Proposed solution the exam, 26. Nov. STK4011/9011

Autumn 2019

Emil Aas Stoltenberg

Exercise 1

A random variable Y has the exponential distribution with parameter α if its density is $\alpha \exp(-\alpha y)$ for $y > 0$, we write $Y \sim \text{Expo}(\alpha)$. The expectation and variance are $\mathrm{E}\, Y = 1/\alpha$ and $\mathrm{Var}\, Y = 1/\alpha^2$. Define

$$a_n = 1 + 1/2 + 1/3 + \cdots 1/n, \quad \text{and} \quad b_n = 1 + 1/2^2 + 1/3^2 + \cdots 1/n^2.$$

We'll use that $a_n \doteq \log n + 0.5772$ and $b_n \to \pi^2/6$ as $n \to \infty$.

(a) $V_1, \ldots, V_n$ independent $\text{Expo}(i/\theta)$. Set $T_n = V_1 + \cdots + V_n$, then

$$\mathrm{E}\, T_n = \mathrm{E}\, V_1 + \mathrm{E}\, V_2 + \cdots \mathrm{E}\, V_n = \theta + \theta/2 + \cdots + \theta/n = a_n \theta,$$

and since the V_i's are independent

$$\mathrm{Var}\, T_n = \sum_{i=1}^n \mathrm{Var}\, V_i = \sum_{i=1}^n \frac{\theta^2}{i^2} = b_n \theta^2.$$

(b) From the expression for $\mathrm{E}\, T_n$ found in (a) we see that $\widehat{\theta} = T_n/a_n$ is an unbiased estimator of θ, that is $\mathrm{E}\, \widehat{\theta} = \theta$ for all values of $\theta > 0$. The variance of this estimator is

$$\mathrm{Var}\, \widehat{\theta} = (1/a_n^2)\, \mathrm{Var}\, T_n = (b_n/a_n^2)\, \theta^2.$$

Since

$$\frac{b_n}{a_n^2} \doteq \frac{\pi^2/6}{(\log n + 0.5772)^2},$$

we see that the variance tends to zero as $n \to \infty$, albeit rather slowly. Since $\widehat{\theta}$ is unbiased for θ, and its variance tends to zero, it follows from Chebyshev's inequality that it is consistent for θ.

(c) The log-likelihood function based on having observed the individual $V_1 = v_1, \ldots, V_n = v_n$ is

$$\ell_n(\theta) = -n \log \theta + \sum_{i=1}^n \log(i) - (1/\theta) \sum_{i=1}^n i v_i.$$

The first derivative (i.e. the score function) is

$$\ell_n'(\theta) = -n/\theta + (1/\theta^2) \sum_{i=1}^n i v_i.$$

Set $\ell_n'(\theta) = 0$ and solve for θ to obtain the maximum likelihood estimator

$$\theta^* = \frac{1}{n} \sum_{i=1}^n i V_i.$$

(d) The maximum likelihood estimator is unbiased

$$\mathrm{E}\, \theta^* = \frac{1}{n} \sum_{i=1}^n i \mathrm{E}\, V_i = \frac{1}{n} \sum_{i=1}^n i(\theta/i) = \theta,$$

and, using the independence of the V_i's, its variance is

$$\mathrm{Var}\, \theta^* = \frac{1}{n^2} \sum_{i=1}^n i^2 \mathrm{Var}\, V_i = \frac{1}{n} \sum_{i=1}^n i^2(\theta^2/i^2) = \frac{\theta^2}{n}.$$

Here it suffices to point out that $\mathrm{Var}\, \theta^*$ tends to zero with speed $1/n$, while $\mathrm{Var}\, \widehat{\theta}$ does it at the rate $1/(\log n)^2$, which is much slower.

One can also look at the ratio of the variances

$$\frac{\operatorname{Var}\widehat{\theta}}{\operatorname{Var}\theta^*} = \frac{nb_n}{a_n^2} = \frac{n\sum_{i=1}^n 1/i^2}{(\sum_{i=1}^n 1/i)^2} = \frac{n^{-1}\sum_{i=1}^n 1/i^2}{(n^{-1}\sum_{i=1}^n 1/i)^2} > 1,$$

because $(n^{-1}\sum_{i=1}^n 1/i)^2 < n^{-1}\sum_{i=1}^n 1/i^2$ by Jensen's inequality (think of a random variable W taking the values $\{1, 1/2, \ldots, 1/n\}$ with equal probability $1/n$, then $\operatorname{E} W = n^{-1}\sum_{i=1}^n 1/i$, and $\operatorname{E} W^2 > (\operatorname{E} W)^2$ because w^2 is strictly convex).

