HW4, Theory of Inference 2016/7

Jonathan Rougier
School of Mathematics
University of Bristol UK

In the following questions, I show marks in square brackets, to give you an idea of the approximate tariff the question would carry in an exam.

1. We proved the Complete Class Theorem for the special case where Ω is finite. Prove the 'if' branch without this constraint. I.e., prove that a decision rule δ is admissible if it is a Bayes rule for some positive prior distribution π. Hint: use proof by contradiction. [10 marks]

Answer. The rule δ is admissible exactly when it is not dominated by any other rule. Suppose that δ is not admissible ('inadmissible'); i.e. there is some other rule δ' which dominates δ, so that

$$R(\delta', \theta) \leq R(\delta, \theta) \quad \text{for all } \theta \in \Omega,$$

with strict inequality for at least one $\theta \in \Omega$. Recollect that R is the risk function,

$$R(\delta, \theta) := \mathrm{E}\{L(\delta(Y), \theta); \theta\} = \sum\nolimits_y L(\delta(y), \theta) \cdot f(y; \theta).$$

We must show that δ cannot be a Bayes rule if it is dominated by δ'. Computing the Bayes loss:

$$\begin{aligned}
\mathrm{E}\{L(\delta(Y), \Theta)\} &= \int \sum\nolimits_y L(\delta(y), \theta) \cdot \mathrm{p}(y, \theta)\, \mathrm{d}\theta \\
&= \int \sum\nolimits_y L(\delta(y), \theta) \cdot f(y; \theta)\, \pi(\theta)\, \mathrm{d}\theta \\
&= \int R(\delta, \theta)\, \pi(\theta)\, \mathrm{d}\theta \\
&> \int R(\delta', \theta)\, \pi(\theta)\, \mathrm{d}\theta && \text{provided that } \pi > 0 \\
&= \mathrm{E}\{L(\delta'(Y), \Theta)\} && \text{reversing the steps,}
\end{aligned}$$

showing that δ cannot be a Bayes rule because δ' has smaller expected loss. We have proved that if δ is inadmissible it is not a Bayes rule. So have also proved the contrapositive that if δ is a Bayes rule then it is not inadmissible, i.e. it is admissible.

2. Consider the model $\mathcal{E} = \{y \in \mathbb{N}, \lambda \in \mathbb{R}_{++}, f\}$, where f is the Poisson probability mass function (and $\mathbb{N} = \{0, 1, \dots\}$). Let the prior for Λ be Gamma(a, b), where a is the 'shape' parameter and b is the 'rate' parameter.[1] Derive the Bayes rule for the point estimate for Λ under quadratic loss. [10 marks]

   In an exam you would be given the Poisson PMF and the Gamma PDF.

   Answer. For the Poisson model, $f(y; \lambda) \propto e^{-\lambda}\lambda^{y}$. For the Gamma prior distribution for Λ, $\pi(\lambda) \propto \lambda^{a-1}\, e^{-b\lambda}$. for parameters $a, b > 0$. The Bayes rule for point estimation under a quadratic loss function is the posterior expectation for Λ. Therefore we need to find the posterior distribution for Λ, and then compute its expectation. By Bayes's theorem, the posterior distribution has kernel

$$\begin{aligned} \pi^*(\lambda) &\propto f(y; \lambda) \cdot \pi(\lambda) \\ &\propto e^{-\lambda}\lambda^{y} \cdot \lambda^{a-1}\, e^{-b\lambda} \\ &= \lambda^{a+y-1} e^{-(b+1)\lambda} \end{aligned}$$

   which is the kernel of a Gamma$(a+y, b+1)$ distribution. The expectation of this distribution is $(a + y)/(b + 1)$, hence

$$\delta^*(y) = \frac{a+y}{b+1}$$

   is the Bayes rule for point estimation under quadratic loss, for a Gamma(a, b) prior distribution.

3. For the loss function given in eq. (3.7a) in the handout, confirm the Bayes rule for the two extreme cases $\kappa \downarrow 0$ and $\kappa \to \infty$. [10 marks]

   Answer. The loss function is

$$L(a, \theta) = |a| + \kappa \cdot (1 - \mathbb{1}_{\theta \in a}),$$

   and hence

$$\mathrm{E}\{L(a, \Theta) \mid Y = y\} = |a| + \kappa \cdot \{1 - \Pr(\Theta \in a \mid Y = y)\}.$$

[1] Remember that 'Λ' is capital 'λ'.

