6262 HOMEWORK 2, VERSION 2

Problem 1. (1) *A sample $X_1, X_2, \ldots, X_n$ is drawn from a distribution $f(x;\theta)$ with parameter $\theta \in \mathbb{R}$. If the loss function is the $0-1$ loss, i.e.*

$$L(\hat{\theta}, \theta) = \begin{cases} 1, & \hat{\theta} \neq \theta \\ 0, & \hat{\theta} = \theta \end{cases},$$

show that any estimator is a minimax estimator. Does this make sense? Can you interpret it?

(2) *If the sample is drawn from $Bernoulli(p)$ with $p \in (0,1)$ and the loss function is*

$$L(\hat{p}, p) = \frac{(\hat{p}-p)^2}{p^3}$$

show that any estimator is a minimax estimator.

Problem 2. *Describe and interpret in your own words what the setup of a minimax and Bayesian estimator means. Why and how is this relevant from the statistical point of view?*

Problem 3. *Assume that $X_1, X_2, \ldots, X_n$ is a sample from a distribution $f(x;\theta)$ and the loss function L. Assume that for a given prior f we know that the Bayesian estimator $\hat{\theta}^f$ satisfies*

$$R(\hat{\theta}^f, \theta) \leq r(f; \hat{\theta}^f) \text{ for any } \theta.$$

Prove that $\hat{\theta}^f$ is a minimax.

Problem 4. *Assume we have a parameter θ taking only two values, $\{1, 2\}$ and for each of these we have distributions given by*

t	-1	1
$p(t;1)$	$1/4$	$3/4$
$p(t;2)$	$3/4$	$1/4$

(1) *We draw a single sample from this distribution. Assume the loss function L is given by the $0-1$ loss.*
   (a) *Find a minimax estimator.*
   (b) *Assume we put a prior distribution on the space of parameters, namely we consider $f(1) = \lambda$ and $f(2) = 1-\lambda$. For this prior distribution find the Bayesian estimator of θ.*
   (c) *Using the Bayesian estimator above, can you find a minimax? Is this the same as the one found above with direct methods?*
   (d) *If we observe the sample $t = 1$, what is the estimation of θ?*

(2) *Now, assume the square loss function $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$,*
   (a) *Find the risk function associated to any estimator $\hat{\theta}(X)$.*
   (b) *Find a minimax estimator and also a Bayesian estimator for the prior $f(1) = \lambda$, $f(2) = 1-\lambda$.*
   (c) *If we observe the sample $t = 1$, what is the estimation of θ?*

(3) *Do the same for the distributions*

t	-1	1
$p(t;1)$	$1/2$	$1/2$
$p(t;2)$	$3/4$	$1/4$

Remark 5. *Problem 4 is a prototype for how one computes in the case the parameter takes a discrete number of values and for each value we have a distribution assigned. This procedure can be extended to the general case, though in that case the problem becomes an optimization problem which in general is hard to solve by hand. For instance, if have multiple values of t in the distribution of $p(x;\theta)$, and we also have many sample points, we need to consider many variables and then the minimization becomes much more complicated. As a message is that from the computational side the Bayesian method is much much simpler and yields faster results, because for a fixed prior we can very easily minimize the average loss (at least in the case of a square loss).*

Problem 6. *Assume we take a sample $X_1, X_2, \dots, X_n$ from $N(\theta, \sigma^2)$. Assume that a prior distribution on θ is $N(a, b^2)$.*

(1) *Find the posterior distribution of θ.*
(2) *Compute the expectation of this posterior distribution and denote it $\hat{\theta}$.*
(3) *Find the risk function associated to this estimator with respect to the square loss function $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$.*
(4) *Find the maximum of this risk over all θ. For which values of a, b is this finite?*
(5) *Try the same thing, this time with respect to the loss function $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^4$.*

Problem 7. *Assume that $X \sim N(\theta, 1)$ and take the estimator $\hat{\theta}_c = cX$ for some constant. Given a loss function L, we say that the estimator $\hat{\theta}$ is better than $\bar{\theta}$ if $R(\hat{\theta}, \theta) \le R(\bar{\theta}, \theta)$ for all choices of $\theta \in \mathbb{R}$.*

(1) *Find the risk associated to $\hat{\theta}$ for the square loss function. Show that θ_1 is better than θ_c for any $c > 1$.*
(2) *Is $\hat{\theta}_{1/2}$ better than θ_1? Comment?*

