Proposed solution to konteeksamen, STK4011

Emil Aas Stoltenberg

January 14, 2020

Exercise 1

We are given the density,

$$f(x;\theta,b) = \frac{x^{1/\theta-1}}{\theta b^{1/\theta}}, \quad \text{for } 0 \le x \le b,$$

and zero otherwise, where $\theta, b > 0$.

(a) The mean of an $X \sim f(x;\theta,b)$ is

$$\mathrm{E}\,X = \int_0^b \frac{x^{1/\theta}}{\theta b^{1/\theta}}\,\mathrm{d}x = \frac{b}{\theta+1},$$

and the variance is

$$\mathrm{Var}\,X = \int_0^b \frac{x^{1/\theta+1}}{\theta b^{1/\theta}}\,\mathrm{d}x - \frac{b^2}{(\theta+1)^2} = \frac{b^2}{2\theta+1} - \frac{b^2}{(\theta+1)^2} = \frac{(\theta b)^2}{(\theta+1)^2(2\theta+1)}.$$

(b) The likelihood of $X_1,\ldots,X_n$ independent from $f(x;\theta,b)$ is

$$L_n(\theta,b;x) = \theta^{-n} b^{-n/\theta} \exp\{(1/\theta-1)\sum_{i=1}^n \log x_i\} I\{\max_{i\le n} x_i \le b\},$$

from which we can use Factorisation theorem to argue that

$$T = (T_1,T_2) = (-\sum_{i=1}^n \log X_i, \max_{i\le n} X_i),$$

is sufficient for (θ,b). For minimal sufficiency we can use Theorem 6.2.13 on p. 281 in C&B. We see that for two different samples $X = (X_1,\ldots,X_n) = (x_1,\ldots,x_n) = x$ and $Y = (Y_1,\ldots,Y_n) = (y_1,\ldots,y_n) = y$, the sufficient statistics $T(x) = T(y)$ if and only if the function

$$(\theta,b) \mapsto \frac{L_n(\theta,b;x)}{L_n(\theta,b;y)} = \exp\{-(1/\theta-1)(T_1(x)-T_1(y)\}\frac{T_2(x)}{T_2(y)},$$

is constant. This shows that T is minimal sufficient.

(c) We see that $-\log X/b \sim \mathrm{Expo}(1/\theta)$, i.e., has density $(1/\theta)\exp(-z/\theta)$, $z > 0$, so that the sum $-\sum_{i=1}^n \log X_i/b$ is $\mathrm{Gamma}(n,1/\theta)$, and consequently

$$T_1 = -\sum_{i=1}^n \log X_i = n\log b + -\sum_{i=1}^n \log X_i/b \sim n\log b + \mathrm{Gamma}(n,1/\theta).$$

The expectation and variance of T_1 is therefore

$$\mathrm{E}\,T_1 = n\log b + n\theta, \quad \text{and} \quad \mathrm{Var}\,T_1 = n\theta^2.$$

For $T_2 = \max_{i\le n} X_i$ we start by noting that the distribution function of $X \sim f(x;\theta,b)$ is

$$F(x;\theta,b) = \frac{1}{\theta b^{1/\theta}}\int_0^x y^{1/\theta-1}\,\mathrm{d}y = (x/b)^{1/\theta}.$$

Then

$$\Pr(T_2 \le t) = \Pr(\max_{i\le n} X_i \le t) = \Pr(X_i \le t)^n = \big(t/b\big)^{n/\theta},$$

so that T_2 has density

$$g_{T_2}(t;\theta,b) = \frac{n}{\theta b^{n/\theta}} t^{n/\theta-1}, \quad 0 < t \le b.$$

From this we see that $g_{T_2}(t;\theta,b) = f(t;\theta/n,b)$, and we can use (a) to find

$$\mathrm{E}\,T_2 = \frac{b}{\theta/n+1}, \quad \text{and} \quad \mathrm{Var}\,T_2 = \frac{1}{n^2}\frac{(\theta b)^2}{(\theta/n+1)^2(2\theta/n+1)}.$$

