STK4011 – Statistical inference. Autumn 2019
Exercises

Emil Aas Stoltenberg

October 7, 2019

Contents

1. Coin tossing

Exercise 1.1. A fair coin is tossed two times. At least one of the tosses came up heads. What is the probability that both came up heads?

Exercise 1.2. In the morning I roll two dice to determine at what time of the day I'll toss my fair coin the first and the second time. Die showing 1 meaning coin flip in the time interval 13:00-13:01, die showing 2 meaning coin flip at 14:00-14:01, and so on. The one minute interval is there in the case that the two dice are equal, giving me time to flip the coin two times. Sometime after 18:01 you're told that at least one of the coins, flipped sometime between 16:00 and 16:01, came up heads. What is the probability that both came up heads?

Exercise 1.3. Helsesøster or bror (school nurse) wants to know the fraction of teenagers who have experienced E, where E is something embarrassing, in fact so embarrassing that the teenagers might not be honest in their 'yes' or 'no' answer when questioned about having experienced E. The school nurse therefore sets up the following anonymisation scheme: Each teenager the school nurse samples are to toss a coin, the outcome of which they keep to themselves. If their coin shows heads, they must answer truthfully. If the coin shows tails, they toss the coin once more. If this second toss shows heads, they answer 'yes', if it shows tails they answer 'no'. Of 17 teenagers 4 answered yes, the remaining ones answered no. The school nurse wants to know the true proportion of 'yes I've experienced E'-teenagers in the population.

(a) State any additional assumptions you feel you need, and provide an unbiased estimator of the true proportion of 'yes'-teenagers. Based on the data given to you by the school nurse, what's your estimate of this proportion?

(b) Explain why the school nurse ought to question more teenagers if (s)he wants a good estimate. That is, explain why, with high probability, it is worthwhile questioning some more teenagers.

(c) Suppose that the school nurse is out of touch with the youth of today, and that the teeneagers are, contrary to the beliefs of the school nurse, fully willing to provide her/him with an honest answer. How much does the school nurse lose by introducing the anonymisation scheme in this situation. *Hint*: Compare the variance of the estimator you found above, with the variance of the estimator you think the school nurse would have used had (s)he known that (s)he were dealing with honest and upright youth.

Exercise 1.4. A natural way to compare estimators is to look at the loss they incur. By 'loss' we mean a non-negative function $L(\theta, \delta)$, where θ is the parameter of the model, and δ is your estimator (or more generally, you chosen action). We are interested in the performance of an estimator δ when it is repeatedly applied. The long-term average loss of using δ is the risk function $R(\theta, \delta)$, defined by

$$R(\theta, \delta) = \mathrm{E}_\theta\, L(\delta, \theta).$$

The problem of comparing estimators is now cast as the problem of comparing their associated risk functions, and we ought to use the estimator that minimises the risk. As this exercise will indicate, this latter statement is not sufficiently precise to be operational.

Suppose $X_1, \ldots, X_n$ are independent Bernoulli random variables with expectation θ. We are to estimate θ under the squared error loss function

$$L(\delta, \theta) = (\delta - \theta)^2.$$

(a) Find the maximum likelihood estimator of θ and, for $n = 10$, sketch its risk function.

(b) Suppose we have some intuition about where on the unit interval the expectation θ might be located, close to a value $0 < \theta_0 < 1$, say. Consider the estimator

$$\delta_1(X) = \bar{X}_n/2 + \theta_0/2,$$

where $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. For $\theta_0 = 1/2$ and $n = 10$, sketch its risk function.

(c) Your task, as was Pierre-Simon Laplace's in 1781 I believe, is to estimate the probability of giving birth to a boy. Which of the two above estimators do you prefer, the maximum likelihood estimator or δ_1 with $\theta_0 = 1/2$?

(d) Consider the estimator

$$\delta_w(X) = w_n \bar{X}_n + (1 - w_n)\frac{1}{2}.$$

Find a function w_n such that the risk function $R(\theta, \delta_{w_n})$ is constant.

(e) For $n = 10$, sketch the risk function of your estimator (that is, draw a line).

(f) Suppose you have absolutely no idea whatsoever about where in the unit interval θ may be located. Which of your three estimators of θ do you prefer?

2. Probabilities, distribution functions, and random variables

Exercise 2.1. Consider the function

$$F(x) = \begin{cases} (1-\theta)x, & 0 \le x < \tau, \\ \theta + (1-\theta)x, & \tau \le x \le 1. \end{cases}, \quad 0 < \theta, \tau < 1. \tag{2.1}$$

(a) Show that F is a distribution function.

(b) Sketch how you would simulate data from F.

(c) Find the expectation and the variance of $X \sim F$.

(d) Find the expected sample size needed to identify τ with probability 1.

(e) Suppose that τ is known and that it does not equal $1/2$. Propose two estimators for θ and compare their variances.

Exercise 2.2. (Pólya urn). An urn contains r red balls and b blue balls. A ball is drawn from the urn at random, its colour is noted, and the ball is returned to the urn along with d more balls of the same colour. This is repeated indefinitely.

(a) What's the probability that the second ball drawn is red?

(b) What's the probability that the k'th ball drawn is red?

(c) What's the probability that the first ball drawn was red, given that the third ball drawn is red?

