

Homeworks of Bayesian Inference
(B004652)

Group 1

Giovanni Papini

Last revision October 11, 2018

Week 1

Exchangeability and stochastic processes

1.1 Exercise 1

1.1.1 Data

- E_1, E_2, E_3, E_4, E_5 , events of simple alternative, exchangeable
- $P(E_2) = \omega_1 = \frac{1}{2}$
- $P(E_3 \wedge E_5) = \omega_2 = \frac{1}{4}$
- $\omega_5 = \frac{\omega_2^5}{\binom{5}{3}} = \frac{\omega_2^5}{\binom{5}{1}} = \frac{1}{30}$

1.1.2 Questions

Compute:

1. $P(E_2 \wedge E_3 \wedge E_4) = \omega_3$
2. $P(E_1 \wedge E_2 \wedge E_3 \wedge E_4) = \omega_4$
3. $P(E_1 \wedge E_2 \wedge \bar{E}_3 \wedge \bar{E}_4 \wedge \bar{E}_5) = \frac{\omega_2^5}{\binom{5}{2}}$

1.1.3 Solutions

First we find ω_1^5 and ω_3^5 :

$$\begin{aligned}\omega_1^5 &= \frac{1}{30} \cdot \binom{5}{1} = \frac{1}{6} \\ \omega_3^5 &= \frac{1}{30} \cdot \binom{5}{3} = \frac{1}{3}\end{aligned}$$

Knowing that

$$\omega_h = \frac{1}{\binom{n}{h}} \sum_{r=h}^n \omega_r^n \binom{r}{h}$$

we can write that

$$\begin{aligned}\omega_1 &= \frac{\omega_1^5 \binom{1}{1} + \omega_2^5 \binom{2}{1} + \omega_3^5 \binom{3}{1} + \omega_4^5 \binom{4}{1} + \omega_5^5 \binom{5}{1}}{\binom{5}{1}} \\ &= \frac{1}{6} \cdot \frac{1}{5} + \frac{2}{5} \omega_2^5 + \frac{1}{5} \cdot \frac{1}{3} \cdot 3 + \frac{1}{5} \cdot 4 \omega_4^5 + \frac{1}{30} \\ &= \frac{8}{30} + \frac{2}{5} \omega_2^5 + \frac{4}{5} \omega_4^5\end{aligned}$$

$$\begin{aligned}\omega_2 &= \frac{\omega_2^5 \binom{2}{2} + \omega_3^5 \binom{3}{2} + \omega_4^5 \binom{4}{2} + \omega_5^5 \binom{5}{2}}{\binom{5}{2}} \\ &= \frac{1}{10} \omega_2^5 + \frac{1}{10} \cdot \frac{1}{10} \cdot 3 + \frac{1}{10} \cdot 6 \omega_4^5 + \frac{1}{30} \\ &= \frac{2}{15} + \frac{1}{10} \omega_2^5 + \frac{3}{5} \omega_4^5\end{aligned}$$

Combining them:

$$\begin{cases} \frac{2}{5} \omega_2^5 + \frac{4}{5} \omega_4^5 = \frac{1}{2} - \frac{8}{30} \\ \frac{1}{10} \omega_2^5 + \frac{3}{5} \omega_4^5 = \frac{1}{4} - \frac{2}{15} \end{cases} \implies \begin{cases} \omega_2^5 = \frac{7}{24} \\ \omega_4^5 = \frac{7}{48} \end{cases}$$

Now we can obtain

$$\begin{aligned}\omega_3 &= \frac{\omega_3^5 \binom{3}{3} + \omega_4^5 \binom{4}{3} + \omega_5^5 \binom{5}{3}}{\binom{5}{3}} &= \frac{1}{3} \cdot \frac{1}{10} + \frac{7}{48} \cdot 4 \frac{1}{10} + \frac{1}{30} = \frac{1}{8} \\ \omega_4 &= \frac{\omega_4^5 \binom{4}{4} + \omega_5^5 \binom{5}{4}}{\binom{5}{4}} &= \frac{7}{48} \cdot \frac{1}{5} + \frac{1}{30} = \frac{1}{16}\end{aligned}$$

