Proposed solution to mandatory assignment STK4011/9011

Autumn 2019

Emil Aas Stoltenberg

Exercise 1

Let $X_1,\ldots,X_n$ be i.i.d. random variables with positive density $f(x)$, distribution function $F(x)$ on some interval, and finite second moment. Set $Z_i = F(X_i)$ for $i=1,\ldots,n$, and let $Z_{(1)} < Z_{(2)} < \ldots < Z_{(n-1)} < Z_{(n)}$ be the order statistics.

(a) Show that $Z_{(i)}$ has density

$$g_i(z) = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} z^{i-1}(1-z)^{n-i}, \quad \text{on } (0,1).$$

We did this is in class, see also page 230 in C&B. We recognise $g_i(z)$ as a Beta$(i, n-i+1)$ density, so that

$$\mathrm{E}\, Z_{(i)} = \frac{i}{n+1}, \quad \text{and} \quad \mathrm{Var}\, Z_{(i)} = \frac{i(n-i+1)}{(n+1)^2(n+2)}.$$

(b) With $n = 100$, find the correlation between $Z_{(17)}$ and $Z_{(18)}$. We can find the joint density $g_{Z_{(17)},Z_{(18)}}(x,y)$. Here is a heuristic argument: For $0 < x < y < 1$ split the interval $[0,1]$ in four parts, $[0,x)$, $[x, x+\mathrm{d}x)$, $[y, y+\mathrm{d}y)$ and $(y,1]$. We can then think of a multinomial experiment with four possible outcomes. Sixteen uniforms falling in $[0,x)$, one in $[x,x+\mathrm{d}x)$, one in $[y,y+\mathrm{d}y)$, and the remaining 82 in $(y,1]$, is an event with probability

$$g_{Z_{(17)},Z_{(18)}}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \frac{100!}{16!1!1!82!} x^{16}(1-y)^{82}\,\mathrm{d}x\,\mathrm{d}y.$$

Then

$$\begin{aligned}
\mathrm{E}\{Z_{(17)}Z_{(18)}\} &= \int_0^1\int_0^y xy g_{Z_{(17)},Z_{(18)}}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{\Gamma(101)}{\Gamma(17)\Gamma(83)}\int_0^1\int_0^y xyx^{16}(1-y)^{82}\,\mathrm{d}x\,\mathrm{d}y = \frac{\Gamma(101)}{\Gamma(17)\Gamma(83)}\frac{1}{18}\int_0^1 y^{19}(1-y)^{82}\,\mathrm{d}y \\
&= \frac{\Gamma(101)}{\Gamma(17)\Gamma(83)}\frac{1}{18}\frac{\Gamma(20)\Gamma(83)}{\Gamma(103)} = \frac{19\times 17}{102\times 101}.
\end{aligned}$$

The covariance is

$$\mathrm{Cov}(Z_{(17)}, Z_{(18)}) = \mathrm{E}\{Z_{(17)}Z_{(18)}\} - \mathrm{E}\,Z_{(17)}\mathrm{E}\,Z_{(18)} = \frac{17}{101}\big(\frac{19}{102} - \frac{18}{101}\big),$$

and the correlation is

$$\mathrm{corr}(Z_{(17)}, Z_{(18)}) = \frac{\mathrm{Cov}(Z_{(17)}, Z_{(18)})}{(\mathrm{Var}\,Z_{(17)}\mathrm{Var}\,Z_{(18)})^{1/2}} = \big(\frac{17\times 83}{18\times 84}\big)^{1/2} = 0.966.$$

(c) Take for convenience $n = 2m+1$, and consider the median $M_n^0 = Z_{((n+1)/2)}$. The density of M_n^0 is

$$g_{(n+1)/2}(z) = \frac{n!}{\{((n-1)/2)!\}^2} z^{(n-1)/2}(1-z)^{(n-1)/2}, \quad z\in(0,1).$$

Transforming the random variable M_n^0, we find the density $f_n(x)$ of $\sqrt{n}(M_n^0 - 1/2)$ (see Nils exercise 1), it is

$$\begin{aligned}
f_n(x) &= g_{(n+1)/2}(1/2 + x/\sqrt{n})\frac{1}{\sqrt{n}} \\
&= \frac{1}{\sqrt{n}}\frac{n!}{\{((n-1)/2)!\}^2}(1/2 + x/\sqrt{n})^{(n-1)/2}(1/2 - x/\sqrt{n})^{(n-1)/2}, \quad x\in(-\sqrt{n}/2, \sqrt{n}/2).
\end{aligned}$$

using Stirling's formula $n! \doteq \sqrt{2\pi n}e^{-n}n^n$,

$$\frac{n!}{\{((n-1)/2)!\}^2} \doteq \frac{\sqrt{n}}{\sqrt{2\pi}}e^{-1}\frac{2^n}{(1-1/n)^n} \doteq \frac{\sqrt{n}2^n}{\sqrt{2\pi}},$$

because $(1-1/n)^n \to e^{-1}$. We also see that

$$(1/2+x/\sqrt{n})^{(n-1)/2}(1/2-x/\sqrt{n})^{(n-1)/2} = \Big(\frac{1}{4}-\frac{x^2}{n}\Big)^{(n-1)/2}$$
$$= 2^{-n}\Big(1-\frac{x^2}{n/4}\Big)^{n/2}\Big(\frac{1}{4}-\frac{x^2}{n}\Big)^{-1/2}.$$

Inserting these two expressions in $f_n(x)$,

$$f_n(x) \doteq \frac{1}{\sqrt{2\pi}}\Big(1-\frac{x^2}{n/4}\Big)^{n/2}\Big(\frac{1}{4}-\frac{x^2}{n}\Big)^{-1/2} \to \frac{1}{\sqrt{2\pi(1/4)}}\exp\big\{-\frac{x^2}{2(1/4)}\big\},$$

as $n\to\infty$.

By Exercise 12 'Convergence of densities' in the Nils exercises (where you are asked to prove Scheffé's lemma), the pointwise convergence $f_n(x)$ to a normal density with variance 1/4 implies convergence in distribution

(1) $$\sqrt{n}(M_n^0-1/2)\overset{d}{\to}\mathrm{N}(0,1/4).$$

(d) From Theorem 5.4.4 p. 229 in C&B we have that the distribution function of the j'th of n order statistics is

$$P(W_{(j)}\le w)=\sum_{k=j}^{n}\binom{n}{k}G(w)^k(1-G(w))^{n-k},$$

when $W_1,\ldots,W_n$ are i.i.d. with distribution function G. The distribution function F is monotonically increasing because f is positive on some interval,

$$P(F^{-1}(Z_{(j)})\le z)=P(Z_{(j)}\le F(z))=\sum_{k=j}^{n}\binom{n}{k}F(z)^k\{1-F(z)\}^{n-k},$$

which shows that $F^{-1}(Z_{(j)})$ and $X_{(j)}$ have the same distribution.

