Compulsory HWA3, Bayesian Statistics

1. (2p) Suppose that we have independent data $X_i \mid \lambda_i \sim \text{Poisson}(\lambda_i)$, $i = 1, ..., n$. The prior of λ_i is $\lambda_i \sim \text{Gamma}(a, b)$ independently. However, we do not know which values a and b should take.

   (a) Suppose that we want to use the full Bayesian approach. Present such idea in detail (as many details as possible).
   Solution: In the full Bayesian approach, we need to assign priors to a and b. Suppose that we assign independent priors $\pi(a)$ and $\pi(b)$, such that their domains are both the positive real line. The posterior is

$$\begin{aligned}\pi(\lambda_1, ..., \lambda_n, a, b \mid x_1, ..., x_n) &\propto f(x_1, ..., x_n \mid \lambda_1, ..., \lambda_n)\left[\prod_{i=1}^{n} \pi(\lambda_i \mid a, b)\right] \pi(a)\, \pi(b)\,. \\ &= \left[\prod_{i=1}^{n} \frac{\lambda_i^{x_i}}{x_i!} \exp\{-\lambda_i\} \cdot \frac{b^a}{\Gamma(a)} \lambda_i^{a-1} \exp\{-b\lambda_i\}\right] \pi(a)\, \pi(b)\, \cdot\end{aligned}$$

   We cannot drop a nor b here because they are also treated as parameters. To sample from the posterior, we can apply MCMC to sample $(\lambda_1, ..., \lambda_n, a, b)$.

   (b) Suppose that we want to use the empirical Bayes approach now. Derive the expression of the estimator of $\{\lambda_i\}$. If you cannot obtain a closed expression, simplify your expression as much as possible.
   Solution: We start with treating a and b as known. The posterior distribution is

$$\pi(\lambda_1, ..., \lambda_n \mid x_1, ..., x_n) \quad \propto \quad \prod_{i=1}^{n} \left[f(x_i \mid \lambda_i)\, \pi(\lambda_i)\right].$$

   From this, we can show that $\lambda_i \mid x_i \sim \text{Gamma}(a + x_i,\ b + 1)$, with the posterior mean

$$\text{E}[\lambda_i \mid x_i] \quad = \quad \frac{a + x_i}{b + 1}$$

   as an estimator of λ_i.
   To estimate a and b from data, the marginal distribution of x is

$$f(x \mid a, b) = \int f(x, \lambda \mid a, b)\, d\lambda \quad = \quad \prod_{i=1}^{n} \int f(x_i, \lambda_i \mid a, b)\, d\lambda_i$$

   where the marginal distribution of x_i is

$$\begin{aligned} & f(x_i \mid a, b) = \int f(x_i, \lambda_i \mid a, b)\, d\lambda_i \\ = & \int \frac{\lambda_i^{x_i}}{x_i!} \exp\{-\lambda_i\} \cdot \frac{b^a}{\Gamma(a)} \lambda_i^{a-1} \exp\{-b\lambda_i\}\, d\lambda_i \\ = & \frac{b^a}{\Gamma(a)\, x_i!} \int \lambda_i^{a+x_i-1} \exp\{-(b+1)\lambda_i\}\, d\lambda_i = \frac{b^a}{\Gamma(a)\, x_i!} \frac{\Gamma(a + x_i)}{(b+1)^{a+x_i}}.\end{aligned}$$

Hence, the log marginal likelihood is

$$\begin{aligned}\log f(x \mid a, b) &= \sum_{i=1}^{n} \log\left[\int f(x_i, \lambda_i \mid a, b)\, d\lambda_i\right] \\ &= na \log b - n \log \Gamma(a) \\ &\quad + \sum_{i=1}^{n} \left[\log \Gamma(a + x_i) - \log(x_i!) - (a + x_i)\log(b + 1)\right].\end{aligned}$$

