Foundations of Statistical Inference, BS2a, Exercises 3

1. *The number of phone calls a man receives in a week follows a Poisson distribution with mean θ. At the start of week 1, the man's opinion about the value of θ corresponds to the gamma distribution*

$$\pi(\theta) = \frac{1}{54}\theta^2 e^{-\theta/3},\ \theta > 0.$$

In the 4 weeks following the start of week 1, the man received 3, 7, 6, and 10 phone calls, respectively. Determine the posterior distribution of θ and the predictive distribution for the number of calls that he will receive in week 5.

The posterior is proportional to

$$\theta^2 e^{-\theta/3}\theta^{3+7+6+10}e^{-4\theta} = \theta^{28}e^{-(13/3)\theta}$$

The posterior has a Gamma $(29, 13/3)$ distribution. The predictive distribution for week 5's calls (x say) is

$$\int_0^\infty \frac{\theta^x}{x!}e^{-\theta}\frac{1}{28!}\left(\frac{13}{3}\right)^{29}\theta^{28}e^{-(13/3)\theta}d\theta = \frac{(28+x)!}{(28)!x!}\left(\frac{13}{16}\right)^{29}\left(\frac{3}{16}\right)^x, x = 0, 1, \ldots$$

2 *Consider a vector of observations X with sampling model $X|\theta \sim f(\cdot|\theta)$ with $\theta \in \mathbb{R}$ and a prior distribution with density π (wrt Lebesgue measure). Consider the loss function for estimating θ*

$$L(\theta, \delta) = h(\theta - \delta), \quad h(u) = e^u - u - 1$$

1. *Show that for all $u \in \mathbb{R}$ $h(u) \geq 0$ and show that h is a convex function.*
2. *Determine the form of the Bayesian estimator and prove that it is almost surely unique as soon as*

$$\int_{\mathbb{R}} [e^\theta + |\theta|]\pi(\theta|x)d\theta < +\infty$$

3. *Assume that $X = (X_1, \cdots, X_n)$ with $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ with prior $\pi(\mu, \sigma) \propto 1/\sigma$, $\theta = (\mu, \sigma^2)$*

   Compute the Bayesian estimator under the loss function

$$L(\theta, \delta) = e^{\mu-\delta} - \mu + \delta - 1 \quad \delta \in \mathbb{R}$$

   Assume that σ is known and do the same thing.

$$h'(u) = e^u - 1 > 0 \quad \text{iff} \quad u > 0$$

Hence $h(u) \geq h(0) = 0$. Moreover $h''(u) = e^u > 0$ so that h is convex.

Posterior risk

$$\rho(\pi, \delta|X) = \int_{\mathbb{R}} (e^{\theta-\delta} - \theta + \delta - 1)\pi(\theta|X)d\theta$$

it is differentiable in δ and

$$\frac{\partial \rho(\pi, \delta|X)}{\partial \delta} = -e^{-\delta}E[e^\theta|X] + 1 = 0 \quad \Leftrightarrow \quad \delta = \log\left(E[e^\theta|X]\right)$$

By convexity of the loss function this is the unique minimizer of the posterior risk when the latter exists, which holds under the given conditions.

If $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ and $\pi(\mu, \sigma) \propto 1/\sigma$ then the posterior distribution on (μ, σ^2) satisfies

$$\pi(\mu, \sigma^2|X) \propto e^{-\frac{n(\mu - \bar{X}_n)^2 + nS_n^2}{2\sigma^2}} \sigma^{-n-1}$$

we can integrate σ (Inverse Gamma$((n-1)/2, (n(\mu - \bar{X}_n)^2 + S_n^2)/2)$)

$$\pi(\mu|X) \propto [n(\mu - \bar{X}_n)^2 + nS_n^2]^{-(n-1)/2}$$

This is a Student distribution. This posterior does not allow for exponential moments and therefore the posterior risk does not exist. Assume now that σ^2 is known then

$$\pi(\mu|X) \propto e^{-\frac{n(\mu - \bar{X}_n)^2}{2\sigma^2}} \equiv \mathcal{N}(\bar{X}_n, \sigma^2/n)$$

