var(L^T\hat{\beta}) = \sigma^2 a^T P_X a \geq 0$ and $Var(L^T\hat{\beta}) = 0$ iff $a^T P_X a = \|P_X a\| = 0$ iff $P_X a = 0$. We showed that $P_X a = 0$ iff $a^T X = 0$, but that would imply that $L = 0$ which contradicts the assumption.

Q5 (June 2015)

Let $X_1, ...X_n$ be a random sample from the exponential distribution on the interval (θ, ∞), which also has scale parameter equal to θ. The density of this exponential distribution is $\frac{e^{-(x-\theta)/\theta}}{\theta}$, where $\theta > 0$ is unknown.
a) Show that $\frac{\bar{X}}{\theta} - 1$ has a gamma distribution with shape parameter n and scale parameter $1/n$. Find MGF of X
Solution:

$$M_X(t) = E(e^{tx}) = \int e^{tx} \frac{1}{\theta} e^{-x/\theta+1} dx = \frac{e}{\theta} \int_\theta^\infty e^{(t-1/\theta)x} dx = \frac{e}{\theta(t-1/\theta)} [e^{(t-1/\theta)x}]_\theta^\infty = \frac{e}{1-t\theta} e^{t\theta} e^{-1} = \frac{e^{t\theta}}{1-t\theta}$$

Then,

$$M_{\frac{\sum X}{n\theta}-1}(t) = E(e^{(\frac{\sum X}{n\theta}-1)t}) = e^{-t} E(e^{\frac{t}{n\theta}X_1} ... e^{\frac{t}{n\theta}X_n}) = e^{-t} M_X(\frac{t}{n\theta})^n = e^{-t} (\frac{e^{\frac{t}{n}}}{1-\frac{t}{n}})^n = (\frac{1}{1-\frac{t}{n}})^n$$

And this is MGF for gamma distribution with shape parameter n and scale parameter $1/n$
b) Show that $\frac{X_{(1)}}{\theta} - 1$ has an exponential distribution with scale parameter $1/n$
Solution:

$$F(x_{(1)}) = P(X_{(1)} < x) = P(min(X_1, ..., X_n) < x) = 1 - P(min(X_1, ..., X_n) > x) = 1 - P(X > x)^n$$

Since $P(X > x) = \int_x^\infty \frac{e^{-(x-\theta)/\theta}}{\theta} dx = e \int_x^\infty \frac{1}{\theta} e^{-x/\theta} = e[-e^{-x/\theta}]_x^\infty = e^{-x/\theta+1}$

$$F(x_{(1)}) = 1 - e^{(-x/\theta+1)n} \Rightarrow f(x_{(1)}) = \frac{n}{\theta} e^{-\frac{nx}{\theta}+n}$$

$$\begin{aligned} M_{X_{(1)}}(t) &= E(e^{tx_{(1)}}) = \int e^{tx} \frac{n}{\theta} e^{-\frac{nx}{\theta}+n} dx = \frac{n}{\theta} \frac{e^n}{t-\frac{n}{\theta}} [e^{(t-\frac{n}{\theta})x}]_\theta^\infty \\ &= \frac{n}{\theta} \frac{e^n}{\frac{n}{\theta}-t} (e^{-n+t\theta}) = \frac{ne^{t\theta}}{n-t\theta} \end{aligned}$$

$$M_{\frac{X_{(1)}}{\theta}-1}(t) = e^{-t} M_{X_{(1)}}(\frac{t}{\theta}) = e^{-t} \frac{ne^t}{n-t} = \frac{n}{n-t} = \frac{1}{1-t/n}$$

c) $\frac{\bar{X}}{\theta} - 1 \sim Gamma(n, \frac{1}{n})$. This does not depend on θ and so it is pivot quantity.
Solution:

$$P(a < \frac{\bar{X}}{\theta} - 1 < b) = P(a+1 < \frac{\bar{X}}{\theta} < b+1) = P(\frac{1}{b+1} < \frac{\theta}{\bar{X}} < \frac{1}{a+1}) = P(\frac{\bar{X}}{b+1} < \theta < \frac{\bar{X}}{a+1})$$

where

$$P(a < \frac{\bar{X}}{\theta} - 1 < b) = \int_a^b \frac{n^n}{\Gamma(n)} x^{n-1} exp(-nx) I_{(0,\infty)}(x) dx = 1 - \alpha$$