1.2 Exercise 2

1.2.1 Data

- Process of simple alternative $\{|E_n|\}$
- $P(E_1) = \omega_1 = \frac{1}{2}$
- $P(E_1 \wedge E_2) = \omega_2 = \frac{1}{4}$
- $P(E_1 \wedge E_2 \wedge E_3) = \omega_3 = \frac{1}{7}$
- $P(E_1 \wedge E_2 \wedge E_3 \wedge E_4) = \frac{3}{28}$

1.2.2 Questions

1. Could the 4 indicators $|E_1|$, $|E_2|$, $|E_3|$ and $|E_4|$ be the starting path of an exchangeable process?
2. Could it continue for at least one step?

1.2.3 Solutions

1. An exchangeable process must satisfy the condition

$$(-1)^{n-h} \Delta^{n-h} \omega_h \geq 0, \forall n, h \leq n$$

Thus we compute

- $(-1)^{4-1} \Delta^{4-1} \omega_1 = (-1) \cdot \Delta^3 \omega_1 = \frac{1}{14} \geq 0$
- $(-1)^{4-2} \Delta^{4-2} \omega_2 = (-1) \cdot \Delta^2 \omega_2 = \frac{1}{14} \geq 0$
- $(-1)^{4-3} \Delta^{4-3} \omega_3 = (-1) \cdot \Delta \omega_3 = \frac{1}{28} \geq 0$
- $(-1)^{4-4} \Delta^{4-4} \omega_4 = (-1) \cdot \omega_4 = \frac{3}{28} \geq 0$

Thus we can affirm that the process is exchangeable.

2. If we consider $n = 5$ we can rewrite

$$\begin{aligned} -\Delta \omega_4 &= \omega_4 - \omega_5 \geq 0 \implies \omega_5 \leq \omega_4 \\ \Delta^2 \omega_3 &= \omega_3 - 2\omega_4 + \omega_5 \geq 0 \implies \omega_5 \geq 2\omega_4 - \omega_3 \\ -\Delta^3 \omega_2 &= \omega_2 - 3\omega_3 + 3\omega_4 - \omega_5 \geq 0 \implies \omega_5 \leq 3\omega_4 - 3\omega_3 + \omega_2 \\ \Delta^4 \omega_1 &= \omega_1 - 4\omega_2 + 6\omega_3 - 4\omega_4 + \omega_5 \geq 0 \implies \omega_5 \geq 4\omega_4 - 6\omega_3 + 4\omega_2 - \omega_1 \end{aligned}$$

And substituting ω_k with their values we obtain a system:

$$\begin{cases} \omega_5 \leq \frac{3}{28} \\ \omega_5 \geq \frac{2}{28} \\ \omega_5 \leq \frac{3}{28} \\ \omega_5 \geq \frac{2}{28} \end{cases} \implies \frac{2}{28} \leq \omega_5 \leq \frac{3}{28}$$

Thus we can affirm that the process could continue.

Week 2

Conjugate priors and posterior distributions

2.1 Exercise 2.3

2.1.1 Data

- $p(x, y, z) \propto f(x, z) g(y, z) h(z)$

2.1.2 Questions

Prove that:

1. $p(x|y, z) \propto f(x, z)$
2. $p(y|x, z) \propto g(y, z)$
3. X and Y conditionally independent, given Z .

2.1.3 Solutions

We know by definition that

$$p(x|y, z) = \frac{p(x, y, z)}{p(y, z)}$$

and also that

$$p(y, z) = \int_{S_X} p(x, y, z) dx \propto \int_{S_X} f(x, z) g(y, z) h(z) dx = g(y, z) h(z) \int_{S_X} f(x, z) dx$$

Where S_X is the support of the r.v. X . Then we can write

$$\begin{aligned} p(x|y, z) &= \frac{f(x, z) g(y, z) h(z)}{g(y, z) h(z) \int_{S_X} f(x, z) dx} \\ &= \frac{f(x, z)}{\int_{S_X} f(x, z) dx} \end{aligned}$$

But $\int_{S_X} f(x, z) dx$ is constant given z , so we can say

$$p(x|y, z) \propto f(x, z)$$

as we wanted to show.

