Advanced Statistical Inference

Assignment 3

Anna Felip and Arnau Garcia

An assessment presented for the course of
Advanced Statistical Inference 2023/24 (MESIO)

Facultat de Matemàtiques i Estadística
Universitat Politècnica de Catalunya
Date of submission: 02/Jan/2024
Teachers: Guadalupe Gómez and Alex Sànchez

Contents

Previous observations

Before starting with the solution of the problems, it is important to note that the theoretical exercises will be presented first, in this latex document. After the latex document the reader will find the pdf document of an rmd with problems that require R code. We think that in this way we get a better presentation, more compact and we can exploit a better theoretical exposure with latex and then practice with R. Anyway, we will comment during the document where the respective solutions are found.

Chapter 1

Problem 1

Let $X_1, \cdots, X_n, Y_1, \cdots Y_n$ be independent random variables where the density of X_i is

$$f_i(x) = \beta_i \phi \exp(-\beta_i \phi x) \quad \text{for } x \geq 0$$

and the density of Y_i is

$$g_i(x) = \beta_i \exp(-\beta_i x) \quad \text{for } x \geq 0$$

where $\beta_1, \cdots, \beta_n$ and ϕ are unknown parameters.

(a) Show that the MLE of $\phi, \hat{\phi}$, based on $X_1, \ldots, X_n, Y_1, \ldots Y_n$ satisfies the equation

$$\frac{n}{\widehat{\phi}} - 2\sum_{i=1}^{n} \frac{R_i}{1+\widehat{\phi} R_i} = 0$$

where $R_i = X_i / Y_i$.

(b) Show that the density of R_i is

$$f_R(x; \phi) = \phi(1+\phi x)^{-2} \quad \text{for } x \geq 0$$

and show that the MLE for ϕ based on $R_1, \ldots, R_n$ is the same as that given in part (a).

(c) Let $\widehat{\phi}_n$ be the MLE in part (b). Find the limiting distribution of $\sqrt{n}\left(\widehat{\phi}_n - \phi\right)$.

(d) Give an estimate of the standard error for the maximum likelihood estimate computed in part (c).

(e) Use the data for $(X_i, Y_i), i = 1, \ldots, 20$ given in the Table below to compute the maximum likelihood estimate of ϕ using either the Newton-Raphson or Fisher scoring algorithm. Find an appropriate starting value for the iterations and justify your choice.

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
0.7	3.8	20.2	2.8	1.1	2.8	15.2	8.8
11.3	4.6	0.3	1.9	1.9	3.2	0.2	7.6
2.1	2.1	0.9	1.4	0.5	8.5	0.7	1.3
30.7	5.6	0.7	0.4	0.8	14.5	0.4	2.2
4.6	10.3	2.3	0.9	1.2	14.4	2.3	4.0

Solution (a): To demonstrate that the MLE of ϕ based on $X_1, \dots, X_n, Y_1, \dots Y_n$ satisfies the equation on the statements we will use the profile likelihood. For our purposes the parameters β_i, for all i, act as a nuisance parameters. Then, we are interested in working with the profile likelihood to estimate the MLE of ϕ, which is:

$$L_{pr}(\phi) = \max_{\beta_i} L(\phi, \beta_i) = L(\phi, \hat{\beta}_{i_{ML}}(\phi)).$$

Then, what we have to do first is to compute $\hat{\beta}_{i_{ML}}(\phi)$, and we do this fixing ϕ and computing the MLE for each β_i. So, because $X_1, \dots, X_n, Y_1, \dots Y_n$ are independent we compute first the likelihood function as:

$$\begin{aligned} L(\beta_i, \phi | X_1, \dots, X_n, Y_1 \dots, Y_n) &= \prod_{i=1}^{n} \beta_i \phi \exp\left(-\beta_i \phi x_i\right) \prod_{i=1}^{n} \beta_i \exp\left(-\beta_i y_i\right) \\ &= \left(\prod_{i=1}^{n} \beta_i^2\right) \phi^n \exp\left(-\sum_{i=1}^{n} \beta_i (\phi X_i + Y_i)\right). \end{aligned}$$

Thus, the log-likelihood is:

$$\begin{aligned} \log(L(\beta_i, \phi | X_1, \dots, X_n, Y_1 \dots, Y_n)) &= 2\sum_{i=1}^{n} \log(\beta_i) + n\log(\phi) - \sum_{i=1}^{n} \beta_i(\phi x_i + y_i) \\ &= 2\sum_{i=1}^{n} \log(\beta_i) + n\log(\phi) - \sum_{i=1}^{n} x_i \beta_i \frac{1 + \phi R_i}{R_i}. \end{aligned}$$

In the last equation we have introduced $R_i = X_i / Y_i$. Now, we compute the derivative of the log-likelihood with respect to β_i and we find the roots:

$$\begin{aligned} &\frac{\partial \log(L(\beta_i, \phi | X_1, \dots, X_n, Y_1 \dots, Y_n))}{\partial \beta_i} = \frac{2}{\beta_i} - x_i \frac{1 + \phi R_i}{R_i} = 0 \iff \\ &\iff \beta_i = 2\frac{R_i}{x_i(1 + \phi R_i)} \end{aligned}$$

And the second derivative of the log-likelihood with respect β_i is:

$$\frac{\partial^2 \log(L(\beta_i, \phi | X_1, \dots, X_n, Y_1 \dots, Y_n))}{\partial \beta_i^2} = -\frac{2}{\beta_i^2}$$

Which is, clearly, negative for every possible value of β_i. Thus, indeed, $\hat{\beta}_{i_{ML}} = 2\frac{R_i}{x_i(1+\phi R_i)}$ is the MLE of β_i. Notice that this MLE is for all the beta parameters, it is $\hat{\beta}_{1_{ML}}, \ldots, \hat{\beta}_{n_{ML}}$ are the MLEs of the parameters $\beta_1, \ldots, \beta_n$. Now, we can compute the MLE of ϕ using the profile likelihood, which can be obtained by plugging in the MLE of β_i in the previously exposed likelihood function.