Similarly,

$$P(c < \frac{X_{(1)}}{\theta} - 1 < d) = P(\frac{\bar{X}}{d+1} < \theta < \frac{\bar{X}}{c+1})$$

where

$$P(c < \frac{X_{(1)}}{\theta} - 1 < d) = \int_c^d nexp(-nx) I_{(0,\theta)}(x) dx = 1 - \alpha$$

Q6 (June 2015)

Part A: Suppose $X_n \sim Beta(\alpha_n, \beta_n)$ with $\alpha_n \to \infty$ and $\alpha_n/\beta_n \to \lambda \in (0,1)$ as $n \to \infty$. Find the limiting distribution of X_n.

Solution:

As $n \to \infty$

$$E(X_n) = \frac{\alpha_n}{\alpha_n + \beta_n} = \frac{\alpha_n/\beta_n}{\alpha_n/\beta_n + 1} \to \frac{\lambda}{\lambda + 1}$$

$$Var(X_n) = \frac{\alpha_n \beta_n}{(\alpha_n + \beta_n)^2(\alpha_n + \beta_n + 1)} \to 0$$

Hence the limiting distribution of X_n is point mass at $\frac{\lambda}{\lambda+1}$

Part B: Let $X \sim N(\theta, 1)$ and consider the problem of testing the null hypothesis $H_0 : \theta = \theta_0$ versus the alternative $H_a : \theta \neq \theta_0$. Consider also the prior distribution that puts mass p_0 on θ_0 and is distributed as $N(\theta_0, \tau^2)$ for $\theta \neq \theta_0$.

a) Show that if p_1 is the posterior mass on θ_0 then $\frac{p_1}{1-p_1} = \frac{p_0}{1-p_0}\sqrt{1+\tau^2}exp[-\frac{\tau^2}{2(1+\tau^2)}(x-\theta_0)^2]$ and the Bayes factor is equal to: $\sqrt{1+\tau^2}exp[-\frac{\tau^2}{2(1+\tau^2)}(x-\theta_0)^2]$.

Hint: You may use the fact that if $X \sim N(\theta, 1)$ and $\theta \sim N(\mu, \tau^2)$ then the marginal density of X is $N(\mu, 1+\tau^2)$.

Solution:

$m(X) = \int f(x \mid \theta)\, f(\theta)\, d\theta = p_0 f(x \mid \theta_0) + (1-p_0)\int f(x \mid \theta)\, f(\theta)\, d\theta = p_0 f(x \mid \theta_0) + (1-p_0)\, N(\theta_0, 1+\tau^2)$

$$\begin{aligned} p_1 = P(\theta = \theta_0 \mid x) &= \frac{P(\theta = \theta_0)P(X \mid \theta = \theta_0)}{m(x)} \\ &= \frac{p_0 \times \frac{1}{\sqrt{2\pi}} exp(-\frac{(x-\theta_0)^2}{2})}{p_0 \times \frac{1}{\sqrt{2\pi}} exp(-\frac{(x-\theta_0)^2}{2}) + (1-p_0)(2\pi(1+\tau^2))^{-\frac{1}{2}} exp(-\frac{1}{2}\frac{(x-\theta_0)^2}{1+\tau^2})} \end{aligned}$$

Then,

$$\begin{aligned} \frac{p_1}{1-p_1} &= \frac{p_0 \times \frac{1}{\sqrt{2\pi}} exp(-\frac{(x-\theta_0)^2}{2})}{(1-p_0)(2\pi(1+\tau^2))^{-\frac{1}{2}} exp(-\frac{1}{2}\frac{(x-\theta_0)^2}{1+\tau^2})} \\ &= \frac{p_0}{1-p_0}\sqrt{1+\tau^2} exp(-\frac{\tau^2(x-\theta_0)^2}{2(1+\tau^2)}) \end{aligned}$$