then

$$E[e^\mu|X] = e^{\bar{X}_n} E[e^{Z\sigma/\sqrt{n}}|Z \sim \mathcal{N}(0,1)] = e^{\bar{X}_n + \frac{\sigma^2}{2n}}$$

and

$$\delta(X) = \bar{X}_n + \frac{\sigma^2}{2n}$$

3. *In order to measure the intensity, θ, of a source of radiation in a noisy environment a measurement X_1 is taken without the source present and a second, independent measurement X_2 is taken with it present. It is known that X_1 is $N(\mu, 1)$ and X_2 is $N(\mu+\theta, 1)$, where μ is the mean noise level. (a) If the prior distribution for μ is $N(\mu_0, 1)$ while the prior for θ is constant.*

- *(i) Show that the posterior is almost surely defined and write down the joint posterior distribution of μ and θ up to a constant of proportionality.*
- *(ii) Hence obtain the posterior marginal distribution of θ.*
- *(iii) The usual estimate of θ is $x_2 - x_1$ explain why $\frac{1}{2}(2x_2 - x_1 - \mu_0)$ might be better.*

(i)

$$\begin{aligned}&\int_{\mathbb{R}^2} e^{-\frac{(X_1-\mu)^2}{2}-\frac{(\mu-\mu_0)^2}{2}} e^{-(X_2-\mu-\theta)^2/2} d\theta d\mu \\ \propto &\int_{\mathbb{R}} e^{-\frac{(X_1-\mu)^2}{2}-\frac{(\mu-\mu_0)^2}{2}} d\mu < +\infty\end{aligned}$$

for all X_1 so that the posterior is well defined.

Joint posterior :

$$\begin{aligned}\pi(\mu,\theta|X_1,X_2) &\propto e^{-\frac{(X_1-\mu)^2}{2}-\frac{(\mu-\mu_0)^2}{2}} e^{-(X_2-\mu-\theta)^2/2} \\ &\propto e^{-\frac{2}{2}\left(\mu-\frac{\mu_0+X_1}{2}\right)^2-\frac{1}{2}(X_2-\mu-\theta)^2}\end{aligned}$$

so that

$$\theta|\mu,X_1,X_2 \sim \mathcal{N}(X_2-\mu,1), \quad \mu|X_1,X_2 \sim \mathcal{N}((\mu_0+X_1)/2,1/2)$$

(ii) The marginal posterior distribution on θ is therefore Gaussian with mean

$$E[\theta|X_1,X_2] = X_2 - \frac{\mu_0+X_1}{2}, \quad V[\theta|X_1,X_2] = 1/2+1 = 3/2$$

Alternative proof:

The posterior of (μ,θ) is

$$q(\mu,\theta;x_1,x_2) \propto \exp\left\{-\frac{1}{2}(\mu-\mu_0)^2 - \frac{1}{2}(x_1-\mu)^2 - \frac{1}{2}(x_2-\mu-\theta)^2\right\}$$

The posterior distribution of θ is proportional to the integral of $q(\mu,\theta;x_1,x_2)$ over μ. Completing the square in the the exponent, first in μ, then in what is left in θ,

$$\begin{aligned}
&[\mu-\mu_0]^2+[x_1-\mu]^2+[x_2-\mu-\theta]^2\\
&\quad=3\mu^2-2\mu(\mu_0+x_1+x_2-\theta)+\theta^2-2\theta x_2+\dots\\
&\quad=3(\mu-(\mu_0+x_1+x_2-\theta)/3)^2-\frac{1}{3}(\mu_0+x_1+x_2-\theta)^2+\theta^2-2\theta x_2+\dots\\
&\quad=3(\mu-(\mu_0+x_1+x_2+\theta)/3)^2+\frac{2}{3}\theta^2-2\theta x_2+\frac{2}{3}\theta(\mu_0+x_1+x_2)+\dots\\
&\quad=3[\mu-(\mu_0+x_1+x_2-\theta)/3]^2+(2/3)[\theta-(2x_2-x_1-\mu_0)/2]^2+\text{const}
\end{aligned}$$

From this factorization we see that the posterior distribution of θ is $\mathrm{N}\big((2x_2-x_1-\mu_0)/2, 3/2\big)$

The posterior mean is $(2x_2-x_1-\mu_0)/2$.