Similarly, we can write

$$\begin{aligned} p(y|x, z) &= \frac{p(x, y, z)}{p(x, z)} \\ &= \frac{f(x, z)g(y, z)h(z)}{\int_{S_Y} f(x, z)h(z)g(y, z)\partial y} \\ &= \frac{g(y, z)}{\int_{S_Y} g(y, z)\partial y} \\ &\propto g(y, z) \end{aligned}$$

To show that $X \perp Y$ given Z we have to prove that $p(y|z, x) = p(y|z)$, so:

$$\begin{aligned} p(y|z) &= \frac{p(y, z)}{p(z)} \\ &= \frac{\int_{S_X} f(x, z)g(y, z)h(z)\partial x}{\int_{S_X} \int_{S_Y} f(x, z)g(y, z)h(z)\partial y \partial x} \\ &= \frac{g(y, z)h(z) \int_{S_X} f(x, z)\partial x}{h(z) \int_{S_X} f(x, z)\partial x \int_{S_Y} g(y, z)\partial y} \\ &= \frac{g(y, z)}{\int_{S_Y} g(y, z)\partial y} \\ &= p(y|x, z) \end{aligned}$$

2.2 Exercise 3.5

2.2.1 Data

- $p(y|\phi) = c(\phi)h(y) \exp(\phi t(y))$
- $p_1(\theta) \dots p_k(\theta)$ conjugate priors
- $\tilde{p}(\theta) = \sum_{k=1}^K \omega_k p_k(\theta)$ where $\omega_k > 0$ and $\sum_k \omega_k = 1$

2.2.2 Questions

1. $p(\theta|y)$ as a function of $p(y|\theta)$ and \tilde{p}
2. Previous question but in the case that $\theta \sim \text{Pois}$ and $p_1 \dots p_k \sim \Gamma$

2.2.3 Solution

For the Bayes rule:

$$\begin{aligned}
p(\theta|y) &= \frac{p(y|\theta) \cdot p(\theta)}{p(y)} \\
&= \frac{p(y|\theta) \cdot \tilde{p}(\theta)}{p(y)} \\
&= \frac{\prod_i p(y_i|\theta) \tilde{p}(\theta)}{p(y)} \\
&= \frac{\prod_i c(\theta) h(y_i) \exp(\theta t(y_i)) \cdot \tilde{p}(\theta)}{p(y)} \\
&= \frac{\prod_i h(y_i) c(\phi)^n \exp(\phi \sum_i t(y_i)) \cdot \sum_k w_k p_k(\theta)}{\int_{S_\theta} \prod_i h(y_i) c(\phi)^n \exp(\phi \sum_i t(y_i)) \sum_k w_k p_k(\theta) d\theta} \\
&= \frac{c(\phi)^n \exp(\phi \sum_i t(y_i)) \cdot \sum_k w_k p_k(\theta)}{\int_{S_\theta} c(\phi)^n \exp(\phi \sum_i t(y_i)) \sum_k w_k p_k(\theta) d\theta}
\end{aligned}$$

In the particular case that $p(y|\theta)$ is a Poisson distribution and p_k are Gamma distribution, we have that

- $t(y) = y$
- $\phi = \log(\theta)$
- $c(\phi) = \exp(e^{-\phi}) = \exp(\theta^{-1})$
- $p_k(\theta) = \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} \theta^{\alpha_k-1} \exp(-\beta_k \theta) = c_k \theta^{\alpha_k-1} \exp(-\beta_k \theta)$

