Statistical Models
Final Exam

Marina Herrera
Stockholm University

June 6, 2020

1 Problem 0

"I, the author of this document, hereby guarantee that I have produced these solutions to this home exam without the assistance of any other person. This means that I have for example not discussed the solutions or the home exam with any other person."

2 Problem 1

Let Y have the following probability mass function

$$P(Y=y;\psi) \propto y\psi^y, \quad y=1,2,3..., \quad \psi \in (0,1)$$

(a) Show that Y belongs to the exponential family? What is the canonical statistics $t(Y)$ and the canonical parameter vector θ?

The probability mass function can be rewritten as:

$$f(y,\psi) \propto ye^{y\log\psi} \tag{1}$$

Which has the form of the exponential family with:

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

Hence

$$f(y,\theta) \propto ye^{y\theta}$$

(b) Derive the norming constant $C(\theta)$.

$$C(\theta) = \sum_{y=1}^{\infty} ye^{y\log\psi} = \sum_{y=1}^{\infty} y\psi^y = \psi + 2\psi^2 + 3\psi^3 + ...$$

We know that

$$\sum_{y=1}^{\infty} y\psi^y = \psi \sum_{y=1}^{\infty} y\psi^{y-1} \Rightarrow \sum_{y=1}^{\infty} y\psi^{y-1} = \frac{d}{d\psi}\sum_{y=1}^{\infty}\psi^y = \frac{d}{d\psi}\frac{1}{1-\psi} = \frac{1}{(1-\psi)^2}$$

$$\iff \sum_{y=1}^{\infty} y\psi^y = \psi\frac{1}{(1-\psi)^2}$$

Which yields to

$$C(\theta) = \frac{e^\theta}{(1-e^\theta)^2} \Rightarrow \log C(\theta) = \theta - 2\log(1-e^\theta)$$

c) Compute $E[Y]$ and $E[Y^2]$

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:

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

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 $C(\theta) =$

$$E[Y] = 1 + \frac{2e^\theta}{1-e^\theta} = 1 + \frac{2\psi}{1-\psi}$$

We also know that:

$$\begin{aligned}
\frac{\partial^2 \log C(\theta)}{\partial^2\theta} &= \frac{\partial}{\partial\theta}\Big(\frac{1}{C(\theta)}\cdot\frac{\partial C(\theta)}{\partial\theta}\Big) \\
&= -\frac{1}{C(\theta)^2}\cdot\Big(\frac{\partial C(\theta)}{\partial\theta}\Big)^2 + \frac{1}{C(\theta)}\cdot\frac{\partial^2 C(\theta)}{\partial^2\theta} \\
&= -\frac{1}{C(\theta)^2}\cdot(C(\theta)E[t])^2 + \frac{1}{C(\theta)}C(\theta)E[t^2] \\
&= E[t^2] - (E[t])^2 = \mathrm{Var}(t) = \mathrm{Var}(Y)
\end{aligned} \tag{3}$$

Hence, in our case:

$$\mathrm{Var}(Y) = \frac{\partial}{\partial\theta}\Big(1 + \frac{2e^\theta}{1-e^\theta}\Big) = \frac{2e^\theta}{(1-e^\theta)^2} = \frac{2\psi}{(1-\psi)^2} \tag{4}$$

d) Let $Y_1, Y_2, ..., Y_n$, $n > 2$, be an iid. sample from the distribution with the probability mass function as given in the statement of the problem. Provide the expression of a minimal sufficient statistic and explain your answer.

First we will compute the **Likelihood function** of the sample:

$$L(\psi, y) = \prod_{i=1}^{n} f(y_i; \psi)$$

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

$$\begin{aligned} \frac{L(\psi; y)}{L(\psi; x)} &= \frac{h(y)}{h(x)} \exp\left\{ \right\} \\ &= \frac{h(y)}{h(x)} \exp\left\{ \right\} \end{aligned} \tag{5}$$

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

$$\left(\right) = \left(\right)$$

3 Problem 2

Let $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4, Y_5)^T$ have a multinomial distribution with probability mass function $P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3, Y_4 = y_4, Y_5 = y_5; p_1, p_2, p_3, p_4)$ given by

$$\frac{n!}{y_1!y_2!y_3!y_4!y_5!} p_1^{y_1} p_2^{y_2} p_3^{y_3} p_4^{y_4} (1 - p_1 - p_2 - p_3 - p_4)^{y_5}$$

with $y_1 + y_2 + y_3 + y_4 + y_5 = n$ $p_1, p_2, p_3, p_4, p_5 \in (0, 1)$ and known n.

a)Show that $\mathbf{Y}$ belongs to the exponential family? What is the canonical statistics $t(Y)$ and the canonical parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^T$?