From a Bayesian point of view this is the Bayesian estimator associated to the quadratic loss function and thus is optimal in this regard.

(b) Compute Jeffrey's prior on (μ, θ) and study the propriety of the posterior distribution.

Jeffrey's for a univariate Gaussian is $\pi(\mu)=1$ under the parameterization (μ_1, μ_2) by independence

$$\pi(\mu_1, \mu_2)=1$$

by invariance to reparameterization and because the Jacobian of the transformation $(\theta, \mu)\to(\mu, \mu+\theta)$ is 1

$$\pi_J(\mu, \theta)=1$$

(c) Compute the frequentist risk, the posterior risk and the integrated risk under Jeffrey's prior and derive the Bayesian estimator associated to the quadratic loss function. Posterior distribution associated to Jeffreys on μ_1, μ_2 is $\mathcal{N}(X_1, 1)\otimes\mathcal{N}(X_2, 1)$

$$\pi(\mu, \theta)\propto e^{-(X_1-\mu)^2/2-(X_2-\mu-\theta)^2/2}$$

so that the marginal posterior distribution of $\theta=\mu_2-\mu_1$ is $\mathcal{N}(X_2-X_1, 2)$

$$\hat{\theta}=X_2-X_1$$

frequentist risk

$$R(\mu_1, \mu_2, \hat{\theta})=V(\hat{\theta}; \theta, \mu_1)=2$$

posterior risk

$$\rho(\pi, \hat{\theta}|X) = E[(\theta - \hat{\theta})^2|X_1, X_2] = V(\theta|X_2, X_1) = 2$$

Integrated risk

$$r(\pi, \hat{\theta}) = \int_{\mathbb{R}^2} 2 d\mu d\theta = +\infty$$

4. *Let $X|\theta \sim f(\cdot|\theta)$ with $\theta \in \Theta \subset \mathbb{R}^d$. Let π be an improper prior.*

(a) *Show that*

$$m(x) = \int_\Theta f(x|\theta)\pi(\theta)d\theta$$

is improper as a measure on $\mathcal{X}$.

(b) *Assume that $X \in \mathcal{X}$ where card$(\mathcal{X}) < +\infty$ i.e. $\mathcal{X}$ is finite. Show that there exists $x \in \mathcal{X}$ such that*

$$\int_\Theta P[X = x|\theta]\pi(\theta)d\theta = +\infty$$

(a) We need to show that

$$\int_\mathcal{X} m(x)dx = +\infty$$

By Fubini's Theorem because all functions are non negative we can write

$$\int_\mathcal{X} m(x)dx = \int_\Theta \int_\mathcal{X} f(x|\theta)dx\pi(\theta)d\theta = \int_\Theta \pi(\theta)d\theta = +\infty$$

Note that the above result holds true if X is discrete where $m(x) = \int_\Theta P[X = x; \theta]\pi(\theta)d\theta$.

$$\sum_\mathcal{X} m(x) = \int_\Theta \left(\sum_\mathcal{X} P[X = x; \theta]\right)\pi(\theta)d\theta.$$

(b) Using question 1 we have that

$$\sum_\mathcal{X} m(x) = +\infty$$

and $\mathcal{X}$ is finite hence there exists $x_0 \in \mathcal{X}$ such that $m(x_0) = +\infty$.

5 Let $(X_1, \cdots, X_n) \overset{iid}{\sim} f(\cdot|\theta)$, $\theta \in \Theta$ where $f(\cdot|\theta)$ is a canonical exponential family

$$f(x|\theta) = w(x)e^{\theta^T B(x) - D(\theta)}, \quad \Theta \subset \mathbb{R}^d$$

Let π be a prior density on Θ with respect to Lebesque measure.
(a) Show that the posterior distribution of θ depends only on $T_n = \sum_{i=1}^n B(X_i)$. Show that this result holds true outside exponential family if T_n is any sufficient statistics for θ.

The likelihood is

$$L_n(\theta) = \prod_i w(X_i)e^{\theta^T T_n - nD(\theta)}$$

so that the posterior is

$$\pi(\theta|X) \propto \pi(\theta)e^{\theta^T T_n - nD(\theta)}$$

thus it depends only on T_n.