(e) We find the Fisher information,

$$-\operatorname{E}\ell_n''(\theta) = -\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^n i\operatorname{E} V_i = \frac{n}{\theta^2},$$

where $\ell_n''(\theta)$ is the second derivative of $\ell_n(\theta)$. The Cramér–Rao lower bound for variances of unbiased estimators of θ is therefore θ^2/n. We see that the variance of the maximum likelihood estimator θ^* that we found in (c) attains this lower bound. It is best unbiased.

(f) Consider the random variable

$$T_n^* = T_{2n} - T_n.$$

Its expectation is

$$\operatorname{E} T_n^* = a_{2n}\theta - a_n\theta \doteq \theta\{(\log(2n) + 0.5772) - (\log(n) + 0.5772)\} = \theta\log 2,$$

Its variance is $\operatorname{Var} T_n^* = (b_{2n} - b_n)\theta^2 > 0$, and

$$0 < b_{2n} - b_n \leq |b_{2n} - \pi^2/6| + |b_n - \pi^2/6| \to 0,$$

or just say that b_n is convergent, hence Cauchy. Write

$$T_n^* - \theta\log 2 = \{T_n^* - \theta(a_{2n} - a_n)\} + \theta\{(a_{2n} - a_n) - \log 2\}.$$

Since $(a_{2n} - a_n) - \log 2 \to 0$, we can appeal to the Cramér–Slutsky rules in order to show that T_n^* converges in probability to $\theta\log 2$. That is, we need to show that $T_n^* - \theta(a_{2n} - a_n)$ converges in probability to 0. But T_n^* is unbiased for $\theta(a_{2n} - a_n)$, and the variance of T_n^* tends to zero, so $T_n^* - \theta(a_{2n} - a_n) \to_p 0$ follows from Chebyshev's inequality.

Exercise 2

The random variable X is binomial (n, p).

(a) Let $\widehat{p} = X/n$. This estimator is unbiased for p, so its risk function under squared error loss is

$$r(\widehat{p}, p) = \operatorname{E}_p(\widehat{p} - p)^2 = \operatorname{Var}_p\widehat{p} = \frac{p(1-p)}{n}.$$

This function is at its largest in $p = 1/2$ where $r(\widehat{p}, 1/2) = 1/(4n)$, and approaches zero when p approaches the 0 or 1.

(b) Give p a uniform prior on $(0, 1)$. The prior mean and variance is then

$$\operatorname{E} p = \int_0^1 p\,\mathrm{d}p = \frac{1}{2}, \quad \text{and} \quad \operatorname{Var} p = \int_0^1 (p - 1/2)^2\,\mathrm{d}p = \frac{1}{12},$$

respectively. This can also be seen from the formula for the expectation and variance given for the beta distribution at the start of the exercise, noting that a beta$(1, 1)$ is a uniform on $(0, 1)$.

Upon observing $X = x$, the posterior distribution, say $\pi(p \mid x)$, is

$$\pi(p \mid x) \propto p^x(1-p)^{n-x} \times 1,$$

which is proportional to a Beta distribution with parameters $(x + 1, n - x + 1)$.

(c) Using the formula for the expectation of the Beta random variable, the posterior mean, i.e. the Bayes solution under squared error loss, is

$$\widehat{p}_B = \operatorname{E}(p \mid x) = \frac{x+1}{(x+1) + (n-x+1)} = \frac{x+1}{n+2}.$$

We can write this estimator as

$$\widehat{p}_B = \frac{n}{n+2}\,\widehat{p} + (1 - \frac{n}{n+2})\frac{1}{2} = w_n\widehat{p} + (1-w_n)(1/2),$$

with $w_n = n/(n+2)$, to see that it is a weighted sum of the empirical mean and the prior expectation. Its risk function is

$$\begin{aligned} r(\widehat{p}_B, p) &= \mathrm{E}_p\,(\widehat{p}_B - p)^2 = \mathrm{Var}\,\widehat{p}_B + (w_n p + (1-w_n)(1/2) - p)^2 \\ &= w_n^2 \mathrm{Var}\,\widehat{p} + (1-w_n)^2(1/2-p)^2 = \frac{1}{(n+2)^2}\{np(1-p) + 4(1/2-p)^2\}, \end{aligned}$$

or, writing it in terms if the risk of $\widehat{p}$,

$$r(\widehat{p}_B, p) = w_n^2 r(\widehat{p}, p) + (1-w_n)^2(1/2-p)^2.$$

From these expressions we see that for p in the vicinity of the prior expectation $1/2$, the Bayes estimator performs better than $\widehat{p}$.