Let $p_5 = 1 - p_1 - p_2 - p_3 - p_4$

We can rewrite the probability mass function as:

$$
\begin{aligned}
&= \frac{n!}{y_1!y_2!y_3!y_4!(n-y_1-y_2-y_3-y_4)!}\exp\{y_1\log p_1 + y_2\log p_2 + y_3\log p_3 + y_4\log p_4 \\
&+ (n-y_1-y_2-y_3-y_4)\log(p_5)\} \\
&= \frac{n!}{y_1!y_2!y_3!y_4!(n-y_1-y_2-y_3-y_4)!}\cdot\exp\{y_1\log p_1 + y_2\log p_2 + y_3\log p_3 + y_4\log p_4 + n\log(p_5) \\
&- y_1\log(p_5) - y_2\log(p_5) - y_3\log(p_5) - y_4\log(p_5)\} \\
&= \frac{n!}{y_1!y_2!y_3!y_4!(n-y_1-y_2-y_3-y_4)!}p_5^n\cdot\exp\left\{y_1\log\left(\frac{p_1}{p_5}\right) + y_2\log\left(\frac{p_2}{p_5}\right) + y_3\log\left(\frac{p_3}{p_5}\right) + y_4\log\left(\frac{p_4}{p_5}\right)\right\}
\end{aligned} \tag{6}
$$

Which has the form of the exponential family with:

$$
\begin{aligned}
&\underline{\textbf{Canonical Parameter:}}\quad \theta = \begin{pmatrix}\theta_1\\ \theta_2\\ \theta_3\\ \theta_4\end{pmatrix} = \begin{pmatrix}\log\frac{p_1}{1-p_1-p_2-p_3-p_4}\\ \log\frac{p_2}{1-p_1-p_2-p_3-p_4}\\ \log\frac{p_3}{1-p_1-p_2-p_3-p_4}\\ \log\frac{p_4}{1-p_1-p_2-p_3-p_4}\end{pmatrix} \\
&\underline{\textbf{Canonical Statistic:}}\quad t(y) = \begin{pmatrix}y_1\\ y_2\\ y_3\\ y_4\end{pmatrix}
\end{aligned} \tag{7}
$$

(b) Derive the norming constant $C(\theta)$

From the canonical parameter vector we can notice that

$$
e^{\theta_i} = \frac{p_i}{1-\sum_{j=1}^4 p_j} \quad \Rightarrow p_i = e^{\theta_i}(1-\sum_{j=1}^4 p_j) \quad \text{for} \quad i = 1,2,3,4
$$

$$
\begin{aligned}
(1-\sum_{j=1}^4 p_j) &= 1 - \sum_{i=1}^4 e^{\theta_i}(1-\sum_{j=1}^4 p_j) \\
&= \left(1-\sum_{j=1}^4 p_j\right)\left(\frac{1}{1-\sum_{j=1}^4 p_j} - \sum_{i=1}^4 e^{\theta_i}\right) \\
1 &= \frac{1}{1-\sum_{j=1}^4 p_j} - \sum_{i=1}^4 e^{\theta_i} \\
\frac{1}{1-\sum\limits_{j=1}^4 p_j} &= 1 + \sum_{i=1}^4 e^{\theta_i}
\end{aligned} \tag{8}
$$

We know that

$$\begin{aligned}\underline{\textbf{Normalizing Constant:}}\quad a(\theta) &= (1-\sum_{j=1}^{4}p_j)^n \\ \underline{\textbf{Norming Constant:}}\quad C(\theta) &= (1-\sum_{j=1}^{4}p_j)^{-n}\end{aligned} \tag{9}$$

$$C(\theta) = \frac{1}{(1-\sum_{j=1}^{4}p_j)^n} = \left(1+\sum_{i=1}^{4}e^{\theta_i}\right)^n$$

Hence,

$$\log C(\theta) = n\log\left(1+\sum_{i=1}^{4}e^{\theta_i}\right)$$

c)Calculate the expected value of the canonical statistics $\boldsymbol{\mu} = E(t(Y))$.