Problem 8. *If we take the loss function of the form*

$$L(\hat{\theta}, \theta) = w(\theta)(\hat{\theta} - \theta)^2 \text{ with } w(\theta) > 0 \text{ for all } \theta$$

and consider a sample $X_1, X_2, \dots, X_n$ from the distribution $f(x|\theta)$ and θ is assumed to follow the distribution $f(\theta)$, show that the Bayes estimator is given by

$$\hat{\theta}^f(\vec{x}) = \frac{\int \theta w(\theta) f(\theta|\vec{x}) d\theta}{\int w(\theta) f(\theta|\vec{x}) d\theta} = \frac{\int \theta w(\theta) f(\vec{x}|\theta) f(\theta) d\theta}{\int w(\theta) f(\vec{x}|\theta) f(\theta) d\theta} \tag{0.1}$$

where $\vec{x} = (x_1, x_2, \dots, x_n)$ while

$$f(\theta|\vec{x}) = \frac{f(\vec{x}|\theta) f(\theta)}{m(\vec{x})} \text{ with } m(\vec{x}) = \int f(\vec{x}|\theta) f(\theta) d\theta \text{ and } f(\vec{x}|\theta) = f(x_1|\theta) f(x_2|\theta) \dots f(x_n|\theta) f(\theta).$$

Problem 9. (1) *Assume $X_1, X_2, \dots, X_n$ is a sample from a Bernoulli random variable with parameter p. Under the assumption that the prior of p is a $Beta(\alpha, \beta)$, find the posteriori distribution of p. Show that this estimator is a Bayes estimator for the square loss function.*
(2) *Find the risk function for the above estimator and a minimax estimator in this case.*
(3) *Under the same assumptions, assume that the loss function is $L(\hat{p}, p) = \frac{(\hat{p} - p)^2}{p(1-p)}$. Find a Bayes rule for this loss function.*
(4) *Can you do the same for the loss function $L(\hat{\theta}, \theta) = \frac{(\hat{p} - p)^2}{p^a (1-p)^b}$ with $a, b \ge 0$ integer numbers?*

Problem 10. *Let X be a Bernoulli with parameter p and assume the prior is uniform on $[0, 1]$.*

(1) *Find the posterior distribution of p and then the Bayes estimator for the square loss function.*
(2) *If in addition, we know that $p \notin (1/3, 2/3)$ and use the prior to be the uniform on the $(0, 1/3) \cup (2/3, 1)$, find the Bayes estimator now, again with respect to the square loss function.*

Problem 11. *This problem is about the* Γ *and* $Beta$ *distributions. For* $\alpha, \beta > 0$*, we set*

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

while

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

(1) $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ *and* $\Gamma(1) = 1$, $\Gamma(n) = (n-1)!$.
(2) *Show that*

$$f_{\alpha,\beta}(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx, \text{ for } x > 0 \\ 0, \text{ for } x < 0 \end{cases}$$

is a density. It is the density of a Gamma distribution with parameters $\alpha, \beta > 0$*, in short* $\Gamma(\alpha, \beta)$.
(3) *Compute the mean and the variance of* $\Gamma(\alpha, \beta)$.
(4) *Show that*

$$B(\alpha+1, \beta) = B(\alpha, \beta)\frac{\alpha}{\alpha+\beta} \tag{0.2}$$

(5) *Show that the following*

$$f_{\alpha,\beta}(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, 0 < x < 1 \\ 0, \text{ otherwise} \end{cases}$$

is a density. This is called the $Beta(\alpha, \beta)$*. Compute the mean and the variance of* $Beta(\alpha, \beta)$.

Problem 12. *Assume we have a sample* $X_1, X_2, \dots, X_n$ *from a Poisson distribution* $f(x;\theta)$ *with* $\theta > 0$*, i.e.* $f(x;\theta) = e^{-\theta}\frac{\theta^x}{x!}$ *for* $x = 0, 1, 2, 3 \dots$*. Assume now that the prior distribution of* θ *is* $\Gamma(\alpha, \beta)$.
(1) *Find the Bayesian estimate of* θ.
(2) *Can you use this for some parameters* $\alpha, \beta > 0$ *to find a minimax estimator?*

Problem 13 (optional)**.** *Assume that we have a sample* $X_1, X_2, \dots, X_n$ *from a given distribution* $f(x;\theta)$ *for which there exists a sufficient statistic* $T = u(X_1, X_2, \dots, X_n)$.
(1) *Show that for any prior* f*, the Bayesian estimator* $\hat{\theta}^f$ *is a function of* T.
(2) *Show that if the loss function is convex and* $\hat{\theta}$ *is an estimator, then* $\tilde{\theta} = \mathbb{E}[\hat{\theta}|T] = \psi(T)$ *has a smaller risk for any value of* θ*. (Hint: You need here Jensen's inequality for the conditional expectation).*
(3) *Show that if the loss function* $L(\hat{\theta}, \theta)$ *is convex in* $\hat{\theta}$*, then any minimax estimator is a function of* T.