Bayesian Statistics III/IV (MATH3361/4071) Michaelmas term 2021

Problem class 1 [a]

Nuisance parameters, the Normal model, and the Normal linear regression with unknown variance

Lecturer: Georgios Karagiannis georgios.karagiannis@durham.ac.uk

[a] Author: Georgios P. Karagiannis.

Nuisance parameters

Exercise 1. ($\star\star$)Assume observable quantities $y = (y_1, ..., y_n)$ forming the available data set of size n. Assume that the observations are drawn i.i.d. from a sampling distribution which is judged to be in the Normal parametric family of distributions $\text{N}(\mu, \sigma^2)$ with unknown mean μ and variance σ^2. We are interested in learning μ and the next outcome $z = y_{n+1}$. We do not care about σ^2. <-story

Assume You specify a Bayesian model <-set-up

$$\begin{cases} y_i|\mu, \sigma^2 \sim \text{N}(\mu, \sigma^2), \text{ for all } i = 1, ..., n & \text{, Statistical model} \\ \mu|\sigma^2 \sim \text{N}(\mu_0, \sigma^2 \frac{1}{\tau_0}) & \text{, prior} \\ \sigma^2 \sim \text{IG}(a_0, k_0) & \text{, prior} \end{cases}$$

1. Show that
$$\sum_{i=1}^{n}(y_i - \theta)^2 = n(\bar{y} - \theta)^2 + ns^2,$$
where $s^2 = \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y})^2$.

2. Show that the joint posterior distribution $\Pi(\mu, \sigma^2|y)$ is such as
$$\mu|y, \sigma^2 \sim \text{N}(\mu_n, \sigma^2 \frac{1}{\tau_n})$$
$$\sigma^2|y \sim \text{IG}(a_n, k_n)$$
with
$$\mu_n = \frac{n\bar{y} + \tau_0\mu_0}{n + \tau_0}; \qquad \tau_n = n + \tau_0; \qquad a_n = a_0 + n$$
$$k_n = k_0 + \frac{1}{2}ns_n^2 + \frac{1}{2}\frac{\tau_0 n(\mu_0 - \bar{y})^2}{n + \tau_0}$$
Hint: It is
$$-\frac{1}{2}\frac{(\mu - \mu_1)^2}{v_1} - \frac{1}{2}\frac{(\mu - \mu_2)^2}{v_2} ... - \frac{1}{2}\frac{(\mu - \mu_n)^2}{v_n} = -\frac{1}{2}\frac{(\mu - \hat{\mu})^2}{\hat{v}} + C$$
where
$$\hat{v} = \left(\sum_{i=1}^{n}\frac{1}{v_i}\right)^{-1}; \quad \hat{\mu} = \hat{v}\left(\sum_{i=1}^{n}\frac{\mu_i}{v_i}\right); \quad C = \frac{1}{2}\frac{\hat{\mu}^2}{\hat{v}} - \frac{1}{2}\sum_{i=1}^{n}\frac{\mu_i^2}{v_i}$$

3. Show that the marginal posterior distribution $\Pi(\mu|y)$ is such as
$$\mu|y \sim \text{T}_1\left(\mu_n, \frac{k_n}{a_n}\frac{1}{\tau_n}, 2a_n\right)$$

Hint-1: If $x \sim \text{IG}(a, b)$, $y = cx$, then $y \sim \text{IG}(a, cb)$.

Hint-2: The definition of Student T is considered as known

4. Show that the predictive distribution $\Pi(z|y)$ is Student T such as

$$z|y \sim \text{T}_1\left(\mu_n, \frac{k_n}{a_n}(\frac{1}{\tau_n}+1), 2a_n\right)$$

Hint-1: Consider that

$$\text{N}(x|\mu_1, \sigma_1^2)\,\text{N}(x|\mu_2, \sigma_2^2) \;=\; \text{N}(x|m, v^2)\,\text{N}(\mu_1|\mu_2, \sigma_1^2+\sigma_2^2)$$

where

$$v^2 = \left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)^{-1}; \quad m = v^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)$$

Hint-2: The definition of Student T is considered as known

Solution.

1. It is

$$\begin{aligned}\sum_{i=1}^{n}(y_i-\theta)^2 &= \sum_{i=1}^{n}\left[(y_i-\bar{y})-(\theta-\bar{y})\right]^2 \\ &= \sum_{i=1}^{n}\left[(y_i-\bar{y})^2+(\theta-\bar{y})^2-2\,(y_i-\bar{y})\,(\theta-\bar{y})\right] \\ &= ns^2+n(\bar{y}-\theta)^2, \text{ where } s^2=\frac{1}{2}\sum_{i=1}^{n}(y_i-\bar{y})^2\end{aligned}$$

