Statistical Models
Exam solutions

June 6, 2020

1 Exam: 2019/05/29

Problem 1
Show that the following distributions belong to the exponential family. Find the canonical statistic and the canonical parameter in the minimal representation for each distribution:

- (a) Poisson distribution $Po(\lambda)$ with $\lambda > 0$.
- (b) Beta distribution.
- (c) Borel distribution.
- (d) Dirichlet distribution.
- (e) What are the minimal sufficient statistics in parts (a)-(d)?

- (a)
  Let $Y \sim Po(\lambda)$, then the probability mass function is given by:

$$\begin{aligned} f(y,\lambda) = P(Y=y) &= \frac{\lambda^y e^{-\lambda}}{y!} \\ &= \frac{1}{y!} e^{-\lambda} e^{y \log \lambda} \end{aligned} \tag{1}$$

  Which has the form of the exponential family with:

$$\begin{aligned} \underline{\textbf{Canonical Parameter:}} &\quad \theta = \log \lambda \Rightarrow \lambda = e^{\theta} \\ \underline{\textbf{Canonical Statistic:}} &\quad t(y) = y \end{aligned} \tag{2}$$

- (b)
  Let $Y \sim Beta(\alpha, \beta)$ then, the probability density function is given by:

$$\begin{aligned} f(y;\alpha,\beta) &= \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1} \quad \text{for} \quad y \in (0,1) \\ &= \frac{1}{B(\alpha,\beta)} \frac{y^{\alpha}}{y} \frac{(1-y)^{\beta}}{(1-y)} \\ &= \frac{1}{B(\alpha,\beta)} \frac{1}{y(1-y)} e^{\alpha \log y + \beta \log(1-y)} \end{aligned} \tag{3}$$

Which has the form of the exponential family with:

$$\begin{aligned} \underline{\textbf{Canonical Parameter:}} \quad & \theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ \underline{\textbf{Canonical Statistic :}} \quad & t(y) = \begin{pmatrix} \log y \\ \log(1-y) \end{pmatrix} \end{aligned} \tag{4}$$

- (c)
  Let $Y \sim Borel(\alpha)$ with probability mass function:

$$\begin{aligned} P(Y = y; \alpha) &= \frac{1}{y!}(\alpha y)^{y-1} e^{-\alpha y} \quad \text{for} \quad y = 1, \ldots \quad \text{with} \quad \alpha \in (0,1) \\ &= \frac{1}{y!}\frac{\alpha^y}{\alpha} y^{y-1} e^{-\alpha y} = \frac{1}{y!}\frac{y^{y-1}}{\alpha} e^{(y \log \alpha - \alpha y)} \\ &= \frac{1}{y!}\frac{y^{y-1}}{\alpha} e^{(y(\log \alpha - \alpha))} \end{aligned} \tag{5}$$

  Which has the form of the exponential family with:

$$\begin{aligned} \underline{\textbf{Canonical Parameter :}} \quad & \theta = \log \alpha - \alpha \\ \underline{\textbf{Canonical Statistic:}} \quad & t(y) = y \end{aligned} \tag{6}$$

- (d)
  Let $Y \sim Dirichlet(\boldsymbol{\alpha})$ where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)^T$ while $\mathbf{y} = (y_1, ..., y_k)$, having probability density function:

$$\begin{aligned} f(\mathbf{y}; \boldsymbol{\alpha}) &= \frac{\Gamma(\sum\limits_{i=1}^{k} \alpha_i)}{\prod\limits_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} y_i^{\alpha_i - 1} = K \cdot \prod_{i=1}^{k} \frac{y_i^{\alpha_i}}{y_i} = K \cdot \prod_{i=1}^{k} \frac{1}{y_i} \exp\{\sum_{i=1}^{k-1} \alpha_i \log y_i + \alpha_k \log y_k\} \\ &= K \cdot \prod_{i=1}^{k} y_i^{-1} \exp\{(\sum_{i=1}^{k-1} \alpha_i \log y_i) + \alpha_k \log(\underbrace{1 - \sum_{i=1}^{k-1} y_i)_{y_k}}\} \end{aligned} \tag{7}$$

  Which has the form of the exponential family with:

$$\begin{aligned} \underline{\textbf{Canonical Parameter:}} \quad & \theta = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{k-1} \\ \alpha_k \end{pmatrix} \\ \underline{\textbf{Canonical Statistic :}} \quad & t(y) = (\log y_1, \ldots, \log y_{k-1}, \log y_k)^T \end{aligned} \tag{8}$$

- (e)
  Is the canonical statistic always minimal sufficient?

- (a) $t(y) = y$
- (b) $t(y) = \begin{pmatrix} \log y \\ \log(1-y) \end{pmatrix}$
- (c) $t(y) = y$
- (d) $t(y) = (\log y_1, \ldots, \log y_{k-1}, \log y_k)^T$

Problem 2 Let Y have a negative binomial distribution with

$$P(Y = y; \pi) = \frac{(y+k-1)!}{y!(k-1)!}\pi^k(1-\pi)^y \quad \text{for} \;\; y = 0, 1, 2, ...$$

where y can be interpreted as the number of failures in Bernoulli trials until k successes with success probability $\pi \in (0, 1)$. Derive the mean of Y and the variance of Y .

We can rewrite the probability mass function as:

$$P(Y = y; \pi) = \frac{(y+k-1)!}{y!(k-1)!}\pi^k \exp\{y \log(1-\pi)\} \tag{9}$$

Which has the form of the exponential family with:

$$\begin{aligned}
\underline{\textbf{Canonical Parameter:}}\;& \theta = \log(1-\pi) \Rightarrow \pi = 1 - e^\theta \Rightarrow e^\theta = 1 - \pi \\
\underline{\textbf{Canonical Statistic :}}\;& t(y) = y \\
\underline{\textbf{Canonical Parameter Space:}}\;& \Theta = (-\infty, 0] \\
\underline{\textbf{Normalizing Constant:}}\;& a(\theta) = \pi^k \\
\underline{\textbf{Norming Constant:}}\;& C(\theta) = \pi^{-k} \\
\underline{\textbf{Constant Factor:}}\;& h(y) = \frac{(y+k-1)!}{y!(k-1)!}
\end{aligned} \tag{10}$$

- (a)
  We will derive the expected value of the random variable Y by considering the first derivative of the $\log C(\theta)$

  By definition we have that:

  $$C(\theta) = \int h(y) \exp\{\theta^T t(y)\} dy \quad \Rightarrow \frac{\partial C(\theta)}{\partial \theta} = C(\theta) \cdot E[t]$$

  Which means that:

  $$\frac{\partial \log C(\theta)}{\partial \theta} = \frac{1}{C(\theta)} C(\theta) \cdot E[t] = E[t] = E[Y]$$

  In our case we have that $C(\theta) = \pi^{-k} \Rightarrow \log C(\theta) = -k \log \pi = -k \log(1 - e^\theta)$

  $$E[Y] = -k\frac{\partial \log(1-e^\theta)}{\partial \theta} = \frac{ke^\theta}{(1-e^\theta)} = k\frac{(1-\pi)}{\pi}$$

- (b)
  We know that:

$$\begin{aligned}
\frac{\partial^2 \log C(\theta)}{\partial^2\theta} &= \frac{\partial}{\partial\theta}\Big(\frac{1}{C(\theta)}\cdot\frac{\partial C(\theta)}{\partial\theta}\Big)\\
&= \frac{\partial}{\partial\theta}\Big(\frac{1}{C(\theta)}\Big)\frac{\partial C(\theta)}{\partial\theta} + \frac{1}{C(\theta)}\cdot\frac{\partial}{\partial\theta}\Big(\frac{\partial C(\theta)}{\partial\theta}\Big)\\
&= -\frac{1}{C(\theta)^2}\cdot\frac{\partial C(\theta)}{\partial\theta}\cdot\frac{\partial C(\theta)}{\partial\theta} + \frac{1}{C(\theta)}\cdot\frac{\partial^2 C(\theta)}{\partial^2\theta}\\
&= -\frac{1}{C(\theta)^2}\cdot\Big(\frac{\partial C(\theta)}{\partial\theta}\Big)^2 + \frac{1}{C(\theta)}\cdot\frac{\partial^2 C(\theta)}{\partial^2\theta}\\
&= -\frac{1}{C(\theta)^2}\cdot(C(\theta)E[t])^2 + \frac{1}{C(\theta)}C(\theta)E[t^2]\\
&= E[t^2] - (E[t])^2 = \mathrm{Var}(t) = \mathrm{Var}(Y)
\end{aligned} \tag{11}$$

  Hence, in our case:

$$\begin{aligned}
\mathrm{Var}(Y) &= \frac{\partial}{\partial\theta}\Big(\frac{ke^\theta}{1-e^\theta}\Big) = k\Big(\frac{e^\theta}{1-e^\theta} + \frac{(e^\theta)^2}{(1-e^\theta)^2}\Big)\\
&= k\Big(\frac{e^\theta(1-e^\theta)+(e^\theta)^2}{(1-e^\theta)^2}\Big) = \frac{ke^\theta}{(1-e^\theta)^2} = k\frac{(1-\pi)}{\pi^2}
\end{aligned} \tag{12}$$

Problem 3
Let $Y = (Y_1, Y_2, Y_3)^T$ have a multinomial distribution with probability mass function given by

$$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3; p_1, p_2) = \frac{n!}{y_1!y_2!y_3!}p_1^{y_1}p_2^{y_2}(1-p_1-p_2)^{y_3} \qquad n = y_1 + y_2 + y_3$$

with $p_1, p_2 \in (0, 1)$ and known n.

