STK4011 – Statistical inference. Autumn 2019

Exercises

Emil Aas Stoltenberg

September 1, 2019

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 then 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 $Y_1, \dots, Y_n$, 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.

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 $X_1, \dots, X_n$ are independent Bernoulli random variables with expectations $p_{1,n}, \dots, p_{n,n}$. 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 $P(0 \leq X_n \leq 1) = P(0 \leq X \leq 1) = 1$ 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 $k = 1, 2, \dots$. 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 $\sqrt{n}(\hat{\theta}_n - \theta)$.