2. I use the Bayes theorem

$$\begin{aligned}\pi(\mu,\sigma^2|y) \propto & f(y|\mu,\sigma^2)\pi(\mu,\sigma^2) = \prod_{i=1}^{n}\text{N}(y_i|\mu,\sigma^2)\text{N}(\mu|\mu_0,\sigma^2\frac{1}{\tau_0})\text{IG}(\sigma^2|a_0,k_0) \\ & \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{\sigma^2}\right)\times\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}}\exp\left(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma^2/\tau_0}\right)\times\left(\frac{1}{\sigma^2}\right)^{a_0+1}\exp\left(-\frac{1}{\sigma^2}k_0\right) \\ & \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\frac{1}{2}+a_0+1}\exp\left(\frac{1}{\sigma^2}\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{1}-\frac{1}{2}\frac{(\mu-\mu_0)^2}{1/\tau_0}\right]-\frac{1}{\sigma^2}k_0\right)\end{aligned}$$

It is

$$-\frac{1}{2}\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{1}-\frac{1}{2}\frac{(\mu-\mu_0)^2}{1/\tau_0} = -\frac{1}{2}\frac{(\mu-\mu_n)^2}{\underbrace{v_n^2}_{=1/\tau_n}}+C_n$$

where

$$
\begin{aligned}
v_n &= \left(\sum_{i=1}^{n}\frac{1}{1}+\frac{1}{1/\tau_0}\right)^{-1} = \frac{1}{n+\tau_0} \implies \tau_n = n+\tau_0\\
\mu_n =& v_n\left(\sum_{i=1}^{n}\frac{y_i}{1}+\frac{\mu_0}{1/\tau_0}\right) \implies \mu_n = \frac{n\bar{y}+\tau_0\mu_0}{n+\tau_0}\\
C_n =& \frac{1}{2}\frac{\mu_n^2}{v_n}-\frac{1}{2}\left(n\sum_{i=1}^{n}y_i^2+\tau_0\mu_0^2\right) = \frac{1}{2}\frac{(n\bar{y}+\tau_0\mu_0)^2}{n+\tau_0}-\frac{1}{2}\left(n\sum_{i=1}^{n}y_i^2+\tau_0\mu_0^2\right)\\
&= \text{...Quest. 1...} = -\frac{1}{2}ns_n^2-\frac{1}{2}\frac{\tau_0 n(\mu_0-\bar{y})^2}{n+\tau_0}
\end{aligned}
$$

So

$$
\begin{aligned}
\pi(\mu,\sigma^2|y) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}+\frac{n}{2}+a_0+1}\exp\left(\frac{1}{\sigma^2}\left[-\frac{1}{2}\frac{(\mu-\mu_n)^2}{1/\tau_n}+C_n\right]-\frac{1}{\sigma^2}k_0\right)\\
&\propto \underbrace{\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}}\exp\left(-\frac{1}{2}\frac{(\mu-\mu_n)^2}{\sigma^2/\tau_n}\right)}_{\propto\mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n)}\times\underbrace{\left(\frac{1}{\sigma^2}\right)^{\overbrace{\frac{n}{2}+a_0}^{=a_n}+1}\exp\left(-\frac{1}{\sigma^2}\overbrace{(k_0-C_n)}^{=k_n}\right)}_{\propto\mathrm{IG}(\sigma^2|a_n,k_n)}\\
&\propto \mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n)\mathrm{IG}(\sigma^2|a_n,k_n)
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_n &= \frac{n\bar{y}+\tau_0\mu_0}{n+\tau_0}\,; & a_n &= \frac{n}{2}+a_0\,;\\
\tau_n &= n+\tau_0\,; & k_n &= k_0+\frac{1}{2}ns_n^2+\frac{1}{2}\frac{\tau_0 n(\mu_0-\bar{y})^2}{n+\tau_0}\,.
\end{aligned}
$$

3. It is

$$
\pi(\mu|y) = \int \pi(\mu,\sigma^2|y)\mathrm{d}\sigma^2 = \int \mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n)\mathrm{IG}(\sigma^2|a_n,k_n)\mathrm{d}\sigma^2
$$

by change of variable $\xi=\sigma^2\frac{1}{2k_n}$, it is

$$
\begin{aligned}
\pi(\mu|y) &= \int \mathrm{N}(\mu|\mu_n,\xi 2k_n\frac{1}{\tau_n}\frac{2a_n}{2a_n})\mathrm{IG}(\xi|\frac{2a_n}{2},\frac{1}{2})\mathrm{d}\xi = \int \mathrm{N}(\mu|\mu_n,\xi\frac{1}{\tau_n}\frac{k_n}{a_n}2a_n)\mathrm{IG}(\xi|\frac{2a_n}{2},\frac{1}{2})\mathrm{d}\xi\\
&= \mathrm{T}_1(\mu|\mu_n,\frac{k_n}{a_n}\frac{1}{\tau_n},2a_n)
\end{aligned}
$$

