Department of Mathematics, Stockholm University

Exam in MT5017, Theory of Statistical Inference, October 27, 2025, 14:00–19:00.

Examiner: Filip Lindskog, lindskog@math.su.se

Allowed aids: None.

Return: Communicated via course forum.

Arguments and computations should be clear and easy to follow.

- - - - - - - - PART 1 - - - - - - - -

Problem 1

Let $X_1, \dots, X_n$ be independent copies of $X \sim f(x;\theta)$, where $f(x;\theta)$ denotes the density function $f(x;\theta) = \theta e^{-\theta x}$ for $x > 0$, and $f(x;\theta) = 0$ for $x \le 0$.

(a) Determine the MLE of θ. (3 p)

(b) Determine the observed Fisher information of θ. (3 p)

(c) Determine the score statistic of θ. (5 p)

(d) State the asymptotic distribution, as $n \to \infty$, of the score statistic of θ and use it to formulate an approximate 95% confidence interval for θ. (9 p)

Problem 2

Consider repeated independent Bernoulli trials with, for each trial, the probability θ of the outcome 1. Let X be the number of trials until and including the first outcome of the value 1. Let $X_1, \dots, X_n$ be independent copies of X.

(a) Determine the MLE of θ. (7 p)

(b) Determine a one-dimensional statistic $h(X_1, \dots, X_n)$ that is sufficient for θ. (8 p)

Problem 3

Consider data $x_1, \dots, x_n$ and a likelihood corresponding to the x_k, $k = 1, \dots, n$, being realizations of independent Poisson(λ)-distributed random variables with probability mass function

$$x \mapsto \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \dots, \quad \lambda > 0.$$

(a) Show that a conjugate prior for λ is given by the Gamma(α, β) distribution with density function

$$\lambda \mapsto \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda > 0, \quad \alpha, \beta > 0$$

with prior mean $\mathrm{E}[\lambda] = \alpha/\beta$. Determine the parameters of the posterior distribution and its mean when the conjugate prior is used. (8 p)

(b) Determine the Jeffreys' prior for λ and determine the corresponding posterior distribution and its mean. (7 p)

- - - - - - - - PART 2 - - - - - - - -

Problem 4

Let $X_1, \dots, X_n$ be independent copies of $X \sim f(x;\theta)$, where $f(x;\theta)$ denotes the density function $f(x;\theta) = \theta e^{-\theta x}$ for $x > 0$, and $f(x;\theta) = 0$ for $x \le 0$. Let $\phi = \mathrm{P}_\theta(X > 1)$, where the subscript θ means that the probability is computed assuming that the parameter value is θ. Let $\widehat{\phi} = \mathrm{P}_{\widehat{\theta}}(X > 1)$, where $\widehat{\theta}$ is the MLE based on $X_1, \dots, X_n$.

(a) Determine σ^2 satisfying $\sqrt{n}(\widehat{\phi} - \phi) \xrightarrow{d} N(0, \sigma^2)$ as $n \to \infty$. (10 p)

(b) Determine an approximate pivot and use it to compute the p-value for testing the null hypothesis is $H_0 : \phi = 1/2$ against the alternative hypothesis is $H_1 : \phi \neq 1/2$, and where $n = 100$ and the observed MLE is $\widehat{\phi} = 0.4$. Express the p-value in terms of given numerical values and the standard normal distribution function $x \mapsto \Phi(x)$. (15 p)

Problem 5

Let $X_1, \dots, X_n$ be independent copies of $X \sim f(x;\theta)$. Let $\widehat{\theta}_g = h(X_1, \dots, X_n)$ be an unbiased estimator of $g(\theta)$, where g is differentiable.

(a) State the Cramér-Rao lower bound applied to the estimator $\widehat{\theta}_g$. (5 p)

(b) Prove the Cramér-Rao lower bound. (14 p)

(c) Let $\widehat{\theta} = \widetilde{h}(X_1, \dots, X_n)$ be an estimator of θ with bias $b(\theta) = \mathrm{E}[\widehat{\theta} - \theta]$. Use the Cramér-Rao lower bound to give a lower bound of the mean squared error $\mathrm{E}[(\widehat{\theta} - \theta)^2]$. (6 p)