Assume that T_n is a sufficient statistic. Then by factorization criterion for some f_1 and h we can write

$$\pi(\theta|X) \propto f(X|\theta)\pi(\theta) = f_1(T_n;\theta)h(x)\pi(\theta) \propto f_1(T_n;\theta)\pi(\theta)$$

thus posterior depends only on T_n.
(b) Let $E(a, b)$ be the distribution of the shifted exponential with density

$$\frac{1}{b}e^{-(x-a)/b}, \qquad x > a$$

where $a \in \mathbb{R}, b > 0$ are parameters. Let $X_1, \ldots, X_n$ be a random sample from the distribution $E(a, b)$.

- *(i) If a is known, derive Jeffrey's prior on b and compute the posterior mean . Show that it is MVUE and attains the Cramer Rao lower bound.*

  If a is known the model is regular and Jeffreys's prior exists. setting $Y = X - a$ then we have an exponential distribution

  $$\ell(b) = -Y/b - \log b, \quad \nabla\ell(b) = Y/b^2 - 1/b, \quad \frac{d^2\ell(b)}{db^2} = -\frac{2Y}{b^3} + \frac{1}{b^2}$$

  so

  $$i(b) = b^{-2}, \quad \pi_J(b) = 1/b$$

Then the posterior is

$$\pi(b|X) \propto e^{-\sum_{i=1}^n Y_i/b} b^{-n-1}$$

which is the density of an $IG(n, \sum_i Y_i)$. The posterior mean is then given by

$$\hat{b} = E[b|X] = \frac{\sum_{i=1}^n (X_i - a)}{n-1}, \qquad \frac{(n-1)\hat{b}}{n} = \tilde{b}$$

$\tilde{b}$ is an unbiased estimator, which is a function of the complete sufficient statistics $\sum_{i=1}^n (X_i - a)$ so it is the MVUE. We have

$$V(\tilde{b}) = V(X_1)/n = b^2 = 1/i(b)$$

so it attains the CR lower bound.

- (ii) *If a is known and a IGamma(α, β) prior is chosen on b , find the posterior mean. Is it an MVUE? Compute the predictive distribution of X_{n+1} .*

  Note that instead of the Gamma we consider the Inverse Gamma, i.e. there was a typo in the exercise sheet.

  Then due to conjugacy

$$\pi(b|X) \propto e^{-\sum_{i=1}^n Y_i/b - \beta/b} b^{-n-1-\alpha} \equiv IG(n+\alpha, \sum_{i=1}^n Y_i + \beta)$$

  and the posterior mean is given by

$$\tilde{b} = \frac{\sum_{i=1}^n (X_i - a) + \beta}{\alpha + n - 1}.$$

  It is not an MVUE as it is biased.

  Predictive posterior distribution is:

$$\begin{aligned} f(x_{n+1}|x_1, \cdots, x_n) &= \mathbf{1}_{x_{n+1} \geq a} \int_0^\infty e^{-(x_{n+1}-a)/b} b^{-1} \pi(b|x_1, \cdots, x_n) db \\ &= \mathbf{1}_{x_{n+1} \geq a} \int_0^\infty e^{-[\sum_{i=1}^{n+1}(x_i - a) + \beta]/b} b^{-\alpha-n-2} db \frac{(\sum_{i=1}^n (x_i - a) + \beta)^{\alpha+n}}{\Gamma(\alpha+n)} \\ &= \mathbf{1}_{x_{n+1} \geq a} \frac{\Gamma(\alpha+n+1)}{\Gamma(\alpha+n)} \frac{(\sum_{i=1}^n (x_i - a))^{\alpha+n}}{\left(\sum_{i=1}^{n+1} (x_i - a)\right)^{\alpha+n+1}} \\ &= \mathbf{1}_{x_{n+1} \geq a} (\alpha+n) \frac{(\sum_{i=1}^n (x_i - a))^{\alpha+n}}{\left(\sum_{i=1}^{n+1} (x_i - a)\right)^{\alpha+n+1}}. \end{aligned}$$