- (a) Show that $\mathbf{Y}$ belongs to the exponential family? What is the canonical statistics $t(\mathbf{Y})$ and the canonical parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$?
- (b) Calculate the norming constant $C(\theta)$.
- (c) Calculate the expected value of the canonical statistics $\boldsymbol{\mu} = E(t(\mathbf{Y}))$.
- (d) Compute the expected Fisher information $I(\boldsymbol{\theta})$ for the canonical parametrization.
- (e) What is the expected Fisher information matrix $I(\boldsymbol{\mu})$ in the mean value parametrization $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$? Present the results in terms of $\boldsymbol{\mu}$.

- (a)
We can rewrite the probability mass function as:

$$\begin{aligned}
&= n!\frac{1}{y_1!y_2!(n-y_1-y_2)!}\exp\{y_1\log p_1 + y_2\log p_2 + (n-y_1-y_2)\log(1-p_1-p_2)\} \\
&= K\cdot\exp\{y_1\log p_1 + y_2\log p_2 + n\log(1-p_1-p_2) - y_1\log(1-p_1-p_2) - y_2\log(1-p_1-p_2)\} \\
&= K_2\cdot\exp\{y_1\log p_1 + y_2\log p_2 - (y_1+y_2)\log(1-p_1-p_2)\} \\
&= K_2\cdot\exp\{y_1\log p_1 + y_2\log p_2 + (y_1+y_2)\log\frac{1}{1-p_1-p_2}\} \\
&= K_2\cdot\exp\{y_1\log p_1 + y_1\log\frac{1}{1-p_1-p_2} + y_2\log p_2 + y_2\log\frac{1}{1-p_1-p_2}\} \\
&= K_2\cdot\exp\{y_1(\log p_1 - \log(1-p_1-p_2)) + y_2(\log p_2 - \log(1-p_1-p_2))\} \\
&= \frac{n!(1-p_1-p_2)^n}{y_1!y_2!(n-y_1-y_2)!}\cdot\exp\{y_1(\log\frac{p_1}{1-p_1-p_2}) + y_2(\log\frac{p_2}{1-p_1-p_2})\}
\end{aligned} \tag{13}$$

Which has the form of the exponential family with:

$$\begin{aligned}
\underline{\textbf{Canonical Parameter:}}\quad & \theta = \begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} = \begin{pmatrix}\log\frac{p_1}{1-p_1-p_2}\\ \log\frac{p_2}{1-p_1-p_2}\end{pmatrix} \\
\underline{\textbf{Canonical Statistic:}}\quad & t(y) = \begin{pmatrix}y_1\\ y_2\end{pmatrix} \\
\underline{\textbf{Canonical Parameter Space:}}\quad & \Theta = (-\infty, 0] \\
\underline{\textbf{Normalizing Constant:}}\quad & a(\theta) = (1-p_1-p_2)^n \\
\underline{\textbf{Norming Constant:}}\quad & C(\theta) = (1-p_1-p_2)^{-n} \\
\underline{\textbf{Constant Factor:}}\quad & h(y) = \frac{n!}{y_1!y_2!(n-y_1-y_2)!}
\end{aligned} \tag{14}$$

- (b)
We can see that

$$e^{\theta_1} = \frac{p_1}{1-p_1-p_2} \qquad \Rightarrow e^{\theta_2} = \frac{p_2}{1-p_1-p_2}$$

$$\Rightarrow p_1 = e^{\theta_1}(1-p_1-p_2); \quad p_2 = e^{\theta_2}(1-p_1-p_2)$$

$$\begin{aligned}
(1-p_1-p_2) &= 1 - e^{\theta_1}(1-p_1-p_2) - e^{\theta_2}(1-p_1-p_2) \\
&= (1-p_1-p_2)(\frac{1}{(1-p_1-p_2)} - e^{\theta_1} - e^{\theta_2}) \\
1 &= \frac{1}{(1-p_1-p_2)} - e^{\theta_1} - e^{\theta_2} \\
\frac{1}{(1-p_1-p_2)} &= 1 + e^{\theta_1} + e^{\theta_2}
\end{aligned} \tag{15}$$

We know that

$$C(\theta) = \frac{1}{(1-p_1-p_2)^n} = (1+e^{\theta_1}+e^{\theta_2})^n$$

Hence,

$$\log C(\theta) = n\log(1+e^{\theta_1}+e^{\theta_2})$$

- (c)
  We know that

$$\boldsymbol{\mu} = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} = \begin{pmatrix}E[t(y_1)]\\ E[t(y_2)]\end{pmatrix} = \begin{pmatrix}E[y_1]\\ E[y_2]\end{pmatrix} = \begin{pmatrix}\frac{\partial \log C(\theta)}{\partial\theta_1}\\ \frac{\partial \log C(\theta)}{\partial\theta_2}\end{pmatrix} = \begin{pmatrix}\frac{ne^{\theta_1}}{1+e^{\theta_1}+e^{\theta_2}}\\ \frac{ne^{\theta_2}}{1+e^{\theta_1}+e^{\theta_2}}\end{pmatrix}$$

- (d)

$$\Rightarrow I(\theta) = E[J(\theta)] = J(\theta) = \frac{-\partial U(\theta)}{\partial\theta} = -\Big(\frac{\partial t}{\partial\theta} - \mu_t(\theta)\Big) = -\Big(-\frac{\partial}{\partial\theta}\mu_t(\theta)\Big) = V_t(\theta) = \mathrm{Var}(\theta)$$

$$\mathrm{Var}(\theta) = \begin{pmatrix}\mathrm{Var}(\theta_1) & \mathrm{Cov}(\theta_1,\theta_2)\\ \mathrm{Cov}(\theta_1,\theta_2) & \mathrm{Var}(\theta_2)\end{pmatrix} = \begin{pmatrix}\frac{\partial^2\log C(\theta)}{\partial^2\theta_1} & \frac{\partial^2\log C(\theta)}{\partial\theta_1\theta_2}\\ \frac{\partial^2\log C(\theta)}{\partial\theta_1\theta_2} & \frac{\partial^2\log C(\theta)}{\partial^2\theta_2}\end{pmatrix}$$

$$\begin{aligned}
\mathrm{Var}(\theta_1) &= \frac{\partial^2\log C(\theta)}{\partial^2\theta_1} = \frac{ne^{\theta_1}}{1+e^{\theta_1}+e^{\theta_2}} - \frac{ne^{\theta_1}e^{\theta_1}}{(1+e^{\theta_1}+e^{\theta_2})^2} = \mu_1(1-\mu_1/n)\\
\mathrm{Var}(\theta_2) &= \frac{\partial^2\log C(\theta)}{\partial^2\theta_2} = \mu_2(1-\mu_2/n)\\
\mathrm{Cov}(\theta_1,\theta_2) &= \frac{\partial^2\log C(\theta)}{\partial\theta_1\theta_2} = \frac{\partial}{\partial\theta_1}\Big(\frac{\partial\log C(\theta)}{\partial\theta_2}\Big) = \frac{\partial}{\partial\theta_1}\Big(\frac{ne^{\theta_2}}{1+e^{\theta_1}+e^{\theta_2}}\Big)\\
&= -\frac{ne^{\theta_1}e^{\theta_2}}{(1+e^{\theta_1}+e^{\theta_2})^2} = -\frac{\mu_1\mu_2}{n}\\
\mathrm{Var}(\theta) &= \begin{pmatrix}\mu_1(1-\mu_1/n) & -\mu_1\mu_2/n\\ -\mu_1\mu_2/n & \mu_2(1-\mu_2/n)\end{pmatrix}
\end{aligned} \tag{16}$$

