Technische Universiteit Delft — Mekelweg 4
Faculteit Elektrotechniek, Wiskunde en Informatica — 2628 CD Delft

Exam Statistical Inference (WI4455)
January 21, 2015, 9.00–12.00

Using books or notes is not allowed at the exam.
Unless stated diferently, always add an explanation to your answer.

1. The risks for five decision processes $\delta_1, \dots, \delta_5$ 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
$0 \le \theta < 1$	10	10	7	6	8
$1 \le \theta < 2$	8	11	8	5	10
$2 \le \theta$	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?

2. Suppose we have a single observation, X, which comes from a distribution with density function f_θ, with $\theta \in \{0, 1\}$ and we want to test

$$H_0 \,:\, f(x) = f_0(x) = 2(1-x)\mathbf{1}_{[0,1]}(x)$$

against

$$H_1 \,:\, f(x) = f_1(x) = 2x\mathbf{1}_{[0,1]}(x)$$

   (a) Using Neyman-Pearson, show that the best critical region for the likelihood ratio test of H_0 versus H_1 is given by $X \ge B$ for some constant B.
   (b) Consider now choosing B using decision theory. Suppose the losses incurred by a type II error is four times the loss of a type I error. Consider decision rules d_B which choose H_1 if $X \ge B$.
      i. Write down the loss function when considering the action space is $\mathcal{A} = \{a_0, a_1\}$ with $a_0 = \{\text{accept } H_0\}$ and $a_1 = \{\text{accept } H_1\}$.
      ii. Calculate the risks $R(0, d_B)$ and $R(1, d_B)$ as functions of B. Use this to find the value of B which gives the minimax rule.
      iii. Calculate the Bayes risk, when the prior probabilities are $1/4$ and $3/4$ for H_0 and H_1 respectively, and find the value of B which gives the Bayes rule.

3. Let $X_1, \dots, X_n$ be independent and identically distributed random variables, each with the $N(\mu, \sigma^2)$-distribution, where $\mu \in \mathbb{R}$ is known and σ^2 is the unknown parameter. Consider

$$T = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}.$$

(a) Compute $I_T(\sigma^2)$, which is the Fisher information about σ^2 based on T.
Hint: Use that $\frac{n-1}{\sigma^2}T \sim Ga\left(\frac{n-1}{2}, \frac{1}{2}\right)$. *The density of the Gamma distribution* $Ga(\alpha, \beta)$ *is given by*

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$

(b) Prove that T is not a sufficient statistic.
Hint: A statistic is sufficient if and only if it preserves the Fisher information.

4. Assume the following hierarchical model:

$$X_i \mid \theta_i \overset{\text{ind}}{\sim} Pois(\theta_i)$$
$$\theta_i \mid \beta \overset{\text{ind}}{\sim} Exp(\beta)$$
$$\pi(\beta) \sim Exp(2)$$

Here $i = 1, \dots, n$.
Derive the Gibbs-sampler for drawing from the posterior of $(\theta_1, \dots, \theta_n, \beta)$.

If $Y \sim Pois(\theta)$, *then* $f_Y(y) = e^{-\theta}\theta^y/y!$, *for* $y = 0, 1, \dots$. *If* $Z \sim Exp(\lambda)$, *then* $f_Z(z) = \lambda e^{-\lambda z} \mathbf{1}_{[0,\infty)}$.

5. Let $\hat{\theta}$ be an unbiased estimator of an unknown parameter $\theta \in \mathbb{R}$. Assuming $\theta \neq 0$, consider the loss function

$$L(\theta, a) = \frac{(a-\theta)^2}{\theta^2}.$$

Assume $0 \leq R(\theta, T) < \infty$ for any estimator T. Show that $\hat{\theta}$ is not minimax.

Hint: Consider the estimator $T = c\hat{\theta}$ *for some particular constant* $c \in (0, 1)$.

6. Let g be a positive integrable function on $(0, \infty)$. Suppose $X_1, \dots, X_n$ are independent random variables with common density f, given by

$$f(x) = \frac{g(x)}{\int_\theta^\infty g(x)dx} \mathbf{1}_{[\theta,\infty)}(x).$$

(a) Show that $X_{(1)} = \min\{X_1, \dots, X_n\}$ is sufficient.

(b) Show that $X_{(1)}$ is minimal sufficient.

7. Suppose

(a) d_0 is extended Bayes: for each $\varepsilon > 0$ there exists a prior π_ε such that

$$r(\pi_\varepsilon, d_0) \leq \varepsilon + \inf_d r(\pi_\varepsilon, d),$$

(b) $R(\theta, d_0)$ is constant for all θ.