Exercise 2.3. Probability measures are continuous. If $A_1 \subset A_2 \subset \cdots$ is a sequence of events, then

$$P(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n). \tag{2.2}$$

And if $A_1 \supset A_2 \supset \cdots$ is a sequence of events, then

$$P(\cap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n). \tag{2.3}$$

Prove (2.2) and (2.3).

Exercise 2.4. Suppose that P is a finitely additive probability measure with the property that if $A_1 \subset A_2 \subset \cdots$ is a sequence of events, then $P(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$. Show that P is countably additive.

Exercise 2.5. Let $F(x)$ be a distribution function. Show that F has at most a countable number of discontinuities, i.e., points x such that $F(x) - F(x-) > 0$, where $F(x-) = \lim_{y\uparrow x} F(x)$.

3. Transformations of random variables

Exercise 3.1. Let X be a standard normal random variable, and set $Y = I\{X \geq c\}$ for constant c.

(a) Find the distribution of Y.

(b) Based on independent draws $Y_1, \ldots, Y_n$, find the maximum likelihood estimator, say $\hat{c}_n$, of c. *Hint:* The maximum likelihood estimator is in this case the estimator your intuition leads you to.

(c) This exercise requires material we have yet to cover in class. Show that

$$\sqrt{n}(\hat{c}_n - c) \overset{d}{\to} \mathrm{N}\big(0, \frac{\Phi(c)(1-\Phi(c))}{\phi(c)^2}\big),$$

as n tends to infinity, where $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ is the standard normal probability density function, and $\Phi(x) = \int_{-\infty}^{x} \phi(y)\,\mathrm{d}y$. *Hint*: Use the delta method. In other words, use the mean value theorem; the fact that $1 - \bar{Y}_n = 1 - n^{-1}\sum_{i=1}^{n} Y_i$ is consistent for $\Phi(c)$; and that $X_n \overset{p}{\to} \eta$, implies $g(X_n) \overset{p}{\to} g(\eta)$, when g is continuous.

Exercise 3.2. Go to the 2014 STK4011 website and do Exercises 1 and 2 ('Transformations of random variables' and 'Transformations of random vectors') from Nils Lid Hjort's 'Exercises and Lecture Notes' `https://www.uio.no/studier/emner/matnat/math/STK4011/h14/exercises_stk4011a.pdf`.

Exercise 3.3. (An exercise from Nils' lecture notes). Let X and Y be independent standard normals, and transform to polar coordinates,

$$X = R\cos\theta, \quad Y = R\sin\theta.$$

Find the distribution of the random length R and the random angle θ, and show that these are independent.

Exercise 3.4. Two more exercises from Nils' lecture notes. Do Exercise 4. Ordering exponentials, and Exercise 5. Ratios of ordered uniforms.

4. Moment generating functions

Exercise 4.1. Let X be a Poisson random variable with mean λ.

(a) Find the moment generating function $M_X(t)$ of X.

(b) Let Y_n be binomial(n, p_n), and assume that np_n tends to λ as $n \to \infty$. Show that the moment generating function of Y_n tends to $M_X(t)$

(c) Suppose $X_1, \ldots, X_n$ are independent Bernoulli random variables with expectations $p_{1,n}, \ldots, p_{n,n}$. Set $Z_n = \sum_{i=1}^{n} X_i$. Assume that $\sum_{i=1}^{n} p_{i,n}$ tends to λ and that $\max_{i\leq n} p_{i,n}$ tends to 0, as $n \to \infty$. Show that the moment generating function of Z_n tends to $M_X(t)$.

Exercise 4.2. Let X be a standard normal random variable and set $Y = \exp(X)$.

(a) Find the distribution of Y. What's the name of this distribution?

(b) Show that all the moments of Y exist.

(c) Show that the random variable Y does not have a moment generating function.

Exercise 4.3. Let X_n be a sequence of random variables with moment generating functions $M_n(t)$, X be a random variable with moment generating function $M(t)$, and suppose that $P(0 \le X_n \le 1) = P(0 \le X \le 1) = 1$ for all n. Suppose that all the moments of X_n converge to the moments of X, that is $\mathrm{E}\, X_n^k \to \mathrm{E}\, X^k$ for $k = 1, 2, \ldots$. Show that $M_n(t) \to M(t)$.

5. Exponential families

Lemma 5.1. *If $\partial g(x, \theta)/\partial\theta$ exists and is continuous in θ for all x and all θ in an open interval S, and if $|\partial g(x, \theta)/\partial\theta| \le k(x)$ for all $\theta \in S$ for some integrable function $k(x)$, and if $\int g(x, \theta)\, \mathrm{d}\nu(x)$ exists on S, then*

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \int g(x, \theta)\, \mathrm{d}\nu(x) = \int \frac{\partial}{\partial\theta} g(x, \theta)\, \mathrm{d}\nu(x).$$

Exercise 5.1. Prove Lemma 5.1. *Hint*: Use first the mean value theorem, then Dominated convergence.

Exercise 5.2. Let $f_\theta(x)$ be a density function, $\theta \in \mathbb{R}^k$, and suppose that $\log f_\theta(x)$ satisfies the conditions of Lemma 5.1 for each θ_j, $j = 1, \ldots, k$.