(d) We found the likelihood function in (a), it is

$$L_n(\theta, b; x) = \theta^{-n} b^{-n/\theta} \exp\{(1/\theta - 1)\sum_{i=1}^n \log x_i\} I\{\max_{i\le n} x_i \le b\},$$

From this we see that the maximum likelihood estimator for b is $\max_{i\le n} X_i = T_2$, and holding $b \ge \max_{i\le n} x_i$ fixed, the log-likelihood is

$$\ell_n(\theta, b; x) = -n\log\theta - \frac{n}{\theta}\log b + (1/\theta - 1)\sum_{i=1}^n \log x_i.$$

The first derivative w.r.t. θ is

$$\begin{aligned}\frac{\partial}{\partial\theta}\ell_n(\theta, b; x) &= -\frac{n}{\theta} + \frac{n}{\theta^2}\log b - \frac{1}{\theta^2}\sum_{i=1}^n \log x_i \\ &= -\frac{n}{\theta} + \frac{n}{\theta^2}\big(\log b - \frac{1}{n}\sum_{i=1}^n \log x_i\big).\end{aligned}$$

Solving $\partial\ell_n(\theta, b; x)/\partial\theta = 0$ for θ, see that the maximum likelihood estimator is

$$\begin{aligned}\widehat{\theta}_{\rm ml} &= \log\widehat{b}_{\rm ml} - \frac{1}{n}\sum_{i=1}^n \log X_i = \log\widehat{b}_{\rm ml} + \frac{1}{n}T_1, \\ \widehat{b}_{\rm ml} &= \max_{i\le n} X_i = T_2\end{aligned}$$

To show consistency, we start with $\widehat{b}_{\rm ml}$. From what we found in (c),

$$\mathrm{E}\,(\widehat{b}_{\rm ml} - b)^2 = \mathrm{Var}\,\widehat{b}_{\rm ml} + \{\mathrm{E}\,(\widehat{b}_{\rm ml} - b)\}^2 = \frac{1}{n^2}\frac{(\theta b)^2}{(\theta/n+1)^2(2\theta/n+1)} + \big(\frac{b}{\theta/n+1} - b\big)^2,$$

which tends to zero as $n\to\infty$. Chebyshev's inequality then gives the result; $\widehat{b}_{\rm ml} \to_p b$ as $n\to\infty$.

To show that $\widehat{\theta}_{\rm ml}$ is consistent for θ we use the Cramér–Slutsky rules several times. First, $\log x$ is a continuous function, so from the continuous mapping theorem we conclude that

$$\log\widehat{b}_{\rm ml} \overset{p}{\to} \log b.$$

Second, by the weak law of large numbers and using that $-\log X/b$ is $\mathrm{Expo}(1/\theta)$,

$$-\frac{1}{n}\sum_{i=1}^n \log X_i = -\log b - \frac{1}{n}\sum_{i=1}^n \log X_i/b \overset{p}{\to} -\log b + \mathrm{E}\,(-\log X_1/b) = -\log b + \theta.$$

Putting the last two displays together, and using the Cramér–Slutsky rules, we conclude that

$$\widehat{\theta}_{\rm ml} \overset{p}{\to} \theta.$$

In the following we set $b = 1$. The distribution function of X is then $F(x;\theta,1) = x^{1/\theta}$. In particular, the median μ is

$$\mu = (1/2)^\theta.$$

(e) The details of this were taken care of in the oblig. With M_n the median of n independent uniforms on $(0,1)$, we set

$$\widetilde{\mu} = F^{-1}(M_n, \theta, 1) = M_n^\theta.$$

and use the delta-method. Here

$$\frac{\mathrm{d}F^{-1}(x,\theta,1)}{\mathrm{d}x} = \theta x^{\theta-1}, \quad \text{and} \quad \frac{\mathrm{d}F^{-1}(1/2,\theta,1)}{\mathrm{d}x} = \frac{2\log(\mu)\mu}{\log(1/2)} = \frac{1}{f(\mu;\theta,1)}.$$

