Solutions to Problem Sheet 6

The *Inverse-Gamma* distribution $\mathrm{IG}(a,b)$ is the distribution on $\mathbb{R}^+$ with density, for $a, b > 0$,

$$x \mapsto x^{-(a+1)} e^{-b/x} \frac{b^a}{\Gamma(a)}.$$

1. **Conjugate Gaussian family in $\mathbb{R}^d$**

   Let $\mathcal{P} = \{P_\theta = \mathcal{N}(\theta, \Sigma), \theta \in \mathbb{R}^d\}$, where Σ is a known invertible variance-covariance matrix. Let Π be the prior distribution $\mathcal{N}(0, \Lambda)$ on θ, with Λ invertible.

   (a) Show that if $(X_1, \ldots, X_n)|\theta \sim P_\theta^{\otimes n}$, the posterior distribution is written as $\theta|(X_1, \ldots, X_n) \sim \mathcal{N}(\theta_X, \Sigma_X)$, with

$$\Sigma_X = (n\Sigma^{-1} + \Lambda^{-1})^{-1}$$
$$\theta_X = n\Sigma_X\Sigma^{-1}\overline{X}.$$

Solution: Let $X|\theta \sim \mathcal{N}(\theta, \Sigma)^{\otimes n}$ and $\theta \sim \Pi = \mathcal{N}(0, \Lambda)$. Bayes' formula gives:

$$f_{\theta|X}(\theta) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i - \theta)^T\Sigma^{-1}(X_i - \theta) - \frac{1}{2}\theta^T\Lambda^{-1}\theta\right\}.$$

It suffices to group the terms in θ to form a new quadratic form in the exponential, of the form $-\frac{1}{2}(\theta - v)^T\Sigma_X^{-1}(\theta - v)$. We are thus looking for v, Σ_X such that, by expanding each quadratic form:

$$\theta^T(n\Sigma^{-1} + \Lambda^{-1})\theta - 2\left(\sum_{i=1}^{n} X_i\right)^T \Sigma^{-1}\theta = \theta^T\Sigma_X^{-1}\theta - 2v^T\Sigma_X^{-1}\theta.$$

By identifying the terms, it suffices to set:

$$\Sigma_X = (n\Sigma^{-1} + \Lambda^{-1})^{-1}$$

and to choose v such that

$$2v^T\Sigma_X^{-1} = 2\left(\sum_{i=1}^{n} X_i\right)^T \Sigma^{-1} = 2n\overline{X}^T\Sigma^{-1},$$

which gives $v = n\Sigma_X\Sigma^{-1}\overline{X}$. Since v corresponds to the posterior mean, $\theta_X = v$.

(b) Is the class of prior distributions $\{\mathcal{N}(0,\Lambda), \Lambda \text{ invertible}\}$ conjugate?

> **Solution:** The family is not conjugate because the posterior distribution does not generally have a mean of zero.

2. **Gaussian family with unknown mean and variance**

Consider the model $\mathcal{P} = \{P_\theta = P_{\mu,\sigma^2} = \mathcal{N}(\mu,\sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$ where we set $\theta = (\mu, \sigma^2)$. We have observations $X = (X_1, \ldots, X_n)$, with distribution $P_\theta^{\otimes n}$ given θ. We define the measure Π:

$$d\Pi(\mu,\sigma^2) = \frac{1}{\sigma^2} d\mu d\sigma^2,$$

where $d\mu d\sigma^2$ is interpreted as $d\mathrm{Leb}_{\mathbb{R}}(\mu) d\mathrm{Leb}_{\mathbb{R}^+}(\sigma^2)$.

(a) Verify that Π is an improper prior "distribution".

> **Solution:** The prior is improper since $\iint \sigma^{-2} d\mu d\sigma^2 = +\infty$.

