Math 6262, Test 2 Solutions

Problem 1. *Assume that $X_1, X_2, \dots, X_n$ comes from a geometric distribution, i.e. $\mathbb{P}(X_i = x) = f(x|\theta) = \theta(1-\theta)^{x-1}$ for $x = 1, 2, \dots$.*

Take the prior for θ as $Beta(\alpha, \beta)$. Find the Bayesian estimator for the square loss function.

Solution. The posterior distribution of θ is given by

$$f(\theta|\vec{x}) = \frac{f(\vec{x}|\theta)f(\theta)}{m(\vec{x})} \text{ where } m(\vec{x}) = \int_0^1 f(\vec{x}|\theta)f(\theta)d\theta$$

Now, with the notation $s = \sum_{i=1}^n x_i$, we write

$$f(\vec{x}|\theta)f(\theta) = \frac{\theta^n(1-\theta)^{s-n}\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} = \frac{\theta^{n+\alpha-1}(1-\theta)^{s-n+\beta-1}}{B(\alpha,\beta)}$$

and thus integrating this with respect to θ, yields that

$$m(\vec{x}) = \frac{B(n+\alpha, s-n+\beta)}{B(\alpha,\beta)}$$

and thus

$$f(\theta|\vec{x}) = \frac{\theta^{n+\alpha-1}(1-\theta)^{s-n+\beta-1}}{B(\alpha+n, s-n+\beta)}$$

which is a $Beta(n+\alpha, s-n+\beta)$. Since the loss function is the square loss, we obtain

$$\hat{\theta}(\vec{x}) = \int \theta f(\theta|\vec{x})d\theta = \frac{\alpha+n}{s+\alpha+\beta}.$$

□

Problem 2. *Let θ be a parameter with values $\{3, 7\}$. Take a single sample X from the distribution given by*

t	0	1
$p(t;3)$	$1/2$	$1/2$
$p(t;7)$	$2/3$	$1/3$

(1) *Find the Bayesian estimator with the prior given by $f(3) = 1/5, f(7) = 4/5$. Do you see anything strange with this estimator?*

(2) *If the loss function is the 0-1 loss given by $L(\hat{\theta}, \theta) = \begin{cases} 1, & \hat{\theta} \neq \theta \\ 0, & \hat{\theta} = \theta \end{cases}$, find the minimax estimator.*

Solution. (1) To find the Bayesian estimator, we compute the Bayesian risk for each of the cases. We denote $\hat{\theta}(0) = x$ and $\hat{\theta}(7) = y$. Then we have the Bayesian risk given by

$$r(f|\hat{\theta}) = \int R(\hat{\theta}, \theta)f(\theta)d\theta = R(\hat{\theta}, 3)/5 + R(\hat{\theta}, 7)4/5$$

while we need to compute the risk. We need to compute the risk function. In the first place an estimator is completely determined by $\hat{\theta}(0) = x$ and $\hat{\theta}(1) = y$. This now leads to

$$R(\hat{\theta}, \theta) = R(x, y, \theta) = \mathbb{E}[L(\hat{\theta}, \theta)] = \mathbb{P}(\hat{\theta} \neq \theta).$$

We do this completing the following table, where $r(x, y)$ denotes the Bayesian risk associated to (x, y):

	$x=3$ $y=3$	$x=3$ $y=7$	$x=3$ $y=ne$	$x=7$ $y=3$	$x=7$ $y=7$	$x=7$ $y=ne$	$x=ne$ $y=3$	$x=ne$ $y=7$	$x=ne$ $y=ne$
$R(\hat{\theta}, 3)$	1/2	1/2	1/2	1	1	1/2	1	1	1
$R(\hat{\theta}, 7)$	1	1/3	1	2/3	0	2/3	1	1/3	1
$r(x, y)$	9/10	11/30	9/10	11/15	1/5	11/15	1	7/15	1

were we denoted $x = ne$ to denote that $x \neq 3, 7$.

The Bayesian estimator in this case predicts $\theta = 7$ in both cases, thus $\hat{\theta}(0) = \hat{\theta}(1) = 7$. This seems a little strange because the value 3 is not seen at all. This is due to our prior which puts too much emphasis on $\theta = 7$ (four times more than $\theta = 3$).

(2) For each of the values of $\theta = 3, 7$ we have the following table.