We maximize such marginal likelihood numerically and obtain the maximizer $\hat{a}$ and $\hat{b}$. Then, we treat our prior as $\lambda_i \sim \text{Gamma}\left(\hat{a}, \hat{b}\right)$. The empirical Bayes estimator of λ_i still uses the posterior mean (derived treating a and b as known) with the plug-in principle (replace a and b by their estimates), and obtain

$$\mathrm{E}\left[\lambda_i \mid x_i, \hat{a}, \hat{b}\right] = \frac{\hat{a} + x_i}{\hat{b} + 1}.$$

2. (2p) Suppose that X_1, ..., X_n are iid $\text{Gamma}\left(4, \theta^{-1}\right)$, where θ^{-1} is the rate parameter. Consider the prior

$$\pi(\theta) = \frac{b_0^{a_0}}{\Gamma(a_0)} \frac{1}{\theta^{a_0+1}} \exp\left\{-\frac{b_0}{\theta}\right\},$$

where $a_0 > 4$ and $b_0 > 4$. It is known that $\mathrm{E}[\theta] = \frac{b_0}{a_0 - 1}$ and $\mathrm{Var}[\theta] = \frac{b_0^2}{(a_0-1)^2(a_0-2)}$.

(a) Find the Bayes estimator of θ under the L_2 loss and the L_1 loss. If you cannot find an closed form expression, explain how you can obtain it or approximate it.
Solution: The posterior satisfies

$$\begin{aligned}\pi(\theta \mid x) \propto f(x \mid \theta)\,\pi(\theta) &\propto \frac{1}{\theta^{4n}} \exp\left\{-\frac{\sum_{i=1}^{n} x_i}{\theta}\right\} \cdot \frac{1}{\theta^{a_0+1}} \exp\left\{-\frac{b_0}{\theta}\right\} \\ &\propto \frac{1}{\theta^{a_0+4n+1}} \exp\left\{-\frac{1}{\theta}\left(b_0 + \sum_{i=1}^{n} x_i\right)\right\} \\ &\sim \text{InvGamma}\left(a_0 + 4n,\ b_0 + \sum_{i=1}^{n} x_i\right).\end{aligned}$$

The Bayes estimator of θ under L_2 loss is the posterior mean $\mathrm{E}[\theta \mid x] = \frac{b_0 + \sum_{i=1}^{n} x_i}{a_0 + 4n - 1}$. The Bayes estimator under L_1 loss is the posterior median, which can be approximated by sampling random numbers from $\text{InvGamma}\left(a_0 + 4n,\ b_0 + \sum_{i=1}^{n} x_i\right)$ and use the sample median to approximate the posterior median.
In order for the above statements to hold, we need an estimator with finite risk. If we take the ubiased estimator $\bar{X}/4$ of θ, it is easy to show that under the L_2 loss, the risk is the variance of the gamma distribution divided by n. Hence, the variance is finite. For the L_1 loss, we know that

$$\left(\mathrm{E}\left[\left|\frac{1}{4}\bar{X} - \theta\right| \mid \theta\right]\right)^2 \leq \mathrm{E}\left[\left(\frac{1}{4}\bar{X} - \theta\right)^2 \mid \theta\right] = \mathrm{Var}\left(\frac{1}{4}\bar{X}\right) < \infty.$$

(b) Consider the L_2 loss. Is the Bayes estimator admissible?
Solution: The posterior risk of an estimator δ satisfies

$$\int L(\theta, \delta)\,\pi(\theta \mid x)\, d\theta = \int L(\theta, \delta_B)\,\pi(\theta \mid x)\, d\theta + (\delta_B - \delta)^T (\delta_B - \delta).$$

Suppose that we can find a set A such that for any $x \in A$, we have $\delta_B(x) = \delta(x)$ and

$$\int L(\theta, \delta) \pi(\theta \mid x) \, d\theta \quad = \quad \int L(\theta, \delta_B) \pi(\theta \mid x) \, d\theta.$$

Then for $x \notin A$, we have

$$\int L(\theta, \delta) \pi(\theta \mid x) \, d\theta \quad > \quad \int L(\theta, \delta_B) \pi(\theta \mid x) \, d\theta,$$

since $(\delta_B - \delta)^T (\delta_B - \delta) \geq 0$ and the equality holds if and only if $\delta_B = \delta$ .This means that δ_B is unique expect on a set with zero probability. Because of such uniqueness, the Bayes estimator is admissible.