Denote by $M_n = X_{((n+1)/2)}$ the median of the $X_1,\ldots,X_n$. To see that M_n is consistent for $\mu = F^{-1}(1/2)$ we can argue as follows: In (c) we found that $\sqrt{n}(M_n^0-1/2)\overset{d}{\to}\mathrm{N}(0,1/4)$. Since the sequence $1/\sqrt{n}$ converges in probability to 0, the Slutsky–Cramér rules give

$$M_n^0-1/2=\frac{1}{\sqrt{n}}\sqrt{n}(M_n^0-1/2)\overset{p}{\to}0.$$

This means that M_n^0 is consistent for 1/2. We could also use that from (a),

$$\mathrm{Var}\,M_n^0=\frac{1}{4(n+2)},$$

so that for any $\varepsilon>0$ we have $P(|M_n^0-1/2|\ge\varepsilon)\le\varepsilon^{-2}\{4(n+2)\}^{-1}\to 0$ by Chebyshev's inequality, because M_n^0 is unbiased. The function F^{-1} is continuous, so by the continuous mapping theorem (see Nils exercises 9(a), 'Convergence in probability'), we have that

$$M_n^0\overset{p}{\to}1/2 \quad\Rightarrow\quad F^{-1}(M_n^0)\overset{p}{\to}F^{-1}(1/2)=\mu.$$

Since $F^{-1}(M_n^0)$ and M_n are equal in distribution, this entails that the sample median M_n is consistent for the population median μ.

The derivative of $F^{-1}(x)$ is

$$\frac{\mathrm{d}}{\mathrm{d}x}F^{-1}(x)=\frac{1}{f(F^{-1}(x))},$$

and evaluated in 1/2 it is $1/f(\mu)$. Combine the convergence in distribution we argued for in (1) with the delta-method,

$$\sqrt{n}(M_n-\mu)=\sqrt{n}(F^{-1}(M_n^0)-F^{-1}(\mu))\overset{d}{\to}1/f(\mu)\mathrm{N}(0,1/4)=\mathrm{N}(0,\tau^2),\quad\text{with}\quad \tau^2=\frac{1}{4f(\mu)^2}.$$

(e) Let μ be the centre point a symmetric distribution with density f. That the density f is symmetric around μ means that $f(\mu - y) = f(\mu + y)$ for all y.

$$\begin{aligned}\mathrm{E}\,X &= \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \mu + \int_{-\infty}^{\infty} (x-\mu) f(x)\,\mathrm{d}x \\ &= \mu + \int_{-\infty}^{\infty} y f(y+\mu)\,\mathrm{d}y = \mu + \int_{-\infty}^{0} y f(y+\mu)\,\mathrm{d}y + \int_{0}^{\infty} y f(y+\mu)\,\mathrm{d}y \\ &= \mu - \int_{0}^{\infty} y f(-y+\mu)\,\mathrm{d}y + \int_{0}^{\infty} y f(y+\mu)\,\mathrm{d}y = \mu.\end{aligned}$$

Moreover,

$$F(\mu) = \int_{-\infty}^{\mu} f(x)\,\mathrm{d}x = \int_{-\infty}^{0} f(y+\mu)\,\mathrm{d}y = \int_{0}^{\infty} f(-y+\mu)\,\mathrm{d}y = 1 - F(\mu),$$

so $F(\mu) = 1/2$. This shows that mean and median is the same in a symmetric distribution. That $\bar{X}_n$ is consistent for the μ follows from the law of large numbers.

We are assuming that $\mathrm{Var}\,X_i = \sigma^2 < \infty$, so the variance of $\bar{X}_n$ is σ^2/n. From (d) we have that the approximate variance of M_n is $1/\{4nf(\mu)^2\}$. If $X_1, \ldots, X_n$ is a random sample from a $\mathrm{N}(\mu, \sigma^2)$ distribution, $\mathrm{Var}\,M_n \approx \pi\sigma^2/(2n)$. Said differently, the asymptotic relative efficiency compared to the mean, is

$$\frac{\tau^2}{\sigma^2} = \frac{\pi}{2} > 1.$$

For the sample median to achieve lower variance than the sample mean, we need a distribution with somewhat heavy tails (but not so heavy that the second moment does not exist, the Cauchy, for example, where the first moment does not exist). Consider the Laplace distribution that is symmetric around μ, with density

$$f(x;\mu,\sigma) = \frac{1}{2\sigma}\exp\Big(-\frac{|x-\mu|}{\sigma}\Big), \quad x \in (-\infty, \infty).$$

The variance of this distribution is $2\sigma^2$. The approximate variance of the median is

$$\frac{1}{4f(x;\mu,\sigma)^2} = \frac{1}{4/(2\sigma)^2} = \sigma^2.$$

This shows that the sample median beats the sample mean in terms of asymptotic relative efficiency when estimating the centre point μ.

Exercise 2

The density of a gamma(a, b) distribution is

$$g(x;\theta) = \frac{b^a}{\Gamma(a)} x^{a-1}\exp(-bx), \quad \text{for } x > 0,$$

and zero otherwise, for positive parameters a and b, and the exponential(λ) density is

$$f(x;\lambda) = \lambda\exp(-\lambda x), \quad \text{for } x > 0,$$

and zero otherwise, for a positive parameter λ.

