MAST30027: Modern Applied Statistics

Week 8 Exercise Sheet

1. Let $X_1, \cdots, X_n$ be a random sample from a $N(\theta, \sigma^2)$ population, and suppose that the prior distribution on θ is $N(\mu, \tau^2)$. Here we assume that σ^2, μ and τ^2 are all known.

   Last week we showed that the posterior pdf of θ, denoted as $p(\theta|\mathbf{x}, \sigma^2, \mu, \tau^2)$, is normal with mean and variance given by $E(\theta|\mathbf{x}) = \frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x} + \frac{\sigma^2}{n\tau^2+\sigma^2}\mu$ and $\mathrm{Var}(\theta|\mathbf{x}) = \frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}$.

   Considering the squared error loss, find the Bayes estimators of θ and $\theta(\theta-1)$.

   Solution: Bayes estimator $\hat{\theta} = E(\theta|\bar{x}, \sigma^2, \mu, \tau^2) = \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2} = \frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x} + \frac{\sigma^2}{n\tau^2+\sigma^2}\mu$. Bayes estimator of $\gamma = \theta(\theta-1)$ is

   $\hat{\gamma} = E(\theta^2|\bar{x}, \sigma^2, \mu, \tau^2) - E(\theta|\bar{x}, \sigma^2, \mu, \tau^2) = \frac{\sigma^2\tau^2}{n\tau^2+\sigma^2} + \left(\frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2 - \frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}$.

2. Consider a random sample X from a Bernoulli distribution with pdf $f(x|\theta) = \theta^x(1-\theta)^{1-x}$. We consider Beta$(1,2)$ as a prior distribution for θ, where Beta$(a,b) \equiv \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$ and Beta(a,b) distribution has mean $= a/(a+b)$ and variance $= ab/[(a+b)^2(a+b+1)]$.

   Under the squared error loss function,

   (a) Find the Bayes estimator of θ.
   (b) Find the risk of the Bayes estimator of θ.
   (c) Find the Bayes risk of the Bayes estimator of θ.

   Solution:

   (a) Under the squared error loss function, the Bayes estimator of θ is a posterior mean. The posterior distribution of θ is

   $$\begin{aligned} f(\theta|x) &\propto P(x|\theta)f(\theta) \\ &\propto \theta^x(1-\theta)^{1-x}\theta^{1-1}(1-\theta)^{2-1} \\ &\propto \theta^{x+1-1}(1-\theta)^{1-x+2-1} \end{aligned}$$

   This is a kernel for Beta$(x+1, 3-x)$ distribution, which has mean $= \frac{x+1}{4}$.

   (b) The risk of the Bayes estimator of θ, $T = \frac{X+1}{4}$, is

   $$\begin{aligned} R_T(\theta) &= E[L(T;\theta)] = E[(\frac{X+1}{4} - \theta)^2] \\ &= \frac{1}{16}\big[E(X^2) + 2(1-4\theta)E(X) + (1-4\theta)^2\big] \\ &= \frac{1}{16}\big[Var(X) + E(X)^2 + 2(1-4\theta)E(X) + (1-4\theta)^2\big] \\ &= \frac{1}{16}\big[\theta(1-\theta) + \theta^2 + 2(1-4\theta)\theta + (1-4\theta)^2\big] \\ &= \frac{1}{16}\big[8\theta^2 - 5\theta + 1\big]. \end{aligned}$$

   (c) The Bayes risk of the Bayes estimator of θ, $T = \frac{X+1}{4}$, is

   $$\begin{aligned} E_\theta[R_T(\theta)] &= E_\theta\big[\frac{1}{16}(8\theta^2 - 5\theta + 1)\big] \\ &= \frac{1}{16}\big[8E(\theta^2) - 5E(\theta) + 1\big] \\ &= \frac{1}{16}\big[8(Var(\theta) + E(\theta)^2) - 5E(\theta) + 1\big] \\ &= \frac{1}{16}\big[8(\frac{1}{18} + \frac{1}{9}) - \frac{5}{3} + 1\big] \\ &= \frac{1}{24}. \end{aligned}$$

We can compute $E(\theta) = \frac{1}{3}$ and $Var(\theta) = \frac{1}{18}$ from $\theta \sim \text{Beta}(1, 2)$.

3. Let $X_1, \cdots, X_n$ be a random sample from an exponential distribution with pdf $f(x|\theta) = \theta e^{-\theta x}$, $x > 0$. (So the mean of X_i is $\frac{1}{\theta}$ instead of θ.) Assume the prior pdf of θ is $p(\theta) = \beta e^{-\beta\theta}$ where β is known.