- (e)
  By the re-parametrization lemma, for $\mu = \mu(\theta)$ we have:

$$I_\mu(\mu) = \Big(\frac{\partial\theta}{\partial\mu}\Big)^T I_\theta(\theta(\mu))\Big(\frac{\partial\theta}{\partial\mu}\Big) = \Big(\Big(\frac{\partial\mu}{\partial\theta}\Big)^{-1}\Big)^T I_\theta(\theta(\mu))\Big(\frac{\partial\theta}{\partial\mu}\Big)$$

$$\frac{\partial\mu}{\partial\theta} = \begin{pmatrix}\frac{\partial\mu_1}{\partial\theta_1} & \frac{\partial\mu_1}{\partial\theta_2}\\ \frac{\partial\mu_2}{\partial\theta_1} & \frac{\partial\mu_2}{\partial\theta_2}\end{pmatrix} = \mathrm{Var}(\theta) = \mathrm{Var}(t) = \mathrm{Var}(t)^T$$

$$I_\mu(\mu) = \mathrm{Var}(t)^{-1}\,\mathrm{Var}(t)\,\mathrm{Var}(t)^{-1} = \mathrm{Var}(t)^{-1}$$

$$\begin{aligned}
I_\mu(\mu) = \mathrm{Var}(t)^{-1} &= \begin{pmatrix}\mu_1(1-\mu_1/n) & -\mu_1\mu_2/n\\ -\mu_1\mu_2/n & \mu_2(1-\mu_2/n)\end{pmatrix}^{-1}\\
&= \frac{1}{\mu_1\mu_2(1-\mu_1/n-\mu_2/n)}\begin{pmatrix}\mu_2(1-\mu_2/n) & \mu_1\mu_2/n\\ \mu_1\mu_2/n & \mu_1(1-\mu_1/n)\end{pmatrix}
\end{aligned} \tag{17}$$

Problem 4
Let $Y_1, Y_2, \dots, Y_n$ be an independent sample from the Weibull distribution with density

$$f(y;\beta) = \beta\alpha y^{\alpha-1}\exp(-\beta y^{\alpha}), \qquad \alpha, \beta > 0$$

where α is assumed to be known.

- (a) Show that the above distribution is a one parameter exponential distribution and find its canonical parameter θ.
- (b) Find the canonical statistics $t(Y_1)$ and compute $E[t(Y_1)], \mathrm{Var}(t(Y_1))$
- (c) Derive the maximum likelihood estimator (MLE) $\hat{\beta}_{MLE}$ for β.
- (d) Find the observed Fisher information $J(\theta)$. What is the expected Fisher information $I(\theta)$?
- (e) What is the asymptotic distribution of $\hat{\beta}_{MLE}$? Specify its parameters.
- (f) Determine the likelihood ratio $L(\beta_0)/L(\hat{\beta}_{MLE})$ when β_0 is a given fixed value.
- (g) Derive the saddle-point approximation for the distribution of $\hat{\beta}_{MLE}$ in a point β_0 without determining the normalization constant.
- (h) Show that the distribution of the statistics $\sum\limits_{i=1}^{n} Y_i^{\alpha}$ belongs to the exponential family. Derive the saddle point approximation of its structural function.

- (a)
  Let $y = (y_1, \dots, y_n)$ then

$$L(y;\beta) = \beta^n \alpha^n \prod_{i=1}^{n} y_i^{\alpha-1} \exp\{-\beta \sum_{i=1}^{n} y_i^{\alpha}\} \tag{18}$$

  Which has the form of the exponential family with:

$$\begin{aligned} \underline{\textbf{Canonical Parameter:}}\ & \theta = -\beta \\ \underline{\textbf{Canonical Statistic :}}\ & t(y) = \sum_{i=1}^{n} y_i^{\alpha} \\ \underline{\textbf{Norming Constant:}}\ & C(\theta) = \beta^{-n} \end{aligned} \tag{19}$$

- (b)

$$t(Y_1) = Y_1^{\alpha} \Rightarrow E[Y_1^{\alpha}] = \frac{\partial \log C_1(\theta)}{\partial\theta} \Rightarrow= C_1(\theta) = \beta^{-1} = -\theta^{-1} \Rightarrow \log C_1(\theta) = -\log(-\theta)$$

$$E[Y_1^{\alpha}] = \frac{\partial - \log(-\theta)}{\partial\theta} = -\frac{1}{\theta} = \frac{1}{\beta}$$

$$\mathrm{Var}(Y_1^\alpha) = \frac{\partial^2 \log C_1(\theta)}{\partial^2\theta} = \frac{\partial}{\partial\theta}\Big(-\frac{1}{\theta}\Big) = \frac{1}{\theta^2} = \frac{1}{\beta^2}$$

- (c)
  We obtain the MLE by computing the score function and setting it's value to zero, i.e $U(\theta) = 0$ where

$$\begin{aligned} l(\beta, y) = \log L(\beta; y) = \log(\prod_{i=1}^{n} f(y_i\beta)) \\ = n\log\alpha + n\log\beta + \log(\prod_{i=1}^{n} y_i^{\alpha-1}) - \beta\sum_{i=1}^{n} y_i^\alpha \end{aligned} \tag{20}$$

  then:

$$U(\theta) = \frac{\partial l(\beta,y)}{\partial\beta} = \frac{n}{\beta} - \sum_{i=1}^{n} y_i^\alpha = 0 \Rightarrow \sum_{i=1}^{n} y_i^\alpha = \frac{n}{\beta}$$
$$\Rightarrow \hat{\beta}_{ML} = \frac{n}{\sum\limits_{i=1}^{n} y_i^\alpha}$$

- (d)
  We know that:

$$I(\beta) = E[J(\beta)] = J(\beta) = \frac{-\partial U(\beta)}{\partial\beta} = -\frac{\partial}{\partial\beta}\Big(\frac{n}{\beta} - \sum_{i=1}^{n} y_i^\alpha\Big) = \frac{n}{\beta^2} = V_t(\beta)$$

- (e)
  By the Central Limit Theorem we have that the asymptotic distribution of the MLE

$$\Rightarrow \sqrt{I(\beta)}(\hat{\beta}_{ML} - \beta) \xrightarrow{D} \mathcal{N}(0,1) \Rightarrow \frac{\sqrt{n}}{\beta}(\hat{\beta}_{ML} - \beta) \xrightarrow{D} \mathcal{N}(0,1)$$

$$\Rightarrow \sqrt{n}(\hat{\beta}_{ML} - \beta) \xrightarrow{D} \mathcal{N}(0,\beta^2) \quad \text{as } n\to\infty$$

  The asymptotic distribution of $\hat{\beta}_{ML}$ is normal with mean β and variance β^2/n

- (f)
  We have that:

$$\begin{aligned} L(y;\beta_0) = \beta_0^n\alpha^n\prod_{i=1}^{n} y_i^{\alpha-1}\exp\{-\beta_0\sum_{i=1}^{n} y_i^\alpha\} \\ L(y;\hat{\beta}_{ML}) = \hat{\beta}_{ML}^n\alpha^n\prod_{i=1}^{n} y_i^{\alpha-1}\exp\{-\hat{\beta}_{ML}\sum_{i=1}^{n} y_i^\alpha\} \end{aligned} \tag{21}$$

$$
\begin{aligned}
\frac{L(\beta_0)}{L(\hat{\beta}_{ML})} &= \frac{\beta_0^n}{\hat{\beta}_{ML}^n}\exp\{-\beta_0\sum_{i=1}^n y_i^\alpha + \hat{\beta}_{ML}\sum_{i=1}^n y_i^\alpha\} \\
&= \frac{\beta_0^n}{\hat{\beta}_{ML}^n}\exp\{(\hat{\beta}_{ML}-\beta_0)\sum_{i=1}^n y_i^\alpha\} \\
&= \left(\frac{\beta_0}{\hat{\beta}_{ML}}\right)^n \exp\{-(\beta_0-\hat{\beta}_{ML})\sum_{i=1}^n y_i^\alpha\} \\
&= \left(\frac{\beta_0}{\hat{\beta}_{ML}}\right)^n \exp\{-(\beta_0-\hat{\beta}_{ML})\frac{n}{\hat{\beta}_{ML}}\} \\
&= \left(\frac{\beta_0}{\hat{\beta}_{ML}}\right)^n e^n\exp\{-\frac{n\beta_0}{\hat{\beta}_{ML}}\}
\end{aligned} \tag{22}
$$

- (g)
  The saddle point approximation is given by (for k = 1):

$$
\begin{aligned}
f(\hat{\beta}_{ML};\beta_0) &\approx (2\pi)^{-k/2}|V_t(\hat{\beta}_{MLE})|^{-1/2}\frac{L(\beta_0)}{L(\hat{\beta}_{ML})} \\
&\approx (2\pi)^{-1/2}|I(\hat{\beta}_{ML})|^{1/2}\left(\frac{\beta_0}{\hat{\beta}_{ML}}\right)^n e^n\exp\{-\frac{n\beta_0}{\hat{\beta}_{ML}}\} \\
&\approx (2\pi)^{-1/2}\sqrt{n/\hat{\beta}^2_{MLE}}\left(\frac{\beta_0}{\hat{\beta}_{ML}}\right)^n e^n\exp\{-\frac{n\beta_0}{\hat{\beta}_{ML}}\} \\
&\approx (2\pi)^{-1/2}\quad n^{1/2}\ \hat{\beta}_{MLE}^{-(n+1)}\quad \beta_0^n\quad e^n\exp\{-\frac{n\beta_0}{\hat{\beta}_{ML}}\}
\end{aligned} \tag{23}
$$

- (h)
  We know that in the mean value parametrization t is the MLE for μ

$$t \sim \mathcal{N}(\mu_t, V_t^{-1}) = \mathcal{N}(\mu_t, V_t) \quad t = \hat{\mu}_{t,ML}$$

