Avd. Matematisk statistik

KTH Matematik

EXAM IN SF2955 COMPUTER INTENSIVE METHODS
WEDNESDAY 27th of May 2009 08.00 – 13.00.

Examiner: Timo Koski, tel. 790 71 34. Email: `tjtkoski@kth.se`

Allowed aids: Formel- och tabellsamling i Matematisk statistik. Formulas in Bayesian Inference. Pocket calculator.

There are six (6) assignments (uppgift). Notation should be explained and defined. Arguments must be detailed enough to make it possible to follow. Numerical answers should be given with the precision of at least 2 significant digits.

Results will be ready before June 12th. You will be informed by email if you specify your email address.

Good luck!

- -

Uppgift 1

A probability distribution for a continuous random variable with the density

$$f(x) = \begin{cases} \frac{\alpha q^{\alpha}}{x^{\alpha+1}} & x > q, \\ 0 & x \leq q, \end{cases} \tag{1}$$

where $q > 0$, $\alpha > 0$, is called a **Pareto distribution** with parameters q and α. The distribution function is thus

$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} 1 - \frac{q^{\alpha}}{x^{\alpha}} & x \geq q, \\ 0 & x \leq q. \end{cases} \tag{2}$$

Let us consider the deterministic (= non-random) exponential function, or

$$X(t) = qe^{\mu t}, \quad q > 0, \mu > 0, \quad t \geq 0.$$

We observe or stop $X(t)$ at an exponentially distributed time $T \in \mathrm{Exp}(\nu)$.

a) Show that the distribution of $X(t)$ at the time of stopping, $X(T)$, has a Pareto distribution with parameters $\alpha = \nu/\mu$ and q. (5 p)

b) Give at least two ways to generate Pareto distributed pseudo-random variables, when you have access to a generator of uniform pseudo-random variables $U \in U(0,1)$. (5 p)

Uppgift 2

Here you are asked to write down some of the ideas of statistical inference by bootstrap. You do not in principle need to do any calculations.

The setting is as follows: the samples $x_1, x_2, \cdots, x_n$ regarded as respective outcomes of I.I.D. random variables $X_1, X_2, \cdots, X_n$, which have the distribution F, where $\theta = t(F)$ is estimated with $\widehat{\theta}(x_1, x_2, \cdots, x_n)$, which we take to be a plug-in estimate.

a) Let the distribution function of $S = \left(\widehat{\theta}(X_1, X_2, \cdots, X_n) - \theta\right)$ be

$$H(z) = P\left(S \leq z\right).$$

Express the *basic bootstrap hypothesis* in terms of $H(z)$. (2 p)

b) We consider next

$$T = \frac{S}{\widehat{\sigma}}$$

where $\widehat{\sigma} = \widehat{\sigma}(X_1, X_2, \cdots, X_n)$ is an estimate of the standard error of $\widehat{\theta}(x_1, x_2, \cdots, x_n)$. Let

$$P\left(T \leq z\right) = G(z). \tag{3}$$

What do we mean by the statement that T is a *pivotal* variable? What is the bootstrap estimate $G^*(z)$ of $G(z)$? (2 p)

c) Let $G(z_\alpha) = \alpha$, where $G(z)$ is the distribution function in (3). Then an exact two-sided confidence interval for θ with degree of confidence $1-\alpha$ is

$$\left(\widehat{\theta}(x_1, x_2, \cdots, x_n) - \widehat{\sigma} z_{1-\alpha/2}, \widehat{\theta}(x_1, x_2, \cdots, x_n) - \widehat{\sigma} z_{\alpha/2}\right).$$

If we do not have full knowledge of $G(z)$, we cannot find $z_{1-\alpha/2}$ and $z_{\alpha/2}$. Since we do not in general know the bootstrap distribution function $G^*(z)$ either, we try to estimate $z^*_{1-\alpha/2}$ and $z^*_{\alpha/2}$ that solve $G^*(z^*_{1-\alpha/2}) = 1-\alpha/2$ and $G^*(z^*_{\alpha/2}) = \alpha/2$. Explain how we can estimate the percentiles $z^*_{1-\alpha/2}$ and $z^*_{\alpha/2}$ using bootstrap samples. Give the resulting confidence interval, too. (4 p)

d) The bootstrap confidence interval found in **c)** is also known as the **bootstrap-t** confidence interval. Explain in some detail the rationale behind this terminology. (2 p)