We are given that $\sqrt{n}(M_n - 1/2) \to_d \mathrm{N}(0, 1/4)$. The delta-method yields

$$\sqrt{n}(\widetilde{\mu} - \mu) \overset{d}{\to} \mathrm{N}(0, \frac{\{\log(\mu)\mu\}^2}{(\log 2)^2}).$$

(f) First, when $b = 1$, the maximum likelihood estimator of θ is simply $\widehat{\theta}_{\rm ml} = -n^{-1}\sum_{i=1}^n \log X_i = n^{-1}T_1$, where $T_1 \sim \text{Gamma}(n, 1/\theta)$. By the invariance property of maximum likelihood estimators,

$$\widehat{\mu}_{\rm ml} = (1/2)^{\widehat{\theta}_{\rm ml}}.$$

To find an expression for the mean of this estimator we use that T_1 is $\text{Gamma}(n, 1/\theta)$,

$$\begin{aligned}\mathrm{E}\,\widehat{\mu}_{\rm ml} &= \mathrm{E}\,(1/2)^{\widehat{\theta}_{\rm ml}} = \mathrm{E}\,e^{n^{-1}\log(1/2)T_1} = \{1 - \frac{\theta}{n}\log(1/2)\}^{-n} \\ &= \{1 - \log(\mu)/n\}^{-n}.\end{aligned}$$

And from $(1 + x/n)^n \to e^x$, we see directly that $\{1 - \log(\mu)/n\}^{-n} \to \mu$ as $n \to \infty$.

(g) To find the Cramér–Rao lower bound we use the hint.

$$\log f(x;\theta,1) = -\log\theta + \frac{1}{\theta}\log x.$$

Differentiating two times,

$$\frac{\partial^2}{\partial\theta^2}\log f(x;\theta,1) = \frac{1}{\theta^2} + \frac{2}{\theta^3}\log X,$$

and the Fisher information is

$$I_\theta(\theta) = -\mathrm{E}\,\frac{\partial^2}{\partial\theta^2}\log f(X;\theta,1) = \frac{1}{\theta^2}.$$

As a function of μ,

$$\theta(\mu) = \frac{\log\mu}{\log(1/2)} \quad \text{then} \quad \frac{\mathrm{d}}{\mathrm{d}\mu}\theta(\mu) = \frac{1}{\mu\log(1/2)}.$$

Using that $I_\mu(\mu) = I_\theta(\theta)(\mathrm{d}\theta/\mathrm{d}\mu)^2$, the expected Fisher information for μ is

$$I_\mu(\mu) = \frac{(\log(1/2))^2}{(\log\mu)^2}\frac{1}{(\mu\log(1/2))^2} = \frac{1}{(\mu\log\mu)^2}.$$

The Cramér–Rao lower bound is therefor

$$\text{CRLB} = \frac{(\mu\log\mu)^2}{n}.$$

(h) To find the limiting distribution of $\sqrt{n}(\widehat{\mu}_{\rm ml} - \mu)$ we can argue directly from likelihood theory

$$\sqrt{n}(\widehat{\mu}_{\rm ml} - \mu) \to_d \mathrm{N}\{0, 1/I_\mu(\mu)\}.$$

We could also use the delta-method once more. First, from likelihood theory and what we found in (g),

$$\sqrt{n}(\widehat{\theta}_{\rm ml} - \theta) \overset{d}{\to} \mathrm{N}(0, \theta^2).$$

Second, the delta-method with the function $\exp(x\log(1/2))$, whose first derivative is $\log(1/2)\exp(x\log(1/2))$,

$$\sqrt{n}(\widehat{\mu}_{\rm ml} - \mu) = \sqrt{n}(e^{\widehat{\theta}_{\rm ml}\log(1/2)} - e^{\theta\log 1/2}) \overset{d}{\to} \mathrm{N}(0, (\log(1/2)e^{\theta\log(1/2)})\theta^2) = \mathrm{N}(0, (\mu\log\mu)^2),$$

because $\log(1/2)e^{\theta\log(1/2)}\theta = \mu\log\mu$. In either case, we see that $\widehat{\mu}_{\rm ml}$ asymptotically attains the CRLB, we say that the maximum likelihood estimator is efficient. Comparing the limiting variance of the two estimators

$$\frac{\text{asymp. var}\ \widetilde{\mu}}{\text{asymp. var}\ \widehat{\mu}_{\rm ml}} = \frac{\{\log(\mu)\mu\}^2/(\log 2)^2}{(\mu\log\mu)^2} = \frac{1}{(\log 2)^2} > 1,$$

since $\log 2 < 1$.