(a) Define $u_j(\theta; x) = \partial \log f_\theta(x)/\partial\theta_j$, and show that

$$\mathrm{E}_\theta\, u_j(\theta; X) = 0, \quad \text{and} \quad \mathrm{E}_\theta\, u_j(\theta; X) u_l(\theta; X) = -\mathrm{E}_\theta \frac{\partial^2}{\partial\theta_j \partial\theta_l} \log f_\theta(X).$$

(b) Suppose that $f_\theta(x)$, $\theta \in \Theta \subset \mathbb{R}^p$ is an exponential family in its natural parametrisation (i.e., $w_j(\theta) = \theta_j$ in Eq. (5.1)). Use (a) to find expressions for the expectation and variance of $t_j(X)\; j = 1, \ldots, p$.

(c) Let X be Gamma(a, b) with density $b^a/\Gamma(a) x^{a-1} \exp(-bx)$. Show that X has an exponential family distribution and find the expectation of $\log X$.

(d) Consider a sequence of independent Bernoulli trials with success probability p. Let X be the number of failures until the first success. Show that X has an exponential family distribution and find the expectation of X.

Exercise 5.3. Let $X_1, \ldots, X_n$ be independent draws from a distribution with density $f_\theta(x)$ of exponential family form, that is

$$f_\theta(x) = h(x) c(\theta) \exp\Big\{\sum_{j=1}^k w_j(\theta) t_j(x)\Big\}, \quad x \in \mathcal{X},\ \theta \in \Theta, \tag{5.1}$$

with the sample space $\mathcal{X}$ not depending on θ; $h(x)$ and $c(\theta)$ are non-negative functions; and the $w_j(\theta), t_j(x)$, $j = 1, \ldots, k \ge 1$, are real valued functions. We suppose that $\Theta \subset \mathbb{R}^p$.

(a) Show that the joint distribution of $X_1, \ldots, X_n$ is also an exponential family.

(b) Define $\tilde{t}_j(x) = \sum_{i=1}^n t_j(x_i)$ for $j = 1, \ldots, k$. For the discrete case, that is when $f_\theta(x) = P_\theta(X = x)$, show that the joint distribution of $\tilde{t}_1(X), \ldots, \tilde{t}_k(X)$ is an exponential family with natural parameters $w_1(\theta), \ldots, w_k(\theta)$.

Exercise 5.4. Let $Y_1, \ldots, Y_n$ be independent Bernoulli trials with success probabilities $p_1, \ldots, p_n$. These are linked to covariates $x_1, \ldots, x_n$, regarded as known constants, through

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}, \quad i = 1, \ldots, n.$$

Show that the joint distribution of $X_1, \ldots, X_n$ is an exponential family.

Exercise 5.5. Let $Y_1, \ldots, Y_n$ be independent $\mathrm{N}(\beta_0 + \beta_1 x_i, 1)$, where $x_1, \ldots, x_n$ are known constants. Show that the joint distribution of $X_1, \ldots, X_n$ is an exponential family.

Exercise 5.6. Let $X_1, \ldots, X_n$ be independent Poisson random variables with mean λ. Show that the joint distribution of X_i, $i = 1, \ldots, n$ is an exponential family, and identify $\tilde{t}(x)$. Find the conditional distribution of $X_1, \ldots, X_n$ given $\tilde{t}(X) = t$. What's noticable about this distribution?

Exercise 5.7. Consider the two-parameter exponential family with $t_1(x) = x$ and $t_2(x) = x^2$, and $h(x) = 1$ on $\{0, 1, 2\}$, and zero outside. Find its natural parameter space and its density.

Exercise 5.8. Consider the one-parameter exponential family $f_\theta(x)$ with $h(x) = 1/x$ on $(0, 1]$, zero elsewhere, and $t(x) = \log x$.

(a) Find its natural parameter space and its density.

(b) Find the distribution of $Y = -\log X$.

(c) Let $X_1, \ldots, X_n$ be independent draws from $f_\theta(x)$. Find the maximum likelihood estimator $\hat{\theta}_n$ of θ and show that it is approximately unbiased.

(d) Find the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$.

6. Sufficiency, ancilliarity and completeness

Exercise 6.1. Suppose $T(X)$ is statistic such that $\theta \mapsto P_\theta(X = x)/P_\theta(T(X) = t)$ is constant for all x. Show that $T(X)$ is sufficient.

Exercise 6.2. Do Exercise 6.9 in Casella and Berger (2002). That is, let $X_1, \ldots, X_n$ be a random sample. For each of the following distributions, find a minimal sufficient statistic for θ.

(a) $f_\theta(x) = (2\pi)^{-1/2} \exp(-(x-\theta)^2/2)$, $x, \theta \in \mathbb{R}$;

(b) $f_\theta(x) = \exp(-(x-\theta))$, $x > \theta, \theta \in \mathbb{R}$;

(c) $f_\theta(x) = \exp(-(x-\theta))/\{1 + \exp(-(x-\theta))\}$, $x, \theta \in \mathbb{R}$;

(d) $f_\theta(x) = 1/\{\pi(1 + (x-\theta)^2)\}$, $x, \theta \in \mathbb{R}$;

(e) $f_\theta(x) = \exp(-|x-\theta|)/2$, $x, \theta \in \mathbb{R}$.

Exercise 6.3. Let $X_1, \ldots, X_n$ be independent draws from the uniform distribution on $(0, \theta)$, $\theta > 0$.

(a) Show that $T(X) = T(X_1, \ldots, X_n) = \max_{i \leq n} X_i$ is minimal sufficient for θ.

(b) Find a so that $\hat{\theta}_{\max} = aT(X)$ is an unbiased estimator of θ, and compute the variance of this estimator. *Hint*: Start by finding the density of $T(X)$.