So we can rewrite the posterior of the first part as

$$\begin{aligned}
p(\theta|y) &= \frac{\exp(\theta^{-1})^n \exp(\log \theta \sum_i y_i) \sum_k w_k c_k \theta^{\alpha_k-1} \exp(-\beta_k \theta)}{\int_{S_\theta} \exp(\theta^{-1})^n \exp(\log \theta \sum_i y_i) \sum_k w_k c_k \theta^{\alpha_k-1} \exp(-\beta_k \theta) d\theta} \\
&= \frac{\exp(n\theta^{-1}) \theta^{\sum_i y_i} \sum_k w_k c_k \theta^{\alpha_k-1} \exp(-\beta_k \theta)}{\int_{S_\theta} \exp(n\theta^{-1}) \theta^{\sum_i y_i} \sum_k w_k c_k \theta^{\alpha_k-1} \exp(-\beta_k \theta) d\theta} \\
&= \frac{\exp(n\theta^{-1}) \sum_k w_k c_k \theta^{\alpha_k + \sum_i y_i} \exp(-\beta_k \theta)}{\int_{S_\theta} \exp(n\theta^{-1}) \sum_k w_k c_k \theta^{\alpha_k + \sum_i y_i} \exp(-\beta_k \theta) d\theta}
\end{aligned}$$

Week 3

Non informative priors

3.1 Exercise 3.10

3.1.1 Data

- $\psi = g(\theta)$ where g is a monotone function
- $h(\cdot) = g^{-1}(\cdot)$, that is $\theta = h(\psi)$
- $p_\theta(\theta) = \text{PDF of } \theta \implies p_\psi(\psi) = p_\theta(h(\psi)) \cdot \left| \frac{dh}{d\psi} \right|$

3.1.2 Questions

1. Let $\theta \sim \text{Beta}(a, b)$ and $\psi = \text{logit}(\theta)$. Obtain p_ψ and plot it for the case $a = b = 1$.
2. Let $\theta \sim \text{Gamma}(a, b)$ and $\psi = \log(\theta)$. Obtain p_ψ and plot it for the case $a = b = 1$.

3.1.3 Solutions

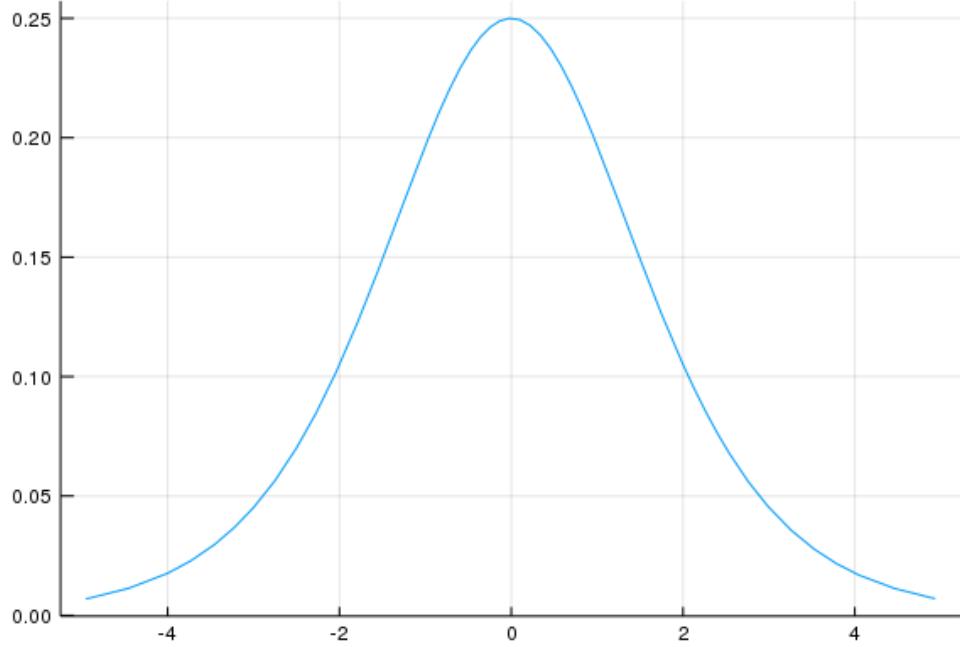
1. The inverse function of $\text{logit}(\cdot)$ is known to be $h(\psi) = \frac{\exp(\psi)}{1 + \exp(\psi)}$, and the derivative w.r.t. ψ of $h(\psi)$ is