(a) The log-likelihood function is

$$\ell_n(\lambda) = n\log\lambda - n\lambda\bar{X}_n,$$

where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Solving

$$n/\lambda - n\bar{X}_n = 0$$

yields the maximum likelihood estimator $\hat{\lambda}_n = 1/\bar{X}_n$. A second derivative test show that the log-likelihood function is everywhere concave, so the solution is unique. The mgf of an exponential with mean $1/\lambda$ is (λ) is $M_X(t) = (1 - t/\lambda)^{-1}$, $t < \lambda$. So when $X_1, \ldots, X_n$ are independent exponentials with mean $1/\lambda$

$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n (1 - \lambda/t)^{-1} = (1 - \lambda/t)^{-n},$$

which is the mfg of a Gamma(n,λ) random variable. Recall that if $Y \sim \text{Gamma}(a,b)$, then

$$\mathrm{E}\,1/Y = \frac{b^a}{\Gamma(a)} \int_0^\infty y^{a-2}\exp(-by) = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a-1)}{b^{a-1}} = \frac{b}{a-1},$$

and similarly we find $\mathrm{E}\,1/Y^2 = b^2/\{(a-2)(a-1)\}$, assuming $a > 2$, which is okay when a is replaced by the sample size n. We can now compute the mean and variance of $\hat{\lambda}_n$.

$$\mathrm{E}_\lambda\,\hat{\lambda}_n = \mathrm{E}_\lambda \frac{1}{X_n} = \mathrm{E}_\lambda \frac{n}{\sum_{i=1}^n X_i} = \mathrm{E}_\lambda \frac{n}{\text{Gamma}(n,\lambda)} = \frac{n}{n-1}\lambda.$$

The variance is the maximum likelihood estimator is

$$\mathrm{Var}_\lambda\,\hat{\lambda}_n = \mathrm{E}\,\hat{\lambda}_n^2 - (\frac{n}{n-1}\lambda)^2 = \frac{n^2\lambda^2}{(n-1)(n-2)} - (\frac{n}{n-1}\lambda)^2 = \frac{\lambda^2 n^2}{(n-1)^2(n-2)}.$$

(b) The Fisher information is

$$-\mathrm{E}_\lambda \frac{\partial^2}{\partial\lambda^2}\ell_n(\lambda) = \frac{n}{\lambda^2},$$

so the Cramér–Rao lower bound for any unbiased estimator of λ is λ^2/n. In (a) we found the expectation of the mle, from which we see that

$$\widetilde{\lambda}_n = \frac{n-1}{n}\hat{\lambda}_n = \frac{n-1}{n}\frac{1}{\bar{X}_n},$$

is an unbiased estimator of λ. The variance of this estimator is

$$\mathrm{Var}_\lambda\,\widetilde{\lambda}_n = \frac{(n-1)^2}{n^2}\mathrm{Var}_\lambda\widehat{\lambda}_n = \frac{\lambda^2}{n-2} > \frac{\lambda^2}{n},$$

so $\widetilde{\lambda}_n$ does not attain the Cramér–Rao lower bound. The statistic $\bar{X}_n$ is complete sufficient by Theorem 6.2.25, p. 289 in C&B. Theorem 7.3.23 p. 347 then tells us that $\widetilde{\lambda}_n$ is the unique best unbiased estimator of its mean. In other words, there does not exist an unbiased estimator of λ attaining the CRLB.

(c) Reparametrise so that the exponentials have density $f(x;\eta) = \eta^{-1}\exp(-x/\eta), x > 0$. Then

$$-\mathrm{E}_\eta \frac{\partial^2}{\partial\eta^2} f(X;\eta) = -\frac{1}{\eta^2} + \frac{2}{\eta^3}\mathrm{E}_\eta\,X = \frac{1}{\eta^2},$$

and the CRLB is η^2/n. The mean $\bar{X}_n$ has expectation η and variance η^2/n.

(d) Use Chebyshev's inequality. For any $\varepsilon > 0$,

$$\Pr_{\lambda_0}\{|\widehat{\lambda}_n - \lambda_0| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2}\{\mathrm{Var}_{\lambda_0}\widehat{\lambda}_n + \text{bias}^2(\widehat{\lambda}_n)\} = \frac{1}{\varepsilon^2}\{\frac{n^2\lambda_0^2}{(n-1)^2(n-2)} + \frac{\lambda_0^2}{(n-1)^2}\},$$

where the right hand side tends to zero as $n \to \infty$. This shows that, under the model assumptions, the maximum likelihood estimator $\widehat{\lambda}_n$ is consistent for λ_0. From the central limit theorem (for example as stated in Nils exercise 16), we know that $\sqrt{n}(\bar{X}_n - 1/\lambda_0)$ converges in distribution to a $\mathrm{N}(0, 1/\lambda_0^2)$. Let $h(x) = 1/x$ and use the delta method. The first derivative of $h(x)$ is $h'(x) = -1/x^2$, then

$$\sqrt{n}(\widehat{\lambda}_n - \lambda_0) = \sqrt{n}\{h(\bar{X}_n) - h(1/\lambda_0)\} \overset{d}{\to} \mathrm{N}(0, h'(1/\lambda_0)^2/\lambda_0^2) = \mathrm{N}(0,\lambda_0^2).$$

(e) The statement of this question in the oblig was unclear. What I was looking for was a so called 'variance stabilising transformation', that is function g that does not involve the unknown parameters, and that ensures that $\sqrt{n}(g(\hat{\lambda}_n) - g(\lambda_0))$ converges in distribution to a standard normal. From the delta method we see such a function must satisfy

$$g'(\lambda_0)^2\lambda_0^2 = 1,$$

so $g(x) = \log x$. Then

$$\sqrt{n}(\log\widehat{\lambda}_n - \log\lambda_0) \overset{d}{\to} \mathrm{N}(0,1).$$

This means that for big enough n, the event

$$-1.96 \leq \sqrt{n}(\log\widehat{\lambda}_n - \log\lambda_0) \leq 1.96,$$

has approximately probability 0.95 of occurring. Moving things around we get

$$\widehat{\lambda}_n \exp(-1.96/\sqrt{n}) \leq \lambda_0 \leq \widehat{\lambda}_n \exp(1.96/\sqrt{n}),$$

so that $[\widehat{\lambda}_n \exp(-1.96/\sqrt{n}), \widehat{\lambda}_n \exp(1.96/\sqrt{n})]$ is an approximate 95% confidence interval for λ.

An exact confidence intervals for λ can be found by noting that when X is exponential(λ), the variable λX is a unit exponential, so

$$\lambda \sum_{i=1}^{n} X_i = \lambda n \bar{X}_n \sim \text{Gamma}(n, 1).$$

Let g_α be the value such that $\int_0^{g_\alpha} \text{Gamma}(n,1)\,\mathrm{d}x = \alpha$, then

$$[g_{0.025}/(n\bar{X}_n), g_{0.975}/(n\bar{X}_n)] = [\hat{\lambda}_n g_{0.025}/n, \hat{\lambda}_n g_{0.975}/n],$$

is a 95% confidence interval for λ_0. Comment on the approximate interval being symmetric, while the exact confidence interval is not.