  , Is asymptotically normal, with density function:

$$f(t;\theta) \approx \frac{1}{(2\pi)^{1/2}\sqrt{V_t}}\exp\Big\{-\frac{(t-\mu_t(\theta)^2)}{2V_t}\Big\} = (2\pi)^{-1/2}\frac{\theta}{\sqrt{n}}\exp\Big\{-\frac{(\sum_{i=1}^n y_i^\alpha + n/\theta)^2)}{2n/\theta^2}\Big\}$$

  While the density function under the null hypothesis is given by:

$$
\begin{aligned}
f(t;\theta_0) &\approx \frac{C(\theta)}{C(\theta_0)}f(t;\theta)\exp\{t(\theta_0-\theta)\} \\
&= \frac{C(\theta)}{C(\theta_0)}\exp\{t(\theta_0-\theta)\}(2\pi)^{-1/2}\frac{\theta}{\sqrt{n}}\exp\Big\{-\frac{(\sum_{i=1}^n y_i^\alpha + n/\theta)^2)}{2n/\theta^2}\Big\}
\end{aligned} \tag{24}
$$

If we now replace θ by $\hat{\theta}_{ML} = -\hat{\beta}_{ML} - \frac{n}{\sum_{i=1}^{n} y_i^{\alpha}} = -\frac{n}{t}$

$$
\begin{aligned}
f(t;\theta_0) &\approx \frac{C(\hat{\theta}_{ML})}{C(\theta_0)} f(t;\hat{\theta}_{ML}) \exp\{t(\theta_0 - \hat{\theta}_{ML})\} \\
&= \frac{C(\hat{\theta}_{ML})}{C(\theta_0)} \exp\{t(\theta_0 - \hat{\theta}_{ML})\}(2\pi)^{-1/2} \frac{-\hat{\theta}_{ML}}{\sqrt{n}} \exp\left\{ \underbrace{\frac{(t-t)^2)}{2t^2}}_{=0} \right\} \\
&= (2\pi)^{-1/2} \frac{-\hat{\theta}_{ML}}{\sqrt{n}} \frac{C(\hat{\theta}_{ML})}{C(\theta_0)} \exp\{t(\theta_0 - \hat{\theta}_{ML})\} \qquad (25) \\
f(t;\beta_0) &\approx (2\pi)^{-1/2} \frac{\hat{\beta}_{ML}}{\sqrt{n}} \frac{\beta_0^n}{\hat{\beta}_{ML}^n} \exp\{t(\hat{\beta}_{ML} - \beta_0)\} \\
&= \beta_0^n \quad (2\pi n)^{-1/2} \quad \hat{\beta}_{ML}^{-n+1} \quad e^n \quad e^{(-t\beta_0)} \\
&= \beta_0^n \quad g(t) \quad e^{(-t\beta_0)}
\end{aligned}
$$

Hence

$$
g(t) \approx (2\pi n)^{-1/2} \quad \hat{\beta}_{ML}^{-n+1} \quad e^n = \frac{e^n}{\sqrt{2\pi n}} \left(\frac{t}{n}\right)^{n-1}
$$

2 Exam: 2018/08/23

Problem 1

Let $Y_1, Y_2, \ldots, Y_n$ be an iid. sample of a Birnbaum-Saunders distributed random variable Y whose density is given by

$$
f(y) = \frac{\sqrt{\frac{y}{\beta}} + \sqrt{\frac{\beta}{y}}}{2\gamma y \sqrt{2\pi}} \exp\left(\frac{-\left(\sqrt{\frac{y}{\beta}} - \sqrt{\frac{\beta}{y}}\right)^2}{2\gamma^2} \right), \quad y, \gamma, \beta > 0
$$

- (a) Derive the expression of the log-likelihood function.
- (b) Calculate the score vector $U(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\gamma, \beta)^T$
- (c) Compute the observed Fisher information matrix $J(\boldsymbol{\theta})$.
- (d) Derive the expression of the profile log-likelihood function for β.
- (e) In this part of Problem 1 we assume that $\beta = 5$. Provide the minimal sufficient statistic for γ and explain your answer.

- (a)
  Likelihood Function:

$$
\begin{aligned}
L(\gamma,\beta;y) = \prod_{i=1}^{n} f(\gamma,\beta;y_i) &= \prod_{i=1}^{n} \frac{\sqrt{\frac{y_i}{\beta}}+\sqrt{\frac{\beta}{y_i}}}{2\gamma y_i\sqrt{2\pi}} \exp\left(\frac{-\left(\sqrt{\frac{y_i}{\beta}}-\sqrt{\frac{\beta}{y_i}}\right)^2}{2\gamma^2}\right) \\
&= (2\gamma\sqrt{2\pi})^{-n} \exp\left(\frac{-\sum_{i=1}^{n}\left(\sqrt{\frac{y_i}{\beta}}-\sqrt{\frac{\beta}{y_i}}\right)^2}{2\gamma^2}\right) \prod_{i=1}^{n} \frac{\sqrt{\frac{y_i}{\beta}}+\sqrt{\frac{\beta}{y_i}}}{y_i}
\end{aligned} \tag{26}
$$

log-Likelihood Function:

$$
\begin{aligned}
l(y) = \log L(y) &= -n\log(2\gamma\sqrt{2\pi}) - \frac{\sum_{i=1}^{n}\left(\sqrt{\frac{y_i}{\beta}}-\sqrt{\frac{\beta}{y_i}}\right)^2}{2\gamma^2} + \log\left(\prod_{i=1}^{n}\frac{\sqrt{\frac{y_i}{\beta}}+\sqrt{\frac{\beta}{y_i}}}{y_i}\right) \\
&= -n\log(2\gamma\sqrt{2\pi}) - \frac{\sum_{i=1}^{n}\left(\sqrt{\frac{y_i}{\beta}}-\sqrt{\frac{\beta}{y_i}}\right)^2}{2\gamma^2} + \sum_{i=1}^{n}\log\left(\sqrt{\frac{y_i}{\beta}}+\sqrt{\frac{\beta}{y_i}}\right) - \log(y_i) \\
&= \sum_{i=1}^{n}\log\left(\sqrt{\frac{y_i}{\beta}}+\sqrt{\frac{\beta}{y_i}}\right) - \log(2\gamma y_i\sqrt{2\pi}) - \frac{\left(\sqrt{\frac{y_i}{\beta}}-\sqrt{\frac{\beta}{y_i}}\right)^2}{2\gamma^2}
\end{aligned} \tag{27}
$$

- (b)
  Score Vector:

  Knowing that $\boldsymbol{\theta} = (\gamma,\beta)^T$,the score vector elements are derived as follows:

$$
\begin{aligned}
U(\boldsymbol{\theta}) = U(\boldsymbol{\theta};t) = \frac{\partial l(\boldsymbol{\theta},t)}{\partial\gamma} &= \sum_{i=1}^{n} -\frac{1}{\gamma} + \frac{\left(\sqrt{\frac{y_i}{\beta}}-\sqrt{\frac{\beta}{y_i}}\right)^2}{\gamma^3} \\
&= -\frac{n}{\gamma} - \frac{2n}{\gamma^3} + \frac{1}{\gamma^3}\sum_{i=1}^{n}\frac{y_i}{\beta}+\frac{\beta}{y_i} \\
\frac{\partial l(\boldsymbol{\theta},t)}{\partial\beta} &= \sum_{i=1}^{n}\frac{1}{2}\Big(-\frac{\sqrt{y_i}}{\beta^{3/2}}+\frac{1}{\sqrt{\beta y_i}}\Big)\Big(\sqrt{\frac{y_i}{\beta}}+\sqrt{\frac{\beta}{y_i}}\Big)^{-1} - \frac{1}{2\gamma^2}\sum_{i=1}^{n}\Big(\frac{-y_i}{\beta^2}+\frac{1}{y_i}\Big) \\
&= \sum_{i=1}^{n}\frac{1}{2}\Big(-\frac{(y_i+\beta)}{\beta\sqrt{\beta y_i}}\Big)\Big(\frac{y_i+\beta}{\sqrt{\beta y_i}}\Big)^{-1} - \frac{1}{2\gamma^2}\sum_{i=1}^{n}\Big(\frac{-y_i}{\beta^2}+\frac{1}{y_i}\Big) \\
&= -\frac{n}{2\beta} - \frac{1}{2\gamma^2}\sum_{i=1}^{n}\Big(\frac{1}{y_i}-\frac{y_i}{\beta^2}\Big)
\end{aligned} \tag{28}
$$