(c) Show that the estimator $\hat{\theta}_{\text{mean}} = 2\bar{X}_n$ is unbiased for θ. Compare the variance of $\hat{\theta}_{\text{mean}}$ with the variance of $\hat{\theta}_{\max}$.

(d) Consider the class of estimators $\delta_a(X) = aT(X)$. Find the value of a, say $a^\star$, that minimises the risk function $R(\theta, \delta_a) = \mathrm{E}_\theta(aT(X) - \theta)^2$. Compare the risk functions of $\delta_{a^\star}$ and $\hat{\theta}_{\max}$.

Exercise 6.4. Let g be a positive and integrable function on $(0, \infty)$. Set $c(\theta)^{-1} = \int_\theta^\infty g(x)\,\mathrm{d}x$, and define $f_\theta(x) = c(\theta)g(x)$ for $x > \theta$ and zero otherwise. Suppose $X_1, \ldots, X_n$ are independent draws from $f_\theta(x)$.

(a) Show that $T(X) = \min_{i \leq n} X_i$ is sufficient for θ.

(b) Show that $T(X)$ is minimal sufficient.

Exercise 6.5. Let $g(x)$ be a positive integrable function on $(-\infty, \infty)$. For $a < b$, Set $c(a, b)^{-1} = \int_a^b g(x)\,\mathrm{d}x$, and define $f_{(a,b)}(x) = c(a, b)g(x)$ for $a < x < b$, and zero otherwise. If $X_1, \ldots, X_n$ are independent from $f_\theta(x)$, show that $(X_{(1)}, X_{(n)}) = (\min_{i \leq n} X_i, \max_{i \leq n} X_i)$ is minimal sufficient for (a, b).

Exercise 6.6. (Lehmann and Casella (1999))Let $X_1, \dots, X_n$ be independent and identically distributed from a contionuous distribution F, that is otherwise unknown. Let $T(X) = (X_{(1)}, \dots, X_{(n)})$ be the order statistics. Show that T is sufficient.

(a) Let $U_1(X) = \sum_{i=1}^n X_i$, $U_2(X) = \sum_{1\leq i<j\leq n}^n X_i X_j$, $U_3(X) = \sum_{1\leq i<j<k\leq n}^n X_i X_j X_k$, and so on, with, $U_n(X) = X_1 \cdots X_n$. Show that $U(X) = (U_1(X), \dots, U_n(X))$ is sufficient.

(b) Let $V_k(X) = X_1^k + \cdots + X_n^k$, and set $V(X) = (V_1(X), \dots, V_n(X))$. Show that V is sufficient.

Exercise 6.7. Let X be a single observation from $\mathrm{N}(0, \theta)$, $\theta > 0$. Show that both X and $|X|$ are sufficient for θ.

(a) Are they both minimal sufficient?

(b) Let U be Bernoulli(1/2), and set $X' = U|X| - (1-U)|X|$. Show that X' has the same distribution as X.

Exercise 6.8. Let $Y_1, \dots, Y_n$ be independent $\mathrm{N}(\beta x_i, 1)$, where $x_1, \dots, x_n$ are fixed constants not all zero.

(a) Show that the least-squares estimator $\hat{\beta} = \sum_{i=1}^n x_i Y_i / \sum_{i=1}^n x_i^2$ is complete and sufficient.

(b) Show that $\hat{\beta}$ and $\sum_{i=1}^n (Y_i - x_i \hat{\beta})^2$ are independent.

Exercise 6.9. Suppose that given λ, the random variables $X_1, \dots, X_n$ are independent Poisson with mean λ, while λ itself stems from an exponential distribution with mean $1/\theta$. Find a minimal sufficient statistic for θ.

Exercise 6.10. For $i = 1, \dots, n$, let ε_i and X_i be independent standard normal random variables. For $\theta \in (0,1)$, set

$$Y_i = \theta X_i + \sqrt{1-\theta^2}\varepsilon_i, \quad i = 1, \dots, n.$$

Suppose we observe the pairs (X_i, Y_i), $i = 1, \dots, n$.

(a) Find a minimal sufficient statistic for θ.

(b) Is your minimal sufficient statistic complete?

(c) Show that $\sum_{i=1}^n X_i^2$ and $\sum_{i=1}^n Y_i^2$ are ancilliary, but that $(\sum_{i=1}^n X_i^2, \sum_{i=1}^n Y_i^2)$ is not.

Exercise 6.11. Let $Y_1, \dots, Y_n$ be independent Bernoulli random variables with success probabilites $p_i = 1/\{1 + \exp(-\beta_0 - \beta_1 x_i)\}$, $i = 1, \dots, n$, for fixed and known x_i. Find a minimal sufficient statistic for $\theta = (\beta_0, \beta_1)$.

Exercise 6.12. (Keener (2011)) Let $X_1, \dots, X_n$ be independent random variables from a Beta (a, b) distribution. Recall that the density of this distribution is

$$f_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad x \in (0,1),$$

for positive parameters a and b. Find a minimal sufficient statistic (i) when a and b vary freely; (ii) when $a = 2b$; and when (iii) $a = b^2$.

Exercise 6.13. Let $X_1, \dots, X_n$ be independent $\mathrm{N}(\theta, \theta^2)$, $\theta > 0$. Find a minimal sufficient statistic for θ, and show that it is not complete. Explain why Theorem 6.2.25 in Casella and Berger (2002, p. 288) does not apply.