(f) The Kullback–Leibler distance of a density f from a density g is

$$\text{KL}(g,f) = \text{E}_g \log(g(X)/f(X)) = \int \log\{g(x)/f(x)\}\, g(x)\,\mathrm{d}x.$$

With the gamma and exponential densities ,

$$\begin{aligned}\text{E}_g \log(g(X;a_0,b_0)/f(X;\lambda)) &= \text{E}_g \log g(X;a_0,b_0) - \text{E}_g\{\log\lambda - \lambda X\} \\ &= \text{E}_g \log g(X;a_0,b_0) - \log\lambda + \lambda\frac{a_0}{b_0},\end{aligned}$$

which is minimised for $\lambda_{\text{lf}} = b_0/a_0$. We are to show that $\widehat{\lambda}_n$ is consistent for the least false value $\lambda_{\text{lf}} = b_0/a_0$. We can find the true expectation and variance of the estimator $\widehat{\lambda}_n$, i.e., the expectation and variance when the $X_1,\ldots,X_n$ are Gamma(a_0,b_0),

$$\text{E}_g\,\widehat{\lambda}_n = \text{E}_g\,\frac{n}{\sum_{i=1}^n X_i} = \text{E}_g\,\frac{n}{\text{Gamma}(na_0,b_0)} = \frac{nb_0}{na_0-1}, \quad \text{and} \quad \text{Var}_g\,\widehat{\lambda}_n = \frac{n^2 b_0^2}{(na_0-1)^2(na_0-2)}.$$

Chebyshev's inequality, for any $\varepsilon > 0$

$$\Pr_g\{|\widehat{\lambda}_n - \lambda_{\text{lf}}| \geq \varepsilon\} \leq \varepsilon^{-2}\{\frac{n^2b_0^2}{(na_0-1)^2(na_0-2)} + \frac{1}{a_0^2(na_0-1)^2}\},$$

which goes to zero as n tends to infinity, so $\widehat{\lambda}_n$ is consistent for $\lambda_{\text{lf}} = b_0/a_0$.

The caveat here is that this proof uses knowledge of the world being gamma (in the exercise I did not specify that you were not to use this knowledge). To avoid using this, however, we could appeal to the consistency theorem presented and proved during the ML-lecture (3. October, I think). For this theorem to go through we need that $\hat{\lambda}_n$ is the unique maximiser of $\ell_n(\lambda)$ and that λ_{lf} is the unique minimiser of KL(g,f), respectively. They are, and many of you verified this. We also need that $\sup_\lambda |n^{-1}\ell_n(\lambda) - \text{E}_g\,\log f(X,\lambda)| \overset{p}{\to} 0$.

Assume that λ lives on the interval $(0,\lambda_{\max}]$, where $\lambda_{\max} < \infty$ is some very big number (this is an okay assumption to make, particularly because it is an assumption about our model, and not about the true distribution g. That being said, we are tacitly making assumptions about g as well here. Which ones?). So when $0 < \lambda \leq \lambda_{\max}$,

$$\begin{aligned}\sup_\lambda |n^{-1}\ell_n(\lambda) - \text{E}_g\,\log f(X,\lambda)| &= \sup_\lambda |n^{-1}\sum_{i=1}^n(\log\lambda - \lambda X_i) - \log\lambda + \lambda\text{E}_g\,X_i| \\ &= \lambda_{\max}|\text{E}_g\,X_i - \bar{X}_n| \overset{p}{\to} 0,\end{aligned}$$

because $\bar{X}_n \overset{p}{\to} \text{E}_g\,X_1 = 1/\lambda_{\text{lf}}$ by the weak law of large numbers (assuming finite first moment $\text{E}_g\,X_1 < \infty$, see the solutions-to-Nils-exercises section in the exercise set for this theorem). From this we conclude, having paid the price of an assumption on λ, that $\hat{\lambda}_n$ converges in probability to λ_{lf} without knowing that the world is gamma.

(g) From the central limit theorem,

$$\sqrt{n}(\bar{X}_n - a_0/b_0) \overset{d}{\to} \text{N}(0, a_0/b_0^2).$$

As above, the delta method with $h(x) = 1/x$,

$$\sqrt{n}(\widehat{\lambda}_n - \lambda_{\rm lf}) = \sqrt{n}\{h(\bar{X}_n) - h(a_0/b_0)\} \overset{d}{\to} {\rm N}(0, h'(a_0/b_0)^2 a_0/b_0^2) = {\rm N}(0, b_0^2/a_0^3).$$

In the notation of the oblig, $\kappa^2 = b_0^2/a_0^3$.

Find a consistent estimator of κ^2 without knowing the world is gamma. First, note that if we knew the world to be gamma, we could use the maximum likelihood estimators $\hat{a}_n$, $\hat{b}_n$ to estimate a, b. These are consistent, and continuous mapping would give $\hat{b}_n^2/\hat{a}_n^3$ as a consistent estimator of κ^2.

We do not know that the world is gamma. This part of the exercise is contained in Example 5.5.25 on p. 243 in C&B, where they use a delta method argument to find the form of the limiting variance. What I had in mind when making the exercise, however, was to show an example of a so called sandwich variance (or sandwich covariance matrix in the multidimensional case). From maximum likelihood theory, see for example Exercise 9.3 in the exercise set, we have that

$$\sqrt{n}(\widehat{\lambda}_n - \lambda_{\rm lf}) = \frac{\lambda_{\rm lf}^2}{\sqrt{n}} \sum_{i=1}^n \big(\frac{1}{\lambda_{\rm lf}} - X_i\big) + o_p(1).$$

The expression on the right is the score function of the exponential model, divided by the second derivative of the exponential log-likelihood, multiplied by $\sqrt{n}$. See Exercise 9.3 in the exercise set. You can verify that

$$\mathrm{Var}\,\Big\{\frac{\lambda_{\rm lf}^2}{\sqrt{n}} \sum_{i=1}^n \big(\frac{1}{\lambda_{\rm lf}} - X_i\big)\Big\} = \frac{b_0^2}{a_0^3} = \kappa^2.$$

This suggests the plug-in estimator

$$\widehat{\kappa}_n^2 = \frac{\widehat{\lambda}_n^4}{n} \sum_{i=1}^n \big(\frac{1}{\widehat{\lambda}_n} - X_i\big)^2 = \frac{1}{\bar{X}_n^4} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

By continuous mapping, $1/\bar{X}_n^4 \overset{p}{\to} b_0^4/a_0^4$. From Nils exercise 10 we know that

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \overset{p}{\to} \mathrm{Var}\, X_1 = \frac{a_0}{b_0^2}.$$

Since $A_n \overset{p}{\to} a$ and $B_n \overset{p}{\to} b$, entails that $A_n B_n \overset{p}{\to} ab$, see Nils exercise 9(c), this gives that

$$\widehat{\kappa}_n^2 = \frac{S_n^2}{\bar{X}_n^4} \overset{p}{\to} \frac{b_0^4}{a_0^4} \frac{a_0}{b_0^2} = \frac{b_0^2}{a_0^3} = \kappa^2.$$

(h) The approximate 95% confidence interval for λ_0 that we found in (e) is

$$[0.17, 0.22].$$

The approximate 95% confidence interval for $\lambda_{\rm lf}$ that we found in (f) is

$$[0.16, 0.23].$$

These are both confidence intervals for the inverse expectation, under both models. So they are comparable. Note, however, that this is not true in general, the least false parameter value might be something very different from what one, naively, sets out to estimate.