(b) Show that $\mathcal{L}(\sigma^2|X)$ is an $\mathrm{IG}(\frac{n-1}{2}, \frac{s}{2})$ distribution, where $s = \sum_{i=1}^n (X_i - \overline{X})^2$. To do this:

i. Show that the integral $\int p_{\mu,\sigma^2}(X)\sigma^{-2} d\mu$ is finite. It thus allows us to define, up to a proportionality constant, a "joint density" of (σ^2, X).

> **Solution:** We write $p_{\mu,\sigma^2}(X)\sigma^{-2}$ by factoring the terms with respect to μ:
>
> $$\begin{aligned} p_{\mu,\sigma^2}(X)\sigma^{-2} &= C(\sigma^2)^{-\frac{n}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2\right\} \\ &= C(\sigma^2)^{-\frac{n}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}\left[n\mu^2 - 2n\overline{X}\mu + n\overline{X^2}\right]\right\} \\ &= C(\sigma^2)^{-\frac{n}{2}-1} \exp\left\{-\frac{n}{2\sigma^2}[\mu-\overline{X}]^2 - \frac{nS^2}{2\sigma^2}\right\}, \end{aligned}$$
>
> where we defined $S^2 := \overline{X^2} - \overline{X}^2$, making $nS^2 = s$. The integral of the preceding quantity with respect to μ is thus finite since it is, up to a constant, the integral of a Gaussian density.

ii. Deduce $\mathcal{L}(\sigma^2|X)$.

> **Solution:** The density of the distribution $\mathcal{L}(\sigma^2|X)$ is proportional, by definition, since we are working with improper priors, to the "joint density" above, i.e.,

$\int p_{\mu,\sigma^2}(X)\sigma^{-2}d\mu$. We then have:

$$\int p_{\mu,\sigma^2}(X)\sigma^{-2}d\mu \propto (\sigma^2)^{-\frac{n}{2}-1}\int \exp\left\{-\frac{n}{2\sigma^2}[\mu-\overline{X}]^2-\frac{nS^2}{2\sigma^2}\right\}d\mu$$
$$\propto (\sigma^2)^{-\frac{n}{2}-1}e^{-\frac{nS^2}{2\sigma^2}}\sqrt{\frac{\sigma^2}{n}}$$
$$\propto (\sigma^2)^{-\frac{n-1}{2}-1}e^{-\frac{nS^2}{2\sigma^2}}.$$

We recognize the density of an $\mathrm{IG}(\frac{n-1}{2},\frac{nS^2}{2})=\mathrm{IG}(\frac{n-1}{2},\frac{s}{2})$ distribution.

iii. Verify also in passing that the posterior distribution $\mathcal{L}((\mu,\sigma^2)|X)$ is well-defined.

Solution: It suffices to verify that the integral $\iint p_{\mu,\sigma^2}(X)\sigma^{-2}d\mu d\sigma^2$ is finite. This follows from the previous calculation since the integral with respect to μ yields a quantity that is integrable in σ^2, as it is proportional to the density of the inverse-gamma distribution established above.

(c) Characterize the posterior distribution $\mathcal{L}(\mu,\sigma^2|X)$ using $\mathcal{L}(\mu|\sigma^2,X)$ and $\mathcal{L}(\sigma^2|X)$.

Solution: The density of the distribution $\mathcal{L}(\mu|\sigma^2,X)$ is the density in μ proportional to the quantity

$$p_{\mu,\sigma^2}(X)\sigma^{-2}\propto \exp\left\{-\frac{n}{2\sigma^2}[\mu-\overline{X}]^2-\frac{nS^2}{2\sigma^2}\right\}\propto \exp\left\{-\frac{n}{2\sigma^2}[\mu-\overline{X}]^2\right\}.$$

We deduce $\mathcal{L}(\mu|\sigma^2,X)=\mathcal{N}(\overline{X},\sigma^2/n)$, and we saw earlier that $\mathcal{L}(\sigma^2|X)=\mathrm{IG}(\frac{n-1}{2},\frac{s}{2})$. Furthermore, we note that the distribution of $\mathcal{L}(\mu,\sigma^2|X)$ has a density which is the product of the densities of $\mathcal{L}(\mu|\sigma^2,X)$ and $\mathcal{L}(\sigma^2|X)$. This is the usual result that the density of (Y,Z) is the product of the densities of $Y|Z$ and Z (simply, all distributions here are conditioned on X).