Exercise 6.14. Let $X_1, \dots, X_n$ be independent $\mathrm{N}(\theta, \sigma^2)$. Show that $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ are independent. *Hint*: Use Basu's theorem.

Exercise 6.15. (Casella and Berger (2002)) Let N be a random variable taking values in $\{1, 2, \dots,\}$ with known probabilities $p_1, p_2, \dots$. Given $N = n$, perform n independent Bernoulli trials $X_1, \dots, X_n$ with success probabilities θ.

(a) Show that $(\sum_{i=1}^N X_i, N)$ is minimal sufficient and that N is ancilliary for θ

(b) Show that $N^{-1}\sum_{i=1}^N X_i$ is unbiased for θ, and find its variance.

Exercise 6.16. (Shao (2003)) Let T and S be two statistics such that $S = g(T)$ for some measurable function g. Show that

(a) If T is complete, then S is complete.

(b) If T is complete and sufficient and g is one-to-one, then S is complete and sufficient.

Exercise 6.17. Let X and Y be independent Poisson random variables with means θ and θ^2, respectively. Find a minimial sufficient statistic. Is the minimal sufficient statistic complete?

Exercise 6.18. Let $X_1, \ldots, X_n$ be independent random variables with density,

$$f_\theta(x) = \frac{1}{\theta}\exp\{-(x-\theta)/\theta\}, \quad \text{for } x > \theta > 0,$$

and zero otherwise. Find a statistic that is minimal sufficient for θ. Is the minimal sufficent statistic complete?

Exercise 6.19. Let $X_1, \ldots, X_n$ be independent exponentials with mean $1/\theta$. Show that $\bar{X}_n$ and $\max_{i\leq n} X_i / \min_{i\leq n} X_i$ are independent.

Exercise 6.20. Let $X_1, \ldots, X_n$ be independent uniform random variables on (a, b). Show that $T(X) = (\min_{i\leq n} X_i, \max_{i\leq n} X_i)$ is sufficient and complete.

Exercise 6.21. Let θ be a real-valued parameter that we are to estimate with a loss function $L(\delta, \theta)$ that is convex in δ for each θ. Let $\delta(X)$ be an unbiased estimator of θ and suppose that T is a sufficient statistic. Consider

$$\delta_2(X) = \mathrm{E}_\theta[\delta_1(X) \mid T].$$

(a) Explain why δ_2 is an estimator, and show that it is unbiased.

(b) Recall that the risk function of an estimator δ is $R(\delta, \theta) = \mathrm{E}_\theta\, L(\delta(X), \theta)$. Show that

$$R(\delta_2, \theta) \leq R(\delta_1, \theta), \quad \text{for all } \theta.$$

What additional assumptions do we need for this inequality to be strict?

(c) Suppose that T is also complete. Show that $R(\delta_2, \theta) \leq R(\delta, \theta)$ for all competitors δ.

(d) Go back to Exercise 6.3(c), and explain what you found in view of the current exercise.

7. Miscellaneous exercises

Exercise 7.1. It's raining at Blindern. Is it raining at Huk? The Norwegian Meteorological Institute have pluviometers, or rain gauges, stationed both places, with daily rain measurements in milimeters available on the website yr.no. I found the data for Blindern and Huk here and here, respectively. Here is the rain data for Blindern and for Huk as text-files, use `read.table(path,sep=";")`.

Denote the rain data for the $n = 248$ days from January 1. to September 5., 2019 by,

$$\begin{aligned} &\text{Blindern}: && Y_{B,1}, \ldots, Y_{B,n}; \\ &\text{Huk}: && Y_{H,1}, \ldots, Y_{H,n}. \end{aligned}$$

In order to predict the weather at Huk by looking out the window at Blindern, we need a statistical model.

(a) All you care about is whether or not it's raining, the amount of rain does not bother you. Sketch how you would estimate Pr(Rain at Huk | Rain at Blindern), and how you would assess the uncertainty of this estimate. Think about what assumptions you're making when using your chosen estimator.

(b) A little rain does not stop you. With the model

$$Y_{H,i} = \beta_0 + \beta_1 Y_{B,i} + \varepsilon_i, \quad i = 1, \ldots, n,$$

where the ε_i are independent mean zero random variables with variance σ^2, we can, after having estimated the parameters, say something about quantities such as $\mathrm{E}\,[Y_H \mid Y_B = y]$, $\Pr(Y_H \leq y \mid Y_B = y)$, and so on. Is this a good model for the phenomenon you are interested in? Why, or why not?

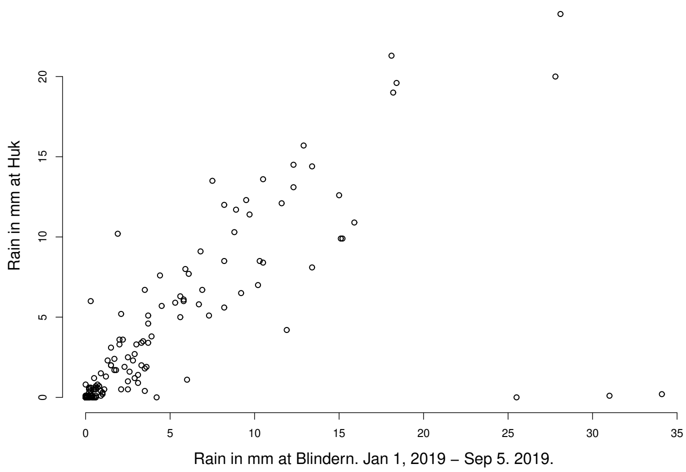

FIGURE 1. Rainfall measured in milimeters at Blindern and Bygdøy weather stations. Data from yr.no, extracted Sept. 5,2019.