4. It is

$$
\begin{aligned}
g(z|y) &= \int f(z|\mu,\sigma^2)\pi(\mu,\sigma^2|y)\mathrm{d}\mu\mathrm{d}\sigma^2 = \int \mathrm{N}(z|\mu,\sigma^2)\mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n)\mathrm{IG}(\sigma^2|a_n,k_n)\mathrm{d}\mu\mathrm{d}\sigma^2\\
&= \int\left[\int \mathrm{N}(z|\mu,\sigma^2)\mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n)\mathrm{d}\mu\right]\mathrm{IG}(\sigma^2|a_n,k_n)\mathrm{d}\sigma^2
\end{aligned}
$$

Normal density is symmetric $\mathrm{N}(z|\mu,\sigma^2)\mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n) = \mathrm{N}(\mu|z,\sigma^2)\mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n)$, and by using the Hint

$$
\int \mathrm{N}(\mu|z,\sigma^2)\mathrm{N}(\mu|\mu_n,\sigma^2/\tau_n)\mathrm{d}\mu = \int \mathrm{N}(\mu|\text{const.},\text{const.})\mathrm{N}\left(z|\mu_n,\sigma^2\left[\frac{1}{\tau_n}+1\right]\right)\mathrm{d}\mu = \mathrm{N}\left(z|\mu_n,\sigma^2\left[\frac{1}{\tau_n}+1\right]\right)
$$

So

$$g(z|y) = \int \mathrm{N}\left(z|\mu_n, \sigma^2\left[\frac{1}{\tau_n}+1\right]\right)\mathrm{IG}(\sigma^2|a_n,k_n)\mathrm{d}\sigma^2$$

by change the variable $\xi = \sigma^2 \frac{1}{2k_n}$, it is

$$g(z|y) = \int \mathrm{N}\left(z|\mu_n, \xi\left[\frac{1}{\tau_n}+1\right]\frac{k_n}{a_n}2a_n\right)\mathrm{IG}(\xi|\frac{2a_n}{2},\frac{1}{2})\mathrm{d}\xi = \mathrm{T}_1\left(z|\mu_n,\left[\frac{1}{\tau_n}+1\right]\frac{k_n}{a_n},2a_n\right)$$

Proper/improper priors

Exercise 2. ($\star\star$)Consider the Bayesian model

$$\begin{cases} x|\sigma & \sim \mathrm{N}(0,\sigma^2) \\ \sigma & \sim \mathrm{Ex}(\lambda) \end{cases}$$

where $\mathrm{Ex}(\lambda)$ is the exponential distribution with mean $1/\lambda$. Show that the posterior distribution is not defined always.

- HINT: Precisely, show that the posterior is not defined in the case that you collect only one observation $x = 0$.

Solution.

It is

$$f(x) \propto \int_{\mathbb{R}_+} \mathrm{N}(x|0,\sigma^2)\mathrm{Ex}(\sigma|\lambda)\mathrm{d}\sigma = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}(x-0)^2)\lambda\exp(-\sigma\lambda)\mathrm{d}\sigma$$
$$f(x=0) \propto \int_0^\infty \frac{1}{\sigma}\exp(-\sigma\lambda)\mathrm{d}\sigma$$

We will use a convergence criteria in order to check if $\int_0^\infty \frac{1}{\sigma}\exp(-\sigma\lambda)\mathrm{d}\sigma = \infty$.

I will use the Limit Comparison Test to check if $\int_0^\infty \frac{1}{\sigma}\exp(-\sigma\lambda)\mathrm{d}\sigma = \infty$. Consider $h(\sigma) = \frac{1}{\sigma}\exp(-\sigma\lambda)$. The function $h(\sigma)$ has an improper behavior at 0, as it is not bounded there. Let $g(\sigma) = \frac{1}{\sigma}$. According to the Limit Comparison Test, it is

$$\lim_{\sigma\to 0^+}\frac{h(\sigma)}{g(\sigma)} = \lim_{\sigma\to 0^+}\frac{\frac{1}{\sigma}\exp(-\sigma\lambda)}{\frac{1}{\sigma}} = 1 \neq 0$$

and

$$\int_0^\infty g(\sigma)\mathrm{d}\sigma = \int_0^\infty \frac{1}{\sigma}\mathrm{d}\sigma = \infty.$$

Therefore, it will be

$$\underbrace{\int_0^\infty h(\sigma)\mathrm{d}\sigma}_{=f(x=0)} = \infty$$

as well.

Conjugate priors

Exercise 3. ($\star\star$)Consider the Bayesian model

$$\begin{cases} x_i|\theta & \overset{\text{IID}}{\sim} \text{Mu}_k(\theta) \\ \theta & \sim \Pi(\theta) \end{cases}$$

where $\theta \in \Theta$, with $\Theta = \{\theta \in (0,1)^k | \sum_{j=1}^k \theta_j = 1\}$ and $\mathcal{X}_k = \{x \in \{0,...,n\}^k | \sum_{j=1}^k x_j = 1\}$.