- (c)
  Fisher Information matrix:

$$J(\boldsymbol{\theta}) = J(\gamma, \beta; y) = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\theta};y)}{\partial^2 \gamma} & \frac{\partial^2 l(\boldsymbol{\theta};y)}{\partial \gamma \beta} \\ \frac{\partial^2 l(\boldsymbol{\theta};y)}{\partial \gamma \beta} & \frac{\partial^2 l(\boldsymbol{\theta};y)}{\partial^2 \beta} \end{pmatrix}$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta};y)}{\partial^2 \gamma} &= \frac{n}{\gamma^2} + \frac{6n}{\gamma^4} - \frac{3}{\gamma^4} \sum_{i=1}^{n} \Big(\frac{y_i}{\beta} + \frac{\beta}{y_i} \Big) \\ \frac{\partial^2 l(\boldsymbol{\theta};y)}{\partial^2 \beta} &= \frac{n}{2\beta^2} - \frac{1}{\beta^3 \gamma^2} \sum_{i=1}^{n} y_i \\ \frac{\partial^2 l(\boldsymbol{\theta};y)}{\partial \gamma \beta} = \frac{\partial}{\partial \gamma} \Big(\frac{\partial}{\partial \beta} \Big) &= \frac{1}{\gamma^3} \sum_{i=1}^{n} \Big(\frac{1}{y_i} - \frac{y_i}{\beta^2} \Big) \end{aligned} \tag{29}$$

- (d)
  Profile log-likelihood function:
  The profile likelihood function $L_p(\beta)$ for β is formed as

  $$L_p(\beta) = L(\hat{\gamma}(\beta), \beta)$$

  Where $\hat{\gamma}(\beta)$ is the MLE of γ when β is regarded as given is obtained by:

  $$\begin{aligned} \frac{\partial l(\boldsymbol{\theta}, t)}{\partial \gamma} &= -\frac{n}{\gamma} - \frac{2n}{\gamma^3} + \frac{1}{\gamma^3} \sum_{i=1}^{n} \frac{y_i}{\beta} + \frac{\beta}{y_i} = 0 \\ \hat{\gamma}(\beta) &= \sqrt{\frac{1}{n} \sum_{i=1}^{n} \Big(\frac{y_i}{\beta} + \frac{\beta}{y_i} \Big) - 2} \end{aligned}$$

  $$\begin{aligned} l_p(\hat{\gamma}(\beta), \beta; y) &= \log L_p(\hat{\gamma}(\beta), \beta; y) \\ &= \sum_{i=1}^{n} \log \Big(\sqrt{\frac{y_i}{\beta}} + \sqrt{\frac{\beta}{y_i}} \Big) - \log(2 y_i \sqrt{2\pi}) - \frac{1}{2} \log \Big(\frac{1}{n} \sum_{i=1}^{n} \Big(\frac{y_i}{\beta} + \frac{\beta}{y_i} \Big) - 2 \Big) \\ &- \frac{\Big(\frac{y_i}{\beta} + \frac{\beta}{y_i} - 2 \Big)}{\frac{2}{n} \sum_{i=1}^{n} \Big(\frac{y_i}{\beta} + \frac{\beta}{y_i} \Big) - 4} \end{aligned} \tag{30}$$

- (e)
  Whenever the parameter $\beta = 5$,

  $$\begin{aligned} L(\gamma; y) &= \prod_{i=1}^{n} \frac{y_i + 5}{y_i \sqrt{5 y_i}} \frac{1}{(2\gamma\sqrt{2\pi})^n} \exp\Bigg(-\frac{1}{2\gamma^2} \sum_{i=1}^{n} \Big(\frac{y_i}{5} + \frac{5}{y_i} - 2 \Big) \Bigg) \\ &= \underbrace{\prod_{i=1}^{n} \frac{\frac{y_i+5}{y_i\sqrt{5y_i}}}{2\sqrt{2\pi}}}_{h(y)} \underbrace{\gamma^{-n} \exp\{n/\gamma^2\}}_{C(\theta)^{-1}} \exp\Bigg\{ -\frac{1}{2\gamma^2} \sum_{i=1}^{n} \Big(\frac{y_i}{5} + \frac{5}{y_i} \Big) \Bigg\} \end{aligned} \tag{31}$$

Which has the form of the exponential family with:

$$\begin{aligned} \underline{\textbf{Canonical Parameter:}}\ \ \theta &= -\frac{1}{2\gamma^2} \\ \underline{\textbf{Canonical Statistic :}}\ \ t(y) &= \sum_{i=1}^{n}\Big(\frac{y_i}{5}+\frac{5}{y_i}\Big) \end{aligned} \tag{32}$$

For finding the minimal sufficient statistic, we will compute the ratio between the likelihood of two samples corresponding to the same distribution.

$$\begin{aligned} \frac{L(\gamma;y)}{L(\gamma;x)} &= \frac{h(y)}{h(x)}\exp\Bigg\{-\frac{1}{2\gamma^2}\sum_{i=1}^{n}\Big(\frac{y_i}{5}+\frac{5}{y_i}\Big)+\frac{1}{2\gamma^2}\sum_{i=1}^{n}\Big(\frac{x_i}{5}+\frac{5}{x_i}\Big)\Bigg\} \\ &= \frac{h(y)}{h(x)}\exp\Bigg\{-\frac{1}{2\gamma^2}\sum_{i=1}^{n}\Big(\frac{y_i}{5}+\frac{5}{y_i}\Big)-\Big(\frac{x_i}{5}+\frac{5}{x_i}\Big)\Bigg\} \end{aligned} \tag{33}$$

The minimal sufficient statistic for γ is $t(y)$. The ratio will be constant with respect to γ if and only if $t(y) = t(x)$, i.e whenever

$$\Big(\frac{y_i}{5}+\frac{5}{y_i}\Big) = \Big(\frac{x_i}{5}+\frac{5}{x_i}\Big)$$

Problem 2
Let Y be a log-normally distributed random variable with density given by:

$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}}\exp\Bigg(-\frac{(\ln y-\mu)^2}{2\sigma^2}\Bigg) \quad y>0, \quad \mu\in R, \sigma>0$$

- (a) Show that Y belongs to the exponential family? What is the canonical statistics $t(y)$ and the canonical parameter vector $\boldsymbol{\theta}$?
- (b) Determine the norming constant $C(\boldsymbol{\theta})$.
- (c) Compute $E(\ln(Y))$ in terms of σ and μ.
- (d) Compute $E((\ln(Y))^2)$ in terms of σ and μ.
- (e) Compute $E((\ln(Y))^3)$ in terms of σ and μ.
- (f) Compute $E((\ln(Y))^4)$ in terms of σ and μ.
- (g) Compute $\mathrm{Var}(\ln(Y))$ in terms of σ and μ.

- (a)

Let $Y \sim \text{Lognormal}(\mu, \sigma^2)$ then the density function can be expressed as:

$$\begin{aligned} f(\mu,\theta;y) &= \frac{1}{y\sigma\sqrt{2\pi}} \exp\Big\{-\frac{1}{2\sigma^2}((\log y)^2 + \mu^2 - 2\ \mu(\log y)\ )\Big\} \\ &= \frac{1}{y\ \sqrt{2\pi}}(\sigma)^{-1} \exp\Big\{-\frac{\mu^2}{\sigma^2}\Big\}\ \exp\Big\{-\frac{1}{2\sigma^2}(\log y)^2 + \frac{\mu}{\sigma^2}(\log y)\Big\} \end{aligned} \tag{34}$$

Which has the form of the exponential family with:

$$\begin{aligned} \underline{\textbf{Canonical Parameter:}}\ \ \theta &= \begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} = \begin{pmatrix}\frac{-1}{2\sigma^2}\\ \frac{\mu}{\sigma^2}\end{pmatrix} \\ &\iff \mu = \theta_2\sigma^2, \quad \sigma^2 = -1/2\theta_1 \\ &\Rightarrow \sigma = 1/\sqrt{-2\theta_1} \quad \iff \mu = -\frac{\theta_2}{2\theta_1} \\ \underline{\textbf{Canonical Statistic:}}\ \ t(y) &= \begin{pmatrix}t_1\\ t_2\end{pmatrix} = \begin{pmatrix}(\log y)^2\\ \log y\end{pmatrix} \end{aligned} \tag{35}$$

- (b)
  The normalizing constant $a(\boldsymbol{\theta})$ is given by

$$\begin{aligned} a(\boldsymbol{\theta}) = C(\boldsymbol{\theta}^{-1}) &= (\sigma)^{-1} \exp\Big\{-\frac{\mu^2}{\sigma^2}\Big\} \\ &= \sqrt{-2\theta_1} \exp\Big\{-\Big(-\frac{\theta_2}{2\theta_1}\Big)^2\Big(-2\theta_1\Big)\Big\} \\ &= \sqrt{-2\theta_1} \exp\Big\{\frac{\theta_2^2}{4\theta_1}\Big\} \end{aligned} \tag{36}$$

  Hence,

$$C(\boldsymbol{\theta}) = \frac{1}{\sqrt{-2\theta_1}} \exp\Big\{-\frac{\theta_2^2}{4\theta_1}\Big\}$$

- (c)
  For computing the expected value of the first element of the canonical statistic we need to derive $\log C(\boldsymbol{\theta})$ before.