Definition of Bayes factor: $\frac{P(H_1|y)}{P(H_0|y)} = \frac{P(H_1)}{P(H_0)} \times$ Bayes Factor. Then Bayes factor is: $\sqrt{1+\tau^2}exp(-\frac{\tau^2(x-\theta_0)^2}{2(1+\tau^2)})$

b) Show that the posterior distribution has density

$$\pi(\theta \mid x) = \begin{cases} p_1 & \text{if}\, \theta = \theta_0 \\ \frac{(1-p_1)\sqrt{1+\tau^2}}{\tau\sqrt{2\pi}} exp[-\frac{1+\tau^2}{2\tau^2}(\theta - \theta_1)^2] & \text{if}\, \theta \neq \theta_0 \end{cases}$$

where $\theta_1 = \frac{x\tau^2 + \theta_0}{1+\tau^2}$

Solution:

$$
\begin{aligned}
P(\theta \mid x) &\propto f(x \mid \theta) f(\theta) \\
&\propto exp(-\frac{1}{2}\{(x-\theta)^2 + (\frac{\theta-\theta_0}{\tau})^2\}) \\
&= exp(-\frac{1}{2}\{\theta^2 - 2x\theta + x^2 + \frac{1}{\tau^2}(\theta^2 - 2\theta\theta_0 + \theta_0^2)\}) \\
&\propto exp(-\frac{1}{2}\{(1+\frac{1}{\tau^2})\theta^2 - 2(x+\frac{\theta_0}{\tau^2})\theta\}) \\
&\propto exp(-\frac{1}{2}(1+\frac{1}{\tau^2})(\theta - \frac{x+\frac{\theta_0}{\tau^2}}{1+\frac{1}{\tau^2}})^2) \\
&\sim N(\frac{x\tau^2+\theta_0}{1+\tau^2}, \frac{\tau^2}{\tau^2+1})
\end{aligned}
$$

Since we put p_1 on θ_0, we have $1-p_1$ mass on posterior when $\theta \neq \theta_0$

c) Show that the Bayes factor is minimized when

$$
\tau^2 = \begin{cases} (x-\theta_0)^2 - 1 & \text{if } \mid x-\theta_0 \mid > 1 \\ 0 & \text{otherwise} \end{cases}
$$

and that the value of the minimum Bayes factor is

$$
\begin{cases} \mid x-\theta_0 \mid exp(\frac{-(x-\theta_0)^2+1}{2}) & \text{if } \mid x-\theta_0 \mid > 1 \\ 1 & \text{if } \mid x-\theta_0 \mid \leq 1 \end{cases}
$$

Solution:

log Bayes Factor $= \frac{1}{2}log(1+\tau^2) - \frac{\tau^2(x-\theta_0)^2}{2(1+\tau^2)}$. Take derivative and set it equal to 0.

$$
\frac{\partial logBF}{\partial \tau^2} = \frac{1}{2(1+\tau^2)} - \frac{(x-\theta_0)^2}{2(1+\tau^2)^2} = 0 \Rightarrow \tau^2 = (x-\theta_0)^2 - 1
$$

Since $\tau^2 \geq 0$, BF is minimzed at

$$
\tau^2 = \begin{cases} (x-\theta_0)^2 - 1 & \text{if } \mid x-\theta_0 \mid > 1 \\ 0 & \text{otherwise} \end{cases}
$$

substitute τ^2_{min} in BF, we get

$$
\text{minimum BF} = \begin{cases} \mid x-\theta_0 \mid exp(\frac{-(x-\theta_0)^2+1}{2}) & \text{if } \mid x-\theta_0 \mid > 1 \\ 1 & \text{if } \mid x-\theta_0 \mid \leq 1 \end{cases}
$$