Hint-1: Mu_k denotes the Multinomial probability distribution with PMF

$$\text{Mu}_k(x|\theta) = \begin{cases} \prod_{j=1}^k \theta_j^{x_j} & \text{, if } x \in \mathcal{X}_k \\ 0 & \text{, otherwise} \end{cases}$$

Hint-2: $\text{Di}_k(a)$ denotes the Dirichlet distribution with PDF

$$\text{Di}_k(\theta|a) = \begin{cases} \frac{\Gamma(\sum_{j=1}^k a_j)}{\prod_{j=1}^k \Gamma(a_j)} \prod_{j=1}^k \theta_j^{a_j-1} & \text{, if } \theta \in \Theta \\ 0 & \text{, otherwise} \end{cases}$$

1. Derive the conjugate prior distribution for θ, and recognize that it is a Dirichlet distribution family of distributions.
2. Verify that the prior distribution you derived above is indeed conjugate by using the definition.

Solution.

1. There are two alternative ways to derive the conjugate prior here.

   (a) [Way (a)] I can factorize the likelihood in a form that the likelihood kernel is a function of a sufficient statistic whose dimension is independent on the sample size n, and then derive the conjugate by substituting the sufficient statistic elements by prior hyper-parameters.
   There are $k-1$ independent parameters in $\text{Mu}_k(\theta)$ because $\sum_{j=1}^k \theta_j = 1$. I consider as parameters $(\theta_1,...,\theta_{k-1})$ and the last one is a function of them as $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$.
   The likelihood is

   $$f(x_{1:n}|\theta) = \prod_{i=1}^n \text{Mu}_k(x_i|\theta) = \prod_{i=1}^n \left[\prod_{j=1}^k \theta_j^{x_{i,j}} \right] = \prod_{j=1}^k \theta_j^{\sum_{i=1}^n x_{i,j}} = \prod_{j=1}^k \theta_j^{x_{*,j}} = \prod_{j=1}^{k-1} \theta_j^{x_{*,j}} \theta_k^{n-x_{*,k}}$$

   where $x_{*,j} = \sum_{i=1}^n x_{i,j}$. So

   $$f(x_{1:n}|\theta) = \prod_{j=1}^{k-1} \theta_j^{x_{*,j}} \left(1 - \sum_{j=1}^{k-1} \theta_j\right)^{n-x_{*,k}} = (1 - \sum_{j=1}^{k-1} \theta_j)^n \exp\left(\sum_{j=1}^{k-1} x_{*,j} \log(\frac{\theta_j}{1-\sum_{j=1}^{k-1}\theta_j})\right)$$

   and the sufficient statistic is

   $$t_n = (n, x_{*,1}, ..., x_{*,k-1})$$

   (b) [Way (b)] Alternatively, we can observe that the sampling space $\mathcal{X}_k$ does not depend on the parameters. So we can show that the sampling distribution is an exponential family of distributions, identify its components, and then derive the conjugate prior.
   There are $k-1$ independent parameters in $\text{Mu}_k(\theta)$ because $\sum_{j=1}^k \theta_j = 1$. I consider as parameters $(\theta_1,...,\theta_{k-1})$ and the last one is a function of them as $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$.

It is

$$\mathrm{Mu}_k(x|\theta) = \prod_{j=1}^{k} \theta_j^{x_j} = \prod_{j=1}^{k-1} \theta_j^{x_j} (1 - \sum_{j=1}^{k-1} \theta_j)^{1-\sum_{j=1}^{k-1} x_j} = (1 - \sum_{j=1}^{k-1} \theta_j) \exp(\sum_{j=1}^{k-1} x_j \log(\frac{\theta_j}{1-\sum_{j=1}^{k-1} \theta_j}))$$

This is the $k-1$ exponential family PDF with

$$u(x) = 1; \qquad g(\theta) = (1 - \sum_{j=1}^{k-1} \theta_j); \qquad c = (1, ..., 1)$$

$$h(x) = (x_1, ...x_{k-1}); \qquad \phi(\theta) = (\log(\frac{\theta_1}{1-\sum_{j=1}^{k-1} \theta_j}), ..., \log(\frac{\theta_{k-1}}{1-\sum_{j=1}^{k-1} \theta_j})),$$

Then either by substituting the sufficient statistics in way (a), or by using the components of the exponential family of distributions in way (b) Let $\tau = (\tau_0, ..., \tau_{k-1})$. It is

$$\begin{aligned}\pi(\theta|\tau) &\propto (1 - \sum_{j=1}^{k-1} \theta_j)^{\tau_0} \exp(\sum_{j=1}^{k-1} \tau_j \log(\frac{\theta_j}{1-\sum_{j=1}^{k-1} \theta_j})) \\ &\propto \prod_{j=1}^{k-1} \theta_j^{\tau_j} (1 - \sum_{j=1}^{k-1} \theta_j)^{\tau_0 - \sum_{j=1}^{k-1} \tau_j} \propto \prod_{j=1}^{k-1} \theta_j^{\tau_j}\, \theta_k^{\tau_0 - \sum_{j=1}^{k-1} \tau_j}\end{aligned}$$

Here, I recognize the Dirichlet distribution with $a_j = \tau_j$ for $j = 1, ..., k-1$ and $a_k = \tau_0 - \sum_{j=1}^{k-1} \tau_j$.