Then d_0 is minimax.

Answers

1. (a) A decision rule δ_i $(1,2,3,4,5)$ is at least as good as δ_1 if

$$R(\theta, \delta_i) \leq R(\theta, \delta_1)$$

for all i. This holds for δ_3 and δ_4.

(b) Inadmissible rules are rules that are dominated by another rule.

- δ_1 is dominated by δ_3 and hence inadmissible.
- δ_2 is best when $\theta \geq 2$ so it cannot be inadmissible.
- δ_3 is only dominated by δ_4 when $0 \leq \theta < 1$, but it is better than δ_4 when $\theta \geq 2$. Hence it cannot be inadmissible.
- δ_4 is best when $0 \leq \theta < 2$, so it cannot be inadmissible.
- δ_5 is dominated by δ_4 and hence inadmissible.

We conclude that δ_2, δ_3 and δ_4 are admissible.

(c) We have

	δ_1	δ_2	δ_3	δ_4	δ_5
$\max_{\theta\geq 0} R(\theta, \delta_i)$	15	11	12	14	14

Hence the minimax rule is given by δ_2.

(d) The Bayes risk for δ_1 is given by $(10 + 8 + 3 \times 15)/5 = 63/5$. Hence we find

	δ_1	δ_2	δ_3	δ_4	δ_5
$5\,r(\pi, \delta_i)$	63	54	51	53	60

So δ_3 is the Bayes rule under this prior (denoted by π).

2. (a) The NP test rejects for large values of

$$\frac{f_1(x)}{f_0(x)} = \frac{x}{1-x}.$$

Now $g(x) = x/(1-x)$ is increasing on $(0,1)$ since $g'(x) = (1-x)^{-2} > 0$. Hence we reject for large values of x.

(b) i. The decision rules we consider are given by

$$d_B(x) = \begin{cases} a_0 & x < B \\ a_1 & x \geq B \end{cases}.$$

ii. The loss function is given by

$$L(\theta, a_0) = \begin{cases} 0 & \theta = 0 \\ 4 & \theta = 1 \end{cases}$$

and

$$L(\theta, a_1) = \begin{cases} 0 & \theta = 1 \\ 1 & \theta = 0 \end{cases}$$

as the loss of a type II error is four times the loss of a type I error.

iii. The risk function is given by

$$R(0, d_B) = E_0 L(0, d_B) = P_0(d_B(X) = a_1)$$
$$= P_0(X \geq B) = \int_B^1 2(1-x)dx = (B-1)^2.$$

and

$$R(1, d_B) = E_1 L(1, d_B) = 4P_1(d_B(X) = a_0)$$
$$= 4P_1(X < B) = 4\int_0^B 2x dx = 4B^2.$$

The minimax rule is determined by $R(0, d_B) = R(1, d_B)$ which gives the equation

$$4B^2 = (1-B)^2.$$

This has one solution in $[0, 1]$, which is given by $B^* = 1/3$. Hence the minimax rule is d_{B^*}.

iv. The Bayes risk is given by

$$r(\pi, d_B) = \frac{1}{4}(B-1)^2 + \frac{3}{4}4B^2$$

which is convex. Hence

$$\frac{r(\pi, d_B)}{dB} = \frac{1}{2}(B-1) + 6B.$$

Equating to zero gives $B = 1/13$.

3. (a) We calculate $I_T(\sigma^2)$. Since, $T = cY$, where $c = \sigma^2/(n-1)$ and $Y \sim Ga\left(\frac{n-1}{2}, \frac{1}{2}\right)$, using a change of variable we find that $f_T(x) = c^{-1} f_Y(x/c)$. Thus

$$L(t; \sigma^2) = \frac{n-1}{\sigma^2} \frac{(1/2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} t^{\frac{n-3}{2}} \left(\frac{n-1}{\sigma^2}\right)^{\frac{n-3}{2}} e^{-\frac{n-1}{2\sigma^2}t}$$

and

$$\log L(t; \sigma^2) = -\frac{n-1}{2\sigma^2}t + \frac{n-1}{2}\log(n-1) - \frac{n-1}{2}\log\sigma^2 + \log\frac{(1/2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} + \frac{n-3}{2}\log t.$$

Consequently,

$$\frac{d^2}{d(\sigma^2)^2}\log L(t; \sigma^2) = \frac{n-1}{2\sigma^4} - \frac{(n-1)t}{\sigma^6}$$

and we obtain

$$I_T(\sigma^2) = -\frac{n-1}{2\sigma^4} + \frac{1}{\sigma^4}(n-1) = \frac{n-1}{2\sigma^4}.$$