Uppgift 3

Let the estimate of the parameter θ be $\widehat{\theta} = \bar{x}$. We make a bootstrap of this estimate with the sample $x_1, x_2, \cdots, x_n$, i.e., we compute $\widehat{\theta}^* = \bar{X}^*$, where $\bar{X}^* = \sum_{i=1}^{n} X_i^*/n$ and $X_1^*, X_2^*, \cdots, X_n^*$ are I.I.D. with $P(X_i^* = x_j) = 1/n$ för $i, j = 1, 2, \cdots n$.

a) Compute $E(\bar{X}^*)$ as a function of $x_1, x_2, \cdots, x_n$. (5 p)

b) Compute $V(\bar{X}^*)$ as a function of $x_1, x_2, \cdots, x_n$. (5 p)

Uppgift 4

Here you are required to demonstrate that Bayes' rule can be understood as a mechanism to generate samples from a posterior distribution by sampling the prior. The insight is implemented in an algorithm called *Bayesian sampling-resampling.* We state first some preliminaries.

Consider a statistical model with the density function, or alternatively, the probability mass function, for data $x \sim f(x \mid \theta)$ and the prior density $\pi(\theta)$. The likelihood function for θ given the data x is

$$l(x;\theta) = f(x \mid \theta).$$

The maximum likelihood estimate of θ given the data x is

$$\widehat{\theta}_{\mathrm{ML}} = \widehat{\theta}_{\mathrm{ML}}(x) = \mathrm{argmax}_{\theta} l(x;\theta).$$

Set

$$M = l\left(x; \widehat{\theta}_{\mathrm{ML}}\right)$$

and

$$\pi_x(\theta) = l(x;\theta)\,\pi(\theta).$$

The **Bayesian sampling-resampling algorithm** is now given by:

1. Generate a random number Θ according to $\pi(\theta)$.
2. Generate a random number U, independently of Θ, which is uniformly distributed on $[0,1]$ and accept the proposal $\Theta = \theta$, if
$$U \leq \frac{\pi_x(\theta)}{M\pi(\theta)}.$$
3. Otherwise generate a new proposal according to 1. and then repeat 2. and keep doing that until we finally accept a proposal.

Call the accepted proposal X. Clearly, Bayesian sampling-resampling is nothing but an instance of **acceptance sampling**.

a) Show that the density of X is the posterior density $\pi(\theta|x)$. *Hint:* Compute $P(X \leq \theta_0) = P(\Theta \leq \theta_0 \mid \Theta \text{ accepted})$. (6 p)

b) What can you say in a qualitative way about the likelihood of those proposals that tend to be accepted ? (1 p)

c) You have obtained n samples of I.I.D. random variables with Bernoulli distribution $\mathrm{Be}(\theta)$ and the samples contain x ones (1), $0 \leq x \leq n$. Give a Bayesian sampling-resampling algorithm for generating samples from the Beta distribution $\mathcal{B}\mathrm{eta}(x+1, n+1-x)$. (3 p)

Uppgift 5

We want to do MCMC for sampling a target distribution $\mathbf{f} = (f_i)_{i=0}^{J}$. **Barker's algorithm** for MCMC is as follows.
$\mathbf{Q} = \left(q_{i|j}\right)_{i=0,j=0}^{J,J}$ is a probability matrix (proposal distribution), i.e.,

$$q_{i|j} = P\left(Y_{n+1} = j \mid X_n = i\right).$$

We assume that $\mathbf{Q}$ is symmetric.