Then, we compute the derivative of the log-profile-likelihood with respect ϕ:

$$\frac{\partial \log(L(\phi, \hat{\beta}_{i_{ML}} | X_1, \ldots, X_n, Y_1 \ldots, Y_n))}{\partial \phi} = \frac{n}{\phi} - \sum_{i=1}^{n} \hat{\beta}_{i_{ML}} x_i = 0 \Longleftrightarrow$$
$$\Longleftrightarrow \frac{n}{\phi} - 2\sum_{i=1}^{n} \frac{R_i}{1+\phi R_i} = 0$$

And then, the MLE for ϕ, say it $\hat{\phi}$, satisfies the equation on the statements.

Note: It is not direct to see that the second derivative of the log-likelihood with respect to ϕ is negative for the value of ϕ that satisfies the equation. And, knowing how is the value of ϕ that satisfies the equation seems to be impossible analytically. Notwithstanding, we know that the MLE of ϕ (whose existence is guaranteed in this case, because both densities belong to the exponential family and the regularity conditions are fulfilled) must satisfy the obtained equation.

SOLUTION (b): In the first part of this section we will be using the Change of Variable Theorem in a way that is different from what is normally used. Here we have a transformation that involves two random variables, $R_i = X_i/Y_i$. It is important to take into account that this transformation is well defined, because Y_i is a non-negative random variable. In addition, of course, we have to take into account that the random variables X_i, Y_i are independent, by hypothesis. Now, observe that we can define the following transformation:

$$\begin{cases} R_i &= \frac{X_i}{Y_i} \\ Z_i &= Y_i \end{cases}$$

Notice that this transformation from X_i, Y_i to R_i, Z_i is bijective. This is clear because Z_i can be mapped directly to Y_i, and for a given Y_i the quotient $R_i = X_i/Y_i$ is monotonic. These properties will allow us to apply the Change of Variable Theorem. The inverse transformation of the transformation defined is

$$\begin{cases} X_i &= R_i Z_i \\ Y_i &= Z_i \end{cases}$$

And the determinant of the Jacobian matrix of the inverse transformation is:

$$\begin{vmatrix} \frac{\partial X_i}{\partial R_i} & \frac{\partial X_i}{\partial Z_i} \\ \frac{\partial Y_i}{\partial R_i} & \frac{\partial Y_i}{\partial Z_i} \end{vmatrix} = \begin{vmatrix} Z_i & R_i \\ 0 & 1 \end{vmatrix} = Z_i$$

And now, using the Change of Variable Theorem and the fact that X_i, Y_i are independent random variables, we can deduce the following:

$$f_{R_i,Z_i}(r,z) = f_{X_i,Y_i}(x,y)J(x,y|r,z) = f_i(x)g_i(y)J(x,y|r,z) = f_i(rz)g_i(z)|z|.$$

Where we have used the notation introduced in the statements for the densities of X_i and Y_i.

And the density of R_i can be computed by marginalizing out Z_i:

$$f_{R_i}(r) = \int_{-\infty}^{+\infty} f_i(rz)g_i(z)|z|dz = \beta_i^2\phi \int_0^{+\infty} \exp\left(-\beta_i z(1+r\phi)\right)zdz = \frac{\phi}{(1+r\phi)^2}.$$

The computation of the last integral is not difficult, but the whole calculus are not added here. What we have done for compute it is: use that the support of the function inside the integral is $(0,+\infty)$, and integrate by parts twice. And as the reader can observe, what we have obtained is, precisely, the expression for the density of R_i presented in the statements.

The second part of this section is show that the MLE for ϕ based on $R_1,\dots,R_n$ is the same as the one given in the section (a). Once we have proof that the density of the random variable R_i is the one on the statements, the problem boils down to calculating the MLE as we have always done.

The likelihood in this case is:

$$L(\phi|R_i) = \frac{\phi^n}{\prod_{i=1}^n (1+\phi R_i)^2}.$$

So the log-likelihood is:

$$\log\left(L(\phi|R_i)\right) = n\log(\phi) - 2\sum_{i=1}^n \log\left(1+\phi R_i\right).$$

And the derivative of the log-likelihood with respect ϕ:

$$\frac{\partial \log\left(L(\phi|R_i)\right)}{\partial\phi} = \frac{n}{\phi} - 2\sum_{i=1}^n \frac{R_i}{1+\phi R_i}$$

So, in conclusion, the MLE of ϕ based on $R_1,\dots,R_n$ has to satisfy the equation

$$\frac{n}{\phi} - 2\sum_{i=1}^n \frac{R_i}{1+\phi R_i} = 0.$$

Which is exactly the same equation than the one given in the section (a).

SOLUTION (c): By the Asymptotic Normality of the MLE Theorem seen in class we know that, assuming several regularity conditions we have:

$$\sqrt{n}(\hat{\phi}_n - \phi) \longrightarrow_D N(0, \mathcal{I}(\phi)^{-1}).$$

Where $\mathcal{I}(\phi)$ denotes the expected the expected Fisher information matrix of one observation. Then, what we are going to do is to compute this expected Fisher information. Now, we compute the second derivative of the log-likelihood of R_i (the first derivative is computed in the previous section) with respect ϕ, now we use only one observation.

$$\frac{\partial^2 \log\left(L(\phi|R_i)\right)}{\partial \phi^2} = -\frac{1}{\phi^2} + 2\frac{R_i^2}{(1+\phi R_i)^2}$$