$$\begin{aligned}\frac{\partial h(\psi)}{\psi} &= \frac{\exp(\psi)(1 + \exp(\psi)) - \exp(2\psi)}{(1 + \exp(\psi))^2} \\ &= \frac{\exp(\psi)}{(1 + \exp(\psi))^2}\end{aligned}$$

So we can write the PDF of ψ as

$$\begin{aligned}p_\psi(\psi) &= \frac{1}{B(a, b)} \left(\frac{\exp(\psi)}{1 + \exp(\psi)} \right)^{a-1} \left(1 - \frac{\exp(\psi)}{1 + \exp(\psi)} \right)^{b-1} \frac{\exp(\psi)}{(1 + \exp(\psi))^2} \\ &= \frac{1}{B(a, b)} \frac{\exp(\psi)^{a-1}}{(1 - \exp(\psi))^{a-1}} \frac{1}{(1 + \exp(\psi))^{b-1}} \frac{\exp(\psi)}{(1 + \exp(\psi))^2} \\ &= \frac{1}{B(a, b)} \frac{\exp(\psi)^a}{(1 + \exp(\psi))^{a+b}}\end{aligned}$$

And in the case that $a = b = 1$ the plot is:



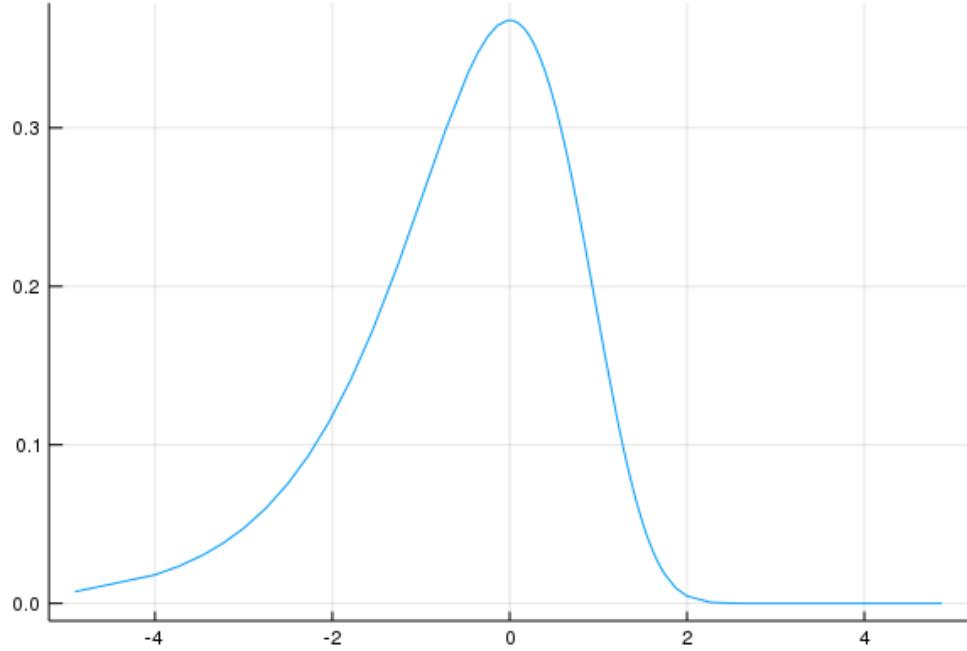
2. The inverse function of $\log(\theta)$ is known to be $h(\psi) = \exp(\psi)$, and the derivative w.r.t. ψ of $h(\psi)$ is

$$\frac{\partial h(\psi)}{\partial \psi} = \exp(\psi)$$

So we can write the PDF of ψ as

$$\begin{aligned} p_\psi(\psi) &= \frac{b^a}{\Gamma(a)} \exp(\psi)^{a-1} \exp(-b \exp(\psi)) \exp(\psi) \\ &= \frac{b^a}{\Gamma(a)} \exp(a\psi - b \exp(\psi)) \end{aligned}$$

