

Solutions Sheet 3

Nina Fischer and Yannick Zelle

November 2, 2021

Exercise 1

Task: Describe the maximum likelihood estimator for the following distributions:

(a)

$$\mathcal{N}(x \mid \mu, \sigma^2)$$

We denote with f_μ the PDF of the normal distribution. By using the Lemma 6.2.3 we can write the log-likelihood function as:

$$\begin{aligned} l(\mu) &= \ln\left(\prod_{i=1}^n f_\mu\right) \\ &= \ln\left(\prod_{i=1}^n e^{a+\eta x_i - \frac{1}{2}\lambda^2 x_i^2}\right) \\ &= \sum_{i=1}^n \ln(e^{a+\eta x_i - \frac{1}{2}\lambda^2 x_i^2}) \\ &= \sum_{i=1}^n -\frac{1}{2}(\log(2\pi) - \log(\lambda^2) + \frac{\mu^2}{\sigma^2}) + \frac{\mu}{\sigma^2}x_i - \frac{1}{2}\lambda^2 x_i^2 \end{aligned}$$

To find μ which maximizes $l(\mu)$ we calculate:

$$\begin{aligned} \frac{\delta l(\mu)}{\delta \mu} &= \sum_{i=1}^n -\frac{\mu}{\sigma^2} + \frac{1}{\sigma^2}x_i \\ &= -n\frac{\mu}{\sigma^2} + \frac{1}{\sigma^2}\sum_{i=1}^n x_i \end{aligned}$$

$l(\mu)$ is maximal if :

$$\begin{aligned}
l(\mu) = 0 &\leftrightarrow 0 = -n \frac{\mu}{\sigma^2} + \frac{1}{\sigma^2} \sum_{i=1}^n x_i \\
&\leftrightarrow \mu = \frac{1}{n} \sum_{i=1}^n x_i
\end{aligned}$$

Since $\frac{\delta^2 l(\mu)}{\delta^2 \mu} = -\frac{1}{\sigma^2} < 0$, with $\mu_o := \frac{1}{n} \sum_{i=1}^n x_i$ we obtain a global maximum for $l(\mu)$ and we chose μ_0 therefore as our estimator.

(b)

Exponential distribution with the probability density function $f(x|\lambda) = \lambda e^{-\lambda x}$ with $\lambda > 0$ and observations $x_i \geq 0$. We have :

$$\begin{aligned}
f(x|\lambda) &= \lambda e^{-\lambda x} \\
&= e^{-\lambda x + \ln(\lambda)}
\end{aligned}$$

and the likelihood function:

$$L(\lambda) = \prod_{i=1}^n f(x_i|\lambda)$$

Therefore the log-likelihood is given by:

$$\begin{aligned}
l(\lambda) &= \ln\left(\prod_{i=1}^n f(x_i|\lambda)\right) \\
&= \sum_{i=1}^n \ln(f(x_i|\lambda)) \\
&= \sum_{i=1}^n \ln(e^{-\lambda x_i + \ln(\lambda)}) \\
&= \sum_{i=1}^n -\lambda x_i + \ln(\lambda)
\end{aligned}$$

With the same argument as in (a) we derive:

$$\begin{aligned}
\frac{\delta l(\lambda)}{\delta \lambda} &= \sum_{i=1}^n x_i + \frac{1}{\lambda} \\
&= \frac{n}{\lambda} - \sum_{i=1}^n x_i
\end{aligned}$$

Then we have :

$$0 = \frac{\delta l(\lambda)}{\delta \lambda} \leftrightarrow \lambda = \frac{n}{\sum_{i=1}^n x_i}$$

We also have

$$\frac{\delta^2 l(\lambda)}{\delta^2 \lambda} = -\frac{1}{\lambda^2} < 0$$

therefore we chose $\lambda_0 = \frac{n}{\sum_{i=1}^n x_i}$ as our estimator.

(c)

Gamma distribution with the probability density function $g(x|\alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{(\alpha-1)} e^{-\lambda x}$ with $\alpha > 0$ and observations $x_1, \dots, x_n \geq 0$. We have :

$$\begin{aligned} g(x|\alpha, \lambda) &= \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{(\alpha-1)} e^{-\lambda x} \\ &= e^{\ln(\frac{1}{\Gamma(\alpha)}) \lambda^\alpha x^{(\alpha-1)} - \lambda x} \end{aligned}$$

and the likelihood function:

$$L(\alpha, \lambda) = \prod_{i=1}^n g(x_i|\alpha, \lambda)$$

Therefore the log-likelihood is given by:

$$\begin{aligned} l(\alpha, \lambda) &= \ln\left(\prod_{i=1}^n g(x_i|\alpha, \lambda)\right) \\ &= \sum_{i=1}^n \ln(g(x_i|\alpha, \lambda)) \\ &= \sum_{i=1}^n \ln(e^{\ln(\frac{1}{\Gamma(\alpha)}) \lambda^\alpha x_i^{(\alpha-1)} - \lambda x_i}) \\ &= \sum_{i=1}^n \ln\left(\frac{1}{\Gamma(\alpha)} \lambda^\alpha x_i^{(\alpha-1)}\right) - \lambda x_i \end{aligned}$$

With the same argument as in (a) we derive:

$$\begin{aligned}\frac{\delta l(\alpha, \lambda)}{\delta \lambda} &= \sum_{i=1}^n \frac{\alpha}{\Gamma(\alpha)} x_i^{(\alpha-1)} \lambda^{(\alpha-1)} \frac{\Gamma(\alpha)}{\lambda^\alpha x_i^{\alpha-1}} \\ &= \sum_{i=1}^n \frac{\alpha}{\lambda} - x_i \\ &= n \frac{\alpha}{\lambda} - \sum_{i=1}^n x_i\end{aligned}$$

Then we have :

$$0 = \frac{\delta l(\alpha, \lambda)}{\delta \lambda} \leftrightarrow \lambda = \frac{n\alpha}{\sum_{i=1}^n x_i}$$

We also have

$$\frac{\delta^2 l(\alpha, \lambda)}{\delta^2 \lambda} = n \frac{\alpha}{\lambda^2} < 0$$

therefore we chose $\lambda_0 = \frac{n\alpha}{\sum_{i=1}^n x_i}$ as our estimator.

Exercise 2

Given: From our observation of the experiment and our prior knowledge we have :

$$\begin{aligned}p(\theta) &= Dir(\theta \mid \alpha) \\ p(\mathcal{D} \mid \theta) &= Mu(x \mid n, \theta)\end{aligned}$$

Show

$$p(\theta \mid \mathcal{D}) = Dir(\theta \mid \alpha + x)$$

Proof. Since the normalizing factor of a Distribution is dependent on the other factor of the distribution we can show the statement by showing :

$$p(\theta \mid \mathcal{D}) \propto Dir(\theta \mid \alpha + x)$$

From the lecture we know:

$$\begin{aligned}
p(\theta \mid \mathcal{D}) &= p(\theta) \cdot p(\mathcal{D} \mid \theta) \\
&= Dir(\theta \mid \alpha) \cdot Mu(x \mid n, \theta) \\
&= Z^{-1} \cdot \prod_{k=1}^K \theta_k^{a_k-1} \cdot Z^{-2} \cdot \prod_{j=1}^K \theta_j^{x_j} \\
&= Z^{-3} \cdot \prod_{k=1}^K \theta_k^{a_k+x_k-1} \\
&\propto Dir(\theta \mid \alpha + x)
\end{aligned}$$

With Z^{-i} being normalizing factors. Therefore we have

$$p(\theta \mid \mathcal{D}) = Dir(\theta \mid \alpha + x)$$

□