So, the expected Fisher information is:

$$\begin{aligned} \mathcal{I}(\phi) = E\left(-\frac{\partial^2 \log\left(L(\phi|R_i)\right)}{\partial \phi^2}\right) = \frac{1}{\phi^2} - 2E\left(\frac{R_i^2}{(1+\phi R_i)^2}\right) = \\ = \frac{1}{\phi^2} - 2\phi\int_0^{+\infty} \frac{x^2}{(1+x\phi)^4} dx = \frac{1}{\phi^2} - 2\frac{1}{3\phi^2} = \frac{1}{3\phi^2} \end{aligned}$$

The computation of the integral in the last expression is not trivial. I have used the *Mathematica* software for compute the integral. Thus, we know that the limiting distribution of $\sqrt{n}(\hat{\phi}_n - \phi)$ is

$$\sqrt{n}(\hat{\phi}_n - \phi) \longrightarrow_D N(0, 3\phi^2).$$

Comment: In order to be able to apply the Asymptotic Normality of the MLE Theorem there are six regularity conditions that have to be fulfilled. In this problem we will not check that all these regularity conditions holds explicitly. Notwithstanding, we observe that the random variables X_i and Y_i belong to the exponential family of distributions, and that R_i comes from a continuous transformation of these variables. Thus, it is quite clear that the conditions will be met.

SOLUTION (d): We have seen in theory class that we can use both $1/\sqrt{\mathcal{I}(\hat{\phi}_n)}$ and $1/\sqrt{I(\hat{\phi}_n)}$ ($I(\hat{\phi}_n)$ denotes the observed Fisher information) as a standard error of the MLE $\hat{\phi}_n$. Since we have computed in the previous section the expected Fisher information we will be using this option. Thus, an estimate of the standard error for the MLE is:

$$\frac{1}{\sqrt{\mathcal{I}(\hat{\phi}_n)}} = \sqrt{3}\hat{\phi}_n.$$

SOLUTION (e): The solution to this problem can be found in the document just below this one. This problem was solved with R markdown to be able to display the code, its results, and the pertinent comments.

Chapter 2

Problem 2

Conjunctivitis is a common infection in children, often occurring in school-aged kids. To assess whether this tends to occur in both eyes rather than in just one eye in urban areas in Catalonia, the following statistical model is proposed.

For a random sample of n infants whose parents reside in urban areas in Catalonia, suppose, for $i = 1, 2, \ldots, n$, that the random variable $X_i = 0$ with probability $(1 - \pi)$ if the i-th infant does not have an the infection, that $X_i = 1$ with probability $\pi(1 - \theta)$ if the i-th infant has the in only one eye, and that $X_i = 2$ with probability $\pi\theta$ if the i-th infant has conjunctivitis in both eyes. Here, $\pi(0 < \pi < 1)$ is the probability that an infant has an infection in at least one eye; that is, π is the prevalence in the studied area of children with an infection in at least one eye. And, since

$$\begin{aligned}
\text{pr}\left(X_i = 2 \mid X_i \geq 1\right) &= \frac{\text{pr}\left[(X_i = 2) \cap (X_i \geq 1)\right]}{\text{pr}\left(X_i \geq 1\right)} \\
&= \frac{\text{pr}\left(X_i = 2\right)}{\text{pr}\left(X_i \geq 1\right)} = \frac{\pi\theta}{\pi} = \theta
\end{aligned}$$

it follows that $\theta(0 < \theta < 1)$ is the conditional probability that an infant has conjunctivitis in both eyes given that this infant has at least one eye that is infected.

(a) Show that a score test statistic $\hat{S}$ for testing $\text{H}_0 : \theta = \theta_0$ versus $\text{H}_1 : \theta \neq \theta_0$ can be written in the form:

$$\hat{S} = \frac{\left(\hat{\theta} - \theta_0\right)^2}{\hat{V}_0(\hat{\theta})}$$

where $\hat{V}_0(\hat{\theta})$ is the estimated variance of $\hat{\theta}$ under the null hypothesis $\text{H}_0 : \theta = \theta_0$.

(b) Suppose that $n = 100$, that $n_0 = 20$ is the number of infants with no eye infections, that $n_1 = 35$ is the number of infants with an eye infection in only one eye, and that $n_2 = 45$ is the number of infants with eye infections in both eyes. Use these data and the score test developed in part (a) to test $\mathrm{H}_0 : \theta = 0.50$ versus $\mathrm{H}_1 : \theta \neq 0.50$ at the $\alpha = 0.025$ significance level. Do these data provide statistical evidence that it is more likely than not that an infant in this region will have both eyes infected once that infant develops an eye infection?

(c) Assuming that $\pi = 0.80$, provide a reasonable value for the smallest value of n, say n^*, required so that the power of a one-sided score test of $\mathrm{H}_0 : \theta = 0.50$ versus $\mathrm{H}_1 : \theta > 0.50$ at the $\alpha = 0.025$ level is at least 0.90 when, in fact, $\theta = 0.60$.

SOLUTION (a): First of all we observe that, because $1 - \pi + \pi(1-\theta) + \pi\theta = 1$ and the structure of our data, it corresponds to a categorical distribution with 3 observations. This is,

$$x \in \{0, 1, 2\} \text{ with } p_0 = 1 - \pi,\ p_1 = \pi(1-\theta) \text{ and } p_2 = \pi\theta$$

Because $\mathrm{pr}(X_i = j) = p_j$ we can write the probability mass function for a given individual as:

$$p^i(x) = p_0^{\mathbb{I}(x=0)} \cdot p_1^{\mathbb{I}(x=1)} \cdot p_2^{\mathbb{I}(x=2)}$$

And so, considering that the variables for every individual are independent we can write the joint probability mass function as:

$$\begin{aligned} p(\mathbf{x}) &= \prod_{i=1}^{n} p^i(x_i) = p_0^{\mathbb{I}(x_1=0)} \cdot p_1^{\mathbb{I}(x_1=1)} \cdot p_2^{\mathbb{I}(x_1=2)} \cdots p_0^{\mathbb{I}(x_n=0)} \cdot p_1^{\mathbb{I}(x_n=1)} \cdot p_2^{\mathbb{I}(x_n=2)} \\ &= p_0^{\sum_{i=1}^n \mathbb{I}(x_i=0)} \cdot p_1^{\sum_{i=1}^n \mathbb{I}(x_i=0)} \cdot p_2^{\sum_{i=1}^n \mathbb{I}(x_i=0)} \end{aligned}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$.