3. (6p) Suppose that we have an iid sample $X_1, \ldots, X_n \sim \text{Poisson}(\theta)$, where $\theta \sim \text{Gamma}(a_0, b_0)$ and $a_0 > 1$. The posterior distribution is $\theta \mid x \sim \text{Gamma}\left(a_0 + \sum_{i=1}^{n} x_i, \ n + b_0\right)$.

   (a) Suppose that there exists a decision rule with finite risk. Show that $\bar{X}$ is the minimax estimator of θ under the loss function $L(\theta, d) = (d - \theta)^2 / \theta$.
   Solution: The posterior risk is

$$\mathrm{E}[L(\theta, \delta) \mid X = x] \quad = \quad \mathrm{E}\left[\left(\delta - \theta^2\right) / \theta \mid X = x\right] = \delta^2 \mathrm{E}\left[\frac{1}{\theta} \mid X = x\right] - 2\delta + \mathrm{E}[\theta \mid X = x],$$

which is minimized at

$$\hat{\delta}(x) \quad = \quad \frac{1}{\mathrm{E}[\theta^{-1} \mid X = x]} = \frac{a_0 - 1 + n\bar{x}}{b_0 + n}.$$

since

$$\begin{aligned}
\mathrm{E}\left[\theta^{-1} \mid X = x\right] &= \int_0^\infty \frac{1}{\theta} \cdot \frac{(b_0 + n)^{a_0 + n\bar{x}}}{\Gamma(a_0 + n\bar{x})} \theta^{a_0 + n\bar{x} - 1} \exp\left\{-(b_0 + n)\theta\right\} d\theta \\
&= \frac{(b_0 + n)^{a_0 + n\bar{x}} \, \Gamma(a_0 + n\bar{x} - 1)}{(b_0 + n)^{a_0 + n\bar{x} - 1} \, \Gamma(a_0 + n\bar{x})} = \frac{b_0 + n}{a_0 + n\bar{x} - 1}.
\end{aligned}$$

Its frequentist risk is

$$\begin{aligned}
R\left(\theta, \hat{\delta}(X)\right) &= \mathrm{E}\left[\frac{\left(\theta - \frac{a_0 - 1 + n\bar{x}}{b_0 + n}\right)^2}{\theta} \mid \theta\right] = \mathrm{E}\left[\frac{\left(\theta - \frac{a_0 - 1}{b_0 + n} - \frac{n}{b_0 + n}\bar{x}\right)^2}{\theta} \mid \theta\right] \\
&= \frac{1}{\theta}\left[\left(\theta - \frac{a_0 - 1}{b_0 + n}\right)^2 - 2\left(\theta - \frac{a_0 - 1}{b_0 + n}\right)\frac{n}{b_0 + n}\theta + \frac{n^2}{(b_0 + n)^2}\left(\frac{\theta}{n} + \theta^2\right)\right] \\
&= \frac{1}{\theta}\left(\theta - \frac{a_0 - 1}{b_0 + n}\right)^2 - 2\left(\theta - \frac{a_0 - 1}{b_0 + n}\right)\frac{n}{b_0 + n} + \frac{n^2}{(b_0 + n)^2}\left(\frac{1}{n} + \theta\right) \\
&= \frac{1}{\theta}\frac{(a_0 - 1)^2}{(b_0 + n)^2} + \left[\frac{n^2}{(b_0 + n)^2} + \frac{b_0 - n}{b_0 + n}\right]\theta + \frac{2na_0 - n}{(b_0 + n)^2} - 2\frac{a_0 - 1}{b_0 + n}.
\end{aligned}$$