Problem 1

$$L(\theta; x_1, \ldots, x_n) = \theta^n e^{-\theta \sum_{k=1}^n x_k}, \quad l(\theta; x_1, \ldots, x_n) = n \log \theta - \theta \sum_{k=1}^n x_k,$$

$$S(\theta; x_1, \ldots, x_n) = n/\theta - \sum_{k=1}^n x_k, \quad \widehat{\theta} = n / \sum_{k=1}^n x_k, \quad I(\theta; x_1, \ldots, x_n) = n/\theta^2.$$

Observed Fisher is $n/\widehat{\theta}^2$. The score statistic is

$$\frac{S(\theta; X_1, \ldots, X_n)}{\sqrt{nJ_X(\theta)}}, \quad J_X(\theta) = \frac{1}{\theta^2}$$

and

$$\frac{S(\theta; X_1, \ldots, X_n)}{\sqrt{nJ_X(\theta)}} = \frac{n/\theta - \sum_{k=1}^n X_k}{\sqrt{n}/\theta} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty.$$

Hence, approximately, with $\overline{X} = (1/n) \sum_{k=1}^n X_k$

$$\begin{aligned}
\gamma &= \mathrm{P}\left(\Phi^{-1}\Big(\frac{1-\gamma}{2}\Big) \le \frac{n - \theta \sum_{k=1}^n X_k}{\sqrt{n}} \le \Phi^{-1}\Big(\frac{1+\gamma}{2}\Big) \right) \\
&= \mathrm{P}\left(\frac{1}{\sqrt{n}} \Phi^{-1}\Big(\frac{1-\gamma}{2}\Big) \le 1 - \theta \overline{X} \le \frac{1}{\sqrt{n}} \Phi^{-1}\Big(\frac{1+\gamma}{2}\Big) \right) \\
&= \mathrm{P}\left(\frac{1}{\overline{X}} - \frac{1}{\overline{X}\sqrt{n}} \Phi^{-1}\Big(\frac{1+\gamma}{2}\Big) \le \theta \le \frac{1}{\overline{X}} + \frac{1}{\overline{X}\sqrt{n}} \Phi^{-1}\Big(\frac{1+\gamma}{2}\Big) \right)
\end{aligned}$$

Hence, an approximate 95% confidence level for θ is the interval

$$\left[\frac{1}{\overline{X}} - \frac{1}{\overline{X}\sqrt{n}} \Phi^{-1}(0.95), \frac{1}{\overline{X}} + \frac{1}{\overline{X}\sqrt{n}} \Phi^{-1}(0.95) \right]$$

Problem 2

The probability mass function for X is $f(x; \theta) = \theta(1-\theta)^{x-1}$. Hence,

$$\begin{aligned}
L(\theta; x_1, \ldots, x_n) &= \theta^n (1-\theta)^{\sum_{k=1}^n x_k - n}, \\
l(\theta; x_1, \ldots, x_n) &= n \log \theta + \Big(\sum_{k=1}^n x_k - n \Big) \log(1-\theta), \\
l'(\theta) &= \frac{n - \theta \sum_{k=1}^n x_k}{\theta(1-\theta)}
\end{aligned}$$

Hence, the MLE is $n / \sum_{k=1}^n x_k$. The factorization theorem says that t is sufficient for θ if and only if $f(x_1, \ldots, x_n; \theta) = g_1(t; \theta) g_2(x_1, \ldots, x_n)$. In the current setting this holds for $t = \sum_{k=1}^n x_k$ and $g_2 = 1$.

Problem 3

$$f(x_1,\dots,x_n \mid \lambda)f(\lambda) = \left(\prod_{k=1}^{n} \frac{\lambda^{x_k}}{x_k!}e^{-\lambda}\right)\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$
$$\propto \lambda^{\alpha+\sum_{k=1}^n x_k - 1}e^{-(\beta+n)\lambda}$$

which implies that the posterior density is that of a Gamma($\alpha + \sum_{k=1}^n x_k, \beta + n$) distribution. Since the Gamma prior distribution implies a Gamma posterior distribution, we have verified that the Gamma prior is a conjugate prior. The posterior mean is $(\alpha + \sum_{k=1}^n x_k)/(\beta+n)$.
The Poisson likelihood and loglikelihood are

$$L(\lambda; x) = \frac{\lambda^x}{x!}e^{-\lambda}, \quad l(\lambda; x) = x\log\lambda - \lambda - \log x!$$