Exercise 2

(a) The two random variables Y_1, Y_2 given by $Y_i = X_i\theta + (1 - X_i)U_i$ have expectation

$$\mathrm{E}\, Y_i = \theta/2 + 1/4.$$

An unbiased estimator θ is therefore

$$\delta_1(Y_1, Y_2) = Y_1 + Y_2 - 1/2.$$

(b) We known that about half of the time $X_1 = 1$ so that $Y_1 = \theta$, and similarly for X_2 and Y_2. Argue conditionally: If $Y_1 = Y_2$, then we know that both Y_1 and Y_2 are both equal to θ, and we can just use their common value as our estimate of θ. If $Y_1 \neq Y_2$, there is a probability 1/3 that one of them equals θ, so we can just pick one, say Y_1. This conditional reasoning leads to the the estimator

$$\delta_2(Y_1, Y_2) = Y_1,$$

which appears much more appealing than δ_1, even though it is not unbiased.

(c) Based on the data $X_i = (Y_i, a_i) = x_i$, the log-likelihood function is

$$\ell_n(\theta) = -\sum_{i=1}^{n} \log \sigma_{a_i} - \frac{1}{2}\sum_{i=1}^{n} \frac{(Y_i - \theta)^2}{\sigma_{a_i}^2} + \text{const.}$$

The score function is

$$\frac{\partial}{\partial\theta}\ell_n(\theta) = \sum_{i=1}^{n} \frac{Y_i - \theta}{\sigma_{a_i}^2}.$$

Set $\partial\ell_n(\theta)/\partial\theta = 0$ and solve to find the maximum likelihood estimator,

$$\widehat{\theta}_{\text{ml}} = \frac{\sum_{i=1}^{n} Y_i/\sigma_{a_i}^2}{\sum_{i=1}^{n} 1/\sigma_{a_i}^2}.$$

(d) Let $a = \sum_{i=1}^{n} a_i$ be the number of experiments performed with instrument 1. The conditional variance of $\widehat{\theta}_{\text{ml}}$ is

$$\mathrm{Var}(\widehat{\theta}_{\text{ml}} \mid (a_1, \ldots, a_n)) = \frac{1}{\sum_{i=1}^{n} 1/\sigma_{a_i}^2} = \frac{1}{(n-a)/\sigma_0^2 + a/\sigma_1^2}.$$

(e) The expected Fisher information is

$$-\mathrm{E}\,\frac{\partial^2}{\partial\theta^2}\ell_n(\theta) = \frac{n}{2}(1/\sigma_0^2 + 1/\sigma_1^2).$$

The Cramér–Rao lower bound for unbiased estimators of θ is therefore

$$\text{CRLB} = \frac{2}{n}\frac{1}{1/\sigma_0^2 + 1/\sigma_1^2} = \frac{1}{n/(2\sigma_0^2) + n/(2\sigma_1^2)}.$$

Can for example look at

$$\frac{\mathrm{Var}(\widehat{\theta}_{\text{ml}} \mid (a_1, \ldots, a_n))}{\text{CRLB}} = \frac{(n/2)\sigma_1^2 + (n/2)\sigma_0^2}{(n-a)\sigma_1^2 + a\sigma_0^2},$$

from which it appears that sometimes the conditional variance of $\widehat{\theta}_{\text{ml}}$ can dip below the CRLB. Hmm? Discuss and *vis forståelse* of what is going on here!