(d) Construct a credible region for μ at level $1-\alpha$.

Solution: To determine a credible region for μ, it suffices to a) determine the distribution $\mathcal{L}(\mu|X)$ and b) select a region deduced, for example, from the posterior quantiles. To determine $\mathcal{L}(\mu|X)$, we proceed as we did for $\mathcal{L}(\sigma^2|X)$: we determine the joint distribution of μ, X by integrating the density of μ,σ^2,X over σ^2, and then deduce the distribution of μ given X as usual. The density of μ, X is proportional to

$$\begin{aligned}\int p_{\mu,\sigma^2}(X)\sigma^{-2}d\sigma^2 &\propto \int (\sigma^2)^{-\frac{n}{2}-1}\exp\left\{-\frac{n}{2\sigma^2}[\mu-\overline{X}]^2 - \frac{nS^2}{2\sigma^2}\right\}d\sigma^2 \\ &\propto \int (\sigma^2)^{-\frac{n}{2}-1}\exp\left\{-\sigma^{-2}\left(\frac{n}{2}[\mu-\overline{X}]^2 + \frac{nS^2}{2}\right)\right\}d\sigma^2 \\ &\propto \left(\frac{n}{2}[\mu-\overline{X}]^2 + \frac{nS^2}{2}\right)^{-\frac{n}{2}},\end{aligned}$$

where we used the expression of the density of an inverse-gamma distribution given in the problem statement to calculate the integral: we express it as $\Gamma(a)/b^a$, with $a = n/2$ and b the term inside the exponent above. The above expression resembles the density of a Student's t-distribution also recalled at the beginning. We write

$$\int p_{\mu,\sigma^2}(X)\sigma^{-2}d\sigma^2 \propto \left(\frac{n-1}{S^2}\frac{[\mu-\overline{X}]^2}{n-1}+1\right)^{-\frac{n}{2}}.$$

Moreover, if Y follows a Student's t-distribution with $n-1$ degrees of freedom, and if g is the density of this distribution, we notice that $Z = aY + \overline{X}$ has density $\frac{1}{a}g\left(\frac{z-\overline{X}}{a}\right)$. We conclude that the above density is the density of the variable

$$Z = \overline{X} + \sqrt{\frac{S^2}{n-1}}Y,$$

where Y follows a Student's t-distribution with $n-1$ degrees of freedom. A credible interval $[a, b]$ therefore satisfies

$$P[Z \in [a,b]] = 1-\alpha,$$

or equivalently

$$P\left[\sqrt{\frac{n-1}{S^2}}(a-\overline{X}) \leq Y \leq \sqrt{\frac{n-1}{S^2}}(b-\overline{X})\right] = 1-\alpha.$$

Since the Student's t-distribution is symmetric, we have $P[|Y| \leq z_\alpha] = 1-\alpha$ if $z_\alpha = q_{1-\alpha/2}^{\text{Student}(n-1)}$. We conclude that

$$I(X) = \left[\overline{X} \pm \sqrt{\frac{S^2}{n-1}}z_\alpha\right]$$

is a credible interval for μ at level $1-\alpha$.

Remark: This credible interval coincides with the confidence interval used in frequentist statistics. Indeed, if $X_1, \ldots, X_n \sim \mathcal{N}(\mu_0, \sigma_0^2)^{\otimes n}$, by definition of the Student's t-distribution, $\sqrt{n-1}(\overline{X}-\mu_0)/\sqrt{S^2}$ follows a Student's t-distribution with $n-1$ degrees of freedom and thus

$$P_{\mu_0,\sigma_0}\left[\mu_0 \in \left[\overline{X} \pm \sqrt{\frac{S^2}{n-1}}z_\alpha\right]\right] = 1-\alpha,$$

so $I(X)$ is also a confidence interval for μ at level $1-\alpha$.