1. $X_n = i$
2. Generate $Y_{n+1} \sim \{q_{i|j}\}_{j=0}^{J}$.
3. Take

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{with probability } \alpha^{B}_{i,Y_{n+1}} \\ i & \text{with probability } 1 - \alpha^{B}_{i,Y_{n+1}}, \end{cases}$$

   where

$$\alpha^{\mathrm{B}}_{i,j} = \frac{f_j}{f_i + f_j}. \tag{4}$$

4. $n + 1 \to n$. Go to 1.

a) What should be regarded as a main difference in the working of Barker's algorithm compared to the Metropolis-Hastings algorithm (with symmetric proposal distribution $\mathbf{Q}$)? Which method seems preferable and why ? (3 p)

Hint: Consider the situation where $Y_{n+1} = j$ and $X_n = i$, and $f_j = f_i$.

b) Find $p_{i|j} = P\left(X_{n+1} = j \mid X_n = i\right)$ for Barker's algorithm, if $i \neq j$. (1 p)

c) Show that Barker's algorithm generates a Markov chain $\{X_n\}_{n \geq 0}$ with the target distribution $\mathbf{f}$ as its stationary distribution. (6 p)

Uppgift 6

We consider two models $\mathcal{M}_p$, $p = 1, 2$, for $Y_1, Y_2, \ldots, Y_n$, that are

$$\mathcal{M}_1 : \quad Y_i = \alpha + \epsilon_i, \quad i = 1, \ldots, , n$$

and

$$\mathcal{M}_2 : \quad Y_i = \alpha + \beta x_i + \epsilon_i \quad i = 1, \ldots, , n.$$

Under both models $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are I.I.D. and $N(0, \sigma)$ (note that this is as in *Formel- och tabellsamling*, σ is the standard deviation), where we assume that $\sigma > 0$ is known. In the model $\mathcal{M}_2$ the variables $x_1, x_2, \ldots, x_n$ are known fixed values (levels) of a regressor variable. In the model $\mathcal{M}_1$ we have thus independent $Y_i \in N(\alpha, \sigma)$ for $i = 1, 2, \ldots, n$. In the model $\mathcal{M}_2$ we have independent $Y_i \in N(\alpha + \beta x_i, \sigma)$ for $i = 1, 2, \ldots, n$.
The joint probability density of $Y_1, Y_2, \ldots, Y_n$ under the model $\mathcal{M}_p$ is denoted by

$$f_{Y_1, Y_2, \ldots, Y_n}\left(y_1, y_2, \ldots, y_n; \theta^{(p)}\right),$$

where $\theta^{(1)} = (\alpha)$ in $\mathcal{M}_1$ and $\theta^{(2)} = (\alpha, \beta)$ in $\mathcal{M}_2$, i.e., the models are nested, so that we may think of $\theta^{(1)}$ as $= (\alpha, 0)$.
We obtain observations $y_1, y_2, \ldots, y_n$ on $Y_1, Y_2, \ldots, Y_n$, respectively, and we want to choose between $\mathcal{M}_1$ and $\mathcal{M}_2$ using these observations.
The **AIC principle of data driven model choice** selects the model that minimizes the AIC criterion

$$Q_{\text{AIC}}\left(\mathcal{M}_p\right) = -\frac{1}{n} \ln f_{Y_1, Y_2, \ldots, Y_n}\left(y_1, y_2, \ldots, y_n; \widehat{\theta}^{(p)}\right) + \frac{p}{n}, \tag{5}$$

where p is the number of parameters in the model $\mathcal{M}_p$ and

$$f_{Y_1, Y_2, \ldots, Y_n}(y_1, y_2, \ldots, y_n; \widehat{\theta}^{(p)})$$

is the joint probability density of $Y_1, Y_2, \ldots, Y_n$ evaluated at the maximum likelihood estimate $\widehat{\theta}^{(p)}$ of $\theta^{(p)}$. In other words we have

$$-\ln f_{Y_1, Y_2, \ldots, Y_n}(y_1, y_2, \ldots, y_n; \widehat{\theta}^{(p)}) = \min_{\theta^{(p)}} \left(-\ln f_{Y_1, Y_2, \ldots, Y_n}(y_1, y_2, \ldots, y_n; \theta^{(p)})\right).$$

a) Verify that

$$-\ln f_{Y_1, Y_2, \ldots, Y_n}(y_1, y_2, \ldots, y_n; \widehat{\theta}^{(2)}) = \frac{n}{2} \ln 2\pi + n \ln \sigma + \min_{\alpha, \beta} \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - (\alpha + \beta x_i)\right)^2. \tag{6}$$