Now, considering $n_j = \sum_{i=1}^n \mathbb{I}(x_i = j)$ for $j = 0, 1, 2$ we can write:

$$p(\mathbf{x}) = p_0^{n_0} \cdot p_1^{n_1} \cdot p_2^{n_2}$$

Let us now write the likelihood in terms of our parameters:

$$L(\pi, \theta | \mathbf{x}) = (1-\pi)^{n_0} \cdot (\pi(1-\theta))^{n_1} \cdot (\pi\theta)^{n_2}$$

And the log-likelihood:

$$\log(L(\pi, \theta|\mathbf{x})) = n_0 \log(1-\pi) + n_1 \log(\pi(1-\theta)) + n_2 \log(\pi\theta)$$

Let us write the score equations to find the maximum likelihood estimator of both parameters.

$$\frac{\partial \log\left(L(\pi, \theta|\mathbf{x})\right)}{\partial \pi} = \frac{-n_0}{1-\pi} + \frac{n_1}{\pi} + \frac{n_2}{\pi} = \frac{-\pi(n_0 + n_1 + n_2) + n_1 + n_2}{\pi(1-\pi)} = 0 \iff$$

$$\iff \hat{\pi}_{ML} = \frac{n_1 + n_2}{n_0 + n_1 + n_2}$$

$$\frac{\partial \log\left(L(\pi,\theta|\mathbf{x})\right)}{\partial\theta} = \frac{-n_1}{1-\theta} + \frac{n_1}{\pi} + \frac{n_2}{\theta} = \frac{-\theta n_1 + (1-\theta)n_2}{\theta(1-\theta)} = 0 \iff$$
$$\iff \hat{\theta}_{ML} = \frac{n_2}{n_1+n_2}$$

We observe that the maximum likelihood estimators of each parameters does not depend on the other one.

Next, we want to compute the Fisher's information matrix of θ.

$$I(\theta|\mathbf{x}) = -\frac{\partial^2 \log\left(L(\pi,\theta|\mathbf{x})\right)}{\partial\theta^2} = -\frac{\partial}{\partial\theta}\left(\frac{-n_1}{1-\theta} + \frac{n_2}{\theta}\right) = \frac{n_1}{(1-\theta)^2} + \frac{n_2}{\theta^2}$$

And the expected Fisher's information matrix of θ.

$$\mathcal{I}(\theta|\mathbf{x}) = E\left(I(\theta|\mathbf{x})\right) = E\left(\frac{n_1}{(1-\theta)^2} + \frac{n_2}{\theta^2}\right) = \frac{1}{(1-\theta)^2}E(n_1) + \frac{1}{\theta^2}E(n_2) \tag{2.1}$$

Let us work on these expected values:

$$E(n_1) = E\left(\sum_{i=1}^{n}\mathbb{I}(x_i=1)\right) = \sum_{i=1}^{n}E\left(\mathbb{I}(x_i=1)\right) = \sum_{i=1}^{n}\mathrm{pr}(x_i=1) = \sum_{i=1}^{n}p_1$$
$$= n\pi(1-\theta)$$

Analogously, we have $E(n_2) = n\pi\theta$.

Going back to 2.1 we have:

$$\mathcal{I}(\theta|\mathbf{x}) = \frac{n\pi}{1-\theta} + \frac{n\pi}{\theta} = \frac{n\pi}{\theta(1-\theta)}$$

We now want to use the theorem stated in page 27 of the notes that ensures the Cramer-Rao bound holds with equality if and only if we can write

$$a(\theta)(W(\mathbf{x}) - \tau(\theta)) = \tau'(\theta)\mathcal{I}^{-1}(\theta|\mathbf{x})S(\theta|\mathbf{x})$$

In our case, $\tau(\theta) = \theta$.

$$\mathcal{I}^{-1}(\theta|\mathbf{x})S(\theta|\mathbf{x}) = \frac{\theta(1-\theta)}{n\pi}\,\frac{-\theta n_1 + (1-\theta)n_2}{\theta(1-\theta)} = \frac{-\theta(n_1+n_2)+n_2}{n\pi}$$

If we substitute π for its MLE:

$$\mathcal{I}^{-1}(\theta|\mathbf{x})S(\theta|\mathbf{x}) = \frac{-\theta(n_1+n_2)+n_2}{n_1+n_2} = \frac{n_2}{n_1+n_2} - \theta = \hat{\theta}_{ML} - \theta$$

Therefore, the Cramer-Rao bound is attained and we can assure that $\hat{V}_0(\hat{\theta}_{ML}) = \mathcal{I}^{-1}(\theta_{ML}|\mathbf{x})$ and so, if we isolate from the previous equality the Score function:

$$S(\theta|\mathbf{x}) = \frac{\hat{\theta}_{ML} - \theta}{\hat{V}_0(\hat{\theta}_{ML})}$$

Finally, we know the Score statistic can be written as:

$$W_S(\theta_0) = \frac{S(\theta_{ML}|\mathbf{x})}{\sqrt{\mathcal{I}(\theta_{ML}|\mathbf{x})}}$$

And substituting with what we have obtained before, taking into account that we are testing $\theta = \theta_0$, and remembering that $\mathcal{I}^{-1}(\theta|\mathbf{x}) = V_0(\theta_{ML})$:

$$W_S(\theta_0) = \frac{(\hat{\theta}_{ML} - \theta_0)\sqrt{\hat{V}_0(\hat{\theta}_{ML})}}{\hat{V}_0(\hat{\theta}_{ML})}$$

Because we want it to behave as a χ^2_1 we square everything and obtain:

$$\hat{S} = \frac{(\hat{\theta}_{ML} - \theta_0)^2}{\hat{V}_0(\hat{\theta}_{ML})}$$

which is what we wanted to prove.