And in the case that $a = b = 1$ the plot is:



3.2 Exercise 3.14

3.2.1 Data

3.2.2 Questions

1. Given $Y_1 \dots Y_n \sim \text{Bernoulli}(\theta)$ find the MLE of θ and $\frac{J(\theta)}{n}$.
2. Find a PDF $p_U(\theta)$ such that $\log p_U(\theta) = l(\theta|\mathbf{y})/n + c$ where c is a constant that does not depend on θ . Compute $-\partial^2 \log p_U(\theta)/\partial^2 \theta$.
3. Obtain a PDF for θ that is proportional to $p_U(\theta)p(\mathbf{y}|\theta)$. Can this be considered a posterior for θ ?
4. Repeat previous points with $Y_1 \dots Y_n \sim \text{Poisson}(\theta)$.

3.2.3 Solutions

1.

$$\begin{aligned}
 \hat{\theta}_{MLE} &= \arg \min_{\theta} l(\theta|\mathbf{y}) \\
 &= \arg \min_{\theta} \log \left(\prod_i \theta^{y_i} (1-\theta)^{1-y_i} \right) \\
 &= \arg \min_{\theta} \sum_i \log (\theta^{y_i} (1-\theta)^{1-y_i}) \\
 &= \arg \min_{\theta} \log(\theta) \sum_i y_i + \log(1-\theta) \sum_i (1-y_i) \\
 &= \arg \min_{\theta} \log(\theta) \sum_i y_i + n \log(1-\theta) - \log(1-\theta) \sum_i (y_i)
 \end{aligned}$$

To find the minimum we compute the zeros of $\partial l(\theta|\mathbf{y})/\partial\theta$

$$\begin{aligned} 0 &= \frac{\partial l(\theta|\mathbf{y})}{\partial\theta} \\ &= \frac{\sum_i y_i}{\theta} - \frac{n}{1-\theta} + \frac{\sum_i y_i}{1-\theta} \\ &= \frac{\sum_i y_i - \theta \sum_i y_i - \theta n + \theta \sum_i y_i}{\theta(1-\theta)} \\ &= \frac{\sum_i y_i - \theta n}{\theta(1-\theta)} \end{aligned}$$

Thus, if $\theta \notin \{0, 1\}$, $\hat{\theta}_{MLE} = \frac{\sum_i y_i}{n}$.

$$\begin{aligned} -\frac{\partial^2 l(\theta|\mathbf{y})}{n\partial^2\theta} &= -\frac{-n\theta(1-\theta) - (\sum_i y_i - n\theta)(1-2\theta)}{n\theta^2(1-\theta)^2} \\ &= -\frac{n\theta^2 - n\theta - \sum_i y_i + 2\theta \sum_i y_i + n\theta - 2n\theta^2}{n\theta^2(1-\theta)^2} \\ &= \frac{n\theta^2 - 2\theta \sum_i y_i + \sum_i y_i}{n\theta^2(1-\theta)^2} \\ &= \frac{\theta^2 - 2\theta \hat{\theta}_{MLE} + \hat{\theta}_{MLE}}{\theta^2(1-\theta)^2} \end{aligned}$$

2. The constraints on $p_U(\theta)$ imply that

$$\begin{aligned} p_U(\theta) &= c \sqrt[n]{\prod_i \theta^{y_i} (1-\theta)^{1-y_i}} \\ &= c \prod_i \theta^{y_i/n} (1-\theta)^{(1-y_i)/n} \\ &= c \theta^{\sum_i y_i/n} (1-\theta)^{\sum_i (1-y_i)/n} \\ &= c \theta^{\sum_i y_i/n} (1-\theta)^{1-\sum_i y_i/n} \end{aligned}$$

where c is the normalization constant.