(d) The minimax perspective dictates that one ought to use the estimator that minimises the maximum risk. The max risk of $\widehat{p}$ is $1/(4n)$.

Where the maximum risk of $\widehat{p}_B$ occurs depends on the sample size n. Solving

$$\frac{\mathrm{d}}{\mathrm{d}p} r(\widehat{p}_B, p) = \frac{1}{(n+2)^2}\{n(1-2p) - 8(1/2-p)\} = 0,$$

for p, gives $p = 1/2$, so $1/2$ is a stationary point. But

$$\frac{\mathrm{d}^2}{\mathrm{d}p^2} r(\widehat{p}_B, p) = \frac{8-2n}{(n+2)^2},$$

so for $n \geq 5$ the risk is concave and the maximum risk is attained in $p = 1/2$. For $p \leq 4$, the risk is convex, and the maximum risk is attained at the boundaries 0 and 1 of the parameter space. This means that

$$\sup_p r(\widehat{p}_B, p) = \begin{cases} \frac{1}{(n+2)^2} & \text{if } n \leq 4, \\ \frac{n}{4(n+2)^2} & \text{if } n \geq 5, \end{cases}$$

In both cases $\sup_p r(\widehat{p}_B, p) < 1/(4n) = \sup_p r(\widehat{p}, p)$, so for any sample size, the Bayes estimator $\widehat{p}_B$ beats $\widehat{p}$ at the minimax game.

Even though it's not part of the exam question, note that $\widehat{p}_B$ is not in general *the* minimax estimator. Recall that in a lecture and in an exercise we met the estimator $p^\star$ given by

$$p^\star = \frac{\sqrt{n}}{\sqrt{n}+1}\widehat{p} + (1 - \frac{\sqrt{n}}{\sqrt{n}+1})\frac{1}{2}.$$

This is the minimax estimator, see Exercise 1.4 in the exercise set. For $n = 4$ we see that $\mathrm{d}^2 r(\widehat{p}_B, p)/\mathrm{d}p^2 = 0$ and $\widehat{p}_B = p^\star$, but this is a special case.

(e) Upon observing $X = x = n$, our updated beliefs about p is the distribution $\pi(p \mid x = n)$, which is a Beta distribution with parameters $(n+1, 1)$.

(f) View the binomial X above as the sum of n independent Bernoulli (p) random variables $Y_1, \ldots, Y_n$. Our probability of the sun rising tomorrow is the posterior probability of observing $Y_{n+1} = 1$. Let $\mathrm{E}_\pi[\cdot]$ denote the expectation with respect to the posterior of p. Then, since $Y_1, \ldots, Y_n, Y_{n+1} \mid p$ are independent Bernoulli(p),

$$\begin{aligned} \Pr(Y_{n+1} = 1 \mid Y_1 = 1, \ldots, Y_n = 1) &= \mathrm{E}_\pi \Pr(Y_{n+1} = 1 \mid Y_1 = 1, \ldots, Y_n = 1, p) \\ &= \mathrm{E}_\pi \Pr(Y_{n+1} = 1 \mid p) = \int_0^1 \Pr(Y_{n+1} = 1 \mid p)\pi(p \mid x = n)\,\mathrm{d}p \\ &= \int_0^1 p\,\pi(p \mid x = n)\,\mathrm{d}p = (n+1)\int_0^1 p^{n+1}\,\mathrm{d}p = \frac{n+1}{n+2}, \end{aligned}$$

where the independence of $Y_1, \ldots, Y_n, Y_{n+1}$ given p is used in the third equality. Our probability of seeing the sun rising tomorrow after having seen it rise for n consecutive days is $(n+1)/(n+2)$.

Exercise 3

Suppose $V_1, \ldots, V_5$ are independent normals with means $\theta_1, \ldots, \theta_5$ and variance one. Set $Z = \sum_{j=1}^5 (V_j - \bar{V})^2$, where $\bar{V} = (1/5)\sum_{j=1}^5 V_j$.

(a) Under $H_0\colon \theta_1 = \cdots = \theta_5$, the random variable Z has a chi-square distribution with 4 degrees of freedom, a χ^2_4. See for example Nils' exercise 8 for a proof. A proposed solution can be found in the exercise set.

A level 0.05 test is: Reject H_0 if

$$Z \geq c_{0.95},$$

where $c_{0.95}$ is the 0.95 quantile of the chi-square distribution with 4 degrees of freedom.

