HW4, Theory of Inference 2015/6

Jonathan Rougier
School of Mathematics
University of Bristol UK

1. Define what it means for a decision rule to be admissible, and illustrate with a diagram containing both admissible and inadmissible decision rules. State Wald's Complete Class Theorem. Explain why an inadmissible decision rule is not necessarily dominated by a Bayes rule. (Hint: a diagram similar to that used in the proof of Complete Class Theorem might help.) [15 marks]

   Answer. As usual, let the model be $Y \sim \{f, \Omega\}$. Define the risk function for decision rule δ as
   $$R(\delta, \theta) := \mathrm{E}\{L(\delta(Y), \theta); \theta\}.$$
   A rule is admissible exactly when it is not dominated by any other rule. That is, δ is admissible exactly when there is no other rule δ' satisfying
   $$R(\delta', \theta) \le R(\delta, \theta) \quad \text{for all } \theta \in \Omega$$
   with a strict inequality for at least one θ. See the diagram at the end. The righthand panel shows an example of an inadmissible rule which is not dominated by a Bayes rule. But I appreciate that the wording of my question was not very clear: "not dominated by every Bayes rule" is what I meant.

2. Give a simple direct proof that all Bayes Rules with strictly positive prior probabilities are admissible (hint: use proof by contradiction). [10 marks]

   Answer. Suppose that δ is a Bayes Rule which is inadmissible. In this case there is another rule δ' which dominates it. Hence, computing the expected loss

$$\begin{aligned}
\mathrm{E}\{L(\delta(Y),\theta)\} &= \mathrm{E}\{\mathrm{E}[L(\delta(Y),\theta) \mid \theta]\} && \text{Law of iterated expectation} \\
&= \mathrm{E}\{R(\delta,\theta)\} && \text{definition of Risk function} \\
&> \mathrm{E}\{R(\delta',\theta)\} && \text{Monotonicity, see below} \\
&= \mathrm{E}\{L(\delta'(Y),\theta)\} && \text{reversing previous steps}
\end{aligned}$$

which is a contradiction, because it shows that δ could not have been a Bayes Rule. The inequality follows from the Monotonicity property of expectations: if $X \geq Y$, then $\mathrm{E}(X) \geq \mathrm{E}(Y)$. In this case, because the prior probabilities are strictly positive, the inequality must be strict, because then there will be positive probability on a θ_j for which $R(\delta, theta_j) > R(\delta', \theta_j)$.

3. In the case of point estimation, describe the behaviour of the Bayes Rule for the loss function

$$L(\tilde{\theta},\theta) = 1 - \mathbb{1}_{|\tilde{\theta}-\theta|\leq\epsilon}$$

in the limit as $\epsilon \downarrow 0$. For convenience, consider the case where $\Omega = \mathbb{R}$; you might find a diagram helpful. [10 marks]

Answer. According to the Bayes Rule Theorem, the Bayes Rule for fixed $\epsilon > 0$ is

$$\begin{aligned}
\theta^*(y) &= \operatorname*{argmin}_{\tilde{\theta}\in\Omega} \mathrm{E}\{1 - \mathbb{1}_{|\tilde{\theta}-\theta|\leq\epsilon} \mid Y = y\} \\
&= \operatorname*{argmin}_{\tilde{\theta}\in\Omega} \big(1 - \Pr\{|\tilde{\theta}-\theta| \leq \epsilon \mid Y = y\}\big) \\
&= \operatorname*{argmax}_{\tilde{\theta}\in\Omega} \Pr\{\tilde{\theta}-\epsilon \leq \theta < \tilde{\theta}+\epsilon \mid Y = y\}
\end{aligned}$$

which is the posterior probability content of a strip of width ϵ centred around $\tilde{\theta}$. This probability is maximised when $\mathrm{p}(\tilde{\theta}-\epsilon \mid y) = \mathrm{p}(\tilde{\theta}+\epsilon \mid y)$, because otherwise it would be better to move $\tilde{\theta}$ slightly to the left or the right. As $\epsilon \downarrow 0$, this strip gets narrower. See the diagram at the end.