(c) Consider the following model. Let $F_{\lambda_j}(y) = 1 - \exp(-\lambda_j y),\ j = B, H$, for $y \geq 0$ and positive parameter λ_j. As (almost) always, let $\Phi(x)$ be the distribution function of the standard normal distribution, and take

$$Y_{j,i} = F_{\lambda_j}^{-1}(\Phi(X_{j,i})), \quad j = H, B,\ i = 1, \dots, n,$$

where

$$\begin{pmatrix} X_{B,i} \\ X_{H,i} \end{pmatrix} \sim \mathrm{N}(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}), \quad i = 1, \dots, n,$$

are independent. Denote $\theta = (\lambda_B, \lambda_H, \rho)$, and derive an expression for the log-likelihood function

$$\ell_n(\theta) = \sum_{i=1}^{n} \log f(y_{B,i}, y_{H,i}; \theta),$$

where $f(y_{B,i}, y_{H,i}; \theta)$ is the density of (Y_B, Y_H).

(d) Are you comfortable with the assumptions the model in (c) is making about rain in Oslo?

(e) If you program the likelihood function you found in (b) and feed it to an optimiser such as `nlm()` in R, things will go astray (the `nlm()`-function is a minimiser, so give it the negative log-likelihood). Why is that? Propose a modification of the model in (c) that takes care of this problem. *Hint*: Fortunately, there are sunny days.

Exercise 7.2. Let X and Y be random variables with $\mathrm{E}\,X = \xi, \mathrm{Var}\,X = \tau^2$, and $\mathrm{E}\,Y = \mu$, $\mathrm{Var}\,Y = \sigma^2$.

$$P(\{|X - \xi| \geq \varepsilon\tau^2\} \cup \{|Y - \mu| \geq \varepsilon\sigma^2\}) \leq \frac{1 + \sqrt{1 - \rho^2}}{\varepsilon^2}.$$

Hint: If X and Y are mean zero random variables with unit variance, and correlation ρ, establish that $\mathrm{E}\,\max\{X^2, Y^2\} \leq 1 + \sqrt{1 - \rho^2}$. Here you may want to use that $\mathrm{E}\,|XY| \leq (\mathrm{E}\,X^2)^{1/2}(\mathrm{E}\,Y^2)^{1/2}$.

Exercise 7.3. Let X be a random variable with probability distribution $P(X = x_j) = p_j$ for $j = 1, \dots, r$, with $r \geq 2$. We define the entropy $H = H(p_1, \dots, p_{r-1})$ of this distribution by

$$H = -\sum_{j=1}^{r} p_j \log(p_j),$$

with $0 \times \log(0) = 0$.

(a) Show that the entropy is maximised when X is uniformly distributed over $\{x_1, \dots, x_r\}$, and that it is minimised when X takes on one value with probability one. *Hint*: Use Jensen's inequality.

(b) Let $X_1, \dots, X_n$ be independent Bernoulli(p) random variables, $0 < p < 1$. Show that $Y = \sum_{i=1}^n X_i$ is Binomial(n, p). Let $B_n(\varepsilon)$ be the set of all Bernoulli(p) sequences of length n, such that $|\sum_{i=1}^n X_i/n - p| < \varepsilon$, for some $\varepsilon > 0$. We'll write $B_n(\varepsilon) = \{|Y/n - p| < \varepsilon\}$. Show that the probability of $B_n(\varepsilon)$ tends to 1 as $n \to \infty$.

(c) Let $H(p)$ be the entropy associated with the Bernoulli(p) distribution. Convince yourself of the following

$$\frac{\text{Bernoulli sequences of length } n \text{ giving } Y = y}{\text{All Bernoulli sequences of length } n} = \binom{n}{y} \exp\{-nH(1/2)\}.$$

Show that

$$|\log P(X_1 = x_1, \dots, X_n = x_n) + nH(p)| \le \varepsilon n |\log p(1-p)|,$$

when $x_1, \dots, x_n$ is a sequence in $B_n(\varepsilon)$.

(d) Define $N_n(\varepsilon)$ to be the number of Bernoulli(p) sequences in $B_n(\varepsilon)$. We are going to show that there is an n_0 such that for all $n \ge n_0$,

$$\exp\{n(H(p) - \varepsilon)\} \le N_n(\eta) \le \exp\{n(H(p) + \varepsilon)\}, \tag{7.1}$$

where $\eta = -\varepsilon/\{2 \log p(1-p)\}$. Write $X^{(n)}$ for Bernoulli sequences of length n, and $x^{(n)} = (x_1, \dots, x_n)$ for zero-one sequences of length n. Show that for $x^{(n)} \in B_n(\eta)$,

$$P(X^{(n)} = x^{(n)}) \le \exp\{-n(H(p) - \varepsilon/2)\}, \quad \text{and} \quad P(X^{(n)} = x^{(n)}) \ge \exp\{-n(H(p) + \varepsilon/2)\}.$$

Use these bounds to show that

$$N_n(\eta) \le \exp\{n(H(p) + \varepsilon/2)\}, \quad \text{and} \quad N_n(\eta) \ge P(B_n(\eta)) \exp\{n(H(p) - \varepsilon/2)\}.$$

Combine this with what you found in (b), and derive (7.1).