(b) Write

$$Z = \sum_{j=1}^5 (V_j - \bar{V})^2 = \sum_{j=1}^5 V_j^2 - 5\bar{V}^2.$$

Let $\bar{\theta} = (1/5)\sum_{j=1}^5 \theta_j$, then $\bar{V} \sim \mathrm{N}(\bar{\theta}, 1/5)$, and $\mathrm{E}\,\bar{V}^2 = 1/5 + \bar{\theta}^2$. Since $\mathrm{E}\,V_j^2 = 1 + \theta_j^2$, the expectation of Z is

$$\begin{aligned}\mathrm{E}\,Z &= \sum_{j=1}^5 \mathrm{E}\,(V_j^2) - 5\mathrm{E}\,(\bar{V}^2) = 5 + \sum_{j=1}^5 \theta_j^2 - (1 + 5\bar{\theta}^2) \\ &= 4 + \sum_{j=1}^5 (\theta_j - \bar{\theta})^2.\end{aligned}$$

Under H_0 the sum $\sum_{j=1}^5 (\theta_j - \bar{\theta})^2 = 0$, so the expectation is 4 as it should be, and $\mathrm{E}\,Z$ is increasing as the 'empirical' variance of the parameters increases.

Consider independent pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ with distribution

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\Big(\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}\Big),$$

with correlation coefficient ρ in $(-1, 1)$. The sample correlation coefficient R_n is

$$R_n = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})\}^{1/2}},$$

and one can show that

$$\sqrt{n}(R_n - \rho) \overset{d}{\to} \mathrm{N}(0, (1 - \rho^2)^2), \quad \text{as } n \to \infty.$$

(c) We are to show that

$$h(\rho) = \frac{1}{2}\log\frac{1 + \rho}{1 - \rho},$$

is a variance stabilising transformation, that is $\sqrt{n}(h(R_n) - h(\rho))$ tends in distribution to a standard normal. From the delta method we know that as long as the first derivative h' of h evaluated in ρ is non-zero,

$$\sqrt{n}(h(R_n) - h(\rho)) \overset{d}{\to} \mathrm{N}(0, h'(\rho)^2(1 - \rho^2)^2),$$

as $n \to \infty$. We therefore need to show that $h'(\rho)$ is equal to plus or minus $1/(1 - \rho^2)$.

$$h'(\rho) = \frac{1}{2}\Big(\frac{1 - \rho}{1 + \rho}\Big)\frac{(1 - \rho) + (1 + \rho)}{(1 - \rho)^2} = \frac{1}{2}\Big(\frac{1 - \rho}{1 + \rho}\Big)\frac{2}{(1 - \rho)^2} = \frac{1}{1 - \rho^2}.$$

(d) For $n = 100$ observed pairs we found $R_n = 0.666$. To find a 95 percent confidence interval, start from

$$-1.96 \leq \sqrt{n}(h(R_n) - h(\rho)) \leq 1.96,$$

note that

$$h^{-1}(x) = \frac{\exp(2x) - 1}{\exp(2x) + 1},$$

and arrive at

$$\frac{\exp\{2(h(R_n) - 1.96/\sqrt{n})\} - 1}{\exp\{2(h(R_n) - 1.96/\sqrt{n})\} + 1} \le \rho \le \frac{\exp\{2(h(R_n) + 1.96/\sqrt{n})\} - 1}{\exp\{2(h(R_n) + 1.96/\sqrt{n})\} + 1}.$$

With $R_n = 0.666 \approx 2/3$, this realisation of this confidence interval is

$$\frac{5\exp(-1.96/5) - 1}{5\exp(-1.96/5) + 1} \le \rho \le \frac{5\exp(1.96/5) - 1}{5\exp(1.96/5) + 1}.$$

(e) We have correlation information for each of five groups, with 100 pairs from each group, and we assume the groups are independent. Let the information be sample correlations coefficients $R_{n,1}, \dots, R_{n,5}$. We want to test $H_0\colon \rho_1 = \cdots = \rho_5$. Let $V_j = h(R_{n,j})$ and $\theta_j = h(\rho_j)$ for $j = 1, 2, 3, 4, 5$. Then the V_j's are approximately $\mathrm{N}(\theta_j, 1/100)$. As in (a), let $Z = \sum_{j=1}^{5}(V_j - \bar{V})^2$, then under the null-hypothesis $100Z$ is approximately distributed as a chi-square with 4 degrees of freedom, and we reject H_0 at approximately level 0.05 if

$$100Z \ge c_{0.95},$$

where $c_{0.95}$ is the quantile from (a), i.e. the 0.95 quantile of the chi-square distribution with 4 degrees of freedom.

