

Homework 2: Conjugate priors and Jeffreys priors

Lecturer: Georgios Karagiannis

georgios.karagiannis@durham.ac.uk

For Formative assessment, submit the solutions of the parts 1, 2, and 3 from the Exercise 1, and the solution of the Exercise 2.

Exercise 1. (★★) Let $x = (x_1, \dots, x_n)$ be observables. Consider a Bayesian model such as

$$\begin{cases} x_i | \lambda & \stackrel{\text{iid}}{\sim} \text{Pn}(\lambda), \forall i = 1, \dots, n \\ \lambda & \sim \Pi(\lambda) \end{cases}$$

Hint-1 Poisson distribution $x \sim \text{Pn}(\lambda)$ has PMF: $\text{Pn}(x|\lambda) = \frac{1}{x!} \lambda^x \exp(-\lambda) 1_{\mathbb{N}}(x)$, where $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\lambda > 0$.

Hint-2 Gamma distribution $x \sim \text{Ga}(a, b)$ has PDF: $\text{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) 1_{(0, \infty)}(x)$, with $a > 0$ and $b > 0$.

Hint-2 Negative Binomial distribution $x \sim \text{Nb}(r, \theta)$ has PMF: $\text{Nb}(x|r, \theta) = \binom{r+x-1}{r-1} \theta^r (1-\theta)^x 1_{\mathbb{N}}(x)$ with $\theta \in (0, 1)$, $r \in \mathbb{N} - \{0\}$, and $\mathbb{N} = \{0, 1, 2, \dots\}$.

1. Compute the likelihood in the aforesaid Bayesian model.
2. Show that the sampling distribution is a member of the exponential family.
3. Specify the PDF of the conjugate prior distribution $\Pi(\lambda)$ of λ , and identify the parametric family of distributions as $\lambda \sim \text{Ga}(a, b)$, with $a > 0$, and $b > 0$. While you are deriving the conjugate prior distribution of λ , discuss which of the prior hyper-parameters can be considered as the ‘strength of the prior information and which can be considered as summarizing the prior information.
4. Compute the PDF of the posterior distribution of λ , identify the posterior distribution as a Gamma distribution $\text{Ga}(\tilde{a}, \tilde{b})$, and compute the posterior hyper-parameters \tilde{a} , and \tilde{b} .
5. Compute the PMF of the predictive distribution of a future outcome $y = x_{n+1}$, identify the name of the resulting predictive distribution, and compute its parameters.

Solution.

1. The likelihood is

$$f(x|\lambda) = \prod_{i=1}^n \text{Pn}(x_i|\lambda) = \left(\prod_{i=1}^n \frac{1}{x_i!} \right) \lambda^{\sum_{i=1}^n x_i} \exp(-n\lambda) \quad (1)$$

2. The k parameter exponential family of distributions has the form

$$\text{Ef}_k(x|u, g, h, \phi, \theta, c) = u(x)g(\theta) \exp\left(\sum_{j=1}^k c_j \phi_j(\theta) h_j(x)\right); \quad x \in \mathcal{X}$$

and if sampling space \mathcal{X} does not depend on θ it is also called regular. So I just need to bring the sampling density distribution in this form. It is

$$\text{Pn}(x|\lambda) = \frac{1}{x!} \lambda^x \exp(-\lambda) 1_{\mathbb{N}}(x) = \frac{1}{x!} \exp(-\lambda) \exp(x \log(\lambda)) 1_{\mathbb{N}}(x)$$

So $\text{Pn}(\lambda)$ is member of the regular 1-parameter exponential family with

$$u(x) = \frac{1}{x!} 1_{\mathbb{N}}(x), \quad g(\lambda) = \exp(-\lambda), \quad h_1(x) = x, \quad \phi_1(\lambda) = \log(\lambda), \quad c_1 = 1.$$

The sampling space \mathcal{X} does not depend on the uncertain parameter λ and hence it is a regular exponential family of distributions.

3. There are two ways to derive the conjugate prior. I will present both.

Way-1 (Theorem in the Handout)

The sampling distribution is member of the 1- regular exponential distribution family, as the density of the sampling density distribution $\text{Pn}(x|\lambda)$ can be written in the form

$$\text{Pn}(x|\lambda) = u(x)g(\lambda) \exp\left(\sum_{j=1}^k c_j \phi_j(\lambda) h_j(x)\right); \quad x \in \mathcal{X}$$

with

$$u(x) = \frac{1}{x!} 1_{\mathbb{N}-\{0\}}(x), \quad g(\lambda) = \exp(-\lambda), \quad h_1(x) = x, \quad \phi_1(\lambda) = \log(\lambda), \quad c_1 = 1.$$