(1 p)

Commment: By (6), the maximum likelihood estimate of $\theta^{(2)} = (\alpha, \beta)$ coincides with the least squares (minsta kvadrat) estimate and we can in the sequel apply section 13 of *Formel- och tabellsamling*. Furthermore, a modification of this argument gives the maximum likelihood estimate of $\theta^{(1)} = (\alpha)$ as the arithmetic mean (**this you need not show**)

$$\widehat{\alpha} = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i. \tag{7}$$

b) Now check that

$$Q_{\text{AIC}}\left(\mathcal{M}_2\right) = \frac{1}{2}\ln 2\pi + \ln\sigma + \frac{1}{2n\sigma^2}\left(S_{yy} - \widehat{\beta}S_{xy}\right) + \frac{2}{n}, \tag{8}$$

where, following section 13.3 on p. 6 of the *Formel- och tabellsamling*,

$$S_{yy} = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2, \tag{9}$$

and

$$S_{xy} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right), \tag{10}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, and

$$\widehat{\beta} = \frac{S_{xy}}{S_{xx}}, \tag{11}$$

where

$$S_{xx} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2. \tag{12}$$

Hint: You may find the result (6) in part **a)** and a formula in section 13.3 on p. 6 of the *Formel- och tabellsamling* useful. In that respect note that β^*_{obs} in the *Formel- och tabellsamling* is $= \widehat{\beta}$ in our notation. You are allowed to use (6) from **a)** even if you have not done **a)**. (1 p)

c) Show that

$$Q_{\text{AIC}}\left(\mathcal{M}_1\right) = \frac{1}{2}\ln 2\pi + \ln\sigma + \frac{1}{2\sigma^2 n}S_{yy} + \frac{1}{n}, \tag{13}$$

where S_{yy} is as defined in (9). (2 p)

d) Show that the AIC criterion is minimized by the model $\mathcal{M}_2$, as soon as

$$\frac{\mid \widehat{\beta} \mid}{\frac{\sigma}{\sqrt{S_{xx}}}} > \sqrt{2}.$$

In other words, the AIC model choice turns out here to imply a statistical test on $\widehat{\beta}$. You are allowed to use **b)** and **c)** even if you have not done these. (3 p)

e) Find the probability

$$P\left(\mathcal{M}_2 \text{ is selected by AIC } \mid \mathcal{M}_1 \text{ is true}\right).$$

What does this tell about the asymptotic consistency of the AIC ? You may solve this even if you have not done **d)**. (3 p)

Hint: You need to regard $\widehat{\beta}$ in (11) as a random variable, i.e., you must take

$$S_{xy} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(Y_i - \bar{Y}\right).$$

Then you may find it useful to consider the formula a) in section 13.1 on p. 5 of *Formel- och tabellsamling.*

SOLUTIONS TO THE EXAM IN SF2955 COMPUTER INTENSIVE METHODS IN MATHEMATICAL STATISTICS
WEDNESDAY 27th of MAY 2009 08.00 a.m. – 13.00 a.m..

Uppgift 1

a) Now, we are asked to compute for any x the distribution

$$P\left(X(T) \leq x\right) = P\left(qe^{\mu T} \leq x\right) = P\left(e^{\mu T} \leq x/q\right) = P\left(T \leq \frac{\ln(x/q)}{\mu}\right)$$

Since $T \in \text{Exp}(\nu)$, $P\left(T \leq \frac{\ln(x/q)}{\mu}\right) = 0$, if $\frac{\ln(x/q)}{\mu} \leq 0$, i.e., $x/q \leq 1$ and $x \leq q$.
Hence for $x > q$ we get from (15)

$$P\left(T \leq \frac{\ln(x/q)}{\mu}\right) = 1 - e^{-\frac{\nu}{\mu}\ln(x/q)} = 1 - \left(\frac{1}{x/q}\right)^{\frac{\nu}{\mu}} = 1 - \left(\frac{q}{x}\right)^{\frac{\nu}{\mu}}. \tag{14}$$