$$\log C(\boldsymbol{\theta}) = -\frac{1}{2}\log(-2\theta_1) - \frac{\theta_2^2}{4\theta_1}$$

$$E[\log(Y)] = E[t_2] = \mu_{t_1}(\theta_2) = \frac{\partial \log C(\boldsymbol{\theta})}{\theta_2} - \frac{\theta_2}{2\theta_1} = \mu$$

- (d)

$$E[(\log(Y))^2] = E[t_1] = \mu_{t_1}(\theta_1) = \frac{\partial \log C(\boldsymbol{\theta})}{\theta_1} = -\frac{1}{2\theta_1} + \Big(\frac{\theta_2}{2\theta_1}\Big)^2 = \sigma^2 + \mu^2 \tag{37}$$

- (e)
  We know that

$$E[(\log(Y))^3] = E[\log(Y) \cdot (\log(Y))^2] = \text{Cov}(\log(Y), (\log(Y))^2) + E[\log(Y)]E[(\log(Y))^2]$$

  Where

$$\text{Cov}(\log(Y), (\log(Y))^2) = \frac{\partial^2 \log C(\boldsymbol{\theta})}{\partial\theta_1\theta_2} = \frac{\partial}{\partial\theta_1}\Big(\frac{\partial \log C(\boldsymbol{\theta})}{\partial\theta_2}\Big) = \frac{\theta_2}{2\theta_1^2} = \frac{\theta_2}{2\theta_1} \cdot \frac{1}{\theta_1} = 2\mu\sigma^2$$

  Hence

$$E[\log(Y) \cdot (\log(Y))^2] = 2\mu\sigma^2 + \mu(\sigma^2 + \mu^2) = \mu\sigma^2(3 + \mu^2)$$

- (f)
  Knowing that

$$\text{Var}((\log(Y))^2) = E[(\log(Y))^4] - E[(\log(Y))^2]^2$$
$$\Rightarrow E[(\log(Y))^4] = \text{Var}((\log(Y))^2) + E[(\log(Y))^2]^2$$

  Where the variance of t_1 is just the first element in the diagonal of the variance-covariance matrix:

$$\text{Var}((\log(Y))^2) = \frac{\partial^2 \log C(\boldsymbol{\theta})}{\partial^2\theta_1} = \frac{1}{2\theta_1^2} - \frac{\theta_2^2}{2\theta_1^3} = \frac{1}{2\theta_1^2}\Big(1 - \frac{\theta_2^2}{\theta_1}\Big) = -\sigma^2(1 + 2\mu^2/\sigma^2)$$

  Which yields to

$$E[(\log(Y))^4] = \mu^2 - \sigma^2(1 + 2\mu^2/\sigma^2)$$

- (g)
  Similarly as in (f) but now we compute the variance of t_2

$$\text{Var}(\log(Y)) = \frac{\partial^2 \log C(\boldsymbol{\theta})}{\partial^2\theta_2} = -\frac{1}{2\theta_1} = \sigma^2$$

Problem 3
Let Y_1 and Y_2 be two independent random variables with $Y1 \sim Bi(n_1; \pi)$ (binomial distribution with parameters n_1 and π) and $Y_2 \sim Bi(n_2; c\pi)$, respectively

- (a) Derive the joint probability mass function of Y_1 and Y_2.
- (b) Prove that the canonical statistic is $t(Y_1; Y_2) = (v, u)^T$ with $v = Y_1$ and $u = Y_1 + Y_2$. Determine the canonical parameter $\boldsymbol{\theta}$.
- (c) Consider the model reduction hypothesis $H_0 : c = 1$. Compute the marginal probability mass function u under H_0, i.e. $g_0(u)$.
- (d) Derive the conditional distribution of v given u under H_0, i.e. $f_0(v|u)$.
- (e) Calculate the p-value of the test from (d) when $n_1 = 12$ and $n_2 = 8$, and $y_1 = 2$ and $y_2 = 4$ are realizations of Y_1 and Y_2, respectively. Is the null hypothesis rejected at significance level 0.1?
- (f) Derive the statistic of the deviance test for the null hypothesis from (d). What is the asymptotic null distribution of this test statistic?
- (g) Perform the deviance test from (f) at significance level 0.1 by using $n_1 = 12$ and $n_2 = 8$, and $y_1 = 2$ and $y_2 = 4$ as realizations of Y_1 and Y_2, respectively.

- (a)
Knowing that Y_1 is independent of Y_2 we can write the joint probability as a product:

$$\begin{aligned}
f(y_1, y_2) &= P(y_1; n_1, \pi)P(y_2; n_2, c\pi) \\
&= \binom{n_1}{y_1}\pi^{y_1}(1-\pi)^{n_1-y_1}\binom{n_2}{y_2}c\pi^{y_2}(1-c\pi)^{n_2-y_2} \\
&= \binom{n_1}{y_1}\binom{n_2}{y_2}\pi^{y_1}c\pi^{y_2}\frac{(1-\pi)^{n_1}}{(1-\pi)^{y_1}}\frac{(1-c\pi)^{n_2}}{(1-c\pi)^{y_2}} \\
&= \binom{n_1}{y_1}\binom{n_2}{y_2}(1-\pi)^{n_1}(1-c\pi)^{n_2}\left(\frac{\pi}{1-\pi}\right)^{y_1}\cdot\left(\frac{c\pi}{1-c\pi}\right)^{y_2} \\
&= \binom{n_1}{y_1}\binom{n_2}{y_2}(1-\pi)^{n_1}(1-c\pi)^{n_2}\exp\left\{y_1\log\left(\frac{\pi}{1-\pi}\right) + y_2\log\left(\frac{c\pi}{1-c\pi}\right)\right. \\
&\quad\left. - y_1\log\left(\frac{c\pi}{1-c\pi}\right) + y_1\log\left(\frac{c\pi}{1-c\pi}\right)\right\} \\
&= \binom{n_1}{y_1}\binom{n_2}{y_2}(1-c\pi)^{n_1}(1-c\pi)^{n_2}\exp\left\{y_1\Big(\log\left(\frac{\pi}{1-\pi}\right) - \log\left(\frac{c\pi}{1-c\pi}\right)\Big)\right. \\
&\quad\left. + (y_1+y_2)\log\left(\frac{c\pi}{1-c\pi}\right)\right\}
\end{aligned} \tag{38}$$

$$= \binom{n_1}{y_1}\binom{n_2}{y_2}(1-c\pi)^{n_1}(1-c\pi)^{n_2}\exp\left\{y_1\log\left(\frac{1-c\pi}{c(1-\pi)}\right)+(y_1+y_2)\log\left(\frac{c\pi}{1-c\pi}\right)\right\}$$

- (b)
Which has the form of the exponential family with:

$$\begin{aligned}
\underline{\textbf{Canonical Parameter:}}\quad & \theta = \begin{pmatrix}\theta_v\\ \theta_u\end{pmatrix} = \begin{pmatrix}\log\frac{1-c\pi}{c-c\pi}\\ \log\frac{c\pi}{1-c\pi}\end{pmatrix}\\
\underline{\textbf{Canonical Statistic:}}\quad & t(y) = \begin{pmatrix} v = & y_1\\ u = & y_1+y_2\end{pmatrix}\\
\underline{\textbf{Norming Constant:}}\quad & C(\theta) = (1-c\pi)^{-n_1}(1-c\pi)^{-n_2}
\end{aligned} \tag{39}$$

- (c)

$$H_0: c=1 \quad \text{against} \quad H_1: c\neq 1 \quad u = y_1+y_2 \overset{H_0}{\sim} Bi(n_1+n_2,\pi)$$

The probability mass function of u under the null hypothesis is given by:

$$\begin{aligned}
g_0(u) &= \binom{n_1+n_2}{u}\pi^u(1-\pi)^{n_1+n_2-u}\\
&= \binom{n_1+n_2}{u}\left(\frac{\pi}{1-\pi}\right)^u(1-\pi)^{n_1+n_2}
\end{aligned} \tag{40}$$

- (d)
We know that $u = y_1+y_2, v = y_1 \iff y_1 = v, y_2 = u-v$

$$f_0(v|u) = \frac{f_0(v,u)}{g_0(u)}$$

$$\begin{aligned}
f_0(v,u) &= ||J((v,u)\to(y_1,y_2))|| \quad f_0(y_1(v,u),y_2(v,u))\\
&= \begin{pmatrix}\frac{\partial y_1}{\partial v} & \frac{\partial y_1}{\partial u}\\ \frac{\partial y_2}{\partial v} & \frac{\partial y_2}{\partial u}\end{pmatrix}\binom{n_1}{v}\binom{n_2}{u-v}(1-\pi)^{n_1+n_2}\exp\left\{u\log\left(\frac{\pi}{1-\pi}\right)\right\}\\
&= \begin{pmatrix}1 & 0\\ -1 & 1\end{pmatrix}\binom{n_1}{v}\binom{n_2}{u-v}(1-\pi)^{n_1+n_2}\exp\left\{y_1\log\left(\frac{\pi}{1-\pi}\right)+y_2\log\left(\frac{\pi}{1-\pi}\right)\right\}
\end{aligned} \tag{41}$$

$$= \frac{\binom{n_1}{y_1}\binom{n_2}{y_2}}{\binom{n_1+n_2}{u}}\frac{(1-\pi)^{n_1+n_2}\left(\frac{\pi}{1-\pi}\right)^u}{(1-\pi)^{n_1+n_2}\left(\frac{\pi}{1-\pi}\right)^u} = \frac{\binom{n_1}{y_1}\binom{n_2}{y_2}}{\binom{n_1+n_2}{u}}$$

Which turns out to be the p.m.f of the hyper-geometric distribution.