If we want to construct an confidence interval for the parameter

$$\mu = \mathrm{E}_g\, X_1,$$

only assuming the second moment to be finite, we use that $\sqrt{n}(\bar{X}_n - \mu) \overset{d}{\to} {\rm N}(0, \mathrm{Var}_g\, X_1)$. And take as our approximate 95% confidence interval for the mean,

$$[\bar{X}_n - 1.96 S_n/\sqrt{n}, \bar{X}_n + 1.96 S_n/\sqrt{n}],$$

where $S_n^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ is a consistent estimator of the variance.

(i) Dependence in time, a model that incorporates the sunny days, and so on.

Here is the R-script that I used to make Figure 1, and to compute the confidence intervals.

```r
data <- read.table("https://www.uio.no/studier/emner/matnat/math/STK4011/h19/weather_blindern.txt",sep=";",header=TRUE)
yy <- data$rain_mm
yy_rain <- yy[yy > 0] # discard sunny days

# Estimate lambda in expo model
lambda_hat <- 1/mean(yy_rain)

# Fit Gamma(a,b) model to make the figure
loglik <- function(params){
  aa <- params[1] ; bb <- params[2]
  ll <- sum( aa*log(bb) - log(gamma(aa)) + (aa - 1)*log(yy_rain) - bb*yy_rain )
  # Obs! Returns the negative
  return(-ll)
}

bb_mom <- mean(yy_rain)/var(yy_rain) # Use MoM as start values
aa_mom <- bb_mom*mean(yy_rain)
fit <- nlm(loglik,c(aa_mom,bb_mom),hessian=TRUE)

#postscript("rain_Blindern_sept18_oct19.eps") # Figure 1 in the oblig
hist(yy_rain,breaks=20,freq=FALSE,ylab="Density", xlab="Precipitation in mm at Blindern Sept. 1, 2018 - Oct. 9, 2019. Sunny days are not included",main="",cex.lab=sqrt(2))
legend(30,0.2,lty = c(1,1),col=c("blue","red"),legend=c("expo","gamma"),cex=sqrt(2),bty="n")
curve(dgamma(x,fit$estimate[1],fit$estimate[2]),add=TRUE,col="red")
curve(dexp(x,lambda_hat),add=TRUE,col="blue")
#dev.off()

# Confidence intervals
nn <- length(yy_rain)
# Assuming exponential world. With CI from exercise (d)
lambda_hat*exp(-1.96/sqrt(nn))
lambda_hat*exp(1.96/sqrt(nn))
# Agnostic confidence interval
kappahat_sq <- (nn-1)/nn*var(yy_rain)/mean(yy_rain)^4
lambda_hat - 1.96*sqrt(kappahat_sq)/sqrt(nn) # 0.1586007
lambda_hat + 1.96*sqrt(kappahat_sq)/sqrt(nn) # 0.2330278
```

Exercise 3

Independent Poisson data $Y_1, \ldots, Y_n$, density given by

$$P_\lambda(Y = y) = \frac{\lambda^y}{y!}\exp(-\lambda), \quad \text{for } y = 0, 1, 2, \ldots,$$

with $\lambda > 0$.

(a) The log-likelihood function is

$$\ell_n(\lambda) = n\bar{y}_n \log\lambda - n\lambda - \sum_{i=1}^{n}\log(y_i!).$$

Solve

$$\frac{n\bar{y}_n}{\lambda} - n = 0,$$

to find the maximum likelihood estimator for λ, which is $\bar{Y}_n$. By the invariance property of maximum likelihood estimators (see Theorem 7.2.10, p. 320 in C&B), the maximum likelihood estimator of $\alpha = \Pr_\lambda(Y = 0) = \exp(-\lambda)$ is

$$\widehat{\alpha}_n = \exp(-\bar{Y}_n).$$

The Fisher information is

$$-\operatorname{E}_\lambda \frac{\partial^2}{\partial \alpha^2} \ell_n(\alpha) = \frac{n}{-\log \alpha} = \frac{n}{\lambda}.$$

so the CRLB for any unbiased estimator of α is λ/n.

Recall that the moment generating function of the Poisson(λ) distribution is

$$\operatorname{E}_\lambda \exp(tY) = \exp\{\lambda(e^t - 1)\},$$

for all t. Since the $Y_1, \ldots, Y_n$ are assumed independent, the expectation of $\widehat{\alpha}_n$ is

$$\begin{aligned}\operatorname{E}_\lambda \widehat{\alpha}_n &= \operatorname{E}_\lambda \exp(-\frac{1}{n}\sum_{i=1}^n Y_i) = \prod_{i=1}^n \operatorname{E}_\lambda \exp(-\frac{1}{n} Y_i) \\ &= \prod_{i=1}^n \exp\{\lambda(e^{-1/n} - 1)\} = \exp\{n\lambda(e^{-1/n} - 1)\},\end{aligned}$$

and similarly, the second moment of the estimator is $\operatorname{E}_\lambda \widehat{\alpha}_n^2 = \exp\{n\lambda(e^{-2/n} - 1)\}$. The variance is then

$$\operatorname{Var}_\lambda \widehat{\alpha}_n = \operatorname{E}_\lambda \widehat{\alpha}_n^2 - \big(\operatorname{E}_\lambda \widehat{\alpha}_n\big)^2 = \exp\{n\lambda(e^{-2/n} - 1)\} - \exp\{2n\lambda(e^{-1/n} - 1)\}.$$

Using l'Hôpital's rule, for $c > 0$ fixed

$$\lim_n n\lambda(e^{-c/n} - 1) = \lim_n \frac{\lambda(e^{-c/n} - 1)}{1/n} = -\lim_n c\lambda e^{-c/n} = -c\lambda.$$

Since $\exp(x)$ is continuous, this gives that

$$\exp\{n\lambda(e^{-2/n} - 1)\} \to \exp(-2\lambda), \quad \text{and} \quad \exp\{2n\lambda(e^{-1/n} - 1)\} \to \exp(-2\lambda),$$

when $n \to \infty$, which entails that $\operatorname{Var}_\lambda \widehat{\alpha}_n \to 0$ and $\operatorname{bias}^2(\widehat{\alpha}_n) = \{\exp\{n\lambda(e^{-1/n} - 1)\} - \exp(-\lambda)\}^2 \to 0$ as $n \to \infty$. So by by Chebyshev's inequality, for all $\varepsilon > 0$,

$$\Pr_\lambda(|\widehat{\alpha}_n - \alpha| \geq \varepsilon) \leq \frac{\operatorname{Var}_\lambda \widehat{\alpha}_n + \operatorname{bias}^2(\widehat{\alpha}_n)}{\varepsilon^2} \to 0, \quad \text{as } n \to \infty,$$

which shows that $\widehat{\alpha}_n$ is consistent for α.