SOLUTION (b): For the data given we obtain the following value of the Score statistic:

$$\hat{S} = \frac{\left(\frac{n_2}{n_1+n_2} - 0.5\right)^2}{\frac{\frac{n_2}{n_1+n_2}\left(1-\frac{n_2}{n_1+n_2}\right)}{n_1+n_2}} \approx 1.27$$

Now, the rejection region of our test is:

$$R_{0.025} = \{\mathbf{x} : \hat{S} \geq \chi^2_1(0.025)\}$$

We have obtained $\hat{S} = 1.27$ which is smaller than $\chi^2_1(0.025) = 5.024$ and therefore we can not reject the null hypothesis and affirm that there's statistical evidence that it is more likely than not that an infant in this region will have both eyes infected once that infant develops an eye infection.

SOLUTION (c): Notice that the formulas for compute the sample size are based on the concept that larger effect sizes, smaller significance levels, and larger sample sizes all contribute to increased statistical power. In the first unit of the course we saw how to compute the sample size when we are testing means via Z-test. In addition, in the literature, and on the internet, we have found similar formulas for computing the sample size in the case of a Bernoulli distribution (remember that in this problem we are working with a generalized Bernoulli distribution). We have decided to use the following formula:

$$n = \frac{(Z_{1-\alpha} - Z_{\beta})^2}{(\theta_1 - \theta_0)^2} V(X) = \frac{(1.96 + 1.28)^2 \theta_0 (1 - \theta_0)}{\pi 0.01} = 328.05.$$

Where, we have substituted the values of $\pi, \theta_0, \theta_1, \beta = 1 - 0.9 = 0.1, \alpha$ given in the hypothesis. And, $Z_{1-\alpha}, Z_{1-\beta}$ are the $1 - \alpha, \beta$, respectively, quantiles of the normal standard distribution.

Then, the smallest value of n for having at least 0.9 power of the test is $n^* = 329$.

Comment: we have used the different explanations and examples in `https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_power/bs704_power_print.html` as a guido for arrive to the formula used.

Chapter 3

Problem 3

The last exercise is intended to explore the Expectation-Maximization (EM) algorithm in the context of a Poisson mixture model. The Poisson distribution is commonly used for count data, making it a suitable choice for various real-world applications.

(a) Consider a Poisson mixture model with two components:

$$f(x; \lambda_1, \lambda_2, \pi) = \pi \cdot \text{Poisson}(x; \lambda_1) + (1 - \pi) \cdot \text{Poisson}(x; \lambda_2)$$

where λ_1, λ_2 are the Poisson rates for the two components, and π is the mixing proportion.

i. Write down the complete expression for the log-likelihood function of the observed data X in terms of λ_1, λ_2, and π.

ii. Derive the expressions for the E-step and M-step updates for λ_1, λ_2, and π in the context of the Poisson mixture model.

(b) Given a sample dataset X :

$$X = (4, 2, 6, 8, 3, 5, 7, 1, 9, 10)$$

Implement the EM algorithm for the Poisson mixture model using the following initial values:

$$\lambda_1^{(0)} = 3, \quad \lambda_2^{(0)} = 7, \quad \pi^{(0)} = 0.5$$

i. Perform the E-step: Calculate the expected values of the latent variables for each data point.

ii. Perform the M-step: Update the parameters λ_1, λ_2, and π using the calculated expected values.

Repeat the E-step and M-step for a total of 3 iterations.

Additional Instructions:

- Provide the detailed calculations for each step.

- Interpret the results after each iteration in terms of how the parameter estimates and likelihood are changing.

SOLUTION (a): We first write down the complete expression for the log-likelihood function of the observed data X in terms of λ_1, λ_2 and π. We assume that we have a sample of size n, $X_1, \ldots, X_n$. Let us denote

$$f_1(x|\lambda_1) = \text{Poisson}\,(x; \lambda_1) = \frac{\exp\,(-\lambda_1)\lambda_1^x}{x!}$$
$$f_2(x|\lambda_2) = \text{Poisson}\,(x; \lambda_2) = \frac{\exp\,(-\lambda_2)\lambda_2^x}{x!}$$

Then, the likelihood of $f(x; \lambda_1, \lambda_2, \pi)$ is

$$L(\lambda_1, \lambda_2, \pi | x_1, \ldots, x_n) = \prod_{i=1}^{n} \left(\pi \frac{\exp\,(-\lambda_1)\lambda_1^{x_i}}{x_i!} + (1-\pi) \frac{\exp\,(-\lambda_2)\lambda_2^{x_i}}{x_i!} \right)$$
$$= \prod_{i=1}^{n} \left(\pi \frac{\exp\,(-\lambda_1)\lambda_1^{x_i} + (1-\pi)\exp\,(-\lambda_2)\lambda_2^{x_i}}{x_i!} \right)$$

And then the log-likelihood is

$$\log\left(L(\lambda_1, \lambda_2, \pi | x_1, \ldots, x_n)\right) = \sum_{i=1}^{n} \log\left(\exp\,(-\lambda_1)\lambda_1^{x_i} + (1-\pi)\exp\,(-\lambda_2)\lambda_2^{x_i}\right) - \sum_{i=1}^{n} \log\,(x_i!)$$