3. **Empirical Bayes and normal distributions**

We are in the framework of the fundamental model $\mathcal{P} = \{P_\theta = \mathcal{N}(\theta, 1), \theta \in \mathbb{R}\}$. We have n i.i.d. observations $X_1, \dots, X_n$ given θ from the distribution P_θ. Let $\Pi = \Pi_\mu = \mathcal{N}(\mu, 1)$ be a prior distribution on θ. We propose to determine μ using an empirical Bayes method.

(a) What does this method consist of?

> **Solution:** It consists of choosing an estimator $\hat{\mu}$ of the parameter μ from the considered class of prior distributions $\{\Pi_\mu, \mu \in \mathbb{R}\}$, and following the Bayesian approach with the prior distribution $\Pi_{\hat{\mu}}$.

(b) We construct $\hat{\mu}$ using the marginal maximum likelihood method. Recall the principle of this method in two lines maximum.

> **Solution:** We form the marginal density of X (which depends on μ, because $\Pi = \Pi_\mu$ depends on it) and $\hat{\mu}$ is the point where the maximum of this density, namely $\mu \mapsto \int p_\theta(X) d\Pi_\mu(\theta)$, is reached.

(c) Show that the marginal distribution of $(X_1, \dots, X_n)$ is that of a Gaussian vector. You may draw inspiration from exercise 5 of PS4.

> **Solution:** *1st method.* We follow the method from exercise 2 of TD5, which allows us to determine the marginal distribution of $X = (X_1, \dots, X_n)$ in the model: it is a $\mathcal{N}(\mu\mathbf{1}, \Sigma)$ distribution with $\Sigma = I_{n\times n} + UU^T$, where U is the column vector composed only of 1s. We write the corresponding density explicitly at the point $X_1, \dots, X_n$ and maximize it with respect to μ. For this, we can use that $\Sigma^{-1} = (I_{n\times n} + UU^T)^{-1} = I_{n\times n} - \lambda UU^T$ with a well-chosen λ. After calculations, we obtain $\hat{\mu} = \overline{X}$.
>
> *2nd method.* We calculate the marginal density of X up to a proportionality constant and maximize it with respect to μ.
>
> $$\begin{aligned} f_X(X_1, \dots, X_n) &= \int \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_i - \theta)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta-\mu)^2} d\theta \\ &\propto e^{-\frac{\mu^2}{2}} \int e^{-\frac{n}{2}\theta^2 + n\overline{X}\theta - \frac{1}{2}\theta^2 + \theta\mu} d\theta \\ &\propto e^{-\frac{\mu^2}{2}} \int e^{-\frac{n+1}{2}\left[\theta - \frac{n\overline{X}+\mu}{n+1}\right]^2} e^{\frac{(n\overline{X}+\mu)^2}{2(n+1)}} d\theta \\ &\propto e^{-\frac{\mu^2}{2} + \frac{(n\overline{X}+\mu)^2}{2(n+1)}} \propto e^{-\frac{n}{2(n+1)}(\mu - \overline{X})^2}. \end{aligned}$$
>
> The maximum in μ of the marginal density of X is therefore reached for $\mu = \overline{X}$.

(d) Deduce that $\hat{\mu} = \overline{X}$. What is the final posterior distribution suggested by the empirical Bayes method?

> **Solution:** The posterior distribution corresponding to the prior $\Pi = \Pi_{\hat{\mu}} = \mathcal{N}(\overline{X}, 1)$ is, by standard calculation, a $\mathcal{N}(\frac{n\overline{X}+\hat{\mu}}{n+1}, \frac{1}{n+1}) = \mathcal{N}(\overline{X}, \frac{1}{n+1})$ distribution. We note that

the posterior distribution corresponding to a choice of μ by the marginal maximum likelihood method is, here, the distribution centered at the maximum likelihood estimate $\overline{X}$ for this model.