In view of (2), we have in (14) a Pareto distribution with parameters $\alpha = \frac{\nu}{\mu}$ and q.

b) We use the inverse function method. If $U \in U(0,1)$, then it is well known that $T = -\frac{1}{\nu}\ln(1-U)$ has the exponential distribution $\text{Exp}(\nu)$. Hence we may by part **a)** generate a Pareto distributed random variable X by

$$X = qe^{-\mu/\nu \ln(1-U)} = q\left(\frac{1}{1-U}\right)^{\mu/\nu}.$$

Note that since $0 < U < 1$, we have $X > q$.
For another application of the inverse method, we note that if $x \geq q$ and $0 < y < 1$, we can solve

$$F(x) = 1 - \frac{q^\alpha}{x^\alpha} = y \tag{15}$$

as

$$x = F^{-1}(y) = \frac{q}{(1-y)^{1/\alpha}}.$$

Hence, if $U \in U(0,1)$, then $X = F^{-1}(U)$ has a Pareto distribution. But, however, the final result seems to amount to more or less the same method as the preceding.

Uppgift 2

a) Let $X_1^*, X_2^*, \cdots, X_n^*$ be a bootstrap sample drawn with replacement from $x_1, x_2, \cdots, x_n$ with uniform probability. The bootstrap hypothesis is that the distribution of $\widehat{\theta}(X_1, X_2, \cdots, X_n) - \theta$ can be well approximated with the distribution of $\widehat{\theta}(X_1^*, X_2^*, \cdots, X_n^*) - \widehat{\theta}(X_1, X_2, \cdots, X_n)$.
Hence, if

$$H(z) = P\left(S \leq z\right)$$

and

$$H^*(z) = P\left(\widehat{\theta}(X_1^*, X_2^*, \cdots, X_n^*) - \widehat{\theta}(X_1, X_2, \cdots, X_n) \leq z\right).$$

Then the bootstrap hypothesis is that

$$H(z) \approx H^*(z).$$

b) If

$$T = \frac{S}{\widehat{\sigma}}$$

is pivotal, then the function $G(z)$ in

$$P\left(T \leq z\right) = G(z)$$

does not depend on the parameter θ. The bootstrap estimate $G^*(z)$ of $G(z)$ is

$$G^*(z) = P\left(\frac{\widehat{\theta}(X_1^*, X_2^*, \cdots, X_n^*) - \widehat{\theta}(X_1, X_2, \cdots, X_n)}{\widehat{\sigma}(X_1^*, X_2^*, \cdots, X_n^*)} \leq z \mid X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\right),$$

where $X_1^*, X_2^*, \cdots, X_n^*$ are bootstrap sample variables drawn with replacement from $x_1, x_2, \cdots, x_n$ with uniform probability.

c) One can use a Monte Carlo method with repeated bootstrap samples. Let

$$\mathbf{x}^{(b)^*} = \left(x_1^{(b)^*}, x_2^{(b)^*}, \cdots, x_n^{(b)^*}\right), \quad b = 1, \ldots, B$$

be B bootstrap samples of $x_1, x_2, \cdots, x_n$. Let

$$\widehat{\theta}(\mathbf{x}^{(b)^*}) = \widehat{\theta}\left(x_1^{(b)^*}, x_2^{(b)^*}, \cdots, x_n^{(b)^*}\right)$$

be the estimate computed using the bth bootsstrap sample for $b = 1, \ldots, B$, and

$$\widehat{\sigma^*} = \widehat{\sigma^*}\left(\mathbf{x}^{(b)^*}\right)$$

be the corresponding estimate of standard error for $b = 1, \ldots, B$. Then we set

$$T^*(b) = \frac{\widehat{\theta}(\mathbf{x}^{(b)^*}) - \widehat{\theta}}{\widehat{\sigma^*}}$$