But, for our purposes in the next part of this section, and also for the next section, it is better to present the complete log-likelihood in a more compact way. The likelihood can also be written as:

$$L(\lambda_1, \lambda_2, \pi | x_1, \ldots, x_n) = \prod_{i=1}^{n} \left(\pi f_1(x_i|\lambda_1) + (1-\pi) f_2(x_i|\lambda_2)\right)$$

So, the log-likelihood is:

$$\log\left(L(\lambda_1, \lambda_2, \pi | x_1, \ldots, x_n)\right) = \sum_{i=1}^{n} (\pi \log\,(f_1(x_i|\lambda_1) + (1-\pi) f_2(x_i|\lambda_2))$$

Now, we do the second part of this section, which is derive the expression for the E-step and the M-step. Let $\gamma_1(x), \gamma_2(x)$ denote the *responsability* of the Poisson models $1, 2$ respectively, over the datapoint x. By the Bayes Theorem we have that

$$\gamma_k(x) = P(K = k|x) = \frac{P(K = k)P(x|K = k)}{P(K = 1)P(x|K = 1) + P(K = 2)P(x|K = 2)} \quad \text{for} \quad k = 1, 2$$

Thus, is clear that $\gamma_2(x) = 1 - \gamma_1(x)$, and then, from now, on we only need to compute $\gamma_1(x)$.

Once we have defined the responsability, we are ready for derive the E-step. We have the following expression:

$$E(\log(L)) = \sum_{i=1}^{n} \left[\gamma_1(x_i)\left\{\log \pi + \log f_1(x|\lambda_1)\right\} + \gamma_2(x_i)\left\{\log(1-\pi) + \log f_2(x|\lambda_2)\right\}\right]$$

Let us denote the last expression as $Q(\pi, \lambda_1, \lambda_2|\pi^0, \lambda_1^0, \lambda_2^0)$. And the responsability in our case is:

$$\gamma_1(x_i) = \frac{\pi f_1(x_i|\lambda_1)}{\pi f_1(x_i|\lambda_1) + (1-\pi)f_2(x_i|\lambda_2)}$$

And, as we said before, $\gamma_2(x_i) = 1 - \gamma_1(x_i)$. Now, we derive the M-step. In the M-step we should differentiate Q with respect the different parameters, set the derivative equal to 0 and solve it. We start with the parameters λ_i, for $j = 1, 2$:

$$\begin{aligned}\frac{\partial Q}{\partial \lambda_j} &= \sum_{i=1}^{n} \frac{\gamma_j(x_i)\frac{\partial f_j(x_i|\lambda_j)}{\partial \lambda_j}}{f_j(x_i|\lambda_j)} = \sum_{i=1}^{n} \frac{\gamma_j(x_i)\frac{\exp(-\lambda_j)\lambda_j^{x_i}}{x_i!}(-1 + \frac{x_i}{\lambda_j})}{\frac{\exp(-\lambda_j)\lambda_j^{x_i}}{x_i!}} \\ &= \sum_{i=1}^{n} \gamma_j(x_i)(-1 + \frac{x_i}{\lambda_j}) = 0.\end{aligned}$$

And isolating λ_j, we obtain:

$$\lambda_j = \frac{\sum_{i=1}^{n} \gamma_j(x_i)x_i}{\sum_{i=1}^{n} \gamma_j(x_i)}.$$

We now differentiate Q with respect the parameter π, we obtain:

$$\begin{aligned}\frac{\partial Q}{\partial \pi} &= \sum_{i=1}^{n} \left(\frac{\gamma_1(x_i)}{\pi} - \frac{1-\gamma_1(x_i)}{1-\pi}\right) = \sum_{i=1}^{n} \frac{\gamma_1(x_i) - \pi}{\pi(\pi-1)} \\ &= \frac{1}{\pi(\pi-1)}\sum_{i=1}^{n} \gamma_1(x_i) - \frac{n\pi}{\pi(\pi-1)} = 0\end{aligned}$$

Isolating π we obtain:

$$\pi = \frac{\sum_{i=1}^{n} \gamma_1(x_i)}{n}.$$

Thus, summarizing, in the M-step we have obtained the following quantities:

$$\lambda_1 = \frac{\sum_{i=1}^{n} \gamma_1(x_i) x_i}{\sum_{i=1}^{n} \gamma_1(x_i)}$$
$$\lambda_2 = \frac{\sum_{i=1}^{n} (1 - \gamma_1(x_i)) x_i}{\sum_{i=1}^{n} (1 - \gamma_1(x_i))}$$
$$\pi = \frac{\sum_{i=1}^{n} \gamma_1(x_i)}{n}$$

SOLUTION (b): The solution can be found in the document just below this one. This problem was solved with R markdown to be able to display the code, its results, and the pertinent comments.

Assignment 3

Anna Felip & Arnau Garcia

SOLUTION Problem 1, section (e): We first introduce the data given in the table:

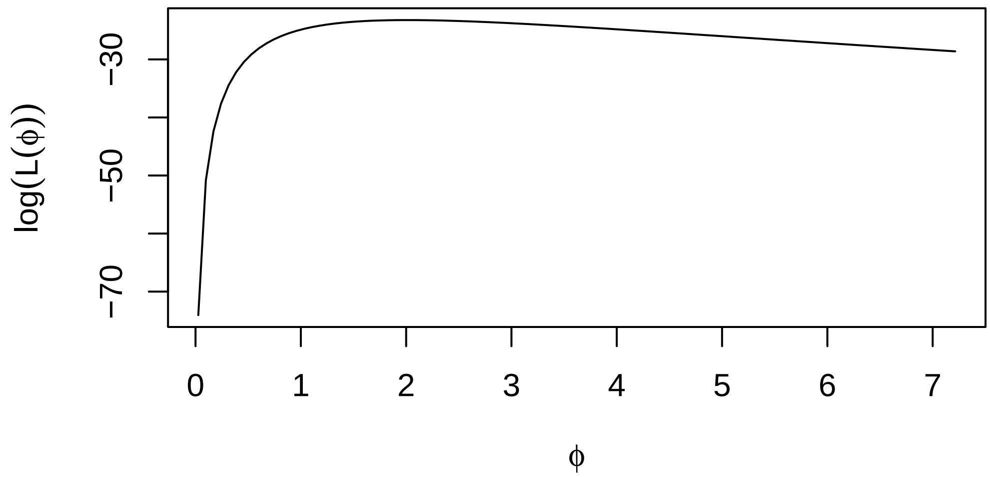
```
x <-c(0.7,11.3,2.1,30.7,4.6,20.2,0.3,0.9,0.7,2.3,1.1,1.9,0.5,0.8,1.2,15.2,
      0.2,0.7,0.4,2.3)
y<-c(3.8, 4.6, 2.1,5.6,10.3,2.8,1.9,1.4,0.4,0.9,2.8,3.2,8.5,14.5,14.4,
     8.8,7.6,1.3,2.2,4.0)
n<-20
```

We try to plot the likelihood in order to get an intuition for a good initial value for the Fisher scoring method. First we program a function of the log-likelihood:

```
loglike <- function(phi, x){
  return(length(x)*log(phi)-2*sum(log(1+phi*x)))
}
```

Now, using what we have seen in class for plotting likelihoods we can plot the likelihood:

```
fred <- function(phi){
    return(apply(as.matrix(phi), 1,loglike, x = x/y))
}
curve(fred, from = min(x/y), to = max(x/y),
    xlab = expression(phi), ylab = expression(log(L(phi))))
```



It seems that, for instance, $\phi = 1.5$ is a reasonable good starting value.

Function for compute the derivative of the log-likelihood:

```
logder <- function(phi, x, n){
  return(n/phi - 2*sum(x/(1+phi*x)))
}
```

Function for compute the inverse of the expected Fisher information (we need the Fisher information for the whole sample):

```
fisher_inv <- function(phi, n){
  return(3*phi^2/n)
}
```

Now, we program the Fisher scoring algorithm:

```
tol<-10e-9
dif<-1
phi0<-1.5
iters<-100
count<-0
while(dif>tol & count<iters){
  phi <- phi0 + fisher_inv(phi0,n)*logder(phi0, x/y, n)
  dif <- abs(phi-phi0)
  phi0<-phi
  count<-count+1
}
```

```
cat("phi value:", phi, "iterations:", count)
```

```
phi value: 2.010169 iterations: 6
```

Then, the Fisher Scoring Algorithm returns the MLE of ϕ as $\hat{\phi}_n = 2.010169$. Now we check whether the output is a good result. We see if the equation that the MLE has to satisfy is satisfied:

```
logder(phi,x/y,n)
```

```
[1] -7.215029e-11
```

Indeed, it is very near to 0. So, it seems that the result is good.

SOLUTION Problem 3, section (b):

We introduce the data:

```
X<-c(4,2,6,8,3,5,7,1,9,10)
```

Setting the initial values given in the statements of the problem:

```
lambda10 <- 3
lambda20 <- 7
pi0 <- 0.5
```

We program a function that computes the complete likelihood. We will use this problem after, for interpret the changing in each iteration of the EM algorithm in the likelihood value.

```
likel <- function(x, l1, l2, p){
  return(prod(p*dpois(x,l1) + (1-p)*dpois(x,l2)))
}
```

Now, we program a function that does the EM algorithm.

```
EM_TwoMixturePoisson <- function(pi0, lambda10, lambda20, X, iters=1000, tol=1e-3){
  dif <- 1
  count <-0
  #printing iteration 0
  cat("Iter", count, ": lambda1=", round(lambda10, 3), ", lambda2=",
```

```r
        round(lambda20, 3), ", pi=", round(pi0, 3),
        ", Likelihood value:",  likel(X, lambda10, lambda20, pi0), "\n")
  while(dif>tol & count<iters){
    #E-step
    f1 <- dpois(X, lambda10)
    f2 <- dpois(X, lambda20)
    gamma <- f1*pi0/(f1*pi0+f2*(1-pi0))
    #M-step:
    pi <- mean(gamma)
    lambda1 <- sum(gamma*X)/sum(gamma)
    lambda2 <- sum((1-gamma)*X)/sum(1-gamma)
    #chechking convergence (we use norm 2)
    dif <- sqrt((lambda1-lambda10)^2 + (lambda2-lambda20)^2)
    #update
    pi0 <- pi
    lambda10 <- lambda1
    lambda20 <- lambda2
    count <- count +1
    #printing results for each iteration
    cat("It", count, ": l1=", round(lambda1, 3), ", l2=",
        round(lambda2, 3), ", pi=", round(pi, 3), ", dif=", dif,
        ", Likel val:",  likel(X, lambda1, lambda2, pi), "\n")
  }
}
```

We use the function in our case. Notice that the statements only demand the repetition of the E and M steps for a total of 3 iterations. Then, using the function we have:

```r
EM_TwoMixturePoisson(pi0, lambda10, lambda20, X, iters=3)
```

```
Iter 0 : lambda1= 3 , lambda2= 7 , pi= 0.5 , Likelihood value: 2.159111e-11
It 1 : l1= 3.117 , l2= 7.257 , pi= 0.424 , dif= 0.2824961 , Likel val: 2.545727e-11
It 2 : l1= 3.048 , l2= 7.181 , pi= 0.407 , dif= 0.1021545 , Likel val: 2.602878e-11
It 3 : l1= 2.977 , l2= 7.124 , pi= 0.392 , dif= 0.09095285 , Likel val: 2.649357e-11
```

As we can see, in each iteration the value of π is decreasing. We can observe that λ_1, λ_2 increase in the first iteration, but then decrease in the following iterations. The likelihood value is increasing, which is what we expect. The estimations obtained after three iterations are $\lambda_1 = 2.977$, $\lambda_2 = 7.124$ and $\pi = 0.392$.