   (a) Show that the posterior pdf of θ given $\mathbf{x} = (x_1, \cdots, x_n)$ is gamma$[(\beta + \sum_{i=1}^n x_i)^{-1}, n + 1]$.
   Solution: Note that in this question we are parameterising the gamma distribution using *scale* and shape parameters (rather than *rate* and shape). The joint pdf of $\mathbf{X} = (X_1, \cdots, X_n)$ and θ is

$$f(\mathbf{x}, \theta) = f(\mathbf{x}|\theta)p(\theta) = (\prod_{i=1}^n \theta e^{-\theta x_i})\beta e^{-\beta\theta} = \beta\theta^n \exp\{-(\beta + \sum_{i=1}^n x_i)\theta\}.$$

   The posterior pdf of θ given $\mathbf{x}$ is

$$\begin{aligned} p(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}, \theta)}{f(\mathbf{x})} = \frac{f(\mathbf{x}|\theta)p(\theta)}{\int_0^\infty f(\mathbf{x}|\theta)p(\theta)d\theta} = \frac{\beta\theta^n \exp\{-(\beta + \sum_{i=1}^n x_i)\theta\}}{\int_0^\infty \beta\theta^n \exp\{-(\beta + \sum_{i=1}^n x_i)\theta\}d\theta} \\ &= \frac{\theta^n \exp\{-(\beta + \sum_{i=1}^n x_i)\theta\}}{\int_0^\infty \theta^n \exp\{-(\beta + \sum_{i=1}^n x_i)\theta\}d\theta} \\ &= \frac{\theta^n e^{-(\beta + \sum_{i=1}^n x_i)\theta}}{[(\beta + \sum_{i=1}^n x_i)^{-1}]^{n+1}\Gamma(n + 1)}, \end{aligned}$$

   noting that $\int_0^\infty \theta^n e^{-(\beta + \sum_{i=1}^n x_i)\theta}d\theta = [(\beta + \sum_{i=1}^n x_i)^{-1}]^{n+1}\Gamma(n + 1)$ by comparing it with a gamma(a, b) pdf. It is easy to see that the posterior pdf is gamma$[(\beta + \sum_{i=1}^n x_i)^{-1}, n + 1]$.

   (b) Using squared error loss, find the Bayes estimator of θ.
   Solution: The mean of a gamma(a, b) pdf is ab. The Bayes estimator of θ is $\hat{\theta} = E(\theta|\mathbf{x}) = (n + 1)(\beta + \sum_{i=1}^n x_i)^{-1}$.

   (c) Using squared error loss, find the Bayes estimator of $\mu = \frac{1}{\theta}$. Then find the risk function of this estimator. Further attempt the associated Bayes risk to see whether it is finite or not. If not finite, what are the possible reason(s) behind this? (*Hint*: no unique answers; think about the appropriateness of the prior.)
   Solution: The Bayes estimator using squared error loss is the posterior mean. The posterior mean of $\mu = \frac{1}{\theta}$ is

$$\begin{aligned} \hat{\mu} &= E(\theta^{-1}|\mathbf{x}) = \int_0^\infty \theta^{-1} \cdot \frac{\theta^n e^{-(\beta + \sum_{i=1}^n x_i)\theta}}{[(\beta + \sum_{i=1}^n x_i)^{-1}]^{n+1}\Gamma(n + 1)}d\theta \\ &= \frac{[(\beta + \sum_{i=1}^n x_i)^{-1}]^n\Gamma(n)}{[(\beta + \sum_{i=1}^n x_i)^{-1}]^{n+1}\Gamma(n + 1)} = \frac{\beta + \sum_{i=1}^n x_i}{n} = \frac{\beta}{n} + \bar{x}. \end{aligned}$$

   Note that μ is the population mean of the sample, so our estimator $\hat{\mu}$ is the sample mean adjusted by adding a bias $\frac{\beta}{n}$ which is related to the prior mean (β^{-1}).
   The risk function is

$$\begin{aligned} R_{\hat{\mu}}(\theta) &= E[(\hat{\mu} - \mu)^2|\theta] \\ &= E[(\frac{\beta}{n} + \bar{X} - \frac{1}{\theta})^2|\theta] \\ &= \frac{\beta^2}{n^2} + \frac{2\beta}{n}E(\bar{X} - \frac{1}{\theta}|\theta) + E[(\bar{X} - \frac{1}{\theta})^2|\theta] \\ &= \frac{\beta^2}{n^2} + \text{Var}(\bar{X}|\theta) \\ &= \frac{\beta^2}{n^2} + \frac{1}{n\theta^2} \end{aligned}$$

   knowing that $\bar{X} \stackrel{d}{=} \text{gamma}((n\theta)^{-1}, n)$ given θ and accordingly $E(\bar{X}_n|\theta) = n(n\theta)^{-1} = \frac{1}{\theta}$ and $\text{Var}(\bar{X}_n|\theta) = n(n\theta)^{-2} = \frac{1}{n\theta^2}$.