2. Well, the posterior is Dirichlet too. It is

$$\pi(\theta|x_{1:n}) = \prod_{i=1}^{n} \mathrm{Mu}_k(x_i|\theta)\mathrm{Di}_k(\theta|a) \propto \prod_{j=1}^{k} \theta_j^{x_{*,j}} \prod_{j=1}^{k} \theta_j^{a_j - 1} = \prod_{j=1}^{k} \theta_j^{x_j + a_{*,j} - 1} \propto \mathrm{Di}_k(\theta|\tilde{a})$$

where $\tilde{a} = (\tilde{a}_1, ..., \tilde{a}_k)$, with $\tilde{a}_j = a_j + x_{*,j}$ for $j = 1, ..., k$. So the posterior is $\theta|x_{1:n} \sim \mathrm{Di}_k(\tilde{a})$.

Jeffreys priors

Exercise 4. ($\star\star$)Consider the trinomial distribution

$$\begin{aligned}p(x, y|\pi, \rho) &= \frac{n!}{x!\,y!\,z!} \pi^x \rho^y \sigma^z, \qquad (x + y + z = n) \\ &\propto \pi^x \rho^y (1 - \pi - \rho)^{n-x-y}.\end{aligned}$$

Specify a Jeffreys' prior for (π, ρ).

HINT: It is $\mathrm{E}(x) = n\pi$, $\mathrm{E}(y) = n\rho$.

Solution.

It is

$$\begin{aligned}\partial^2 L/\partial\pi^2 &= -x/\pi^2 - z/(1 - \pi - \rho)^2 \\ \partial^2 L/\partial p^2 &= -y/p^2 - z/(1 - \pi - \rho)^2 \\ \partial^2 L/\partial\pi\partial p &= -z/(1 - \pi - \rho)^2\end{aligned}$$

and

$$
\begin{aligned}
I(\pi, \rho \,|\, x, y, z) &= -\mathrm{E}\begin{pmatrix} -x/\pi^2 - z/(1-\pi-\rho)^2 & -z/(1-\pi-\rho)^2 \\ -z/(1-\pi-\rho)^2 & -y/p^2 - z/(1-\pi-\rho)^2 \end{pmatrix} \\
&= \begin{pmatrix} n/\pi + n/(1-\pi-\rho) & n/(1-\pi-\rho) \\ n/(1-\pi-\rho) & n/\rho + n/(1-\pi-\rho) \end{pmatrix}
\end{aligned}
$$

Because $\mathrm{E}(x) = n\pi$, $\mathrm{E}(y) = np$, $\mathrm{E}(z) = n(1-\pi-\rho)$, and

$$
\begin{aligned}
\det I(\pi, \rho \,|\, x, y, z) &= (n/\pi + n/(1-\pi-\rho))(n/\rho + n/(1-\pi-\rho)) - (n/(1-\pi-\rho))^2 \\
&= ... \\
&= n\{\pi\rho(1-\pi-\rho)\}^{-1}
\end{aligned}
$$

So the Jeffrey's prior is

$$
p(\pi, \rho) \propto \pi^{-\frac{1}{2}} \rho^{-\frac{1}{2}} (1-\pi-\rho)^{-\frac{1}{2}}
$$

Exercise 5. ($\star\star$)Consider the Bayesian model

$$
\begin{cases} x_i|\theta & \stackrel{\text{iid}}{\sim} \text{Ga}(\alpha, \beta), \ \forall i = 1, ..., n \\ (\alpha, \beta) & \sim \Pi(\alpha, \beta) \end{cases}
$$

where $\text{Ga}(a, \beta)$ is the Gamma distribution with expected value α/β. Specify a Jeffrey's prior for $\theta = (\alpha, \beta)$.

Hint-1: Gamma distr.: $x \sim \text{Ga}(a, b)$ has pdf $f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) 1_{(0,+\infty)}(x)$, and Expected value $\mathrm{E}_{\text{Ga}}(x|a, b) = \frac{a}{b}$

Hint-2: You may also need that the second derivative of the logarithm of a Gamma function is the 'polygamma function of order 1'. Ie,

- $F^{(0)}(\alpha) = \frac{\mathrm{d}}{\mathrm{d}\alpha} \log(\Gamma(a))$
- $F^{(1)}(\alpha) = \frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} \log(\Gamma(a))$

Hint-3: You may leave your answer in terms of function $F^{(1)}(\alpha)$.