and estimate G^* by the empirical distribution G_B^* of $T^*(b)$, i.e,

$$G_B^*(z) = \frac{\text{number of } T^*(b) \leq z}{B}.$$

Then we solve and $z_{\alpha/2}^*$ from the equation

$$\frac{\alpha}{2} = \frac{\text{number of } T^*(b) \leq z_{\alpha/2}^*}{B},$$

and similarly for $z_{1-\alpha/2}^*$. A practical way to do this is to find the order variables $T^*(b)_{(k)}$ of the bootstrapped ratios $T^*(b)$, i.e.,

$$T^*(b)_{(1)} < T^*(b)_{(2)} < \ldots T^*(b)_{(B)}$$

and take $z^*_{B,\alpha/2}$ as $T^*(b)_{(k)}$ where $k = \lfloor (B+1)\alpha/2 \rfloor$ and $z^*_{B,1-\alpha/2}$ as $T^*(b)_{(l)}$, where $k = \lfloor (B+1)/(1-\alpha/2) \rfloor$. Then the boostrapped confidence interval is

$$\left(\widehat{\theta}(x_1, x_2, \cdots, x_n) - \widehat{\sigma} z^*_{B,1-\alpha/2}, \widehat{\theta}(x_1, x_2, \cdots, x_n) - \widehat{\sigma} z^*_{B,\alpha/2}\right).$$

d) If $X_1, X_2, \cdots, X_n$ are I.I.D $N(\theta, \sigma)$ and θ is estimated with $\widehat{\theta}(x_1, x_2, \cdots, x_n) = \bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, and $\widehat{\sigma} = s/\sqrt{n}$, where

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2},$$

then

$$T = \frac{S}{\widehat{\sigma}}$$

is a pivotal variable, t-distributed with $n-1$ degrees of freedom, or $T \in \mathbf{t}(n-1)$. Hence the confidence interval

$$\left(\widehat{\theta}(x_1, x_2, \cdots, x_n) - \widehat{\sigma} z_{1-\alpha/2}, \widehat{\theta}(x_1, x_2, \cdots, x_n) - \widehat{\sigma} z_{\alpha/2}\right).$$

is an exact t-interval, i.e., $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are respective percentiles of $\mathbf{t}(n-1)$.

a) Since $P(X_i^* = x_j) = 1/n$ för $i = 1, 2, 3, \cdots, n$, $j = 1, 2, \cdots, n$, which means that the random variables X_i^* have a uniform discrete distribution, we get

$$E(\bar{X}^*) = E(\frac{1}{n}\sum_{i=1}^n X_i^*) = \frac{1}{n}\sum_{i=1}^n E(X_i^*) = E(X_1^*) = \sum_{j=1}^n x_j P(X_1^* = x_j) = \frac{1}{n}\sum_{j=1}^n x_j = \bar{x}.$$

ANSWER **a)**: $\underline{E(\bar{X}^*) = \bar{x}.}$

b) Since $X_1^*, X_2,^*, \cdots, X_n^*$ are independent, we get

$$V(\bar{X}^*) = V(\frac{1}{n}\sum_{i=1}^n X_i^*) = \frac{1}{n^2}\sum_{i=1}^n V(X_i^*) = \frac{1}{n}V(X_1^*) = \frac{1}{n}E\left((X_1^* - E(X_1^*))^2\right) =$$

$$= \frac{1}{n}E\left((X_1^* - \bar{x})^2\right) = \frac{1}{n}\sum_{j=1}^n (x_j - \bar{x})^2 \frac{1}{n} = \frac{1}{n^2}\sum_{j=1}^n (x_j - \bar{x})^2.$$

ANSWER **b)**: $\underline{V(\bar{X}^*) = \frac{1}{n^2}\sum_{j=1}^n (x_j - \bar{x})^2}$

Uppgift 4

a) We compute $P\left(\Theta \leq \theta_0 \mid \Theta \text{ accepted}\right)$. We have

$$P\left(\Theta \leq \theta_0 \mid \Theta \text{ accepted}\right) = \frac{P(\Theta \leq \theta_0 \quad \& \quad \Theta\text{accepted})}{P(\Theta \text{ accepted})}.$$