We know that

$$\boldsymbol{\mu} = \begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \mu_3\end{pmatrix} = \begin{pmatrix}E[t(y_1)]\\ E[t(y_2)]\\ E[t(y_3)]\\ E[t(y_4)]\end{pmatrix} = \begin{pmatrix}E[y_1]\\ E[y_2]\\ E[y_3]\\ E[y_4]\end{pmatrix} = \begin{pmatrix}\frac{\partial\log C(\theta)}{\partial\theta_1}\\ \frac{\partial\log C(\theta)}{\partial\theta_2}\\ \frac{\partial\log C(\theta)}{\partial\theta_3}\\ \frac{\partial\log C(\theta)}{\partial\theta_4}\end{pmatrix} = \begin{pmatrix}\frac{ne^{\theta_1}}{1+\sum_{i=1}^{4}e^{\theta_i}}\\ \frac{ne^{\theta_2}}{1+\sum_{i=1}^{4}e^{\theta_i}}\\ \frac{ne^{\theta_3}}{1+\sum_{i=1}^{4}e^{\theta_i}}\\ \frac{ne^{\theta_4}}{1+\sum_{i=1}^{4}e^{\theta_i}}\end{pmatrix}$$

d) Compute the expected Fisher information $I(\boldsymbol{\theta})$ in the canonical parametrization.

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

$$= \begin{pmatrix}\mathrm{Var}(\theta_1) & & & \\ \mathrm{Cov}(\theta_1,\theta_2) & \mathrm{Var}(\theta_2) & \dots & \\ \mathrm{Cov}(\theta_3,\theta_3) & \mathrm{Cov}(\theta_3,\theta_2) & \mathrm{Var}(\theta_3) & \\ \mathrm{Cov}(\theta_4,\theta_1) & \mathrm{Cov}(\theta_4,\theta_2) & `\,\mathrm{Cov}(\theta_4,\theta_3) & \mathrm{Var}(\theta_4)\end{pmatrix} = \begin{pmatrix}\frac{\partial^2\log C(\theta)}{\partial^2\theta_1} & & & \\ \frac{\partial^2\log C(\theta)}{\partial\theta_1\theta_2} & \frac{\partial^2\log C(\theta)}{\partial^2\theta_2} & & \\ \frac{\partial^2\log C(\theta)}{\partial\theta_3\theta_1} & \frac{\partial^2\log C(\theta)}{\partial\theta_3\theta_2} & \frac{\partial^2\log C(\theta)}{\partial^2\theta_3} & \\ \frac{\partial^2\log C(\theta)}{\partial\theta_4\theta_1} & \frac{\partial^2\log C(\theta)}{\partial\theta_4\theta_2} & \frac{\partial^2\log C(\theta)}{\partial\theta_4\theta_3} & \frac{\partial^2\log C(\theta)}{\partial^2\theta_4}\end{pmatrix}$$

$$\operatorname{Var}(\theta_i) = \frac{\partial^2 \log C(\theta)}{\partial^2 \theta_i} = \frac{ne^{\theta_i}}{1+\sum_{j=1}^{4} e^{\theta_j}} - \frac{ne^{\theta_i}e^{\theta_i}}{(1+\sum_{j=1}^{4} e^{\theta_j})^2} = \mu_i(1-\mu_i/n)$$

$$\begin{aligned}\operatorname{Cov}(\theta_i,\theta_j) &= \frac{\partial^2 \log C(\theta)}{\partial\theta_i\theta_j} = \frac{\partial}{\partial\theta_i}\Big(\frac{\partial \log C(\theta)}{\partial\theta_j}\Big) = \frac{\partial}{\partial\theta_i}\Big(\frac{ne^{\theta_j}}{1+\sum_{m=1}^{4} e^{\theta_m}}\Big) \\ &= -\frac{ne^{\theta_i}e^{\theta_j}}{(1+\sum_{m=1}^{4} e^{\theta_m})^2} = -\frac{\mu_i\mu_j}{n} \quad \text{for } i \neq j \quad i,j = 1,2,3,4\end{aligned} \tag{10}$$

$$\Rightarrow I(\theta) = \operatorname{Var}(\theta) = \begin{pmatrix} \mu_1(1-\mu_1/n) & & & \\ -\mu_2\mu_1/n & \mu_2(1-\mu_2/n) & & \\ -\mu_3\mu_1/n & -\mu_3\mu_2/n & \mu_3(1-\mu_3/n) & \\ -\mu_4\mu_1/n & -\mu_4\mu_2/n & -\mu_4\mu_3/n & \mu_4(1-\mu_4/n) \end{pmatrix}$$

e) What is the expected Fisher information matrix $I(\boldsymbol{\mu})$ in the mean value parametrization $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)^T$? Present the results in terms of $\boldsymbol{\mu}$.