According to the Bayes rule theorem, a Bayes rule will minimise this expression over $a \in \mathcal{A} = 2^\Omega$. Let a be fixed. $\kappa \downarrow 0$ diminishes the second term relative to the first. Thus in the limit the Bayes rule minimises the first term, $|a|$, which leads to $a = \emptyset$. On the other hand, $\kappa \to \infty$ diminishes the first term relative to the second. Thus in the limit the Bayes rule minimises the second term, which leads to $a = \Omega$.

4. Prove the result in eq. (3.8) in the handout. [10 marks]

Answer. In this case, the Bayes Rule Theorem states that

$$\delta^*(y) = \operatorname*{argmin}_{a \in \{H_0, H_1\}} \mathrm{E}\{L(a, \Theta) \mid Y = y\}.$$

This rule will choose H_1 when

$$\mathrm{E}\{L(H_1, \Theta) \mid Y = y\} < \mathrm{E}\{L(H_0, \Theta) \mid Y = y\}.$$

The lefthand side equals

$$\begin{aligned} L(H_1, \theta_0) \cdot \Pr(\Theta = \theta_0 \mid Y = y) &+ L(H_1, \theta_1) \cdot \Pr(\Theta = \theta_1 \mid Y = y) \\ &= \ell_0 \cdot \Pr(\Theta = \theta_0 \mid Y = y) \\ &= \ell_0 \cdot \frac{f(y; \theta_0)\, \pi_0}{\Pr(Y = y)}, \end{aligned}$$

the last line by Bayes's Theorem. Likewise, the righthand equals

$$\ell_1 \cdot \frac{f(y; \theta_1)\, \pi_1}{\Pr(Y = y)}.$$

Therefore this rule will choose H_1 when

$$\ell_0 \cdot \frac{f(y; \theta_0)\, \pi_0}{\Pr(Y = y)} < \ell_1 \cdot \frac{f(y; \theta_1)\, \pi_1}{\Pr(Y = y)},$$

or, equivalently, when

$$\frac{f(y; \theta_0)}{f(y; \theta_1)} < \frac{\ell_1 \cdot \pi_1}{\ell_0 \cdot \pi_0},$$

as stated in the notes (first case of (3.8)). The second and third cases follow immediately. For fixed (ℓ_0, ℓ_1), the righthand side (c in the notes) can attain any positive value by varying $\pi_0 \in (0, 1)$. This shows, by the Complete Class Theorem, that the set of admissible rules is exactly the set of rules of the form $f(y; \theta_0)/f(y; \theta_1) \lesseqgtr c$ for some $c > 0$.

5. Prove that for hypothesis testing, the Bayes rule for the zero-one loss func-

tion is to select the hypothesis with the largest posterior probability. [10 marks]

Answer. The zero-one loss function is

$$L(H, \theta) = 1 - \mathbb{1}_{\theta \in H}.$$

According to the Bayes Rule Theorem, the Bayes Rule is

$$\begin{aligned} \delta^*(y) &= \operatorname*{argmin}_{H \in \mathcal{H}} \mathrm{E}\{1 - \mathbb{1}_{\theta \in H} \mid Y = y\} \\ &= \operatorname*{argmin}_{H \in \mathcal{H}} \big(1 - \Pr\{\theta \in H \mid Y = y\}\big) \\ &= \operatorname*{arg\,max}_{H \in \mathcal{H}} \Pr\{\theta \in H \mid Y = y\}, \end{aligned}$$

as claimed in the question.

6. Consider a 2D parameter space partitioned into two non-degenerate hypotheses, $\Omega = \Omega_0 \cup \Omega_1$. On a single figure, sketch three set estimates: $\delta_1(y^{\text{obs}})$ accepts H_0, $\delta_2(y^{\text{obs}})$ rejects H_0, and $\delta_3(y^{\text{obs}})$ is undecided about H_0. Redraw your picture for the case where Ω_0 is composite degenerate. [10 marks]

Answer.