Exam Statistical Inference (WI4455)
January 23, 2020, 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. Let g be a density function supported on $[0,1]$ (i.e. its density is zero outside $[0,1]$). Denote $\mu = \int x g(x) \mathrm{d}\, x$ and $\sigma^2 = \int (x-\mu)^2 g(x) \mathrm{d}\, x$. *In this exercise we consider g fixed and known.*

   Let $\theta \in [0,1]$ and define the probability density

$$f(x \mid \theta) = \begin{cases} \theta + (1-\theta) g(x) & \text{if } x \in [0,1] \\ 0 & \text{if } x \notin [0,1] \end{cases}.$$

   Assume $X_1, \dots, X_n$ are independent, each with density f.

   (a) [1 pt]. Give an expression for the score-function (this is the derivative of the log-likelihood).

   (b) [1 pt]. Show that $\mathrm{E}_\theta X_1 = \Psi_\mu(\theta)$, with $\Psi_\mu(\theta) = \mu + (1-\mu)\theta$.

   (c) [2 pt]. Define the estimator $\hat{\Theta}_n$ by the relation $\bar{X}_n = \Psi_\mu(\hat{\Theta}_n)$. Derive an expression for $\hat{\Theta}_n$ and show it is unbiased for θ.

   (d) [3 pt]. Using the central limit theorem, give the limiting distribution of $\sqrt{n}(\hat{\Theta}_n - \theta)$ under $\mathbb{P}_\theta$ and show that if $\mu < 1$

$$\lim_{n\to\infty} \mathbb{P}_\theta(\hat{\Theta}_n < 0) = 0.$$

   (e) [2 pt]. In the remainder of the exercise we assume $n = 1$, so just one observation X_1. Show that the maximum likelihood estimator is given by

$$\hat{\Theta}_{\text{MLE}} = \begin{cases} 0 & \text{if} \quad g(X_1) > 1 \\ 1 & \text{if} \quad g(X_1) < 1 \end{cases}$$

   (f) [2 pt]. We shift perspective to the Bayesian view. Hence we consider $X_1 \mid \Theta = \theta \sim f(x \mid \theta)$ and employ a prior on the parameter θ that is supported on $[0,1]$ with density denoted by f_Θ. Show that the posterior density satisfies

$$f_{\Theta|X_1}(\theta \mid x) = \frac{f(x \mid \theta) f_\Theta(\theta)}{\pi_1 + (1-\pi_1) g(x)} \mathbf{1}_{[0,1]}(\theta).$$

where π_1 is the prior mean.

(g) [2 pt]. Suppose the prior on Θ is taken to be the uniform distribution on $[0,1]$. Express the posterior mean in terms of $g(X_1)$.

2. Let $\theta \in (0,\infty)$ be an unknown parameter and X be a random variable such that $\mathrm{E}_\theta X = \theta$ and $\mathrm{var}_\theta X = \nu(\theta)$, where $\nu(\theta)$ is known and specified below. Consider estimation of θ by a decision rule within the class $\mathcal{D}$ defined by

$$\mathcal{D} = \{d_a(X) = aX,\ a \in (0,1]\}.$$

Assume squared error loss, that is, $L(\theta, d_a) = (\theta - aX)^2$.

(a) [3 pt]. For $\nu(\theta) = \theta^2$, calculate the risk function of d_a, and show that there is a value of a which is optimal, no matter the value of θ.

(b) [1 pt]. Show that d_1 is inadmissible (for the given loss-function). *Hint: consider also* $d_{1/2}$.

(c) [3 pt]. Suppose $\nu(\theta) = \theta^k$ where k is a positive integer. Show that the Bayes risk of the decision rule d_a is given by $a^2\, k! + 2(a-1)^2$, when Θ has prior density $f_\Theta(\theta) = e^{-\theta}\mathbf{1}_{[0,\infty)}(\theta)$. In addition, compute the Bayes decision rule.

You can use the fact that $\int_0^\infty x^n e^{-x} \mathrm{d}\, x = n!$ *for positive integers* n.

(d) [1 pt]. Suppose again that $\nu(\theta) = \theta^2$. Are the minimax rule and Bayes rule (that you derived in part (c)) the same?

(e) [1.5 pt]. Show that the Bayes rule does not depend on the chosen prior distribution on Θ if $\nu(\theta) = \theta^2$.