4. In the case of hypothesis testing, show that if the loss function is zero-one

then the Bayes Rule is to select the hypothesis with the largest posterior probability. [5 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

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

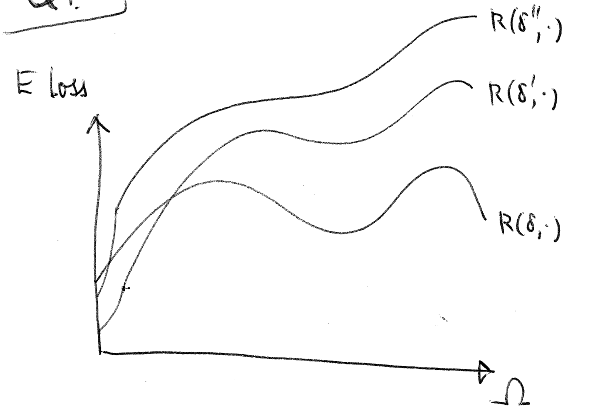
as claimed in the question.

5. An analyst at a hospital has 'failed to reject' the hypothesis that mortality rates from emergency surgery are the same on weekdays and weekend-days. He claims that this proves that there is no excess mortality at the weekend. What do you make of his claim? [15 marks]

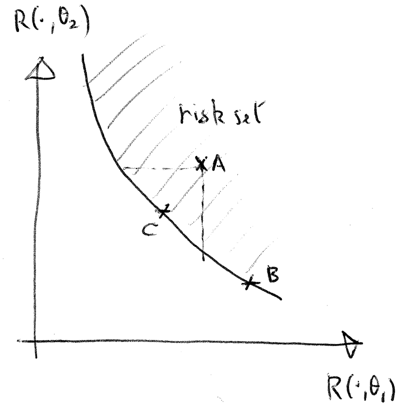
   Answer. Not much, to be honest. What he has actually found is that $C(y)$ intercepts $H_0 : \mu_W = \mu_D$, where the parameter space contains (μ_W, μ_D), the mortality rates for weekdays and weekend-days, and maybe other parameters as well. Consider six different ways this could occur (see diagram at the end)—would you draw the same conclusion as the analyst from all of these? Wouldn't it have been so much more informative if he had drawn the picture?!

   We would also need more details about the nature of the set estimator C. For example, what is its level, and how is it constructed?

Q1.

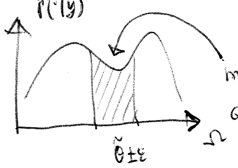


δ'' is dominated by δ' but not by δ.

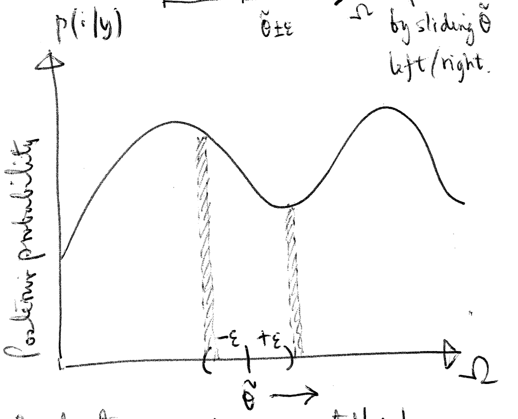


A is inadmissible, being dominated by C. But A is not dominated by B. Both B and C are Bayes rules.

Q3.



make as large as possible by sliding $\tilde\theta$ left/right.

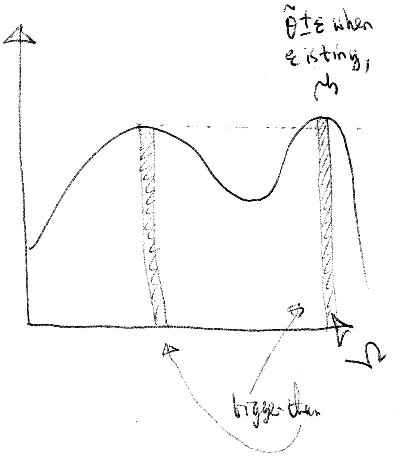


optimal outcome when the two columns have the same height.