- (e)
Knowing that $n_1 = 12, y_1 = 2, n_2 = 8, y_2 = 4$, we would like to know how extreme is v_{obs}.

The p-value area can be represented by the following sum:

$$\sum_{\{v: f(v|u_{\text{obs}}) \leq f(v_{\text{obs}}|u_{\text{obs}})\}} f(v|u_{\text{obs}}) = \sum_{\{v: f(v|u_{\text{obs}}) \leq f(2|6)\}} f(v|6)$$

$$\iff v = 2, \quad u = 6, \quad v = 0, 1, 2, \ldots$$

$$f_0(v_{obs}|u_{obs}) = f_0(2|6) = \frac{\binom{12}{2}\binom{8}{4}}{\binom{20}{6}} = \frac{12!}{2!10!}\frac{8!}{4!4!}\frac{6!14!}{20!} = \frac{77}{646} = 0.1191$$

In order for us to compute the probability of observing some value of v more extreme than what we already have, we need to find those values of v for which $\{v : f(v|u_{\text{obs}}) \leq f(v_{\text{obs}}|u_{\text{obs}})\}$ holds true.

We have that:

$$\begin{aligned}
f_0(0|6) &= \frac{6!14!}{20!}\binom{12}{0}\binom{8}{6} = \frac{7}{9690} = 0.000 \\
f_0(1|6) &= \frac{6!14!}{20!}\binom{12}{1}\binom{8}{5} = \frac{28}{1615} = 0.017 \\
f_0(2|6) &= \frac{6!14!}{20!}\binom{12}{2}\binom{8}{4} = \frac{77}{646} = 0.1191 \\
f_0(3|6) &= \frac{6!14!}{20!}\binom{12}{3}\binom{8}{3} = \frac{308}{969} = 0.317 \\
f_0(4|6) &= \frac{6!14!}{20!}\binom{12}{4}\binom{8}{2} = \frac{231}{646} = 0.3575 \\
f_0(5|6) &= \frac{6!14!}{20!}\binom{12}{5}\binom{8}{1} = \frac{264}{1615} = 0.1634 \\
f_0(6|6) &= \frac{6!14!}{20!}\binom{12}{6}\binom{8}{0} = \frac{77}{3230} = 0.023
\end{aligned} \tag{42}$$

Which yields to

$$p - \text{value} = f_0(0|6) + f_0(1|6) + f_0(2|6) + f_0(6|6) = 0.161$$

$$\Rightarrow 0.161 > \alpha = 0.1 \Rightarrow \text{We cannot reject the null hypothesis}$$

- (f)

Consider that

$$\begin{aligned}
L(\pi_1,\pi_2) &= \binom{n_1}{y_1}\binom{n_2}{y_2}(1-\pi_1)^{n_1}(1-\pi_2)^{n_2}\left(\frac{\pi_1}{1-\pi_1}\right)^{y_1}\cdot\left(\frac{\pi_2}{1-\pi_2}\right)^{y_2} \\
l(\pi_1,\pi_2) = \log L(\pi_1,\pi_2) &= K + n_1\log(1-\pi_1) + n_2\log(1-\pi_2) + y_1\log\left(\frac{\pi_1}{1-\pi_1}\right) + y_2\log\left(\frac{\pi_2}{1-\pi_2}\right) \\
\frac{\partial l(\pi_1,\pi_2)}{\partial\pi_1} &= -\frac{\pi_1 n_1 + y_1}{\pi_1(1-\pi_1)} = 0 \quad \Rightarrow \hat{\pi}_1 = \frac{y_1}{n_1} \\
\frac{\partial l(\pi_1,\pi_2)}{\partial\pi_2} &= -\frac{\pi_2 n_2 + y_2}{\pi_2(1-\pi_2)} = 0 \quad \Rightarrow \hat{\pi}_2 = \frac{y_2}{n_2}
\end{aligned} \tag{43}$$

Under the null hypothesis $H_0 : c = 1$ we have that:

$$\begin{aligned}
L(\pi_0) = f_0(u,v) &= \binom{n_1}{v}\binom{n_2}{u-v}(1-\pi)^{n_1+n_2}\left(\frac{\pi}{1-\pi}\right)^{y_1+y_2} \\
l(\pi_0) = \log L(\pi_0) &= K + (n_1+n_2)\log(1-\pi) + (y_1+y_2)\log\left(\frac{\pi}{1-\pi}\right) \\
\frac{\partial l(\pi_0)}{\partial\pi_0} &= -\frac{n_1+n_2}{1-\pi} + \frac{y_1+y_2}{\pi} + \frac{y_1+y_2}{1-\pi} = 0 \\
&\Rightarrow \hat{\pi}_0 = \frac{y_1+y_2}{n_1+n_2}
\end{aligned} \tag{44}$$

If we now substitute $\pi_0 \to \hat{\pi}_0$ and $\pi \to \hat{\pi}$, the likelihood ratio test is given by

$$W = 2\log\frac{L(\hat{\pi})}{L(\hat{\pi}_0)} = -2\{\log L(\hat{\pi}_0) - \log L(\hat{\pi})\} \xrightarrow{H_0} \chi^2_{df} = \chi^2_1$$

Where the number of restrictions under the null hypothesis correspond to the degrees of freedom.

$$\begin{aligned}
W &= 2\log\left\{\frac{\binom{n_1}{y_1}\binom{n_2}{y_2}(1-\hat{\pi}_1)^{n_1}(1-\hat{\pi}_2)^{n_2}\left(\frac{\hat{\pi}_1}{1-\hat{\pi}_1}\right)^{y_1}\cdot\left(\frac{\hat{\pi}_2}{1-\hat{\pi}_2}\right)^{y_2}}{\binom{n_1}{v}\binom{n_2}{u-v}(1-\hat{\pi}_0)^{n_1+n_2}\left(\frac{\hat{\pi}_0}{1-\hat{\pi}_0}\right)^{y_1+y_2}}\right\} \\
&= 2\log\left\{\frac{(1-\hat{\pi}_1)^{n_1-y_1}\ (1-\hat{\pi}_2)^{n_2-y_2}\ \ \hat{\pi}_1^{y_1}\ \ \hat{\pi}_2^{y_2}}{(1-\hat{\pi}_0)^{(n_1+n_2)-(y_1+y_2)}\ \ \hat{\pi}_0^{(y_1+y_2)}}\right\}
\end{aligned} \tag{45}$$

- (g)

We have that by using $n_1 = 12$ and $n_2 = 8$, and $y_1 = 2$ and $y_2 = 4$, the deviance test statistic gives:

$$\begin{aligned}
W &= 2\log\left\{\frac{(1-\hat{\pi}_1)^{n_1-y_1}\ (1-\hat{\pi}_2)^{n_2-y_2}\ \ \hat{\pi}_1^{y_1}\ \ \hat{\pi}_2^{y_2}}{(1-\hat{\pi}_0)^{(n_1+n_2)-(y_1+y_2)}\ \ \hat{\pi}_0^{(y_1+y_2)}}\right\} \\
&= 2\log\left\{\frac{(1-\frac{y_1}{n_1})^{n_1-y_1}\ (1-\frac{y_2}{n_2})^{n_2-y_2}\ \ \frac{y_1}{n_1}^{y_1}\ \ \frac{y_2}{n_2}^{y_2}}{(1-\frac{y_1+y_2}{n_1+n_2})^{(n_1+n_2)-(y_1+y_2)}\ \ \frac{y_1+y_2}{n_1+n_2}^{(y_1+y_2)}}\right\} \\
&= 2\log\left\{\frac{(1-\frac{2}{12})^{12-2}\ (1-\frac{4}{8})^{8-4}\ \ \frac{2}{12}^{2}\ \ \frac{4}{8}^{4}}{(1-\frac{6}{20})^{(12+8)-(6)}\ \ \frac{6}{20}^{(6)}}\right\} \\
&= 2\log\left\{\frac{(\frac{5}{6})^{10}\ (\frac{1}{2})^{4}\ \ \frac{2}{12}^{2}\ \ \frac{4}{8}^{4}}{(\frac{7}{10})^{(14)}\ \ \frac{3}{10}^{(6)}}\right\} \\
&= 2\log(3.54441) = 2.53
\end{aligned} \tag{46}$$

At a significance level of $\alpha = 0.1$ and knowing that $\chi^2_{1,0.9} = 2.71 \Rightarrow 2.53 < 2.71$ We reject the null hypothesis.