Using the chain rule

$$\begin{aligned}\frac{1}{\lambda} &= -\frac{\partial^2}{\partial \lambda^2} \log f(y; \lambda) = -\{\frac{\partial^2}{\partial \alpha^2} \log f(y; \alpha) e^{-2\lambda} - \frac{\partial}{\partial \alpha} \log f(y; \alpha) e^{-\lambda}\} \\ &= -e^{-2\lambda} \frac{\partial^2}{\partial \alpha^2} \log f(y; \alpha).\end{aligned}$$

Taking expectations on both sides and moving things around yields the Fisher information for α,

$$-\operatorname{E} \frac{\partial^2}{\partial \alpha^2} \log f(y; \alpha) = \frac{1}{\lambda e^{-2\lambda}},$$

because the score function has expectation zero. The CRLB for unbiased estimators of $\alpha = \exp(-\lambda)$ is $\lambda e^{-2\lambda}/n$.

(b) For $i = 1, \ldots, n$, let $Z_i = 1$ if $Y_i = 0$, and $Z_i = 0$ otherwise (there was a typo in the oblig here). The random variables $Z_1, \ldots, Z_n$ are then independent Binomial with success probability $\alpha = \exp(-\lambda)$, thus $\operatorname{E} Z_i = \Pr(Y_i = 0) = e^{-\lambda} = \alpha$, so that $\widetilde{\alpha}_n = n^{-1} \sum_{i=1}^n Z_i$ is an unbiased estimator for α, and $\operatorname{Var} Z_i = \alpha(1 - \alpha)$. By the central limit theorem

$$\sqrt{n}(\widetilde{\alpha}_n - e^{-\lambda}) \overset{d}{\to} \mathrm{N}(0, \kappa_2^2), \quad \text{with} \quad \kappa_2^2 = \alpha(1 - \alpha) = e^{-\lambda} - e^{-2\lambda}$$

We also have that

$$\sqrt{n}(\bar{Y}_n - \lambda) \overset{d}{\to} \mathrm{N}(0, \lambda).$$

Use the delta method to find

$$\sqrt{n}(\widehat{\alpha}_n - e^{-\lambda}) = \sqrt{n}(e^{-\bar{Y}_n} - e^{-\lambda}) \overset{d}{\to} \mathrm{N}(0, e^{-2\lambda}\lambda),$$

In the notation of the oblig $\kappa_1^2 = e^{-2\lambda}\lambda$. The asymptotic relative efficiency is

$$\frac{\kappa_1^2}{\kappa_2^2} = \frac{e^{-\lambda}\lambda}{1-e^{-\lambda}} = \frac{\lambda}{e^{\lambda}-1},$$

which is always smaller than 1 because $e^{\lambda} \geq 1+\lambda$ for $\lambda > 0$. The estimator $\widehat{\alpha}_n$ is therefore to be preferred above $\widetilde{\alpha}_n$.

(c) The joint density of $Y_1,\ldots,Y_n$ is

$$P_\lambda(Y_1 = y_1,\ldots,Y_n = y_n) = \frac{1}{y_1!\cdots y_n!}\exp(-n\lambda)\exp\{\log(\lambda)\sum_{i=1}^n y_i\}.$$

From this we see that $\sum_{i=1}^n Y_i$ is sufficient by the Factorisation theorem, p. 276 in C&B; and complete by Theorem 6.2.25, p. 288 in the same book. The Rao–Blackwell theorem (p. 342 in C&B) yields that

$$\alpha^\star = \mathrm{E}_\lambda\,\{\widetilde{\alpha} \mid \sum_{i=1}^n Y_i\},$$

is the best unbiased estimator of α. Since $\sum_{i=1}^n Y_i$ is sufficient and complete, uniqueness of $\alpha_n^\star$ follows from Theorem 7.3.23, p. 347 in C&B.

To compute this estimator we can proceed in (at least) two different manners. (i) A sum $\sum_{i=1}^k Y_i$ of independent Poisson rv's, is Poisson$(k\lambda)$. By Bayes theorem and independence of the Y_i's,

$$\begin{aligned}P_\lambda(Z_i = 1 \mid \sum_{j=1}^n Y_j = w) = P_\lambda(Y_i = 0 \mid \sum_{j=1}^n Y_j = w) &= \frac{P_\lambda(\sum_{j=1}^n Y_j = w \mid Y_i = 0)P(Y_i = 0)}{P_\lambda(\sum_{j=1}^n Y_j = w)}\\ &= \frac{P_\lambda(\sum_{j\neq i} Y_j = w)P(Y_i = 0)}{P_\lambda(\sum_{i=1}^n Y_i = w)} = \big(1-\frac{1}{n}\big)^w.\end{aligned}$$

By linearity of conditional expectation and since the Y_i's are i.i.d., the estimator is

$$\alpha^\star = \big(1-\frac{1}{n}\big)^{\sum_{i=1}^n Y_i}.$$

(ii) Here is a slightly more involved way of doing this, but where you learn some more along the way. In general, if $Y_1,\ldots,Y_n$ are independent Poisson random variables with means $\lambda_1,\ldots,\lambda_n$, then

$$(Y_1,\ldots,Y_n) \mid (\sum_{i=1}^n Y_i = w) \sim \text{Multinomial}\big(w, \frac{\lambda_1}{\gamma},\ldots,\frac{\lambda_n}{\gamma}\big),$$

where $\gamma = \sum_{i=1}^n \lambda_i$ To see this, use that $W_n \coloneqq \sum_{i=1}^n Y_i \sim \text{Poisson}(\gamma)$, then

$$\begin{aligned}\Pr(Y_1 = y_1,\ldots,Y_n = y_n \mid W_n = w) &= \frac{\Pr(W_n = w \mid Y_1 = y_1,\ldots,Y_n = y_n)P(Y_1 = y_1,\ldots,Y_n = y_n)}{P(W_n = w)}\\ &= \frac{\frac{1}{y_1!\cdots y_n!}\lambda_1^{y_1}\cdots\lambda_n^{y_n}e^{-\sum_{i=1}^n\lambda_i}}{\frac{1}{w!}\gamma^w e^{-\gamma}} = \frac{w!}{x_1!\cdots x_n!}(\lambda_1/\gamma)^{y_1}\cdots(\lambda_n/\gamma)^{y_n}.\end{aligned}$$

This means that

$$Y_j \mid (W_n = w) \sim \text{Binomial}(w, \lambda_j/\gamma).$$

Back to the estimator,

$$\mathrm{E}\,\{Z_j \mid W_n = w\} = \Pr(Y_j = 0 \mid W_n = w) = \big(1-\frac{1}{n}\big)^w,$$

and we get $\alpha_n^\star = (1-1/n)^{\sum_{i=1}^n Y_i}$ as above.