Now, we run the function using as a maximum iteration number arguments the one by default in the function (1000), and we see if the algorithm converges.

```
EM_TwoMixturePoisson(pi0, lambda10, lambda20, X)
```

```
Iter 0 : lambda1= 3 , lambda2= 7 , pi= 0.5 , Likelihood value: 2.159111e-11
It 1 : l1= 3.117 , l2= 7.257 , pi= 0.424 , dif= 0.2824961 , Likel val: 2.545727e-11
It 2 : l1= 3.048 , l2= 7.181 , pi= 0.407 , dif= 0.1021545 , Likel val: 2.602878e-11
It 3 : l1= 2.977 , l2= 7.124 , pi= 0.392 , dif= 0.09095285 , Likel val: 2.649357e-11
It 4 : l1= 2.911 , l2= 7.077 , pi= 0.379 , dif= 0.08181663 , Likel val: 2.686965e-11
It 5 : l1= 2.85 , l2= 7.036 , pi= 0.367 , dif= 0.07296227 , Likel val: 2.71691e-11
It 6 : l1= 2.797 , l2= 7.001 , pi= 0.357 , dif= 0.06449607 , Likel val: 2.740395e-11
It 7 : l1= 2.75 , l2= 6.969 , pi= 0.348 , dif= 0.05661604 , Likel val: 2.758576e-11
It 8 : l1= 2.709 , l2= 6.941 , pi= 0.34 , dif= 0.04942847 , Likel val: 2.772501e-11
It 9 : l1= 2.673 , l2= 6.917 , pi= 0.334 , dif= 0.04296655 , Likel val: 2.78307e-11
It 10 : l1= 2.643 , l2= 6.895 , pi= 0.328 , dif= 0.03721867 , Likel val: 2.791033e-11
It 11 : l1= 2.617 , l2= 6.877 , pi= 0.323 , dif= 0.03214755 , Likel val: 2.796996e-11
It 12 : l1= 2.594 , l2= 6.861 , pi= 0.319 , dif= 0.02770215 , Likel val: 2.801437e-11
It 13 : l1= 2.575 , l2= 6.847 , pi= 0.315 , dif= 0.0238252 , Likel val: 2.80473e-11
It 14 : l1= 2.558 , l2= 6.835 , pi= 0.312 , dif= 0.02045798 , Likel val: 2.807165e-11
It 15 : l1= 2.544 , l2= 6.825 , pi= 0.309 , dif= 0.01754329 , Likel val: 2.808958e-11
It 16 : l1= 2.531 , l2= 6.816 , pi= 0.307 , dif= 0.01502723 , Likel val: 2.810276e-11
It 17 : l1= 2.521 , l2= 6.808 , pi= 0.305 , dif= 0.01286017 , Likel val: 2.811243e-11
It 18 : l1= 2.512 , l2= 6.802 , pi= 0.304 , dif= 0.01099716 , Likel val: 2.811951e-11
It 19 : l1= 2.505 , l2= 6.796 , pi= 0.302 , dif= 0.009397989 , Likel val: 2.812469e-11
It 20 : l1= 2.498 , l2= 6.792 , pi= 0.301 , dif= 0.00802704 , Likel val: 2.812847e-11
It 21 : l1= 2.493 , l2= 6.788 , pi= 0.3 , dif= 0.00685298 , Likel val: 2.813122e-11
It 22 : l1= 2.488 , l2= 6.784 , pi= 0.299 , dif= 0.005848416 , Likel val: 2.813323e-11
It 23 : l1= 2.484 , l2= 6.781 , pi= 0.298 , dif= 0.004989511 , Likel val: 2.813469e-11
It 24 : l1= 2.48 , l2= 6.779 , pi= 0.297 , dif= 0.004255596 , Likel val: 2.813576e-11
It 25 : l1= 2.478 , l2= 6.777 , pi= 0.297 , dif= 0.003628805 , Likel val: 2.813653e-11
It 26 : l1= 2.475 , l2= 6.775 , pi= 0.296 , dif= 0.003093736 , Likel val: 2.81371e-11
It 27 : l1= 2.473 , l2= 6.773 , pi= 0.296 , dif= 0.002637132 , Likel val: 2.813751e-11
It 28 : l1= 2.471 , l2= 6.772 , pi= 0.296 , dif= 0.002247608 , Likel val: 2.81378e-11
It 29 : l1= 2.47 , l2= 6.771 , pi= 0.295 , dif= 0.001915395 , Likel val: 2.813802e-11
It 30 : l1= 2.468 , l2= 6.77 , pi= 0.295 , dif= 0.001632123 , Likel val: 2.813818e-11
It 31 : l1= 2.467 , l2= 6.769 , pi= 0.295 , dif= 0.001390628 , Likel val: 2.813829e-11
It 32 : l1= 2.466 , l2= 6.768 , pi= 0.295 , dif= 0.001184781 , Likel val: 2.813837e-11
It 33 : l1= 2.465 , l2= 6.768 , pi= 0.295 , dif= 0.001009342 , Likel val: 2.813843e-11
It 34 : l1= 2.465 , l2= 6.767 , pi= 0.295 , dif= 0.0008598378 , Likel val: 2.813848e-11
```

We can observe that the algorithm has converged after 34 iterations, and the estimations are $\lambda_1 = 2.465$, $\lambda_2 = 6.767$ and $\pi = 0.295$.