3. Consider the following hierarchical model

$$\begin{aligned} X_1, \ldots, X_n \mid \Theta = \theta &\overset{\text{ind}}{\sim} Pois\,(\theta) \\ \Theta &\sim Ga\,(\alpha, \beta)\,, \end{aligned}$$

where $Ga\,(\alpha, \beta)$ denotes the Gamma-distribution with parameters α and β. That is,

$$f_\Theta(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \mathbf{1}_{[0,\infty)}(\theta),$$

where Γ denotes the Gamma function. Recall that under the specified model $\mathrm{P}_\theta(X_i = x) = e^{-\theta} \frac{\theta^x}{x!}$, when $x \in \{0, 1, \ldots\}$.

(a) [3 pt]. Show that the posterior distribution of Θ is a Gamma distribution. Specify its parameters.

(b) [2 pt]. Show that the marginal density of $X = (X_1, \ldots, X_n)$ equals

$$f_X(x) = \frac{\beta^\alpha \Gamma(\alpha + s)}{\Gamma(\alpha) \prod_{i=1}^n (x_i!)(\beta + n)^{\alpha+s}},$$

where $s = \sum_{i=1}^n x_i$.

(c) [3 pt]. Assume $\alpha = 2$ and that we further endow β with a prior distribution with density $p(\beta) = e^{-\beta} \mathbf{1}_{[0,\infty)}(\beta)$. Give the steps of the Gibbs sampler for sampling from the posterior distribution of (θ, β).

4. Suppose $X \sim Pois\,(\theta)$.

(a) [2 pt].Verify that $\varphi(X)$ is an unbiased estimator for $e^{-3\theta}$ if

$$\sum_{k=0}^{\infty} \varphi(k) \frac{\theta^k}{k!} = e^{-2\theta}.$$

(b) [2 pt]. Prove that $(-2)^X$ is UMVU for θ. *Hint: You may use the trivial fact that X is a complete and sufficient statistic for θ.*

Solutions

1. (a) The loglikelihood is given by

$$\ell(\theta) = \sum_{i=1}^{n} \log f(x_i \mid \theta).$$

Hence

$$s(\theta) = \sum_{i=1}^{n} \frac{1 - g(x)}{\theta + (1-\theta)g(x)}.$$

(b)

$$\mathbb{E}_\theta = \theta + (1-\theta)\int xg(x)\,\mathrm{d}\,x = \theta + (1-\theta)\mu = \mu + (1-\mu)\theta.$$

(c) We have $(1-\mu)\hat{\Theta}_n + \mu = \bar{X}_n$. This gives

$$\hat{\Theta}_n = \frac{\bar{X}_n - \mu}{1-\mu}.$$

We then have

$$\mathbb{E}_\theta \, \hat{\Theta}_n = \frac{\mathbb{E}_\theta \bar{X}_n - \mu}{1-\mu} = \frac{\theta + (1-\theta)\mu - \mu}{1-\mu} = \theta.$$

(d) Note that

$$\operatorname{var}_\theta \hat{\Theta}_n = \frac{\sigma^2}{n(1-\mu)^2}.$$

By the CLT we have

$$\sqrt{n}(\hat{\Theta}_n - \theta) \xrightarrow{\mathrm{w}} N(0, \sigma^2/(1-\mu)^2).$$

Hence

$$\mathbb{P}_\theta(\hat{\Theta}_n < 0) = \mathbb{P}_\theta\left(\sqrt{n}\frac{1-\mu}{\sigma}(\hat{\Theta}_n - \theta) < -\sqrt{n}\frac{1-\mu}{\sigma}\theta\right) \approx \Phi\left(-\sqrt{n}\frac{1-\mu}{\sigma}\theta\right).$$

(e) The likelihood is $L(\theta) = \theta + (1-\theta)g(x)$. Hence $L'(\theta) = 1 - g(x)$. So if $g(x) > 1$ the likelihood is decreasing and the maximiser is at 0; if $g(x) < 1$, then the likelihood is increasing and then the maximiser is at 1.

(f) The posterior density satisfies

$$f_{\Theta|X}(\theta \mid x) \propto f_\Theta(\theta) f(x \mid \theta).$$

The normalising constant is obtained by integrating the RHS over θ and equals henceforth $\pi_1 + (1-\pi_i)g(x)$.

(g) As $\pi_1 = 1/2$ the posterior mean equals

$$\frac{\int_0^1 \theta^2 \mathrm{d}\,\theta + g(X_1)\int_0^1 \theta(1-\theta)\mathrm{d}\,\theta}{1/2 + g(X_1)/2} = \frac{1/3 + g(X_1)/6}{1/2 + g(X_1)/2} = \frac{2+g(X_1)}{3+3g(X_1)}.$$

2. (a) First note that

$$R(\theta, d_a) = \mathrm{E}_\theta(d_a - \theta)^2 = \mathrm{E}_\theta(aX - \theta)^2.$$

Using the bias-variance decomposition of the Mean-Squared-Error, this is seen to be equal to

$$\left(\mathrm{E}_\theta(aX-\theta)\right)^2 + \mathrm{var}_\theta(aX) = (a-1)^2\theta^2 + a^2\theta^2 = (2a^2 - 2a + 1)\theta^2.$$

This is a strictly convex function in a and minimised for $a = 1/2$.