Since the sampling space \mathcal{X} of the sampling distribution does not depend on the unknown parameter λ , (Theorem 20 from the Handout) the conjugate prior is

$$\begin{aligned} \pi(\lambda) &\propto g(\lambda)^{\tau_0} \exp(c_1 \tau_1 \phi_1(\lambda)) \\ &= \exp(-\lambda \tau_0) \exp(\tau_1 \log(\lambda)) \\ &= \lambda^{\tau_1} \exp(-\lambda \tau_0) \\ &\propto \text{Ga}(\lambda|a, b), \text{ for } a = \tau_1 + 1, \ b = \tau_0 \end{aligned} \tag{2}$$

So the conjugate prior is $\lambda \sim \text{Ga}(\lambda|a, b)$ with $a > 0$ and $b > 0$.

Way-2 (Theorem in the Handout)

The likelihood can be written as

$$f(x|\lambda) = \prod_{i=1}^n \text{Pn}(x_i|\lambda) = \underbrace{\lambda^{\sum_{i=1}^n x_i} \exp(-n\lambda)}_{=k(t(x)|\lambda)} \underbrace{\left(\prod_{i=1}^n \frac{1}{x_i!}\right)}_{=\rho(x)} \tag{3}$$

where a kernel of the likelihood is $k(t(x)|\lambda) = \lambda^{\sum_{i=1}^n x_i} \exp(-n\lambda)$, with sufficient statistics $t(x) = (n, \sum_{i=1}^n x_i)$, and $\rho(x) = \left(\prod_{i=1}^n \frac{1}{x_i!}\right)$ is the residual term of it. The dimensionality of the sufficient statistic $t(x)$ does not depend on the sample size n , and the observables are iid. Hence, (Theorem 12 in the Handout) the conjugate prior results as the aforesaid likelihood kernel from (3) where the sufficient statistics are replaced by a priori hyper-parameters $\tau = (\tau_0, \tau_1)$, such as

$$\pi(\lambda) \propto k(\tau|\lambda) = \lambda^{\tau_1} \exp(-\tau_0 \lambda) \propto \text{Ga}(\lambda|a, b), \text{ for } a = \tau_1 + 1, \ b = \tau_0 \tag{4}$$

where I recognize the kernel of the Gamma distribution. So the conjugate prior is $\lambda \sim \text{Ga}(a, b)$ with $a > 0$ and $b > 0$.

In (2) and (4), as strength of the prior information can be considered the parameter τ_0 (and hence b) because it substitutes the sample size n in the likelihood (1). In (2) and (4), as prior information summary can be considered the parameter τ_1 (and hence a) because it substitutes the summary $\sum_{i=1}^n x_i$ in the likelihood (1).

4. According to the definition, the posterior PDF can be computed via the Bayes theorem

$$\begin{aligned}\pi(\lambda|x) &\propto f(x|\lambda)\pi(\lambda) \propto \prod_{i=1}^n \text{Pn}(x_i|\lambda)\text{Ga}(\lambda|a, b) \\ &\propto \left(\prod_{i=1}^n \frac{1}{x_i!}\right) \lambda^{\sum_{i=1}^n x_i} \exp(-n\lambda) \times \frac{b^a}{\Gamma(a)} \lambda^{a-1} \exp(-\lambda b) \\ &\propto \lambda^{\sum_{i=1}^n x_i + a - 1} \exp(-\lambda(n+b)) \\ &\propto \text{Ga}(\lambda | \sum_{i=1}^n x_i + a, n+b)\end{aligned}$$

So the posterior distribution is $\lambda|x \sim \text{Ga}(\tilde{a}, \tilde{b})$, $\tilde{a} = \sum_{i=1}^n x_i + a$, $\tilde{b} = n + b$.

- Alternatively, we could use the Theorem in the Lecture notes stating the properties of the Conjugate priors... I.e. $\lambda|x \sim \text{Ga}(\sum_{i=1}^n x_i + (\tau_1 + 1), n + (\tau_0))$ –It is up to you...