8. Convergence concepts

Exercise 8.1. Let $X_1, X_2, \dots$ be independent Bernoulli variables with success probabilities $p_1, p_2, \dots$. We shall investigate when

$$Z_n = \frac{\sum_{i=1}^n (X_i - p_i)}{B_n} \xrightarrow{d} \mathrm{N}(0, 1),$$

where $B_n = \{\sum_{i=1}^n p_i(1-p_i)\}^{1/2}$. Show, using mgf's, that this happens if and only if $\sum_{i=1}^\infty p_i = \infty$. Show also that this condition is equivalent to $B_n \to \infty$. This means that the cases $p_i = 1/i$ and $p_i = 1/i^2$, for example, are very different.

Exercise 8.2. Suppose X_n is Beta($1/n, 1/n$) and X is Bernoulli($1/2$). Show that $X_n \xrightarrow{d} X$. What if X_n is Beta($a/n, b/n$)? *Hint*: See Nils exercise 12.

Exercise 8.3. Suppose X_n is uniformly distributed on $\{1/n, 2/n, \dots, 1\}$. Show that X_n converges in distribution to X, where X is uniform$(0, 1)$. Does $X_n \xrightarrow{p} X$?

Exercise 8.4. A few counterexamples.

(a) Make an example where $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$, but $X_n + Y_n$ does not converge in distribution to $X + Y$.

(b) Make an example where X_n converges to 0 in probability, but $\mathrm{E}\, X_n$ does not converge to 0.

(c) Let Z be uniform$(0, 1)$, and set $X_1 = 1, X_2 = I_{[0,1/2)}(Z), X_3 = I_{[1/2,1)}(Z), X_4 = I_{[0,1/4)}(Z), X_5 = I_{[1/4,1/2)}(Z), \dots$, and so on. Find the probability limit of X_n. Does X_n converge almost surely to this limit?

Exercise 8.5. Let $X_1, X_2, \dots$ be i.i.d. with density $f(x) = ax^{-(a+1)}$ for $x \in (1, \infty)$, and zero otherwise.

(a) For what values of $a > 0$ is it true that $X_n/n \xrightarrow{p} 0$?

(b) For what values of $a > 0$ and $r > 0$ is it true that $\mathrm{E}\, X_n^r/n \to 0$?

(c) For what values of $a > 0$ is it true that $X_n/n \to 0$ almost surely? *Hint*: Use the Borel–Cantelli lemma.

Exercise 8.6. Let $Y_1, \ldots, Y_n$ be independent Poisson variables with density $P_\theta(X = x) = e^{-\theta}\theta^x/x!$, $x = 0, 1, 2, \ldots$, and let Z_n be the proportion of zeros observed, $Z_n = n^{-1}\sum_{i=1}^n I\{X_i = 0\}$. Show that

$$\sqrt{n}\begin{pmatrix}\bar{X}_n - \theta \\ Z_n - e^{-\theta}\end{pmatrix} \overset{d}{\to} \mathrm{N}_2\Big(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\theta & -\theta e^{-\theta} \\ -\theta e^{-\theta} & e^{-\theta}(1-e^{-\theta})\end{pmatrix}\Big).$$

Exercise 8.7. Consider independent observations $Y_1, \ldots, Y_n$ from a normal distribution with expectation μ and variance σ^2.

(a) Write down the log-likelihood function $\ell_n(\mu, \sigma)$, and derive formulae for the maximum-likelihood estimators, say $(\hat{\mu}_n, \hat{\sigma}_n^2)$. Identify also the exact distributions of $\hat{\mu}_n$ and $\hat{\sigma}_n^2$. *Hint*: The maximum-likelihood estimators are the estimators solving $\partial \ell_n(\mu,\sigma)/\partial\mu = 0$ and $\partial\ell_n(\mu,\sigma)/\partial\sigma = 0$.

(b) Show that

$$\sqrt{n}\begin{pmatrix}\hat{\mu}_n - \mu \\ \hat{\sigma}_n - \sigma\end{pmatrix} \overset{d}{\to} \mathrm{N}_2\Big(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma^2 & 0 \\ 0 & \sigma^2/2\end{pmatrix}\Big).$$

Hint: You can show this by computing it in a straight forward manner, using tools that we covered in class. But, if you want to, take a look at the variance of the score function (the first derivative of the log-likelihood) and its inverse; look back at Exercise 5.2; and Taylor-expand the score function around the true values of the parameters. We'll do all this on Thursday.

(c) Consider the parameter $\gamma = \mu/\sigma$, sometimes called the normalised mean. With $\hat{\gamma}_n = \hat{\mu}_n/\hat{\sigma}_n^2$, show that

$$\sqrt{n}(\hat{\gamma}_n - \gamma) \overset{d}{\to} \mathrm{N}(0, 1 + \gamma^2/2).$$

(d) Let now $h(x) = \sqrt{2}\log(x/\sqrt{2} + \sqrt{1 + x^2/2})$. Show that

$$\sqrt{n}(h(\hat{\gamma}_n) - h(\gamma)) \overset{d}{\to} \mathrm{N}(0, 1).$$

9. Finding estimators

Exercise 9.1. Let $X_1, \ldots, X_n$ be independent Gamma(a, b) random variables with density

$$f_{(a,b)}(x) = \frac{b^a}{\Gamma(a)} x^{a-1}\exp\{-bx\}, \quad x > 0. \tag{9.1}$$

(a) Write down the log-likelihood function $\ell_n(a, b)$.

