SB2.1 Foundations of Statistical Inference

Sheet 3 — MT22

Section A

1. The risks for five decision processes $\delta_1, \ldots, \delta_5$ depend on the value of a positive-valued parameter θ. The risks are given in the table below

	δ_1	δ_2	δ_3	δ_4	δ_5
$0 \leq \theta < 1$	10	10	7	6	8
$1 \leq \theta < 2$	8	11	8	5	10
$2 \leq \theta$	15	11	12	14	14

(a) Which decision procedures are at least as good as δ_1 for all θ ?

(b) Which decision procedures are admissible?

(c) Which is the minimax procedure?

(d) Suppose θ has a uniform distribution on $[0, 5]$. Which is the Bayes procedure and what is the Bayes risk for that procedure?

Solution:

(a)

$$R(\theta, \delta_1) \geq R(\theta, \delta_3) \text{ and } R(\theta, \delta_4)$$

for all $\theta > 0$ so δ_3, δ_4 are at least as good as δ_1.

(b) δ_1 and δ_5 are inadmissible procedures, so $\delta_2, \delta_3, \delta_4$ are admissible.

(c) The minimax procedure chooses δ to minimize

$$\max_{\theta} R(\theta, \delta)$$

Values of the interior maximum for $\delta_1, \ldots, \delta_5$ are 15,11,12,14 and 14, so δ_2 is the minimax solution.

George Deligiannidis: deligian@stats.ox.ac.uk

(d) In the Bayes procedure we minimize

$$\int_0^5 R(\theta, \delta) \cdot \frac{1}{5} d\theta$$

Values of 5 $\times$ the integral are 63,54,51,53, 60 so δ_3 is the Bayes procedure. The Bayes risk is 51/5.

George Deligiannidis: deligian@stats.ox.ac.uk

Section B

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 $\pi(\theta)$. Consider the loss function for estimating θ

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

   (a) Show that for all $u \in \mathbb{R}$ $h(u) \geq 0$ and show that h is a convex function.

   (b) Determine the form of the Bayes estimator $\delta(X)$ and prove that it is almost surely unique as soon as

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

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

   Compute the marginal posterior $\pi(\mu|X)$ up to proportionality. Is the Bayes estimator $\hat{\mu}_{\text{Bayes}}$ under the loss function $L(\mu,\delta) = h(\mu-\delta)$ well defined? (here δ is an estimator of μ)

   (d) Now assume that σ is known and that we are using a flat prior for μ. Find the Bayes estimator.

   [Hint: It might be useful to look-up the definition of the inverse Gamma distribution (see Wikipedia) and of the Student distribution (see Wikipedia)]

3. Remember Question 3 in PS2. *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.* In PS2, we showed that if the prior distribution for μ is $N(\mu_0,1)$ while the prior for θ is constant then the marginal posterior for θ is Gaussian $N((2x_2 - x_1 - \mu_0)/2, 3/2)$.

   (a) Suppose Y_1, Y_2 are independent Gaussian random variables with respective means μ_1, μ_2 and variances 1. What is the Fisher information matrix for (μ_1,μ_2)?

   (b) Compute Jeffrey's prior on (μ,θ). Is the posterior proper? *[Hint: it might be helpful to remember how reparametrization acts on Jeffrey's prior. See Wikipedia.]*

   (c) Compute the marginal posterior distribution of θ under Jeffrey's prior. Hence derive the Bayes estimator associated to the quadratic loss function and compute its frequentist risk, the posterior risk and integrated risk.

George Deligiannidis: deligian@stats.ox.ac.uk

4. Let $E(a,b)$ be the distribution from Question 5 on PS2, 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 $E(a,b)$. *[Hint: It might be useful to look-up the definition of the inverse Gamma distribution (see Wikipedia)*

(a) If a is known, derive Jeffrey's prior on b and compute the posterior mean $\hat{b}$. Show that $\tilde{b} = (n-1)\hat{b}/n$ is MVUE and attains the Cramer Rao lower bound.

(b) If a is known and an inverse $\Gamma(\alpha, \beta)$ prior is chosen for b , find the posterior mean. Is it an MVUE? Compute the predictive distribution of X_{n+1} .

(c) If b is known and a prior density π_a on a is considered. Show that the posterior distribution 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 proper and compute it.