Under the gamma prior, the integrated risk is

$$\begin{aligned}
\int R\left(\theta, \hat{\delta}(X)\right) \pi(\theta) \, d\theta &= \frac{b_0}{a_0 - 1}\frac{(a_0 - 1)^2}{(b_0 + n)^2} + \left[\frac{n^2}{(b_0 + n)^2} + \frac{b_0 - n}{b_0 + n}\right]\frac{a_0}{b_0} + \frac{2na_0 - n}{(b_0 + n)^2} - 2\frac{a_0 - 1}{b_0 + n} \\
&= \frac{(a_0 - 1) b_0 + (2a_0 - 1) n}{(b_0 + n)^2} + \left[\frac{n^2}{(b_0 + n)^2} + \frac{b_0 - n}{b_0 + n}\right]\frac{a_0}{b_0} - 2\frac{a_0 - 1}{b_0 + n}.
\end{aligned}$$

We take $b_0 \to n/m$, then

$$\int R\left(\theta, \hat{\delta}(X)\right) \pi(\theta)\, d\theta = \frac{(2a_0 - 1)\, m^{-1} + (2a_0 - 1)}{(m^{-1}+1)^2\, n} - \frac{2a_0 - 2}{(m^{-1}+1)\, n}.$$

As $m \to \infty$, we have

$$\int R\left(\theta, \hat{\delta}(X)\right) \pi(\theta)\, d\theta \to \frac{1}{n}.$$

The frequentist risk of $\hat{\theta} = \bar{X}$ is

$$R(\theta, \bar{X}) = \mathrm{E}\left[\frac{(\theta - \bar{X})^2}{\theta} \mid \theta\right] = \frac{\mathrm{E}\left[(\theta - \bar{X})^2 \mid \theta\right]}{\theta} = \frac{1}{n}.$$

Hence, $\bar{X}$ is minimax.

(b) By Doob's theorem, what can you establish regarding consistency of posterior distribution?
Solution: It is easy to see that the Poisson model is identified, since if we have

$$\frac{\theta_1^x}{x!} \exp(-\theta_1) = \frac{\theta_2^x}{x!} \exp(-\theta_2) \text{ for all } x,$$

we must have $x \log \theta_1 - \theta_1 = x \log \theta_2 - \theta_2$. For $x = 0$, we already need to $\theta_1 = \theta_2$. Hence, the Doob's theorem says that there exists a Θ_0 with prior probability $\mathrm{P}(\Theta_0) = 1$ such that, for every $\theta_0 \in \Theta_0$,

$$\lim_{n\to\infty} \mathrm{P}(\theta \in O \mid X_1, ..., X_n) = 1, \text{ almost surely under } \mathrm{P}_{\theta_0}$$

for any open set O with $\theta_0 \in O$.

(c) By Doob's theorem, what can you establish regarding consistency of posterior mean?
Solution: We have shown that the Poisson model is identified. Note also that, under the Gamma prior,

$$\int \theta \pi(\theta)\, d\theta = \frac{a_0}{b_0} < \infty.$$

Hence, the Doob's theorem says that there exists a Θ_0 with prior probability $\mathrm{P}(\Theta_0) = 1$ such that, for every $\theta_0 \in \Theta_0$,

$$\lim_{n\to\infty} \mathrm{E}[\theta \mid X_1, ..., X_n] = \theta_0, \text{ almost surely under } \mathrm{P}_{\theta_0}.$$

(d) Approximate the posterior distribution by Bernstein-Von Mises theorem.
Solution: The log-likelihood is

$$\log f(x \mid \theta) = \sum_{i=1}^n x_i \log(\theta) - \log\left(\prod_{i=1}^n x_i!\right) - n\theta.$$

The Fisher information is

$$\mathcal{I}(\theta) = -\frac{n}{\theta}.$$

It means that the posterior is approximately

$$\sqrt{n}\left(\theta - \hat{\theta}\right) \mid \approx N\left(0, \hat{\theta}\right)$$

This is equivalent as saying

$$\theta \mid x \approx N\left(\hat{\theta}, \frac{\hat{\theta}}{n}\right).$$

(e) Approximate the posterior probability $\mathrm{P}(\theta \geq 1 \mid x)$.
Solution: $\mathrm{P}(\theta \geq 1 \mid x) \approx \mathrm{P}\left(N\left(\hat{\theta}, \frac{\hat{\theta}}{n}\right) \geq 1 \mid x\right)$.