Hence we get Fisher information $I(\lambda; x) = x/\lambda^2$ and expected Fisher $J(\lambda) = \mathrm{E}[I(\lambda; X)] = \mathrm{E}[X]/\lambda^2 = 1/\lambda$. Consequently, we get the Jeffreys' prior $f^J(\lambda) \propto \lambda^{-1/2}$ and the posterior

$$f(\lambda \mid x_1,\dots,x_n) \propto \lambda^{\sum_{k=1}^n x_k}e^{-\beta\lambda}\lambda^{-1/2} = \lambda^{1/2+\sum_{k=1}^n x_k - 1}e^{-n\lambda}$$

which we identify as a Gamma($1/2 + \sum_{k=1}^n x_k, n$) posterior distribution. The posterior mean is

$$\frac{1/2 + \sum_{k=1}^n x_k}{n} = \overline{x} + \frac{1}{2n}$$

Problem 4

$$L(\theta; x_1,\dots,x_n) = \theta^n e^{-\theta\sum_{k=1}^n x_k}, \quad l(\theta; x_1,\dots,x_n) = n\log\theta - \theta\sum_{k=1}^n x_k,$$

$$S(\theta; x_1,\dots,x_n) = n/\theta - \sum_{k=1}^n x_k, \quad \widehat{\theta} = n/\sum_{k=1}^n x_k, \quad I(\theta; x_1,\dots,x_n) = n/\theta^2.$$

$\phi = g(\theta) = \mathrm{P}_\theta(X > 1) = e^{-\theta}$. We know that $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, J(\theta)^{-1})$. Here $J(\theta) = 1/\theta^2$. Since

$$g(\widehat{\theta}) \approx g(\theta) + g'(\theta)(\widehat{\theta} - \theta)$$

we get $\sqrt{n}(g(\widehat{\theta}) - g(\theta)) \xrightarrow{d} N(0, g'(\theta)^2 J(\theta)^{-1})$. Here

$$g'(\theta)^2 J(\theta)^{-1} = e^{-2\theta}\theta^2 = (-\phi\log\phi)^2$$

and an approximate pivot is

$$T_1(\phi) = \frac{\widehat{\phi} - \phi}{-\phi\log\phi/\sqrt{n}}$$

An application of the continuous mapping theorem shows that also

$$T_2(\phi) = \frac{\widehat{\phi} - \phi}{-\widehat{\phi}\log\widehat{\phi}/\sqrt{n}}$$

is an approximate pivot. A level γ approximate confidence interval is

$$\Big\{\phi : \Phi^{-1}\Big(\frac{1-\gamma}{2}\Big) \le T_k(\phi) \le \Phi^{-1}\Big(\frac{1+\gamma}{2}\Big)\Big\}, \quad k = 1, 2.$$

The p-value is the number $1-\gamma$ so that $T_k(\phi_0)$, with $\phi_0 = 1/2$, is a boundary point of the confidence interval. Since $\widehat{\phi} = 0.4 < 0.5 = \phi_0$ this means $1 - \gamma = 2\Phi(T_k(\phi_0))$:

$$2\Phi(T_1(\phi_0)) \approx 0.004 \quad \text{and} \quad 2\Phi(T_2(\phi_0)) \approx 0.006.$$

Problem 5

(a)-(b): see e.g. the book. For (c) note that $\widehat{\theta}$ is an unbiased estimator of $\theta + b(\theta)$. Hence,

$$\mathrm{var}(\widehat{\theta}) \ge \frac{(1 + b'(\theta))^2}{nJ(\theta)} = \frac{(1 + b'(\theta))^2}{n\,\mathrm{E}[S(\theta; X)^2]}$$

and

$$\begin{aligned}
\mathrm{E}[(\widehat{\theta} - \theta)^2] &= \mathrm{E}[(\widehat{\theta} - (\theta + b(\theta)) + (\theta + b(\theta)) - \theta)^2] \\
&= \mathrm{var}(\widehat{\theta}) + b(\theta)^2 + 2\,\mathrm{E}[\widehat{\theta} - (\theta + b(\theta))]b(\theta) \\
&\ge b(\theta)^2 + \frac{(1 + b'(\theta))^2}{n\,\mathrm{E}[S(\theta; X)^2]}.
\end{aligned}$$