Problem 4

Let Y be a random variable with probability mass function given by

$$f(y) = \frac{(y+k-1)!}{y!(k-1)!}\pi^k(1-\pi)^y \qquad y = 0, 1, 2, \ldots, \quad \pi \in (0, 1)$$

and known integer k.

- (a) Show that Y belongs to the exponential family and compute its canonical statistics $t(Y)$ as well as canonical parameter $\boldsymbol{\theta}$.
- (b) Determine the norming constant $C(\boldsymbol{\theta})$.
- (c) Compute $\mu = E(Y)$.
- (d) Show that this distribution satisfies the demands for use as ingredient in a generalized linear model. Find the canonical link function.
- (e) Let $Y_1, \ldots, Y_n$ be independent observations with density of Y_i given by

  $$f(y_i) = \frac{(y_i+k-1)!}{y_i!(k-1)!}\pi_i^k(1-\pi_i)^{y_i} \qquad y_i = 0, 1, 2, \ldots, \quad \pi \in (0, 1)$$

  and known integer k. Consider the canonical link function and the linear predictor $\eta_i = \alpha + \beta x_i$ where x_i is a deterministic variable. Derive the likelihood equation system for α and β.
- (f) Find an expression of the deviance, and provide an expression of the square deviance residuals of the generalized linear model from part (e).

- (a)-(b)
  The probability mass function of Y can be rewritten as:

$$f(y;\theta) = \frac{(y+k-1)!}{y!(k-1)!}\pi^k \exp\{y\log(1-\pi)\}$$

  Which has the form of the exponential family with:

$$\begin{aligned}
\underline{\textbf{Canonical Parameter:}} &\ \theta = \log(1-\pi) \Rightarrow \pi = 1-e^\theta \Rightarrow e^\theta = 1-\pi \\
\underline{\textbf{Canonical Statistic:}} &\ t(y) = y \\
\underline{\textbf{Canonical Parameter Space:}} &\ \Theta = (-\infty, 0] \\
\underline{\textbf{Norming Constant:}} &\ C(\theta) = \pi^{-k} = (1-e^\theta)^{-k} \\
\underline{\textbf{Constant Factor:}} &\ h(y) = \frac{(y+k-1)!}{y!(k-1)!}
\end{aligned} \tag{47}$$

- (c)
  We will derive the expected value of the random variable Y by considering the first derivative of the $\log C(\theta)$

  By definition we have that:

$$C(\theta) = \int h(y)\exp\{\theta^T t(y)\}dy \quad \Rightarrow \frac{\partial C(\theta)}{\partial\theta} = C(\theta)\cdot E[t]$$

  Which means that:

$$\frac{\partial \log C(\theta)}{\partial\theta} = \frac{1}{C(\theta)}C(\theta)\cdot E[t] = E[t] = E[Y]$$

  In our case we have that $C(\theta) = \pi^{-k} \Rightarrow \log C(\theta) = -k\log\pi = -k\log(1-e^\theta)$

$$E[Y] = -k\frac{\partial \log(1-e^\theta)}{\partial\theta} = \frac{ke^\theta}{(1-e^\theta)} = k\frac{(1-\pi)}{\pi}$$

- (d)
  Our sample distribution satisfies the main assumptions for a generalized linear model: (i) the distribution of our sample belongs to the exponential family and it is parametrized by a single parameter $\theta \in R$, (ii) the canonical statistic is linear in each observation y, i.e $t(y) = y$.

  The generalized linear model consists usually in three blocks that determines these models.

  - (i) **Linear predictor:** collection of covariates which we would like to use to model some parameters in our model.

$$\eta = \mathbf{x}^T\beta, \quad \dim(\beta) = k < n$$

- **(ii) Distribution type:** This part is related with the exponential family distribution because a GLM is determined with respect to the exponential family distribution.
- **(iii) Link function:** connect the mean value of the canonical statistic which we have in the exponential family to the new parameter which we denote by η which we will like to model by using the independent covariate, also known as some factors in our model.

$$\eta = g(\mu)$$

If the function $g(\mu)$ is of an specific way, we will have a specific type of a GLM specially if it can be made a transformation $\eta = g(\mu) = g(\mu(\theta)) = \theta$. Where θ is the canonical parameter in the exponential family distribution. Then this function is called *canonical link*.

,

In our case, from previous results we know that

$$\mu_t(\theta) = \frac{ke^\theta}{(1-e^\theta)} \iff e^\theta = \frac{\mu(1-e^\theta)}{k} \Rightarrow \frac{\mu}{k} - \frac{\mu e^\theta}{k} - e^\theta = 0 \Rightarrow e^\theta = \frac{\mu}{\mu+k}$$

Hence, the link function $g(\mu)$ can be obtained by taking the inverse of the mean value as a function of the parameter θ, which gives us the **canonical link function**:

$$g(\mu) = \theta = \log\left(\frac{\mu}{\mu+k}\right)$$

- (e)

We have that the **linear predictor** is given by

$$\eta_i = \alpha + \beta x_i$$

The **likelihood Equation** in the matrix form is given by

$$X^T(y - \mu(\beta)) = 0$$

$$\Rightarrow \underbrace{\begin{pmatrix} 0 & X_{11} \\ 0 & X_{12} \\ \vdots & \vdots \\ 0 & X_{1n} \end{pmatrix}^T}_{X} \left(\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{y} - \underbrace{\begin{pmatrix} \mu_1(\beta) \\ \mu_2(\beta) \\ \vdots \\ \mu_n(\beta) \end{pmatrix}}_{\mu(\beta)} \right) = 0 \Rightarrow X \cdot \mu(\beta) - X^T y = 0$$

$$\sum_{i=1}^{n} \mu_i(\beta) X_{1i} - X^T y = 0 \Rightarrow \sum_{i=1}^{n} \frac{ke^{\alpha+\beta X_i}}{1+e^{\alpha+\beta X_i}} X_{1i} - X^T y$$

- (f)

  Residuals are used in order to construct goodness of fit analysis of a model. In the case of the GLM we use the deviance, instead of the classical residuals which are different between observations y and each mean (which we model) $\hat{\mu}$ we define the residual in another way.

$$D = 2\{\log L(y; y) - \log L(\mu(\hat{\beta}); y)\}$$

  Considering that we have independence between the elements, we can rewrite this in the following way

$$2\sum_{i=1}^{n} \log L(y_i; y_i) - \log L(\mu_i(\hat{\beta}); y) = \sum_{i=1}^{n} D_i^2$$

  where D_i^2 is always positive.

  From previous results we have that

$$\mu = \frac{ke^{\theta}}{1-\theta} = \frac{k(1-\pi)}{\pi} \iff 1-\pi = \frac{\mu}{k}\pi \Rightarrow 1 = \pi(\frac{\mu+k}{k}) \iff \pi = \frac{k}{\mu+k}$$

  The **Likelihood function:**

$$\begin{aligned} L(\mu; y) &= \prod_{i=1}^{n} f(y_i; \pi(\mu_i)) \\ &= \prod_{i=1}^{n} \frac{(y_i+k-1)!}{y_i!(k-1)!} \pi_i^k (1-\pi_i)^{y_i} \\ &= \prod_{i=1}^{n} h(y_i)\Big(\frac{k}{\mu_i+k}\Big)^k \Big(\frac{\mu_i}{\mu_i+k}\Big)^{y_i} \end{aligned} \tag{48}$$

  Where $\mu_i = \mu(\alpha + \beta x_i)$,

$$\begin{aligned} D &= 2\{\log L(y; y) - \log L\Big(\begin{pmatrix} \mu(\hat{\alpha} + \hat{\beta}x_1) \\ \vdots \\ \mu(\hat{\alpha} + \hat{\beta}x_n) \end{pmatrix}; y\Big)\} \\ &= 2\Big\{\sum_{i=1}^{n} \log h(y_i) + \sum_{i=1}^{n} k \log\Big(\frac{k}{y_i+k}\Big) + \sum_{i=1}^{n} y_i \log\Big(\frac{y_i}{y_i+k}\Big) \\ &- \sum_{i=1}^{n} \log h(y_i) - \sum_{i=1}^{n} k \log\Big(\frac{k}{\mu(\hat{\alpha} + \hat{\beta}x_i) + k}\Big) - \sum_{i=1}^{n} y_i \log\Big(1 - \frac{k}{\mu(\hat{\alpha} + \hat{\beta}x_i) + k}\Big)\Big\} \\ &= 2\sum_{i=1}^{n} \log\Bigg\{\frac{\Big(\frac{k}{y_i+k}\Big)^k \Big(\frac{y_i}{y_i+k}\Big)^{y_i}}{\Big(\frac{k}{\mu(\hat{\alpha}+\hat{\beta}x_i)+k}\Big)^k \Big(1 - \frac{k}{\mu(\hat{\alpha}+\hat{\beta}x_i)+k}\Big)^{y_i}}\Bigg\} = \sum_{i=1}^{n} D_i^2 \end{aligned} \tag{49}$$