(d) The point estimates of $\alpha = \exp(-\lambda)$ are

$\widehat{\alpha}_n$	$\widetilde{\alpha}_n$	$\alpha_n^\star$
0.338	0.374	0.336

(e) We are going to treat the x_i's as fixed and known covariates. Let $\theta = (\beta_0, \beta_1)$, and consider the Poisson regression model where $Y_1, \dots, Y_n$ are independent with means

$$\lambda_i = \exp(\beta_0 + \beta_1 x_i), \quad \text{for } i = 1, \dots, n. \tag{2}$$

The log-likelihood function is

$$\ell_n(\beta_0, \beta_1) = \sum_{i=1}^n \{y_i(\beta_0 + \beta_1 x_i) - \exp(\beta_0 + \beta_1 x_i) - \log(y_i!)\},$$

the score function are

$$\frac{\partial}{\partial \beta_0} \ell_n(\beta_0, \beta_1) = \sum_{i=1}^n \{y_i - \exp(\beta_0 + \beta_1 x_i)\},$$
$$\frac{\partial}{\partial \beta_1} \ell_n(\beta_0, \beta_1) = \sum_{i=1}^n x_i\{y_i - \exp(\beta_0 + \beta_1 x_i)\}.$$

The observed information matrix is

$$J_n(\beta_0, \beta_1) = -\frac{\partial^2}{\partial \beta \partial \beta^{\mathrm{t}}} \ell_n(\beta_0, \beta_1) = \sum_{i=1}^n \lambda_i \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}.$$

Since $\lambda_i > 0$, this matrix is positive definite provided $n^{-1} \sum_{i=1}^n x_i^2 > \bar{x}_n^2$, or equivalently, $\sum_{i=1}^n (x_i - \bar{x}_n)^2 > 0$. We can see this from the definition. Let $a = (a_1, a_2)^{\mathrm{t}}$ be any non-zero vector in $\mathbb{R}^2$, then under our assumption

$$\begin{aligned} \frac{1}{n} a^{\mathrm{t}} J_n(\beta_0, \beta_1) a &= \frac{1}{n} \sum_{i=1}^n \lambda_i (a_1 + a_2 x_i)^2 \\ &\geq \max_{i \leq n}(\lambda_i)\{a_1^2 + 2a_1 a_2 \bar{x}_n + a_2^2 n^{-1} \sum_{i=1}^n x_i^2\} > \max_{i \leq n}(\lambda_i)(a_1 + a_2 \bar{x}_n)^2 \geq 0, \end{aligned}$$

for any vector a, and $J(\beta_0, \beta_1)$ is positive definite. Then $-\ell_n(\beta_0, \beta_1)$ is everywhere convex, so $\ell_n(\beta_0, \beta_1)$ is everywhere concave, and the maximisers of $\ell_n(\beta_0, \beta_1)$ are unique.

(f) The likelihood of the data can be written on exponential family form

$$L_n(\theta; y) = \prod_{i=1}^n \frac{1}{y_i!} \lambda_i^{y_i} \exp(-\lambda_i) = h(y) c(\beta_0, \beta_1) \exp\big(\beta_0 \sum_{i=1}^n y_i + \beta_1 \sum_{i=1}^n x_i y_i\big), \tag{3}$$

where

$$h(y) = \frac{1}{y_1! \cdots y_n!}, \quad \text{and} \quad c(\beta_0, \beta_1) = \exp(n\beta_0 + \beta_1 n \bar{x}_n).$$

According to Theorem 6.2.25 on p. 288 in C&B, this shows that

$$T = (\sum_{i=1}^n Y_i, \sum_{i=1}^n x_i Y_i),$$

is complete and sufficient as long as the parameter space contains an open set in $\mathbb{R}^2$, a very natural assumption to make in this model. If $\mu^\star$ is an estimator such that

$$R(\mu^\star, \mu) = \mathrm{E}\, L(\mu^\star, \mu) \leq \mathrm{E}\, L(\delta, \mu) = R(\delta, \mu),$$

for every other estimator δ, then $\mu^\star$ must be a function of a complete and sufficient statistic, or else we could find an estimator improving on $\mu^\star$ by Rao–Blackwellisation (see Exercise 6.21 in the exercise set). But from the likelihood in (3) we see that the maximum likelihood estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are functions of $T = (\sum_{i=1}^n Y_i, \sum_{i=1}^n x_i Y_i)$.

(g) The parameter estimates along with associated approximate 95% confidence intervals are given in Table 1.

	β_0	β_1
Estimates	-0.140	0.016
Approx. 95% CIs	[-0.617, 0.337]	[-0.016, 0.048]

Table 1. Estimates and approximate 95% confidence intervals for the Poisson model of Exercise 3g).

Since the approximate confidence interval for β_1 includes zero, it appears that the insights of 50 cent in 2003 do not apply to Oslo in 2019.

(h) In this exercise you are asked to fit a type of model that is often referred to as a fixed effects model. The parameter estimates along with confidence intervals are in Table 2.

	Estimates	Approx. 95% CIs
β	0.011	[-0.020, 0.043]
Monday	0.670	[0.241, 1.098]
Tusday	0.856	[0.341, 1.370]
Wednesday	0.643	[0.235, 1.051]
Thursday	0.680	[0.256, 1.104]
Friday	0.776	[0.288, 1.269]
Saturday	1.626	[0.737, 2.516]
Sunday	1.258	[0.543, 1.972]

Table 2. Estimates and approximate 95% confidence intervals for the Poisson model of Exercise 3h).

The estimate of β is not significant on the 5% level. The approximate confidence intervals of Tuesdays and Fridays are overlapping, so this simple test does not indicate a difference. But there are other tests.