Solution. It is $\pi(\alpha, \beta) \propto \sqrt{\det(\mathscr{I}(\alpha, \beta))} \propto \sqrt{\det(\mathscr{I}_1(\alpha, \beta))}$ where

$$
\mathscr{I}_1(\alpha, \beta) = -\mathrm{E}_{\mathrm{F}(x|\alpha,\beta)} \begin{bmatrix} \frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} \log(f(x|\alpha, \beta)) & \frac{\mathrm{d}^2}{\mathrm{d}\alpha \mathrm{d}\beta} \log(f(x|\alpha, \beta)) \\ \frac{\mathrm{d}^2}{\mathrm{d}\alpha \mathrm{d}\beta} \log(f(x|\alpha, \beta)) & \frac{\mathrm{d}^2}{\mathrm{d}\beta^2} \log(f(x|\alpha, \beta)) \end{bmatrix}, \text{ with}
$$

So

$$
\begin{aligned}
f(x|\alpha, \beta) = \text{Ga}(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(a)} x^{\alpha-1} \exp(-\beta x) \Longrightarrow \\
\log(f(x|\alpha, \beta)) = a \log(\beta) - \log(\Gamma(\alpha)) - \beta x + (\alpha - 1) \log(x)
\end{aligned}
$$

So

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\alpha}\log(f(x|\alpha,\beta)) &= \log(\beta) - \frac{\mathrm{d}}{\mathrm{d}\alpha}\log(\Gamma(\alpha)) + \log(x) \\
\frac{\mathrm{d}}{\mathrm{d}\alpha^2}\log(f(x|\alpha,\beta)) &= -\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\log(\Gamma(\alpha)) = -\digamma^{(1)}(\alpha) \\
\frac{\mathrm{d}}{\mathrm{d}\beta}\log(f(x|\alpha,\beta)) &= \frac{\alpha}{\beta} - x \\
\frac{\mathrm{d}^2}{\mathrm{d}\beta^2}\log(f(x|\alpha,\beta)) &= -\frac{\alpha}{\beta^2} \\
\frac{\mathrm{d}^2}{\mathrm{d}\alpha\mathrm{d}\beta}\log(f(x|\alpha,\beta)) &= \frac{1}{\beta}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{E}_{\mathrm{Ga}(a,b)}\left(\frac{\mathrm{d}}{\mathrm{d}\alpha^2}\log(f(x|\alpha,\beta))\right) &= -\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\log(\Gamma(\alpha)) = -\digamma^{(1)}(\alpha) \\
\mathrm{E}_{\mathrm{Ga}(a,b)}\left(\frac{\mathrm{d}^2}{\mathrm{d}\beta^2}\log(f(x|\alpha,\beta))\right) &= -\frac{\alpha}{\beta^2} \\
\mathrm{E}_{\mathrm{Ga}(a,b)}\left(\frac{\mathrm{d}^2}{\mathrm{d}\alpha\mathrm{d}\beta}\log(f(x|\alpha,\beta))\right) &= \frac{1}{\beta}
\end{aligned}
$$

Hence

$$
\mathscr{I}_1(\alpha,\beta) = -\mathrm{E}_{\mathrm{Ga}(a,b)}\begin{bmatrix} -\digamma^{(1)}(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & -\frac{\alpha}{\beta^2} \end{bmatrix} = \begin{bmatrix} \digamma^{(1)}(\alpha) & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{bmatrix}
$$

Therefore

$$
\pi(\alpha,\beta) \propto \sqrt{\det(\mathscr{I}(\alpha,\beta))} \propto \sqrt{\det(\mathscr{I}_1(\alpha,\beta))} = \sqrt{\digamma^{(1)}(\alpha)\frac{\alpha}{\beta^2} + \frac{1}{\beta^2}} = \frac{1}{\beta}\sqrt{\digamma^{(1)}(\alpha)\alpha + 1}
$$

..

Exercise 6. ($\star\star$)Consider the Bayesian model

$$
\begin{cases} x_i|\theta & \stackrel{\text{iid}}{\sim} \mathrm{Ex}(\theta),\ \forall i = 1, ..., n \\ \theta & \sim \mathrm{Ga}(a,b) \end{cases}
$$

Hint-1: The PDF of $x \sim \mathrm{G}(a,b)$ is $\mathrm{Ga}(x|a,b) = \frac{b^a}{\Gamma(a)}x^{a-1}\exp(-bx)1_{(0,+\infty)}(x)$

Hint-2: The PDF of $x \sim \mathrm{Ex}(\theta)$ is $\mathrm{Ex}(x|\theta) = \mathrm{Ga}(x|1,\theta)$

1. Show that the parametric model is member of the Exponential family, and the sufficient statistic for a sample of observables $x = (x_1, ..., x_n)$.
2. Show that the posterior distribution θ given x is Gamma and compute its parameters.
3. Show that the predictive distribution $G(z|x)$ of a future z given $x = (x_1, ..., x_n)$, has PDF

$$
g(z|x) = \frac{a^*(b^*)^{a^*}}{(z+b^*)^{a^*+1}}1(x \geq 0)
$$

Solution.