Exercise 4

Start with (X, Y) being a pair of independent standard normals, so their joint density is

$$f_0(x, y) = \frac{1}{2\pi}\exp\{-\frac{1}{2}(x^2 + y^2)\},$$

for x, y on the real line.

(a) Transform to polar coordinates

$$X = R\cos\theta \quad \text{and} \quad Y = R\sin\theta,$$

with $\theta \in [-\pi, \pi]$. The Jacobian of this transformation is

$$J = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix},$$

with determinant $\det(J) = r\cos^2\theta + r\sin^2\theta = r$. The density of (R, θ) is then

$$\begin{aligned} g_0(r, \theta) &= \frac{1}{2\pi}\exp\{-\frac{1}{2}(r^2\cos^2\theta + r^2\sin^2\theta)\}|\det(J)| \\ &= \frac{1}{2\pi}r\exp\{-r^2/2\}, \quad \text{for } r > 0 \text{ and } \theta \in [-\pi, \pi]. \end{aligned}$$

which is the product of a uniform on $[-\pi, \pi]$ and the Raleigh density $h_0(r) = r\exp\{-\frac{1}{2}r^2\}$ for $r > 0$. This shows that R and θ are independent.

(b) Suppose R has density $h(r)$ rather than $h_0(r)$. Let $k(r, \theta) = (r\cos\theta, r\sin\theta) = (x, y)$, then

$$k^{-1}(x, y) = (\sqrt{x^2 + y^2}, \arctan(x/y)),$$

with Jacobian

$$J = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ \frac{1}{y(1+(x/y)^2)} & -\frac{x}{y^2(1+(x/y)^2)} \end{pmatrix}.$$

The determinant is

$$\det(J) = -\frac{1}{\sqrt{x^2 + y^2}}\Big(\frac{(x/y)^2}{1 + (x/y)^2} + \frac{1}{1 + (x/y)^2}\Big) = -\frac{1}{\sqrt{x^2 + y^2}}.$$

Then

$$\begin{aligned} f(x, y) = g_0(\sqrt{x^2 + y^2}, \arctan(x/y)) &= h(\sqrt{x^2 + y^2})\frac{1}{2\pi}|\det(J)| \\ &= h(\sqrt{x^2 + y^2})\frac{1}{h(\sqrt{x^2 + y^2})}\frac{1}{2\pi}. \end{aligned}$$

(c) Consider the generalisation of Raleigh-distribution given by

$$h(r) = \frac{\gamma}{2} r^{\gamma-1} \exp\{-r^\gamma/2\}, \quad \text{for } r > 0.$$

Note that with $\gamma = 2$ we are back to the Raleigh-density $h_0(r)$. Here are two ways to test $H_0\colon \gamma = 2$. Given a sample of independent pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, transform to polar coordinates. The marginal log-likelihood of $R_1 = (X_1^2 + Y_1^2)^{1/2}, \ldots, R_n = (X_n^2 + Y_n^2)^{1/2}$ is

$$\ell_n(\gamma) = n \log \gamma + (\gamma - 1) \sum_{i=1}^n \log(r_i) - \frac{1}{2} \sum_{i=1}^n r_i^\gamma - n \log(2).$$

We can find the maximiser $\widehat{\gamma}$ of this function numerically, for example using the `nlm()` function in R.

Test 1: Maximum likelihood theory tells us that $\sqrt{n}(\widehat{\gamma} - \gamma) \to_d \mathrm{N}(0, \kappa^2)$, where $1/\kappa^2 = -\mathrm{E}\, \ell_n''(\gamma)$ is the Fisher information, and we let $\widehat{\kappa}^2$ be a consistent estimator of the inverse Fisher information. Under H_0 the random variable $Z_n = \sqrt{n}(\widehat{\gamma} - 2)/\widehat{\kappa}$ is approximately $\mathrm{N}(0, 1)$, and we reject H_0 at approximately level α if

$$Z_n < \Phi^{-1}(\alpha/2) \quad \text{or} \quad Z_n > \Phi^{-1}(1 - \alpha/2).$$

Test 2: We still denote by $\widehat{\gamma}$ the maximum likelihood estimator of γ. We know from Ch. 10.3 in the C&B-book that

$$W_n = 2\{\ell_n(\widehat{\gamma}) - \ell_n(2)\} \xrightarrow{d} \chi_1^2, \qquad \text{as } n \to \infty.$$

Let $w_{1-\alpha}$ be the $(1 - \alpha)$ quantile of the chi-square distribution with one degree of freedom, and reject H_0 approximately at level α if $W_n \geq w_{1-\alpha}$.