(i) This type of model is often called a random effect model. To find the likelihood we must integrate out the parameters $\gamma_1, \dots, \gamma_7$. By independence, for $j = 1, \dots, 7$,

$$L_{n_j}(\beta, \eta) = \int_0^\infty \{\prod_{i=1}^{n_j} \frac{1}{y_{i,j}} \gamma_j^{y_{i,j}} e^{\beta x_i y_{i,j}} e^{-\gamma_j e^{x_i \beta}}\} \, \eta e^{-\eta\gamma_j} \, \mathrm{d}\gamma_j$$
$$= \eta \frac{\Gamma(\sum_{i=1}^{n_j} y_{i,j} + 1)}{\prod_{i=1}^{n_j} y_{i,j}!} \prod_{i=1}^{n_j} \Big(\frac{e^{x_i\beta}}{\eta + e^{x_i\beta}}\Big)^{y_{i,j}} \frac{1}{\eta + e^{x_i\beta}},$$

and the full likelihood is $L_n(\beta, \eta) = L_{n_1}(\beta, \eta) \cdots L_{n_7}(\beta, \eta)$ by independence of the γ_j's. The full log-likelihood is

$$\ell_n(\beta, \eta) = \ell_{n_1}(\beta, \eta) + \cdots + \ell_{n_7}(\beta, \eta)$$
$$= 7 \log \eta + \sum_{j=1}^{7} \big[\{\log \Gamma(\sum_{i=1}^{n_j} y_{i,j} + 1) - \sum_{i=1}^{n_j} \log(y_{i,j}!)\} + \sum_{i=1}^{n_j} \{y_{i,j} x_i \beta - (y_{i,j} + 1) \log(\eta + e^{x_i\beta})\}\big],$$

where $\ell(\beta, \eta) = \log L(\beta, \eta)$ as usual. Maximising this log-likelihood gives the parameter estimates in Table 3.

	η	β
Estimates	1.078	0.012
Approx. 95% CIs	[0.130, 2.026]	[-0.020, 0.043]

Table 3. Estimates and approximate 95% confidence intervals for the Poisson model of Exercise 3i).

Here is a R-script for exercises 3d), 3g), 3h), and 3i).

```r
data <- read.table("https://www.uio.no/studier/emner/matnat/math/STK4011/h19/tweetcount_temp.txt",sep=";")
head(data)

yy <- data$violence
xx <- data$avg_temp
nn <- length(yy)

# Exercise 3d)
# Three different point estimators
yy_bar <- mean(yy)
exp(-yy_bar) # \hat{\alpha}_n = 0.3376013
mean(yy == 0) # \tilde{\alpha}_n # 0.3742331
(1 - 1/nn)^{sum(yy)} # \alpha_n^{\star} # 0.3364741


# Exercise 3e)

loglik1 <- function(params){
  beta0 <- params[1] ; beta1 <- params[2]

  ll <- sum( yy*(beta0 + beta1*xx) - exp(beta0 + beta1*xx) - log(gamma(yy+1)) )
  # Obs! Returns negative
  return(-ll)
}

fit1 <- nlm(loglik1,c(-0.14,0.02),hessian=TRUE)


fit1$estimate # -0.14025679  0.01592917
# Approximate 95 percent CIs
J_hat <- fit1$hessian/nn
fit1$estimate - 1.96*sqrt(diag(solve(J_hat)))/sqrt(nn) # -0.61723524 -0.01614002
fit1$estimate + 1.96*sqrt(diag(solve(J_hat)))/sqrt(nn) # 0.33672166 0.04799835



# Exercise 3h)
day <- data$day_of_week
day_num <- 0:1*163
for(ii in 1:163){
  if(day[ii] == "Monday"){day_num[ii] <- 1}
  if(day[ii] == "Tuesday"){day_num[ii] <- 2}
  if(day[ii] == "Wednesday"){day_num[ii] <- 3}
  if(day[ii] == "Thursday"){day_num[ii] <- 4}
  if(day[ii] == "Friday"){day_num[ii] <- 5}
  if(day[ii] == "Saturday"){day_num[ii] <- 6}
  if(day[ii] == "Sunday"){day_num[ii] <- 7}
}
day_num

loglik_fix <- function(params){
  beta <- params[1] ; gammas <- params[2:8]
  ll_aux <- 0*1:7
  for(jj in 1:7){
```

```
    ll_aux[jj] <- sum( log( dpois(yy[day_num==jj],gammas[jj]*exp(beta*xx[day_num == jj]) ) ) )
  }
  ll <- sum(ll_aux)
  return(-ll)
}


fit_fix <- nlm(loglik_fix,c(0.016,rep(exp(-0.14),7)),hessian=TRUE)
fit_fix$estimate

for(pp in 1:8){
  lower <- fit_fix$estimate[pp] - 1.96*sqrt(diag(solve(fit_fix$hessian/nn))[pp])/sqrt(nn)
  upper <- fit_fix$estimate[pp] + 1.96*sqrt(diag(solve(fit_fix$hessian/nn))[pp])/sqrt(nn)
  cat(lower,upper,"\n")
}
# -0.02046302 0.04314277
# 0.2414227 1.097854
# 0.3409762 1.370064
# 0.2346463 1.051126
# 0.2557105 1.10354
# 0.2884926 1.263888
# 0.7367713 2.515736
# 0.5433132 1.971829


# Exercise 3i)

loglik_random <- function(params){
  eta <- params[1] ; beta <- params[2] ;
  ll_aux <- 0*1:7
  for(jj in 1:7){
    ll_aux[jj] <- log(eta) + log(gamma( sum(yy[day_num==jj]) + 1  )) - sum(log( gamma(yy[day_num==jj] + 1) ) ) + sum( beta*yy[day_num==jj]*xx[day_num == jj] ) - (sum(yy[day_num==jj]) + 1)*log(eta + sum(exp(beta*xx[day_num == jj])))

  }
  ll <- sum(ll_aux)
  return(-ll)
}

fit_random <- nlm(loglik_random,c(1/exp(-0.14),0.2),hessian=TRUE)
fit_random$estimate # 1.07834837 0.01151055

for(pp in 1:2){
  lower <- fit_random$estimate[pp] - 1.96*sqrt(diag(solve(fit_random$hessian/nn))[pp])/sqrt(nn)
  upper <- fit_random$estimate[pp] + 1.96*sqrt(diag(solve(fit_random$hessian/nn))[pp])/sqrt(nn)
  cat(lower,upper,"\n")
}
# 0.1303514 2.026345
# -0.02031253 0.04333364
```