4. **Empirical Bayes and Poisson distributions**

Let $\mathcal{P} = \{P_\theta = \mathcal{P}(\theta), \theta > 0\}$. We have observations $X_1, \ldots, X_n$ i.i.d. with distribution P_θ given θ. Let $\Pi = \Pi_\lambda = \mathcal{E}(\lambda)$ be a prior distribution on θ. We propose to determine λ using an empirical Bayes method.

(a) Show that the marginal distribution of X_1 is a geometric distribution with parameter $\lambda/(\lambda + 1)$.

Solution: The marginal density of X_1 is:

$$\begin{aligned} f_{X_1}(x_1) &= \int p_\theta(x_1) d\Pi_\lambda(\theta) \\ &= \int \frac{\theta^{x_1}}{x_1!} e^{-\theta} \lambda e^{-\lambda\theta} d\theta \\ &= \frac{\lambda}{x_1!} \int \theta^{x_1} e^{-(\lambda+1)\theta} d\theta \\ &= \frac{\lambda}{x_1!} \frac{\Gamma(x_1+1)}{(\lambda+1)^{x_1+1}} = \frac{\lambda}{(\lambda+1)^{x_1+1}}, \end{aligned}$$

where we used the expression for the density of a Gamma distribution (cf. TD2) to calculate the last integral, and the fact that $\Gamma(p + 1) = p!$ for any integer p. This matches the density of a geometric distribution with parameter $\lambda/(\lambda + 1)$.

(b) Calculate the marginal density of $(X_1, \ldots, X_n)$ as a function of λ.

Solution: Similarly, the marginal density of $(X_1, \ldots, X_n)$ evaluated at the observed points is:

$$\begin{aligned} \int \prod_{i=1}^n p_\theta(X_i) d\Pi_\lambda(\theta) &= \int \frac{\theta^{n\overline{X}}}{\prod_{i=1}^n X_i!} e^{-n\theta} \lambda e^{-\lambda\theta} d\theta \\ &= \frac{\lambda}{\prod_{i=1}^n X_i!} \int \theta^{n\overline{X}} e^{-(n+\lambda)\theta} d\theta \\ &= \frac{\lambda}{\prod_{i=1}^n X_i!} \frac{\Gamma(n\overline{X}+1)}{(\lambda+n)^{1+n\overline{X}}}. \end{aligned}$$

(c) Deduce that $\hat{\lambda}^{EB} = 1/\overline{X}$, then the final posterior distribution suggested by the empirical Bayes method.

Solution: We deduce that $\hat{\lambda}^{EB}$ maximizes the preceding quantity with respect to λ. It thus suffices to find the point where the maximum of the function

$$\psi(\lambda) = \frac{\lambda}{(\lambda + n)^{1+n\overline{X}}}$$

is reached. We set the derivative to zero and find $\hat{\lambda}^{EB} = 1/\overline{X}$, after verifying that it is indeed a maximum (the derivative is positive and then negative, so this is the case). The final posterior distribution suggested by the method is therefore $\Pi_{\hat{\lambda}^{EB}}[\cdot|X]$. We calculate the posterior density for any fixed λ:

$$f_{\theta|X}(\theta) \propto \theta^{n\overline{X}} e^{-(n+\lambda)\theta}.$$

Thus $\Pi_\lambda[\cdot|X]$ is a Gamma$(n\overline{X} + 1, n + \lambda)$ distribution. We conclude:

$$\Pi_{\hat{\lambda}^{EB}}[\cdot|X] = \text{Gamma}\left(n\overline{X} + 1, \frac{n\overline{X} + 1}{\overline{X}}\right).$$

We note that the posterior mean is $\overline{X}$ (using the expectation formula of a Gamma distribution recalled in TD2). It thus coincides here (as in the previous exercise) with the maximum likelihood estimator for θ in this model.