(b) Suppose a is known. Find the maximum likelihood estimator $\hat{b}_n$ of b. Why is it unique?

(c) Find the limiting distribution of $\sqrt{n}(\bar{X}_n - a/b)$. Use this to find the limiting distribution of $\sqrt{n}(\hat{b}_n - b)$. Propose an approximate 95% confidence interval for b.

(d) Suppose now that both a and b are unknown. Write down the score functions $u_1(a, b, x) = \partial \log f_{(a,b)}(x)/\partial a$ and $u_2(a, b, x) = \partial \log f_{(a,b)}(x)/\partial b$, and show that

$$\mathrm{E}\begin{pmatrix}u_1(a, b, X) \\ u_2(a, b, X)\end{pmatrix} = 0.$$

Find also the variance matrix $J(a, b)$ of $(u_1(a, b, X), u_2(a, b, X))^{\mathrm{t}}$. *Hint*: Use what you found in Exercise 5.2.

(e) Explain why the maximiser $(\hat{a}_n, \hat{b}_n)$ of $\ell_n(a, b)$ is unique. Download the rain data for Blindern from Exercise 7.1, and remove the days with no rain. Fit a Gamma(a, b) model to these data. Provide point estimates and standard errors of your estimators. Here is a sketch of how to do this in R (there surely are other ways to go about this).

```
path <- "https://www.uio.no/studier/emner/matnat
/math/STK4011/data/rain_blindern2019_tom5sep.txt"
rain <- read.table(path,sep=";")
yy_full <- rain$rain_mm ;yy <- yy_full[yy_full>0] ; nn <- length(yy)
loglik <- function(params){
aa <- params[1] ; bb <- params[2]
ll <- # the log-likelihood here
```

```
return(ll)}
min_loglik <- function(params){ # nlm() is a minimiser
return(-loglik(params)) }
fit <- nlm(min_loglik,c(start_aa,start_bb),hessian=TRUE) # provide start values for nlm
```

Exercise 9.2. Consider the model

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

for independent $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, and fixed constants $x_1, \ldots, x_n$.

(a) Write down the log-likelihood function of this model.

(b) Find its score function $(u_1(\beta, \sigma^2, Y), u_2(\beta, \sigma^2, y))^{\mathrm{t}}$, and show that $\sum_{i=1}^n (u_1(\beta, \sigma^2, Y_i), u_2(\beta, \sigma^2, Y_i))^{\mathrm{t}}$ can be expressed in terms of a chi-square and normal random variable, and that these are independent. *Hint*: The third central moment of a normal distribution is zero.

(c) Find the observed information matrix $J_n(\beta, \sigma^2)$.

(d) Suppose σ^2 is known. Derive a 95 percent confidence interval for β. *Hint*: Start by identifying the distribution of $\hat{\beta}_n$.

Exercise 9.3. Let $X_1, \ldots, X_n$ be i.i.d. from a distribution with density $f_{\theta_0}(x)$, where $\theta_0 \in \Theta \subset \mathbb{R}$, and θ_0 denotes the true parameter value. Let $U_n(\theta) = \sum_{i=1}^n \partial \log f_\theta(X_i)/\partial\theta$. The maximum likelihood estimator is the value $\hat{\theta}_n$ of θ such that $U_n(\hat{\theta}_n) = 0$. By Taylor's theorem, assuming it is applicable, there exists a (random) value $\tilde{\theta}_n$ between $\hat{\theta}_n$ and θ_0 such that

$$0 = U_n(\hat{\theta}_n) = U_n(\theta_0) + U_n'(\theta_0)(\hat{\theta}_n - \theta_0) + \frac{1}{2} U_n''(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0)^2, \tag{9.2}$$

where $U_n'(\theta) = \partial U_n(\theta)/\partial\theta$ and $U_n''(\theta) = \partial^2 U_n(\theta)/\partial\theta^2$. Under what conditions do you expect this to lead to

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathrm{N}(0, 1/J(\theta_0)),$$

with $J(\theta_0)$ the Fisher information matrix evaluated in the true value.

References

Billingsley, P. (1995). *Probability and Measure. Third Edition.* Wiley.

Casella, G. and Berger, R. L. (2002). *Statistical Inference. Second Edition.* Duxbury Pacific Grove, CA.

Ferguson, T. S. (1996). *A course in large sample theory.* Chapman & Hall/CRC.

Hjort, N. L. and Fenstad, G. (1992). On the last time and the number of times an estimator is more than ε from its target value. *The Annals of Statistics*, 20:469–489.

Keener, R. W. (2011). *Theoretical statistics: Topics for a core course.* Springer, New York.

Lehmann, E. L. and Casella, G. (1999). *Theory of point estimation. Second edition.* Springer, New York.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses. Third edition.* Springer, New York.

Romano, J. P. and Siegel, A. F. (1986). *Counterexamples in Probability and Statistics.* Wadsworth & Brooks/Cole.

Schervish, M. J. (1995). *Theory of Statistics.* Springer, New York.

Shao, J. (2003). *Mathematical Statistics. Second Edition.* Springer.

Shiryaev, A. N. (1990). *Probability. Second Edition.* Springer.

Van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press.

Young, G. A. and Smith, R. L. (2005). *Essentials of Statistical Inference.* Cambridge University Press.