(c) From now on, all expectations are conditional on the observed sequence $(a_1, \ldots, a_n)$. We assume that $n = 100$ and $a = 50 = n/2$. Let $Y_{0,1} \ldots, Y_{0,n/2}$ and $Y_{1,1}, \ldots, Y_{1,n/2}$ be measurements from instrument 0 and 1, respectively. Let $\bar{Y}_0 = n^{-1}\sum_{i=1}^{n/2} Y_{0,i}$ and $\bar{Y}_1 = n^{-1}\sum_{i=1}^{n/2} Y_{1,i}$. We want to test

$$H_0\colon \xi = 0, \quad \text{vs.} \quad H_A\colon \xi > 0.$$

Under H_0, from independence

$$\bar{Y}_0 - \bar{Y}_1 \sim \mathrm{N}(0, 2\frac{\sigma_0^2 + \sigma_1^2}{n}).$$

Here is a level α test: Reject H_0 if

$$\bar{Y}_0 - \bar{Y}_1 \geq (2\frac{\sigma_0^2 + \sigma_1^2}{n})^{1/2}\Phi^{-1}(1 - \alpha).$$

(d) Consider

$$Z_n = \frac{\sum_{i=1}^{n/2}(Y_{0,i} - Y_{1,i})^2}{\sigma_0^2 + \sigma_1^2} \sim \chi^2_{n/2}.$$

A level α test rejects $H_0\colon \xi_1 = \cdots = \xi_{n/2}$ if Z_n is bigger than the $1-\alpha$ quantile of the chi-square distribution with $n/2$ degrees of freedom. The power of this test in points such as $\xi_1 = \cdots = \xi_{49} \neq \xi_{50}$ is probably not very good.

Exercise 3

The regression model

$$Y_i = \beta x_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ independent standard normal random variables, and $x_1, \ldots, x_n$ are fixed and known constants. We are to estimate β under squared error loss $L(\delta, \beta) = (\delta - \beta)^2$.

(a) The maximum likelihood estimator, which is also the least squares estimator, is

$$\widehat{\beta}_{\rm ml} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

Writing

$$\widehat{\beta}_{\rm ml} = \beta + \frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2}, \quad \text{we see that} \quad \widehat{\beta}_{\rm ml} \sim \mathrm{N}\{\beta, 1/(\sum_{i=1}^n x_i^2)\}.$$

It is then immediate that the risk equals the variance,

$$R(\widehat{\beta}_{\rm ml}, \beta) = \frac{1}{\sum_{i=1}^n x_i^2}.$$

(b) The CRLB is equal to $1/(\sum_{i=1}^n x_i^2)$, and we conclude that $\widehat{\beta}_{\rm ml}$ the UMVU estimator.

(c) We know, or you can show via the Factorisation theorem that when the $x_1, \ldots, x_n$ are fixed, $\widehat{\beta}_{\rm ml}$ is sufficient for β. As Bayesians it must therefore be the same information in knowing independent

$$Y_1 \sim \mathrm{N}(\beta x_1, \sigma^2), \ldots, Y_1 \sim \mathrm{N}(\beta x_1, \sigma^2) \quad \text{and} \quad \widehat{\beta}_{\rm ml} \sim \mathrm{N}\{\beta, 1/(\sum_{i=1}^n x_i^2)\}.$$

The ‘posterior $\propto$ likelihood $\times$ prior’-calculation can therefore be carried out treating $\widehat{\beta}_{\rm ml}$ as our observation. It then follows directly from the prior-posterior results given at the start of the exercise that the posterior of β given the data is

$$\beta \mid \widehat{\beta}_{\rm ml} \sim \mathrm{N}\big(m, v^2\big),$$

where

$$m = \frac{\tau^2 \widehat{\beta}_{\rm ml}}{(1/\sum_{i=1}^n x_i^2) + \tau^2} \quad \text{and} \quad v^2 = \frac{\tau^2 (1/\sum_{i=1}^n x_i^2)}{(1/\sum_{i=1}^n x_i^2) + \tau^2} = \frac{\tau^2}{1 + \tau^2 \sum_{i=1}^n x_i^2}.$$

(d) The Bayes estimator under the squared error loss function is the posterior mean

$$\widetilde{\beta}_{\rm b} = \mathrm{E}\,(\beta \mid \widehat{\beta}_{\rm ml}) = \frac{\tau^2 \widehat{\beta}_{\rm ml}}{(1/\sum_{i=1}^n x_i^2) + \tau^2}.$$

We see that when the prior variance τ^2 tends to infinity,

$$\widetilde{\beta}_{\rm b} \to \widehat{\beta}_{\rm ml}, \quad \text{as } \tau^2 \to \infty,$$

This suggests using Blyth's theorem to get at proof of the admissibility of $\widehat{\beta}_{\rm ml}$. An admissible constant risk estimator is clearly minimax. Had it not been, it would not have been admissible.