(b) When $a = 1/2$, the risk equals $\theta^2/2$, whereas for $a = 1$ we get risk θ^2. As $\theta > 0$ this implies d_1 is inadmissible (for the given loss-function).

(c) The Bayes risk is obtained by weighting the risk with respect to the prior. Hence

$$\begin{aligned} r(f_\Theta, d_a) &= \int R(\theta, d_a) f_\Theta(\theta)\mathrm{d}\,\theta \\ &= \int_0^\infty \left((a-1)^2\theta^2 + a^2\theta^k\right) e^{-\theta}\mathrm{d}\,\theta \\ &= 2(a-1)^2 + a^2 k! \end{aligned}$$

The Bayes rule follow upon minimising the Bayes risk over $a \in (0,1]$. Setting the derivative with respec to a to zero gives $2a - 2 + ak! = 0$ and hence $a = 2/(2+k!)$. The Bayes rule is henceforth given by

$$d_{\text{Bayes}}(X) = \frac{2}{2+k!}X.$$

(d) First note that the result of exercise (a) says that $d_{1/2}$ is minimax. From exercise (c) it follows that the Bayes rule and minimax rule are the same.

(e) If $k = 2$ we have

$$r(f_\Theta, d_a) = \left((a-1)^2 + a^2\right) \int_0^\infty \theta^2 f_\Theta(\theta) \mathrm{d}\,\theta$$

and in that case the Bayes rule is the same for all priors.

3. (a) We have

$$L(\theta \mid X) = \prod_{i=1}^n e^{-\theta} \frac{\theta^{X_i}}{X_i!} \propto e^{-n\theta} \theta^{\sum_{i=1}^n X_i}.$$

Hence

$$p(\theta \mid X) \propto L(\theta \mid X) f_\Theta(\theta) \propto \theta^{a + \sum_{i=1}^n X_i - 1} e^{-(b+n)}.$$

Hence the posterior for Θ is $Ga\left(a + \sum_{i=1}^n X_i, b+n\right)$.

(b) We have

$$\begin{aligned} f_X(x) &= \int f_{X|\Theta}(x \mid \theta) f_\Theta(\theta) \mathrm{d}\,\theta \\ &= \frac{b^a}{\Gamma(a)} \frac{1}{\prod_{i=1}^n x_i!} \int \theta^{a+s-1} e^{-(b+n)\theta} \mathrm{d}\,\theta \\ &= \frac{b^a \Gamma(a+s)}{\Gamma(a) \prod_{i=1}^n (x_i!)(b+n)^{a+s}} \end{aligned}$$

(c) We use Bayesian notation

$$p(\theta, \beta \mid x) \propto \underbrace{e^{-n\theta}\theta^s}_{\text{likelihood}} \times \underbrace{b^2 \theta e^{-b\theta}}_{\text{prior on } \theta \mid b} \times \underbrace{e^{-b}}_{\text{prior on } b}.$$

The Gibbs sampler consists of two steps

- Updating θ given b (and x): this amounts to drawing from the $Ga\,(2+s, b+n)$ distribution.
- Updating b given θ (and x): this amounts to drawing from a density proportional to $b^2 e^{-(\theta+1)b}$, i.e. drawing from the $Ga\,(3, \theta+1)$ distribution.

4. (a) The estimator $\varphi(X)$ should satisfy

$$\mathrm{E}_\theta\, \varphi(X) = \sum_{k=0}^\infty \varphi(k) e^{-\theta} \frac{\theta^k}{k!} = e^{-\theta} \sum_{k=0}^\infty \varphi(k) \frac{\theta^k}{k!} = e^{-3\theta}.$$

Hence, we should have

$$\sum_{k=0}^{\infty} \varphi(k) \frac{\theta^k}{k!} = e^{-2\theta}.$$

This is exactly when $\varphi(k) = (-2)^k$.

(b) The estimator

$$\varphi(X) = (-2)^X$$

is unbiased for $e^{-3\theta}$ and depends on the complete and sufficient statistic X. The result follows from the Lehmann-Scheffé theorem. This is a terrible estimator!