1. The parametric model is

$$\text{Ex}(x|\theta) = \theta \exp(-\theta x) 1(x \geq 0)$$

It is member of the exponential family

$$\text{Ef}_1(x|u, g, h, c, \phi, \theta, c) = u(x) g(\theta) \exp(\sum_{j=1}^{k} c_j \phi_j(\theta) h_j(x))$$

with $u(x_{1:n}) = 1$, $g(\theta) = \theta$, $c_1 = -1$, $\phi_1(\theta) = \theta$, $h_1(x) = x$. The sufficient statistic is $t_n = (n, \sum_{i=1}^{n} x_i)$.

2. I can get the posterior by using the Bayes theorem

$$\begin{aligned}\pi(\theta|x) \propto f(x|\theta)\pi(\theta|a,b) \qquad &\propto \prod_{i=1}^{n} \text{Ex}(x_i|\theta)\text{Ga}(\theta|a,b) \\ \propto \theta^n \exp(-\theta \sum_{i=1}^{n} x_i)\theta^{a-1}\exp(-\theta b) \ &\propto \theta^{a+n-1}\exp(-\theta(\sum_{i=1}^{n} x_i + b)) \\ \propto \text{Ga}(\theta|\underbrace{a+n}_{=a^*}, \underbrace{b + \sum_{i=1}^{n} x_i}_{=b^*})&\end{aligned}$$

3. By using the definition of the predictive distribution, it is ...

$$\begin{aligned}g(z|x) &= \int_{\mathbb{R}_+} f(z|\theta)\pi(\theta|x)\text{d}\theta \ \overset{z\geq 0}{=} \int_{\mathbb{R}_+} \theta \exp(-\theta z) \frac{(b^*)^{a^*}}{\Gamma(a^*)}\theta^{a^*-1}\exp(-\theta b^*)\text{d}\theta \\ &= \frac{(b^*)^{a^*}}{\Gamma(a^*)}\int_{\mathbb{R}_+} \theta^{a^*+1-1}\exp(-\theta(z+b^*))\text{d}\theta \ = \frac{(b^*)^{a^*}}{\Gamma(a^*)}\frac{\Gamma(a^*+1)}{(z+b^*)^{a^*+1}} = \frac{a^*(b^*)^{a^*}}{(z+b^*)^{a^*+1}} \\ &= \frac{a^*(b^*)^{a^*}}{(z+b^*)^{a^*+1}}\end{aligned}$$

Further practice

From the exercise sheet, have a look at Exercises 26, 31, 47, 6, and 50.

A About Nuisance parameters

Assume observable quantities $y = (y_1, ..., y_n)$. Assume that the sampling distribution is $\mathrm{d}F(y|\theta)$ labeled by an unknown parameter $\theta \in \Theta$. Let $\theta = (\phi, \lambda)^\top$ with $\phi \in \Phi$ and $\lambda \in \Lambda$. Assume You are interested in learning parameter $\phi \in \Phi$, and You are not interested in learning the unknown parameter $\lambda \in \Lambda$; but both ϕ, λ are parts of the statistical model parameterisation. The unknown quantity $\lambda \in \Lambda$ is called nuisance parameter. We an call $\phi \in \Phi$ parameter of interest.

Note 7. In Bayesian Stats, learning (or quantifying uncertainty about) parameter of interest ϕ under the presence of a nuisance parameter $\lambda \in \Lambda$ is performed according to the Bayesian paradigm as usual: You specify a prior $\mathrm{d}\Pi(\phi, \lambda)$ with PDF/PMF $\pi(\phi, \lambda) = \pi(\phi|\lambda)\pi(\lambda)$ on the joint space of ALL Your unknown parameters $\theta = (\phi, \lambda)^\top$; you compute the joint posterior distribution $\mathrm{d}\Pi(\theta|y)$ of $\theta = (\phi, \lambda)^\top$ via the Bayesian theorem. Reasonably, Your posterior degree of believe about the parameter of interest ϕ given the data $y = (1_1, ..., y_n)$ is given through the marginal posterior distribution $\mathrm{d}\Pi(\phi|y)$.