	$x=3$ $y=3$	$x=3$ $y=7$	$x=3$ $y=ne$	$x=7$ $y=3$	$x=7$ $y=7$	$x=7$ $y=ne$	$x=ne$ $y=3$	$x=ne$ $y=7$	$x=ne$ $y=ne$
$R(\hat{\theta},3)$	1/2	1/2	1/2	1	1	1/2	1	1	
$R(\hat{\theta},7)$	1	1/3	1	2/3	0	2/3	1	1/3	1
$\bar{R}(x,y)$	1	1/2	1	1	1	2/3	1	1	1

From this table it is clear that the minimax estimator (completely determined by x, y) is $\hat{\theta}(0) = 3$ and $\hat{\theta}(1) = 7$.

$\square$

Problem 3. *If $X \sim Bin(n,p)$ and the loss function given by*

$$L(\hat{p}, p) = \frac{(\hat{p} - p)^2}{p^2}.$$

show that $\hat{p} = 0$ is the unique minimax rule.

Solution. The risk function is this case is given by

$$R(\hat{p}, p) = \mathbb{E}[\frac{(\hat{p}(X_1, X_2, \dots, X_n) - p)^2}{p^2}] =$$

If $\hat{p} \equiv 0$, then $R(\hat{p}, p) = 1$. If $\hat{p}$ is not identically 0, then its mean is not going to be 0 and thus we will have $\mathbb{E}[\hat{\mathbb{P}}(X_1, \dots, X_n)] > 0$ which leads to

$$\mathbb{R}(\hat{p}, p) \geq \frac{(\mathbb{E}[\hat{p}] - p)^2}{p^2}$$

and as p approaches 0 this converges to $+\infty$. Therefore we have that

$$\bar{R}(\hat{p}) = \sup_{p \in (0,1)} R(\hat{p}, p) = \begin{cases} 1, & \hat{p} \equiv 0 \\ +\infty, & \hat{p} \not\equiv 0. \end{cases}$$

This shows that $\bar{R}(\hat{p})$ is actually finite only for $\hat{p} \equiv 0$.

$\square$

Problem 4. *Assume $X_1, X_2, \dots, X_n$ is a sample of iid from the distribution $f(x|\theta)$ with θ following a prior $F(\theta)$. Now define for $\vec{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$,*

$$f(\theta|\vec{x}) = \frac{f(\vec{x}|\theta) f(\theta)}{m(\vec{x})} \text{ with } f(\vec{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \text{ and } m(\vec{x}) = \int f(\vec{x}|\theta) f(\theta) d\theta.$$

(1) *If $L(\hat{\theta}, \theta)$ is a loss function and we define*

$$R(\hat{\theta}, \theta) = \int L(\hat{\theta}(\vec{x}), \theta) f(\vec{x}|\theta) dx \text{ and } r(f, \hat{\theta}) = \int R(\hat{\theta}, \theta) f(\theta) d\theta,$$

show that

$$r(f, \hat{\theta}) = \int r(f|\vec{x}) m(x) dx \text{ where } r(f|\vec{x}) = \int L(\hat{\theta}(\vec{x}), \theta) f(\theta|\vec{x}) d\theta$$

(2) *Show that if we define*

$$\hat{\theta}(\vec{x}) = argmin_{z \in \mathbb{R}} \int L(z, \theta) f(\theta|\vec{x}) d\theta \tag{1}$$

then $\hat{\theta}$ is the Bayesian estimator.

(3) *In the case $\theta \geq 1$ and the loss function given by $L(\hat{\theta}, \theta) = \frac{\hat{\theta}}{\theta} - 1 - \ln\left(\frac{\hat{\theta}}{\theta}\right)$, show that the Bayesian estimator is given by*

$$\hat{\theta}(\vec{x}) = \frac{1}{\int \frac{1}{\theta} f(\theta|\vec{x}) d\theta}. \tag{2}$$

Solution. (1) These we did in class. The proof is based on direct writing and exchange of integration. The details,

$$\begin{aligned} r(f,\hat{\theta}) &= \int R(\hat{\theta},\theta)f(\theta)d\theta = \int\int L(\hat{\theta},\theta)f(\vec{x}|\theta)f(\theta)dx\,d\theta \\ &= \int\int L(\hat{\theta},\theta)f(\theta|\vec{x})m(\vec{x})d\theta\,dx \\ &= \int r(f|\vec{x})m(\vec{x})dx. \end{aligned}$$