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

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

$$\frac{\partial\mu}{\partial\theta} = \operatorname{Var}(\theta) = \operatorname{Var}(t) = \operatorname{Var}(t)^T$$

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

$$I_\mu(\mu) = \operatorname{Var}(t)^{-1} = \frac{1}{|\operatorname{Var}(t)|}\tilde{\operatorname{Var}}(t) \tag{11}$$

4 Problem 3

Let Y_1 and Y_2 be two independent random variables with $Y_1 \sim Gamma(\alpha_1, \beta_1)$ (Gamma distribution with shape parameter $\alpha_1 > 0$ and rate parameter $\beta_1 > 0$) and $Y_2 \sim Gamma(\alpha_2, \beta_2)$ whose densities are given by:

$$f(y_1; \beta_1) = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} y_1^{\alpha_1 - 1} \exp\{-\beta_1 y_1\} \quad y_1 > 0, \quad \alpha_1, \beta_1 > 0$$

and

$$f(y_2; \beta_2) = \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} y_2^{\alpha 2 - 1} \exp\{-\beta_2 y_2\} \quad y_2 > 0, \quad \alpha_2, \beta_2 > 0$$

a) Derive the joint probability mass function of Y_1 and Y_2.

Knowing that Y_1 is independent of Y_2 we can write the joint density as a product:

$$\begin{aligned} f(y_1, y_2) &= Pf(y_1, \beta_1) P(y_1, \beta_2) \\ &= \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1 - 1} y_2^{\alpha 2 - 1} \exp\{-\beta_1 y_1 - \beta_2 y_2\} \end{aligned} \tag{12}$$

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 $\boldsymbol{\theta}$

The joint density function can be rewritten as:

$$\begin{aligned} &= \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1 - 1} y_2^{\alpha 2 - 1} \exp\{-\beta_1 y_1 - \beta_2 y_2 - \beta_2 y_1 + \beta_2 y_1\} \\ &= \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1 - 1} y_2^{\alpha 2 - 1} \exp\{-\beta_2(y_1 + y_2) + y_1(\beta_2 - \beta_1)\} \end{aligned} \tag{13}$$

Which has the form of the exponential family with:

$$\begin{aligned} \underline{\textbf{Canonical Parameter:}} \quad \theta &= \begin{pmatrix} \theta_v \\ \theta_u \end{pmatrix} = \begin{pmatrix} \beta_2 - \beta_1 \\ -\beta_2 \end{pmatrix} \\ \underline{\textbf{Canonical Statistic:}} \quad t(Y_1, Y_2) &= \begin{pmatrix} v = & y_1 \\ u = & y_1 + y_2 \end{pmatrix} \end{aligned} \tag{14}$$

c) The aim is to test the model reduction hypothesis:

$$H_0 : \psi = 0 \quad \text{against} \quad H_1 : \psi \neq 0$$

with $\psi = \beta_2 - \beta_1$ Calculate the marginal probability mass function $f_0(u)$ and specify the conditional distribution $f_0(v|u)$ under H_0

$$H_0 : \beta_2 = \beta_1 \quad \text{against} \quad H_1 : \beta_2 \neq \beta_1 \quad u = y_1 + y_2 \overset{H_0}{\sim} \Gamma(\alpha_1 + \alpha_2, \beta)$$

Note that under the null hypothesis we will now switch to write the probability mass function with parameter β as we consider $\beta_1 = \beta_2$

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

$$g_0(u) = f(u; \beta) = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} u^{(\alpha_1+\alpha_2)-1} \exp\{-\beta u\} \tag{15}$$

We know that

$$u = y_1 + y_2, v = y_1 \iff y_1 = v, y_2 = u - v$$

The conditional distribution is given by:

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

$$\begin{aligned} f_0(v, u) &= ||J((v, u) \to (y_1, y_2))|| \quad f_0(y_1(v, u), y_2(v, u)) \\ &= \det \begin{pmatrix} \frac{\partial y_1}{\partial v} & \frac{\partial y_1}{\partial u} \\ \frac{\partial y_2}{\partial v} & \frac{\partial y_2}{\partial u} \end{pmatrix} \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1} y_2^{\alpha 2-1} \exp\{-\beta(y_1 + y_2)\} \\ &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1} y_2^{\alpha 2-1} \exp\{-\beta(y_1 + y_2)\} \end{aligned} \tag{16}$$

$$\begin{aligned} f_0(v|u) &= \frac{\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1} y_2^{\alpha 2-1} \exp\{-\beta u\}}{\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} u^{(\alpha_1+\alpha_2)-1} \exp\{-\beta u\}} \\ &= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{v^{\alpha_1-1}(u - v)^{\alpha 2-1}}{u^{(\alpha_1+\alpha_2)-1}} \end{aligned} \tag{17}$$

> d) Calculate the p-value of the test from (c) i$f\alpha_1 = 5$ and $\alpha_2 = 1$, and the realizations of Y_1 and Y_2 are $y_1 = 8$ and $y_2 = 4$, respectively. Is the null hypothesis rejected at significance level 0.1?

We would like to know how extreme is v_{obs}. The p-value area can be represented by the following sum:

$$\int_{\{v: f(v|u_{\text{obs}}) \leq f(v_{\text{obs}}|u_{\text{obs}})\}} f(v|u_{\text{obs}}) = \int_{\{v: f(v|u_{\text{obs}}) \leq f(8|12)\}} f(v|12)$$

$$\iff v = 8, \quad u = 12, \quad v > 0$$

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. Given that $v : f(v|u_{\text{obs}})$ is a non-decreasing function of v, this will be true for: $0 < v \leq 8$.

We have that:

$$f_0(v|u_{obs}) = f_0(v|12) = \frac{\Gamma(6)}{\Gamma(5)\Gamma(1)}\frac{v^4}{12^5} = 6\frac{v^4}{12^5}$$

$$\int_0^8 6\frac{v^4}{12^5} d\ v = 0.15802 > 0.10$$

We cannot reject the null hypothesis.

> e) Derive the statistic of the deviance test for the null hypothesis from (c). What is the asymptotic null distribution of this test statistic?

Consider that

$$\begin{aligned} L(\beta_1, \beta_2) = f(y_1, y_2) &= \frac{\beta_1^{\alpha_1}\beta_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1 - 1} y_2^{\alpha 2 - 1} \exp\{-\beta_1 y_1 - \beta_2 y_2\} \\ l(\beta_1, \beta_2) = \log L(\beta_1, \beta_2) &= K + \alpha_1 \log \beta_1 + \alpha_2 \log \beta_2 - \beta_1 y_1 - \beta_2 y_2 \\ \frac{\partial l(\beta_1, \beta_2)}{\partial \beta_1} &= \frac{\alpha_1}{\beta_1} - y_1 = 0 \quad \Rightarrow \hat{\beta}_1 = \frac{\alpha_1}{y_1} \\ \frac{\partial l(\beta_1, \beta_2)}{\partial \beta_2} &= \frac{\alpha_2}{\beta_2} - y_2 = 0 \quad \Rightarrow \hat{\beta}_2 = \frac{\alpha_2}{y_2} \end{aligned} \tag{18}$$

Under the null hypothesis $H_0 : \beta_1 = \beta_2$ we have that:

$$\begin{aligned} L(\beta_0) = f_0(u, v) &= \frac{y_1^{\alpha_1 - 1} y_2^{\alpha 2 - 1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \beta^{\alpha_1 + \alpha_2} \exp\{-\beta(y_1 + y_2)\} \\ l(\beta_0) = \log L(\beta_0) &= K + (\alpha_1 + \alpha_2) \log \beta - \beta(u) \\ \frac{\partial l(\beta_0)}{\partial \beta_0} &= \frac{\alpha_1 + \alpha_2}{\beta} - u = 0 \\ \Rightarrow \hat{\beta}_0 &= \frac{\alpha_1 + \alpha_2}{u} = \frac{\alpha_1 + \alpha_2}{y_1 + y_2} \end{aligned} \tag{19}$$