(d) We continue to assume the prior $\pi(a) \propto 1$. Show that for $z > 0$ and all $b > 0$, the posterior probability $\Pi(n|a - X_{(1)}| > z \big| X_1, \cdots, X_n)$ goes to zero as z goes to infinity uniformly in n. Compute the frequentist risk of $\hat{a}$ (the posterior mean) under the quadratic loss.

[Hint: note that in the expression $\Pi(n|a - X_{(1)}| > z \big| X_1, \cdots, X_n)$ *it is* a *which is the random variable since we are conditioning on* $X_1, \cdots, X_n$*]*

(e) **Optional, not to be marked**. Suppose now that both a and b are unknown. As in PS2, define $T_1(X) = X_{(1)}$ and $T_2(X) = \sum_{i=1}^n X_i - X_{(1)}$ and show that the posterior distribution depends only on (T_1, T_2). Show that the family of priors

$$\pi(a|b) \propto e^{\alpha(a-\beta)/b}\mathbf{1}_{a<z_1}, \quad \pi(b) \propto b^{-m-1}e^{-c/b}\mathbf{1}_{b>0}$$

is conjuguate.

5. Let $X_i \overset{ind.}{\sim} \mathcal{N}(\mu_i, \sigma^2)$, $i \leq p$ and denote $\mu = (\mu_1, \cdots, \mu_p)^T$ Consider the quadratic loss function. σ^2 is known.

   (a) Show that for all real functions h of $(X_1, \cdots, X_n)$ continuously differentiable and satisfying

   $$\forall j \quad \mathbb{E}\left(\left|\frac{\partial h}{\partial x_j}(X)\right|\right) < +\infty, \quad \text{and} \quad \mathbb{E}[h^2(X)] < +\infty$$

   we have

   $$\mathbb{E}\left(h(X)(X_j - \mu_j)\right) = \sigma^2 \mathbb{E}\left(\frac{\partial h}{\partial x_j}(X)\right)$$

   (b) Find a such that

   $$\tilde{\mu}_{JS} = \left(1 - \frac{a}{\sum_j X_j^2}\right) X$$

   dominates strictly the maximum likelihood estimator

   (c) Denote $\vec{\mu} = (\mu_1, \ldots, \mu_p)$ the point defined by $\mu_j = \mu$ if $j \leq p_0$ and $\mu_j = 0$ if $j > p_0$ and show, using Jensen's inequality that

   $$\mathbb{E}\left[\frac{1}{\sum_j X_j^2}; \vec{\mu}\right] \geq \frac{1}{p\sigma^2 + \mu^2 p_0}$$

   (d) Deduce that if p_0 and μ are fixed

   $$\text{limsup}_{p \to +\infty} R(\tilde{\mu}_{JS}, \vec{\mu}) < +\infty$$

   (e) Why is shrinkage so important when p is large?

Section C

6. 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 $\mathrm{card}(\mathcal{X}) < +\infty$ i.e. $\mathcal{X}$ is finite. Show that there exists $x \in \mathcal{X}$ such that
   $$\int_\Theta P[X = x|\theta]\pi(\theta)d\theta = +\infty$$

George Deligiannidis: deligian@stats.ox.ac.uk

7. 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 Lebesgue measure.

(a) 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 Lebesgue measure.

(b) 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.

(c) 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 $\hat{b}$. Show that $\tilde{b} = (n-1)\hat{b}/n$ is MVUE and attains the Cramer Rao lower bound.

(ii) If a is known and a $Gamma(\alpha, \beta)$ prior is chosen on $1/b$, find the posterior mean. Is it an MVUE? Compute the predictive distribution of X_{n+1} .

(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.

(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. Do we have a Bernstein von Mises property as described in the lectures? Compute frequentist risk of a and $\hat{a}$ (the posterior mean) under the quadratic loss.

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

$$\pi(a|b) \propto e^{\alpha(a-\beta)/b}\mathbf{1}_{\mathbf{a<z_1}}, \quad \pi(\mathbf{b}) \propto \mathbf{b^{-m-1}e^{-c/b}1_{b>0}}$$

is conjuguate.

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.

George Deligiannidis: deligian@stats.ox.ac.uk