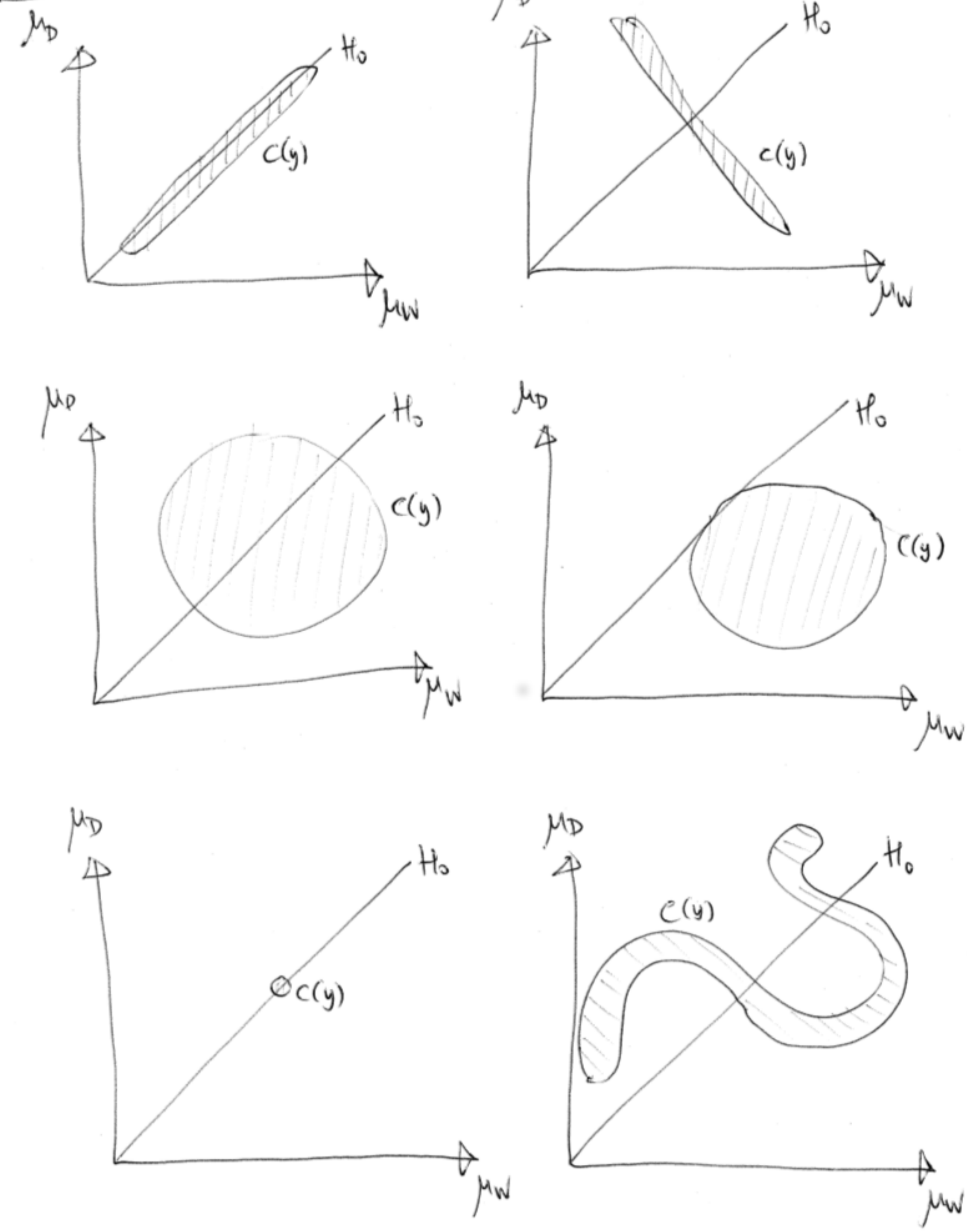
expected loss | y goes up, because probability that leaves $\Pr(\theta \in \tilde\theta \pm \varepsilon \mid y)$ is ~~smaller~~ larger than probability that enters

$p(\cdot|y)$

argmin at $\tilde\theta \pm \varepsilon$ when ε is tiny,



Q5


μ_D
H_0
$C(y)$
μ_W
μ_D
H_0
$C(y)$
μ_W
μ_D
H_0
$C(y)$
μ_W
μ_D
H_0
$C(y)$
μ_W
μ_D
H_0
$C(y)$
μ_W
μ_D
H_0
$C(y)$
μ_W