(e) The risk function of the Bayes estimator is

$$R(\widetilde{\beta}_{\rm b}, \beta) = \frac{\tau^2/(\sum_{i=1}^n x_i^2)}{(1 + \tau^2 \sum_{i=1}^n x_i^2)^2} + \big(\frac{\tau^2 \beta}{(1/\sum_{i=1}^n x_i^2) + \tau^2} - \beta\big)^2.$$

Which in $\beta = 0$ equals

$$R(\widetilde{\beta}_{\rm b}, 0) = \frac{1}{(\sum_{i=1}^n x_i^2)} \frac{\tau^4}{(1 + \tau^2 \sum_{i=1}^n x_i^2)^2}$$

From this we see that

$$R(\widetilde{\beta}_{\rm b}, 0) \leq R(\widehat{\beta}_{\rm ml}, 0), \quad \text{provided} \quad \sum_{i=1}^n x_i^2 \geq 1.$$

We now assume that

$$\begin{pmatrix} x_i \\ \varepsilon_i \end{pmatrix} \sim {\rm N}_2\big(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\big),$$

independently for $i = 1, \ldots, n$.

(f) In this random covariate model the maximum likelihood estimator is as above, but is biased

$$\begin{aligned}
{\rm E}\,\widehat{\beta}_{\rm ml} &= \beta + {\rm E}\,\frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2} = \beta + {\rm E}\,{\rm E}\,\Big\{\frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2} \mid (x_1, \ldots, x_n)\Big\} \\
&= \beta + {\rm E}\,\frac{1}{\sum_{i=1}^n x_i^2} \sum_{i=1}^n x_i {\rm E}\,\{\varepsilon_i \mid (x_1, \ldots, x_n)\} = \beta + {\rm E}\,\frac{1}{\sum_{i=1}^n x_i^2} \sum_{i=1}^n x_i^2 \rho \\
&= \beta + \rho.
\end{aligned}$$

Since we believe our colleague that $\rho = 1/2$, this shows that

$$\widehat{\beta}_{1/2} = \widehat{\beta}_{\rm ml} - 1/2,$$

is unbiased for β.

(g) We can start by finding the distribution of $\widehat{\beta}_{1/2}$. Its expectation is β and its variance is

$$\begin{aligned}
{\rm Var}\,\widehat{\beta}_{1/2} &= {\rm Var}\,\widehat{\beta}_{\rm ml} = {\rm E}\,{\rm Var}\{\widehat{\beta}_{\rm ml} \mid (x_1, \ldots, x_n)\} + {\rm Var}\,{\rm E}\,\{\widehat{\beta}_{\rm ml} \mid (x_1, \ldots, x_n)\} \\
&= {\rm E}\frac{1}{\sum_{i=1}^n x_i^2} = \frac{1/2}{n/2 - 1} = \frac{1}{n-2},
\end{aligned}$$

where we use that $\sum_{i=1}^n x_i^2 \sim \chi_n^2 = {\rm Gamma}(n/2, 1/2)$. We conclude that

$$\widehat{\beta}_{1/2} \sim {\rm N}\{\beta, 1/(n-2)\}.$$

Since this variance tends to zero as $n \to \infty$, we get consistency from Chebyshev's inequality. An exact 95% confidence interval for β is

$$[\widehat{\beta}_{\rm ml} - 1/2 - 1.96/(n-2)^{1/2}, \widehat{\beta}_{\rm ml} - 1/2 + 1.96/(n-2)^{1/2}].$$