- (iii) *If b is known and a prior density π_a on a is considered. Show that the posterior distribution on b depends only on $X_{(1)} = \min X_i$, compute the posterior mean associated with the prior $\pi(a) \propto 1$. Show that the posterior is defined and compute it.*

  If b is known the likelihood is

$$L_n(a) = \mathbf{1}_{X_{(1)} \geq a} e^{an/b} e^{-\sum_i X_i/b}$$

  so that $X_{(1)}$ is a sufficient statistics and by the above question the posterior depends only on $X_{(1)}$.

$$\pi(a|X) \propto \mathbf{1}_{X_{(1)} \geq a} e^{an/b}$$

  and

$$\int_{-\infty}^{X_{(1)}} e^{an/b} da = \frac{b}{n} e^{X_{(1)} n/b} < +\infty$$

  so the posterior is defined and is a truncated negative exponential.

- (iv) *Show that for all $X_{(1)}$ and all $b > 0$ $\Pi(n|a - X_{(1)}| > z|X_1, \cdots, X_n)$ goes to zero as z goes to infinity uniformly in n. Compute the posterior mean of a and its frequentist risk.*

$$\Pi(n|a - X_{(1)}| > z|X_1, \cdots, X_n) = \Pi(n(a - X_{(1)}) \leq -z|X) = \int_{-\infty}^{-z/n + X_{(1)}} \pi(a|X) da$$
$$= e^{n(-z/n + X_{(1)})/b} e^{-X_{(1)} n/b} = e^{-z/b} \to 0$$

  as z goes to infinity uniformly in n as it does not depend on n. The posterior distribution of $n(a - X_{(1)})$ is negative Exponential.

  Posterior mean

$$\hat{a} = E[a|X] = X_{(1)} + \frac{1}{n} E[n(a - X_{(1)})|X] = X_{(1)} - \frac{b}{n}$$

  Frequentist risk

$$R(a, b, \hat{a}) = E[(a - \hat{a})^2; a, b] = V(X_{(1)}) = \frac{b^2}{n^2}$$

- (v) *Suppose now that both a and b are unknown. Define $T_1(X) = X_{(1)}$ and $T_2(X) = \sum_{i=1}^n \left(X_i - X_{(1)}\right)$ and show that the posterior distribution depends only on (T_1, T_2). Show that the family of priors $\pi(a, b) = \pi(a|b)\pi(b)$ with*

  $$\pi(a|b) \propto e^{\alpha(a-\beta)/b}\mathbf{1}_{a<z}, \quad \pi(b) \propto b^{-m-1}e^{-c/b}\mathbf{1}_{b>0}$$

  is conjugate to this model.

  The likelihood can be written as

  $$\begin{aligned} L_n(a,b) &= \mathbf{1}_{X_{(1)}\geq a}b^{-n}e^{-\sum_{i=1}^n(X_i-a)/b} \\ &= \mathbf{1}_{X_{(1)}\geq a}b^{-n}e^{-\sum_{i=1}^n(X_i-X_{(1)})/b}e^{-n(X_{(1)}-a)/b} \\ &= \mathbf{1}_{a\leq T_1}b^{-n}e^{-T_2/b}e^{-n(T_1-a)/b}. \end{aligned}$$

  so that (T_1, T_2) is a sufficient statistic for (a, b) and the posterior distribution depends only on them. The specified family of priors is

  $$\pi(a,b) \propto \mathbf{1}_{a<z}\mathbf{1}_{b>0}b^{-m-1}e^{(\alpha a-\gamma)/b},$$

  where we wrote $\gamma = \alpha\beta + c$. Under the specified family of priors we have

  $$\begin{aligned} \pi(a,b|T_1,T_2) &\propto \mathbf{1}_{a\leq T_1}b^{-n}e^{(na-nT_1-T_2)/b}\mathbf{1}_{a<z}\mathbf{1}_{b>0}b^{-m-1}e^{(\alpha a-\gamma)/b} \\ &\propto \mathbf{1}_{a<\min\{z,T_1\}}\mathbf{1}_{b>0}b^{-m-n-1}e^{((\alpha+n)a-nT_1-T_2-\gamma)/b} \end{aligned}$$

  so it belongs to the same family.