(b) Let $X = (X_1, \ldots, X_n)$. We calculate the Fisher Information $I_X(\sigma^2)$. The log-likelihood function is

$$\log L(x;\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Hence

$$\frac{d^2}{d(\sigma^2)^2}\log L(x;\sigma^2) = \frac{n}{2\sigma^4} - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^6}$$

and

$$I_X(\sigma^2) = \frac{n}{2\sigma^4}.$$

From $I_T(\sigma^2) < I_X(\sigma^2)$ we derive that T is not a sufficient statistic.

4. Using Bayesian notation

$$f(x_1,\ldots,x_n,\theta_1,\ldots,\theta_n,\beta) = \prod_{i=1}^{n}\left(e^{-\theta_i}\frac{\theta_i^{x_i}}{x_i!}\beta e^{-\beta\theta_i}\right)2e^{-2\beta}.$$

Hence

$$f(\theta_i \mid \theta_{-i},\beta,x_1,\ldots,x_n) \propto \theta_i^{x_i}e^{-(1+\beta)\theta_i},$$

which shows that each θ_i can be simulated conditional on β and x_i by drawing from the $Ga(x_i+1, 1+\beta)$-distribution. For updating β, note that

$$f(\beta \mid \theta_1,\ldots,\theta_n,x_1,\ldots,x_n) \propto \beta^n e^{-\beta\sum\theta_i}e^{-2\beta},$$

which show that β can be simulated from the $Ga(n+1, 2+\sum\theta_i)$-distribution.

5. We consider an estimator T given by $T = c\,\hat\theta$. In this case the risk function is

$$\begin{aligned} R(\theta, c\hat\theta) &= E_\theta\left[\frac{(c\hat\theta-\theta)^2}{\theta^2}\right] = E_\theta\left[\frac{(c\hat\theta - c\theta + c\theta - \theta)^2}{\theta^2}\right] \\ &= (1-c)^2 + c^2 R(\theta,\hat\theta). \end{aligned}$$

The double product term vanishes since $\hat\theta$ is unbiased. We look for a constant c such that

$$\sup_{\theta\in\mathbb{R}} R(\theta, c\hat\theta) < \sup_{\theta\in\mathbb{R}} R(\theta,\hat\theta),$$

which is equivalent to

$$\sup_{\theta\in\mathbb{R}} R(\theta,\hat\theta) > \frac{(1-c)^2}{1-c^2} = \frac{1-c}{1+c} =: g(c). \tag{1}$$

Since $g : [0,1] \to [0,1]$ is continuous and decreasing, (1) holds for all $c \in (0,1)$ if $\sup_\theta R(\theta,\hat\theta_n) > 1$. If $\sup_\theta R(\theta,\hat\theta_n) \le 1$, then there exists a c^* such that (1) holds for all $c \ge c^*$.

6. (a) The joint density is zero unless all $x_i \geq \theta$. This is equivalent to saying that $x_{(1)} > \theta$. Hence

$$f(x_1, \dots, x_n \mid \theta) = \mathbf{1}_{[\theta,\infty)}(X_{(1)}) \prod_{i=1}^{n} \frac{g(x_i)}{\int_\theta^\infty g(x)dx}.$$

Sufficiency follows from the factorisation theorem.

(b) Suppose

$$f(x_1, \dots, x_n \mid \theta) = f(y_1, \dots, y_n \mid \theta)\psi(x_1, \dots, x_n, y_1, \dots, y_n).$$

then the region where the two functions are zero must agree. Hence this implies $x_{(1)} = y_{(1)}$ from which minimality follows.

7. This is theorem 2.2 from YS.

Suppose $R(\theta, d_0) = C$. Suppose d_0 is not minimax. Then there exists a rule d' for which $\sup_\theta R(\theta, d') < C$. So let $\sup_\theta R(\theta, d') = C - \varepsilon$ for some $\varepsilon > 0$. As d_0 is extended Bayes, we can find a prior π_ε such that

$$r(\pi_\varepsilon, d_0) < \inf_d r(\pi_\varepsilon, d) + \varepsilon/2 \leq r(\pi_\varepsilon, d') + \varepsilon/2 \leq C - \varepsilon + \varepsilon/2 = C - \varepsilon/2.$$

Now since d_0 is an equaliser rule $r(\pi_\varepsilon, d_0) = C$ and we have reached a contradiction.