The Bayes risk is

$$\begin{aligned} E(R_{\hat{\mu}}(\theta)) &= \int_0^\infty (\frac{\beta^2}{n^2} + \frac{1}{n\theta^2})\beta e^{-\beta\theta} d\theta \\ &= \infty. \end{aligned}$$

The Bayes estimator is supposed to minimise the Bayes risk, which suggests that the Bayes risk is infinite for any estimator of μ, making it a poor way of choosing an estimator. None-the-less we have a result that tells us the posterior mean minimises the Bayes risk, and in this case the posterior of $\mu = 1/\theta$ does have a well defined mean. The seeming contradiction arises because our theorem implicitly assumes the Bayes risk is finite (and so has a well defined infimum), which does not hold in this case. We can in fact *define* the *generalised Bayes estimator under squared loss* to be the posterior mean, to avoid this problem

Getting back to our particular problem, the reason that the Bayes risk is infinite is that the prior $p(\theta) = \beta e^{-\beta\theta}$ puts too much probability for θ around 0. Thus, if for example we let $p(\theta) = \frac{\beta^3}{2}\theta^2 e^{-\beta\theta}$ for $\theta > 0$, a gamma$(\beta^{-1}, 3)$ pdf, then we can use the same procedure as above to find out that the Bayes estimator of μ is $\tilde{\mu} = \frac{\beta + \sum_{i=1}^n x_i}{n+2}$, with the risk $R_{\tilde{\mu}}(\theta) = \frac{(\beta\theta - 2)^2 + n}{(n+2)^2\theta^2}$ and a finite Bayes risk $\frac{\beta^2}{2(n+2)}$.

(d) Using squared error loss, find the Bayes estimator of $\xi = 2^{-\theta}$.

Solution: The Bayes estimator of $\xi = 2^{-\theta}$ is

$$\begin{aligned} \hat{\xi} &= E(2^{-\theta}|\mathbf{x}) \\ &= E[e^{(-\ln 2)\theta}|\mathbf{x}) \\ &= \frac{1}{(1 - [\beta + \sum_{i=1}^n x_i]^{-1}(-\ln 2))^{n+1}} = \frac{1}{(1 + [\beta + \sum_{i=1}^n x_i]^{-1} \ln 2)^{n+1}} \end{aligned}$$

knowing that $E(e^{t\theta})$ is the mgf of θ, and for gamma(a, b) the mgf is $\frac{1}{(1-at)^b}$.

4. Show that using the loss function

$$L(t;\theta) = \begin{cases} 0, & |t - \theta| < \epsilon \\ c, & o/w \end{cases}$$

where ϵ is very small and c is large, results in a Bayes estimator approximately equal to the mode of the posterior distribution.

Moreover, show that if the prior is uniform over an interval containing the MLE, then the posterior mode *is* the MLE.

Solution: For a given value $\mathbf{x}$ of $\mathbf{X}$, the Bayes estimator $T = t(\mathbf{X})$ minimises

$$\begin{aligned} \mathbb{E}_\theta(L(T;\theta)|\mathbf{X} = \mathbf{x}) &= \int c I_{\{|T-\theta| \geq \epsilon\}} p_{\theta|\mathbf{X}}(\theta) d\theta \\ &= c - c\int_{T-\epsilon}^{T+\epsilon} p_{\theta|\mathbf{X}}(\theta) d\theta \\ &\approx c - 2c\epsilon p_{\theta|\mathbf{X}}(T) \text{ for small } \epsilon \end{aligned}$$

We minimise this by making $p_{\theta|\mathbf{X}}(T)$ as large as possible. That is, T is the mode of $p_{\theta|\mathbf{X}}$.

If the $p_\theta(\theta) = aI_A(\theta)$ for some interval A, then

$$p_{\theta|X}(\theta) \propto p_{X|\theta}(x) I_A(\theta).$$

Thus, if $\theta \in A$ then $p_{\theta|X}(\theta) \propto p_{X|\theta}$, so the value of θ that maximises the LHS also maximises the RHS. But the RHS is the likelihood at θ, and by definition the value of θ that maximises the RHS is the maximum likelihood estimator.