Problem Sheet 3: Advanced Bayesian Statistics

1. **Conditional Distributions**

   For a pair of random variables (X, Y), knowledge of the marginal distributions of X and Y is not sufficient to characterize the joint distribution of (X, Y).

   (a) Give an example of two different bivariate distributions admitting the same marginal distributions. *[Hint: For example, construct two variables with $\mathcal{N}(0, 1)$ laws such that the pair does not follow a Gaussian distribution; or, construct a non-uniform distribution on $[0, 1]^2$ whose marginals are uniform on $[0, 1]$.]*

   (b) Show that knowledge of the conditional distributions $X|Y$ and $Y|X$ is sufficient to characterize the distribution of (X, Y). *[Assume we are in the density framework; one might first seek to write the marginal densities using an integral involving the conditional densities $f(x|y)$ and $f(y|x)$.]*

2. **Risk Types (Review)**

   Let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ be a statistical model, ℓ a given loss function on $\Theta \times \Theta$, and Π a prior distribution on Θ.

   (a) For an arbitrary estimator $T(X)$, recall the definitions of:

      i. The risk function $R(\theta, T)$ of T.

      ii. The Bayes risk $R_B(\Pi, T)$ of T for the prior Π.

   (b) Recall the definitions of an **admissible** estimator, a **Bayes estimator** for Π, and a **minimax** estimator.

3. **Mixtures**

   A two-component mixture distribution is defined as a distribution Q of the form:

   $$Q = (1 - \rho)Q_0 + \rho Q_1,$$

   where Q_0 and Q_1 are two probability distributions and $\rho \in [0, 1]$. We assume Q_0 and Q_1 have densities denoted q_0 and q_1 with respect to a reference measure μ.

   (a) Show that a random variable Y with distribution Q can be obtained by the following scheme:

   $$Z \sim \mathcal{B}ern(\rho), \quad Y|Z \sim Q_Z.$$

   (b) Let $\mathcal{P} = \{P_\theta = \mathcal{N}(\theta, 1), \theta \in \mathbb{R}\}$. We place a prior Π on θ of the form of a two-component mixture $(1-\rho)Q_0+\rho Q_1$ with $Q_0 = \mathcal{N}(0, 1)$, $Q_1 = \mathcal{N}(5, 1)$, and $\rho > 0$. We have observations $X_1, \ldots, X_n|\theta \sim P_\theta^{\otimes n}$.

      i. Determine the density of the posterior distribution $\theta|X_1, \ldots, X_n$.

ii. Show that the posterior distribution is again a two-component mixture. Determine the new corresponding weight $\rho_n(X)$.

(c) In the context of coin toss observations, we suspect the coins might be biased with a probability 2/3 of obtaining heads. We propose the following modeling: $X_1, \dots, X_n|\theta$ i.i.d. $\mathcal{B}ern(\theta)$ and $\theta \sim (1-\rho)Q_0 + \rho Q_1$, with $Q_0 = \mathcal{B}eta(2,4)$ and $Q_1 = \mathcal{B}eta(3,3)$.

i. Justify the choice of the prior distribution.

ii. Answer the same questions as in part (b). *[Hint: You may use the expression of the Beta function. Recall that $\Gamma(p+1) = p!$.]*

(d) Show that the fact observed in examples (b) and (c)—that the posterior distribution is still a two-component mixture—is a general phenomenon, and give the expression of $\rho_n(X)$ as a function of the problem data.

4. **Bayes Estimators**

Let X be a random variable with distribution $\mathcal{N}(\theta, 1)$ given θ, and let $\Pi = \mathcal{N}(0, \sigma^2)$ be the prior on θ, for fixed $\sigma^2 > 0$. Let

$$L(\theta, T) = (\theta - T)^2$$

be the quadratic loss function.

(a) Give the Bayes estimator of θ for the loss function L.

(b) Assume $\sigma^2 \le 1$ and consider the weighted loss function:

$$L_w(\theta, T) = \exp\left\{\frac{3\theta^2}{4}\right\}(\theta - T)^2.$$

Determine the associated Bayes estimator. Calculate its Bayes risk.

5. **Minimaxity**

(a) Prove that if an admissible estimator has constant risk, it is minimax.

(b) Let T be a Bayes estimator for Π and let $R_B(\Pi, T)$ be its Bayes risk.

i. Assume $R(\theta, T) \le R_B(\Pi, T)$ for all $\theta \in \Theta$. Show that T is minimax.

ii. In particular, recover the result that a Bayes estimator of constant risk is minimax.

6. **Quantiles**

Let X be a real random variable with distribution P_θ given θ, where θ is a real number with prior law Π. Assume P_θ and Π have densities with respect to the Lebesgue measure, denoted $dP_\theta = p_\theta(x)dx$ and $d\Pi(\theta) = \pi(\theta)d\theta$. Let the loss function be, for $k_1 > 0, k_2 > 0$:

$$\ell(\theta, T) = \begin{cases} k_2(\theta - T), & \text{if } \theta > T, \\ k_1(T - \theta), & \text{if } \theta \le T. \end{cases}$$

Show that the Bayes estimator for the loss function ℓ is a quantile of the posterior distribution $\mathcal{L}(\theta|X)$. Express this quantile as a function of k_1 and k_2.

7. **The Hodges Phenomenon**

Consider, for $X_1, \dots, X_n$ i.i.d. with law $\mathcal{N}(\theta, 1)$ where $\theta \in \mathbb{R}$, the estimator $T_n = \overline{X}$ as well as the modified estimator:

$$S_n = \begin{cases} T_n & \text{if } |T_n| \ge n^{-1/4} \\ 0 & \text{if } |T_n| < n^{-1/4} \end{cases}$$

(a) Calculate the quadratic risk $R(\theta, T_n) = E_\theta[(T_n - \theta)^2]$ of T_n for all real θ.

(b) Show that for all real θ,

$$E_\theta[(S_n - \theta)^2] \geq \theta^2 \mathbb{P}\left[\left|\mathcal{N}\left(\theta, \frac{1}{n}\right)\right| \leq \frac{1}{n^{1/4}}\right].$$

(c) Deduce that there exists a constant $c > 0$ such that

$$\sup_{\theta \in \left[-\frac{1}{n^{1/4}}, \frac{1}{n^{1/4}}\right]} n E_\theta[(S_n - \theta)^2] \geq c\sqrt{n}.$$

(d) Compare the maximal risk of T_n to that of S_n. Is the latter "uniformly good"?

8. **Bayesian Estimator of the Loss Function**

In the decision-theoretic framework, we choose a decision rule T to minimize the loss $L(\theta, T)$. However, since θ is unknown, the actual loss $L(\theta, T)$ is a random variable (with respect to the posterior distribution). We can therefore consider the problem of *estimating the loss value itself*.

Let standard notation be:

- $\theta \in \Theta$: the parameter.
- $T \in \mathcal{D}$: the decision (estimator of θ).
- $L(\theta, T)$: the loss function for estimating θ.

We introduce a "meta-loss" function $\tilde{L}$ to evaluate an estimator $\hat{L}$ of the loss $L(\theta, T)$:

$$\tilde{L}(\text{True Value}, \text{Estimate}) = \left(L(\theta, T) - \hat{L}\right)^2.$$

Question: What is the Bayes estimator of the loss $L(\theta, T)$ under this quadratic meta-loss $\tilde{L}$?