Note 8. To summarize; Specify the Bayesian model as: <sum-up

$$\begin{cases} y|\overbrace{\phi, \lambda}^{=\theta} \sim F(y|\overbrace{\phi, \lambda}^{=\theta}) & \text{, the statistical model} \\ (\underbrace{\phi, \lambda}_{=\theta}) \sim \Pi(\underbrace{\phi, \lambda}_{=\theta}) & \text{, the prior model} \end{cases}$$

The joint posterior of θ given y is $\mathrm{d}\Pi(\theta|y) = \mathrm{d}\Pi(\lambda|y, \phi)\mathrm{d}\Pi(\phi|y)$ is with PDF/PMF

$$\pi(\overbrace{\phi, \lambda}^{=\theta}|y) = \frac{f(y|\overbrace{\phi, \lambda}^{=\theta})\pi(\overbrace{\phi, \lambda}^{=\theta})}{f(y)} = \underbrace{\frac{f(y|\phi, \lambda)\pi(\lambda|\phi)}{f(y|\phi)}}_{=\pi(\lambda|y,\phi)}\underbrace{\frac{f(y|\phi)\pi(\phi)}{f(y)}}_{=\pi(\phi|y)} = \pi(\lambda|y, \phi)\pi(\phi|y)$$

The (marginal) likelihood $f(y|\phi)$ of y given ϕ is

$$f(y|\phi) = \underbrace{\int_\Lambda f(y|\overbrace{\phi, \lambda}^{=\theta})\mathrm{d}\Pi(\lambda|\phi)}_{=\mathrm{E}_{\Pi(\lambda|\phi)}(f(y|\phi,\lambda)|\phi)} = \begin{cases} \int_\Lambda f(y|\phi, \lambda)\pi(\lambda|\phi)\mathrm{d}\lambda & \text{, if } \lambda \text{ cont} \\ \sum_{\forall \lambda \in \Lambda} f(y|\phi, \lambda)\pi(\lambda|\phi) & \text{, if } \lambda \text{ discr} \end{cases}$$

The PDF/PMF $\pi(\phi|y)$ of marginal posterior $\mathrm{d}\Pi(\phi|y)$ of ϕ is

$$\pi(\phi|y) = \underbrace{\int_\Lambda \pi(\overbrace{\phi, \lambda}^{=\theta}|y)\mathrm{d}\lambda}_{=\mathrm{E}_{\Pi(\lambda|y)}(\pi(\phi|y,\lambda))} \qquad \text{or equivalently} \qquad \pi(\phi|y) = \frac{f(y|\phi)\pi(\phi)}{f(y)}$$

The predictive distribution $\mathrm{d}G(z|y)$ of the next outcome $z = (y_{n+1}, ...y_{n+m})$ given y has pdf/pmf

$$g(z|y) = \int f(y|\overbrace{\phi, \lambda}^{=\theta})\mathrm{d}\Pi(\overbrace{\phi, \lambda}^{=\theta}|y)$$

and the marginal likelihood $f(y)$ is

$$f(y) = \int f(y|\overbrace{\phi, \lambda}^{=\theta})\pi(\overbrace{\phi, \lambda}^{=\theta})\mathrm{d}\phi\mathrm{d}\lambda$$

B Criteria for integrals

General: Let integrable functions $f(x)$, and $g(x)$ for $x \geq a$.

Let

$$0 \leq f(x) \leq g(x), \quad \text{for } x \geq a$$

Then

$$\int_a^\infty g(x)\mathrm{d}x < \infty \implies \int_a^\infty f(x)\mathrm{d}x < \infty$$
$$\int_a^\infty f(x)\mathrm{d}x = \infty \implies \int_a^\infty g(x)\mathrm{d}x = \infty$$

Type I: Let integrable functions $f(x)$, and $g(x)$ for $x \geq a$, and let $g(x)$ be positive.

Let

$$\lim_{n\to\infty} \frac{f(x)}{g(x)} = c$$

Then

- If $c \in (0, \infty)$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \iff \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = 0$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \implies \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = \infty$:
$$\int_a^\infty f(x)\mathrm{d}x = \infty \implies \int_a^\infty g(x)\mathrm{d}x = \infty$$

Type II: Let integrable functions $f(x)$, and $g(x)$ for $a < x \leq b$, and let $g(x)$ be positive.

Let

$$\lim_{n\to a^+} \frac{f(x)}{g(x)} = c$$

Then

- If $c \in (0, \infty)$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \iff \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = 0$:
$$\int_a^\infty g(x)\mathrm{d}x < \infty \implies \int_a^\infty f(x)\mathrm{d}x < \infty$$
- If $c = \infty$:
$$\int_a^\infty f(x)\mathrm{d}x = \infty \implies \int_a^\infty g(x)\mathrm{d}x = \infty$$

Note: A useful test function is

$$\int_0^\infty \left(\frac{1}{x}\right)^p \mathrm{d}x \quad \begin{cases} < \infty & , \text{ when } p > 1 \\ = \infty & , \text{ when } p \leq 1 \end{cases}$$