5. According to the definition, the predictive PMF is

$$\begin{aligned}g(y|x) &= \int_{(0,\infty)} f(y|\lambda)\pi(\lambda|x)d\lambda = \int_{(0,\infty)} \text{Pn}(y|\lambda)\text{Ga}(\lambda|\tilde{a}, \tilde{b})d\lambda \\ &= \int_{(0,\infty)} \frac{1}{y!} \lambda^y \exp(-\lambda) \mathbf{1}_{\mathbb{N}-\{0\}}(y) \frac{\tilde{b}^{\tilde{a}}}{\Gamma(\tilde{a})} \lambda^{\tilde{a}-1} \exp(-\lambda\tilde{b})d\lambda \\ &= \frac{1}{y!} \frac{\tilde{b}^{\tilde{a}}}{\Gamma(\tilde{a})} \mathbf{1}_{\mathbb{N}-\{0\}}(y) \int_{(0,\infty)} \lambda^{y+\tilde{a}-1} \exp(-\lambda(\tilde{b}+1))d\lambda \\ &= \frac{1}{y!} \frac{\tilde{b}^{\tilde{a}}}{\Gamma(\tilde{a})} \frac{\Gamma(y+\tilde{a})}{(\tilde{b}+1)^{y+\tilde{a}}} \mathbf{1}_{\mathbb{N}-\{0\}}(y) = \frac{1}{y!} \left(\frac{\tilde{b}}{\tilde{b}+1}\right)^{\tilde{a}} \left(\frac{1}{\tilde{b}+1}\right)^y \frac{\Gamma(y+\tilde{a})}{\Gamma(\tilde{a})} \mathbf{1}_{\mathbb{N}-\{0\}}(y) \\ &= \frac{1}{y!} \left(\frac{\tilde{b}}{\tilde{b}+1}\right)^{\tilde{a}} \left(\frac{1}{\tilde{b}+1}\right)^y \frac{(y+\tilde{a}-1)(y+\tilde{a}-2)\cdots(\tilde{a})\Gamma(\tilde{a})}{\Gamma(\tilde{a})} \mathbf{1}_{\mathbb{N}-\{0\}}(y) \\ &= \frac{1}{y!} \left(\frac{\tilde{b}}{\tilde{b}+1}\right)^{\tilde{a}} \left(\frac{1}{\tilde{b}+1}\right)^y (y+\tilde{a}-1)(y+\tilde{a}-2)\cdots(\tilde{a}) \mathbf{1}_{\mathbb{N}-\{0\}}(y) \\ &= \frac{1}{y!} \left(\frac{\tilde{b}}{\tilde{b}+1}\right)^{\tilde{a}} \left(\frac{1}{\tilde{b}+1}\right)^y \frac{(y+\tilde{a}-1)!}{(\tilde{a}-1)!} \mathbf{1}_{\mathbb{N}-\{0\}}(y) = \frac{(y+\tilde{a}-1)!}{(\tilde{a}-1)!y!} \left(\frac{\tilde{b}}{\tilde{b}+1}\right)^{\tilde{a}} \left(\frac{1}{\tilde{b}+1}\right)^y \mathbf{1}_{\mathbb{N}-\{0\}}(y) \\ &= \binom{y+\tilde{a}-1}{\tilde{a}-1} \left(\frac{\tilde{b}}{\tilde{b}+1}\right)^{\tilde{a}} \left(1 - \frac{\tilde{b}}{\tilde{b}+1}\right)^y \mathbf{1}_{\mathbb{N}-\{0\}}(y) = \text{Nb}(y|\tilde{a}, \frac{\tilde{b}}{\tilde{b}+1})\end{aligned}$$

where $\tilde{a} = \sum_{i=1}^n x_i + a$, $\tilde{b} = n + b$.

Exercise 2. (**) Assume observation x sampled from a Maxwell distribution with density

$$f(x|\theta) = \sqrt{\frac{2}{\pi}} \theta^{3/2} x^2 \exp(-\frac{1}{2}\theta x^2).$$

Find the Jeffreys prior density for the parameter θ .

Solution. It is

$$\begin{aligned}
 f(x|\theta) &= \sqrt{\frac{2}{\pi}} \theta^{3/2} x^2 \exp(-\frac{1}{2} \theta x^2) \implies \\
 \log(f(x|\theta)) &= \log(\sqrt{\frac{2}{\pi}} x^2) + \frac{3}{2} \log(\theta) - \frac{1}{2} \theta x^2 \implies \\
 \frac{d}{d\theta} \log(f(x|\theta)) &= \frac{3}{2} \frac{1}{\theta} - \frac{1}{2} x^2 \implies \\
 \frac{d^2}{d\theta^2} \log(f(x|\theta)) &= -\frac{3}{2} \frac{1}{\theta^2} \implies \\
 \underbrace{-E(\frac{d^2}{d\theta^2} \log(f(x|\theta)))}_{=I(\theta)} &= \frac{3}{2} \frac{1}{\theta^2} \implies \\
 \pi(\theta) &\propto \sqrt{I(\theta)} \propto \frac{1}{\theta}
 \end{aligned}$$

Hence, we take $\pi(\theta) \propto \frac{1}{\theta}$.