If we now substitute $\beta_0 \rightarrow \hat{\beta}_0$ and , $\beta_2 \rightarrow \hat{\beta}_2$, the likelihood ratio test is given by

$$W = 2 \log \frac{L(\hat{\beta}_1, \beta_2)}{L(\hat{\beta}_0)} = -2\{\log L(\hat{\beta}_0) - \log L(\hat{\beta}_1, \hat{\beta}_2)\} \xrightarrow{H_0} \chi^2_{df} = \chi^2_1$$

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

$$
\begin{aligned}
W &= 2\log\left\{\frac{\frac{\hat{\beta}_1^{\alpha_1}\hat{\beta}_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1} y_2^{\alpha 2-1}\exp\{-\hat{\beta}_1 y_1 - \hat{\beta}_2 y_2\}}{\frac{y_1^{\alpha_1-1} y_2^{\alpha 2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\hat{\beta}_0^{\alpha_1+\alpha_2}\exp\{-\hat{\beta}_0(y_1+y_2)\}}\right\} \\
&= 2\log\left\{\frac{\hat{\beta}_1^{\alpha_1}\hat{\beta}_2^{\alpha_2}\exp\{-\alpha_1-\alpha_2\}}{\hat{\beta}_0^{\alpha_1+\alpha_2}\exp\{-(\alpha_1+\alpha_2)\}}\right\} \\
&= 2\log\left\{\frac{\hat{\beta}_1^{\alpha_1}\hat{\beta}_2^{\alpha_2}}{\hat{\beta}_0^{\alpha_1+\alpha_2}}\right\}
\end{aligned} \tag{20}
$$

f) Perform the deviance test from (e) at significance level 0.1 by using that $\alpha_1 = 200$ and $\alpha_2 = 300$, and the realizations of Y_1 and Y_2 are $y_1 = 145$ and $y_2 = 239$, respectively.

$$
\begin{aligned}
W &= 2\log\left\{\frac{\hat{\beta}_1^{\alpha_1}\hat{\beta}_2^{\alpha_2}}{\hat{\beta}_0^{\alpha_1+\alpha_2}}\right\} \\
&= 2\log\left\{\frac{\left(\frac{\alpha_1}{y_1}\right)^{\alpha_1}\left(\frac{\alpha_2}{y_2}\right)^{\alpha_2}}{\left(\frac{\alpha_1+\alpha_2}{y_1+y_2}\right)^{\alpha_1+\alpha_2}}\right\} \\
&= 2\log\left\{\frac{\left(\frac{200}{145}\right)^{200}\left(\frac{300}{239}\right)^{300}}{\left(\frac{200+300}{145+239}\right)^{200+300}}\right\} \\
&= 1.0592
\end{aligned} \tag{21}
$$

At a significance level of $\alpha = 0.1$ and knowing that $\chi^2_{1,0.9} = 2.71 \Rightarrow 1.0592 < 2.71$ We reject the null hypothesis.

5 Problem 4

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

$$f(y;\pi) = (1-\pi)^{y-1}\pi \quad y = 1, 2, ..\pi \in (0,1)$$

a) Show that Y belongs to the exponential family and compute its canonical statistics $t(Y)$ as well as canonical parameter θ.

We can rewrite the probability mass function as:

$$f(y;\pi) = \frac{\pi}{1-\pi}\exp\{y\log(1-\pi)\}$$

Which has the form of the exponential family with:

$$\begin{aligned} \underline{\textbf{Canonical Parameter:}} \ & \theta = \log(1-\pi) \Rightarrow \pi = 1 - e^{\theta} \Rightarrow e^{\theta} = 1 - \pi \\ \underline{\textbf{Canonical Statistic:}} \ & t(y) = y \end{aligned} \tag{22}$$

b) Derive the norming constant $C(\theta)$.

$$\underline{\textbf{Constant Factor:}} \ C(\theta) = \left(\frac{\pi}{1-\pi}\right)^{-1} = \frac{1-\pi}{\pi} = \frac{e^{\theta}}{1-e^{\theta}} \tag{23}$$

$$\Rightarrow \log C(\theta) = \theta - \log(1-e^{\theta})$$

c) Compute $\mu = E(Y)$.

$$E[Y] = \frac{\partial \log C(\theta)}{\partial \theta} = 1 + \frac{e^{\theta}}{1-e^{\theta}} = \frac{1}{1-e^{\theta}}$$

d) Show that this distribution satisfies the demands for use as ingredient in a generalized linear model. Find the canonical link function

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

$$\mu_t(\theta) = \frac{1}{1-e^{\theta}} \iff e^{\theta} = \mu$$

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**:

$$g(\mu) = \theta = \log\Big(\mu\Big)$$

e) Let $Y_1, ..., Y_n$ be independent observations with probability mass function of Y_i given by be independent observations with probability mass function of Y_i given by

$$f(y_i; \pi_i) = \frac{\pi_i}{1-\pi_i} \exp\{y_i \log(1-\pi_i)\} \quad y = 1, 2, ..., \pi_i \in (0,1)$$

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.