(2) This is relatively straightforward in the sense that for each $\vec{x}$ we actually have for any other estimator $\tilde{\theta}$, by definition, we get

$$\int L(\hat{\theta}(\vec{x}),\theta)f(\theta|\vec{x})d\theta \le \int L(\tilde{\theta}(\vec{x}),\theta)f(\theta|\vec{x})d\theta$$

which integrated leads to

$$r(f|\hat{\theta}) \le r(f|\tilde{\theta}).$$

(3) We need to find the estimator using (1) and thus first define

$$h(z) = \int L(z,\theta)f(\theta|\vec{x})\theta = \int \left(\frac{z}{\theta} - 1 - \ln\left(\frac{z}{\theta}\right)\right) f(\theta|\vec{x})d\theta$$

Taking the derivative with respect to θ leads to

$$\int \frac{1}{\theta}f(\theta|\vec{x})d\theta - \frac{1}{z}\int f(\theta|\vec{x})d\theta = 0$$

and since $f(\theta|\vec{x})d\theta = 1$ ($f(\theta|\vec{x})$ is the posterior density) , this yields $z = \frac{1}{\int \frac{1}{\theta}f(\theta|\vec{x})d\theta}$, which gives the expression of $\hat{\theta}$ from (2).

□

Problem 5. *Assume we have a sample $X_1, X_2, \ldots, X_n$ from a Poisson distribution $f(x;\theta)$ with $\theta > 0$, i.e. $f(x;\theta) = e^{-\theta}\frac{\theta^x}{x!}$ for $x = 0,1,2,3\ldots$. Assume now that the prior distribution of θ is $\Gamma(\alpha,\beta)$ and the loss function is*

$$L(\hat{\theta},\theta) = \frac{(\hat{\theta}-\theta)^2}{\theta}.$$

(1) *Find the Bayesian estimate of θ.*
(2) *Can you use this for some parameters $\alpha,\beta > 0$ to find a minimax estimator?*

Solution. (1) This is very similar to the homework problem, the only difference is the loss function. At any rate, the main difference is that now, since the loss function is not the square loss, the Bayesian estimator is not directly the mean of the posterior distribution. We need to use a little bit of work here to carry this out. Thus, for instance using the Bayesian theory (in fact the previous problem) we need to find the estimator by minimizing

$$h(z) = \int L(z,\theta)f(\theta|\vec{x})d\theta = \int \frac{(z-\theta)^2}{\theta}f(\theta|\vec{x})d\theta.$$

Differentiating and taking the derivative equal to 0 leads to

$$\hat{\theta}(\vec{x}) = \frac{\int f(\theta|\vec{x})d\theta}{\int \frac{f(\theta|\vec{x})}{\theta}d\theta} = \frac{1}{\int \frac{f(\theta|\vec{x})}{\theta}d\theta}.$$

Now, with the problem at hand, with the notation $s = \sum_{i=1}^n x_i$,

$$f(\theta|\vec{x}) = \frac{\theta^s e^{-n\theta}\theta^{\alpha-1}e^{-\beta\theta}}{\theta^s e^{-n\theta}\theta^{\alpha-1}e^{-\beta\theta}} = \frac{\theta^{s+\alpha-1}e^{-(n+\beta)\theta}}{\Gamma(s+\alpha)/(\beta+n)^{s+\alpha}}$$

Thus,

$$\hat{\theta}(\vec{x}) = \frac{1}{\int \frac{1}{\theta} f(\theta|\vec{x}) d\theta} = \frac{\Gamma(s+\alpha)/(\beta+n)^{s+\alpha}}{\int \theta^{s+\alpha-1-1} e^{-\beta\theta}} = \frac{\Gamma(s+\alpha)/(\beta+n)^{s+\alpha}}{\Gamma(s+\alpha-1)/(n+\beta)^{s+\alpha-1}} = \frac{s+\alpha-1}{\beta+n}$$

(2) We need to compute here the risk function for this estimator. Thus we compute

$$R(\hat{\theta}, \theta) = \frac{\mathbb{E}[(\hat{\theta}-\theta)^2]}{\theta} = \frac{E[(s-n\theta+\alpha-1-\beta\theta)^2]}{\theta(\beta+n)^2} = \frac{n\theta+(\alpha-1-\beta\theta)^2}{\theta(\beta+n)^2}$$

To get a minimax estimator here we need to figure out the values of α, β for which this does not depend on θ. This is possible if we set $\beta = 0$ and $\alpha = 1$. Thus the minimax estimator in this case is $\frac{s}{n}$ which is the same as $\bar{x}$. However technically this is correct, we do have to complain about the issue with $\beta = 0$ which does not satisfy the requirement from the beginning that $\alpha > 0, \beta > 0$, thus this is not really a consistent choice.

$\square$