$$\begin{aligned} -\frac{\partial^2 \log p_U(\theta)}{\partial^2\theta} &= -\frac{\partial^2 l(\theta|\mathbf{y})/n + c}{\partial^2\theta} \\ &= -\frac{\partial^2 l(\theta|\mathbf{y})}{\partial^2\theta}/n \\ &= \frac{\theta^2 - 2\theta \hat{\theta}_{MLE} + \hat{\theta}_{MLE}}{\theta^2(1-\theta)^2} \end{aligned}$$

3. Such a PDF would have the form

$$\begin{aligned} p(\theta|\mathbf{y}) &= c \cdot p_U(\theta) p(\mathbf{y}|\theta) \\ &= c \theta^{\sum_i y_i/n} (1-\theta)^{1-\sum_i y_i/n} \cdot \theta^{\sum_i y_i} (1-\theta)^{\sum_i (1-y_i)} \\ &= c \theta^{\sum_i y_i(1+\frac{1}{n})} (1-\theta)^{\sum_i (1-y_i)(1+\frac{1}{n})} \\ &= c \theta^{\sum_i y_i(1+\frac{1}{n})} (1-\theta)^{(n-\sum_i y_i)(1+\frac{1}{n})} \end{aligned}$$

Where $c = \int_{S_\theta} p(\theta, \mathbf{y}) d\theta$ to guarantee that the PDF is proper. We can observe that the obtained PDF is a Beta distribution with parameters $\sum_i y_i(1 + \frac{1}{n}) + 1$ and $(n - \sum_i y_i)(1 + \frac{1}{n}) + 1$. It is a posterior because it is the product of a prior and a conditioned probability. Moreover, it is a case of conjugate prior.

4. Same steps: simplify the log-likelihood, find the first derivative and constrain to zero.

$$\begin{aligned}
 \hat{\theta}_{MLE} &= \arg \min_{\theta} l_{\text{Poisson}}(\theta | \mathbf{y}) \\
 &= \arg \min_{\theta} \log \left(\prod_i \frac{\exp(-\theta) \theta^{y_i}}{y_i!} \right) \\
 &= \arg \min_{\theta} \sum_i \log \frac{\exp(-\theta) \theta^{y_i}}{y_i!} \\
 &= \arg \min_{\theta} -n\theta + \log \theta \sum y_i - \sum \log y_i!
 \end{aligned}$$

$$\begin{aligned}
 0 &= \frac{\partial l(\theta | \mathbf{y})}{\partial \theta} \\
 &= -n + \frac{\sum_i y_i}{\theta}
 \end{aligned}$$

Thus $\hat{\theta}_{MLE} = \sum_i y_i / n$.

$$\begin{aligned}
 -\frac{\partial^2 l(\theta | \mathbf{y})}{n \partial^2 \theta} &= \frac{\sum_i y_i}{n \theta^2} \\
 &= \frac{\hat{\theta}_{MLE}}{\theta^2}
 \end{aligned}$$

In this case the PDF p_U would be

$$\begin{aligned}
 p_U(\theta) &\propto \sqrt[n]{\prod_i \exp(-\theta) \theta^{y_i}} \\
 &= \exp(-\theta) \theta^{\sum_i y_i / n}
 \end{aligned}$$

Thus

$$-\frac{\partial^2 \log p_U(\theta)}{\partial^2 \theta} = \frac{\sum_i y_i}{n \theta^2}$$

While the posterior is

$$\begin{aligned}
 p(\theta | \mathbf{y}) &\propto \exp(-\theta) \theta^{\sum_i y_i / n} \prod_i \exp(-\theta) \theta^{y_i} \\
 &= \exp(-(n+1)\theta) \theta^{\sum_i y_i (1 + \frac{1}{n})}
 \end{aligned}$$

This time the posterior is a Gamma distribution with parameters $\sum_i y_i (1 + \frac{1}{n})$ and $n+1$.