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

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

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

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

f) Find an expression of the deviance, and provide an expression of the squared deviance residuals of the generalized linear model from part (e).

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

$$\mu = \frac{1}{1-e^{\theta}} = \frac{1}{\pi} \iff \pi = \frac{1}{\mu}$$

The **Likelihood function:**

$$\begin{aligned} L(\mu; y) &= \prod_{i=1}^{n} f(y_i; \pi(\mu_i)) \\ &= \prod_{i=1}^{n} \frac{\pi_i}{1-\pi_i} \exp\{y_i \log(1-\pi_i)\} \\ &= \prod_{i=1}^{n} \frac{\pi_i}{1-\pi_i} (1-\pi_i)^{y_i} \\ &= \prod_{i=1}^{n} \frac{1}{\mu-1} \Big(1 - \frac{1}{\mu_i}\Big)^{y_i} \end{aligned} \tag{24}$$

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

$$\begin{aligned} D &= 2\{\log L(y; y) - \log L\Big(\begin{pmatrix} \mu(\hat{\alpha} + \hat{\beta}x_1) \\ \vdots \\ \mu(\hat{\alpha} + \hat{\beta}x_n) \end{pmatrix}; y\Big)\} \\ &= 2\Big\{\sum_{i=1}^{n} \log h(y_i) + \sum_{i=1}^{n} y_i \log\Big(1 - \frac{1}{\mu_i}\Big) - \sum_{i=1}^{n} \log(\mu - 1) - \sum_{i=1}^{n} \log h(y_i) \\ &- \sum_{i=1}^{n} y_i \log\Big(1 - \frac{1}{\mu(\hat{\alpha} + \hat{\beta}x_i)}\Big) + \sum_{i=1}^{n} \log(\mu(\hat{\alpha} + \hat{\beta}x_i) - 1)\Big\} \\ &= 2\sum_{i=1}^{n} \log\left\{\frac{\Big(1 - \frac{1}{\mu_i}\Big)^{y_i} (\mu(\hat{\alpha} + \hat{\beta}x_i) - 1)}{\Big(1 - \frac{1}{\mu(\hat{\alpha}+\hat{\beta}x_i)}\Big)^{y_i} (\mu_i - 1)}\right\} = \sum_{i=1}^{n} D_i^2 \end{aligned} \tag{25}$$

6 Problem 6

Let data be a sample of size $n > 1$ from $\mathcal{N}(\mu, \sigma^2)$. Find the profile likelihood functions for μ and for σ^2.

If we consider a sample of independent and normally distributed random variables, the **Likelihood function** is obtained by:

$$\begin{aligned} L(\mu, \sigma^2; X) &= \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) \\ &= (2\pi)^{-n/2} \sigma^{-n} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\} \end{aligned} \tag{26}$$

the **log-Likelihood function** is obtained by:

$$l(\mu, \sigma^2; X) = \log L(\mu, \sigma^2; X) = -\frac{n}{2} \log 2\pi - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \tag{27}$$

The MLE for σ^2 and μ are given by:

$$\begin{aligned} \hat{\mu} &= \frac{\sum_{i=1}^{n} x_i}{n} \\ \hat{\sigma}^2(\mu) &= \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} \end{aligned} \tag{28}$$

The profile likelihood function $L_p(\mu)$ for μ is formed as

$$L_p(\mu) = L(\hat{\sigma}^2(\mu), \mu) = (2\pi \Big(\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} \Big))^{-n/2} \exp\{-\frac{1}{2\Big(\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}\Big)} \sum_{i=1}^{n} (x_i - \mu)^2\}$$

While

$$L_p(\mu) = L(\mu(\hat{\sigma}^2), \hat{\sigma}^2) = (2\pi)^{-n/2} \sigma^{-n} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \Big(\sum_{i=1}^{n} x_i \Big)/n)^2\}$$