We start with the denominator.

$$P(\Theta \text{ accepted}) = P\left(U \leq \frac{\pi_x(\theta)}{M\pi(\theta)}\right) = \int P\left(U \leq \frac{\pi_x(\theta)}{M\pi(\theta)} \middle| \Theta = \theta\right) \pi(\theta) d\theta =$$

$$(U \text{ and } \Theta \text{ are independent and as } \quad U \in U(0,1)) =$$

$$\int P\left(U \leq \frac{\pi_x(\theta)}{M\pi(\theta)}\right) \pi(\theta) d\theta = \int \frac{\pi_x(\theta)}{M\pi(\theta)} \pi(\theta) d\theta$$

$$= \int \frac{\pi_x(\theta)}{M} d\theta = \frac{1}{M} \int \pi_x(\theta)\, d\theta = \frac{1}{M} \int l(x;\theta)\, \pi(\theta)\, d\theta. \tag{16}$$

Next, in the same manner,

$$P(\Theta \leq \theta_0 \quad \& \quad \Theta \text{ accepted}) =$$

$$= \int P\left(\Theta \leq \theta_0 \quad \& \quad U \leq \frac{\pi_x(\theta)}{M\pi(\theta)} \middle| \Theta = \theta\right) \pi(\theta) d\theta = \int^{\theta_0} P\left(U \leq \frac{\pi_x(\theta)}{M\pi(\theta)}\right) \pi(\theta) d\theta =$$

$$\int^{\theta_0} \frac{\pi_x(\theta)}{M\pi(\theta)} \pi(\theta) d\theta = \frac{1}{M} \int^{\theta_0} \pi_x(\theta)\, d\theta = \frac{1}{M} \int^{\theta_0} l(x;\theta)\, \pi(\theta)\, d\theta.$$

Thus, from (16) we get

$$P(\Theta \leq \theta_0 \mid \Theta \text{ accepted}) = \frac{P(\Theta \leq \theta_0 \quad \& \quad \Theta \text{ accepted})}{P(\Theta \text{ accepted})} =$$

$$\frac{\frac{1}{M} \int^{\theta_0} l(x;\theta)\, \pi(\theta)\, d\theta}{\frac{1}{M} \int l(x;\theta)\, \pi(\theta)\, d\theta} = \frac{\int^{\theta_0} l(x;\theta)\, \pi(\theta)\, d\theta}{\int l(x;\theta)\, \pi(\theta)\, d\theta}$$

If we differentiate w.r.t θ_0 we get with a change of variable $\theta_0 \mapsto \theta$

$$\frac{d}{d\theta} P(\Theta \leq \theta \mid \Theta \text{ accepted}) = \frac{l(x;\theta)\, \pi(\theta)\, d\theta}{\int l(x;\theta)\, \pi(\theta)\, d\theta}$$

and this is nothing but the formula for the posterior density

$$\pi(\theta|x) = \frac{f(x \mid \theta) \cdot \pi(\theta)}{\int f(x \mid \theta) \cdot \pi(\theta)\, d\theta}.$$

Thus the density of X is the posterior density $\pi(\theta|x)$.

b) It holds by the given construction that

$$U \leq \frac{\pi_x(\theta)}{M\pi(\theta)} \Leftrightarrow U \leq \frac{\pi_x(\theta)}{M\pi(\theta)} = \frac{l(x;\theta)\, \pi(\theta)}{l\left(x;\widehat{\theta}_{\mathrm{ML}}\right) \pi(\theta)}$$

$$\Leftrightarrow$$

$$U \leq \frac{l(x;\theta)}{l\left(x;\widehat{\theta}_{\mathrm{ML}}\right)}.$$

Hence if a proposed θ has a high likelihood, it will have a high probability of being accepted.

c) With n samples of I.I.D. random variables with Bernoulli distribution Be (θ) and contain x ones (1), $0 \leq x \leq n$ we get

$$f(x|\theta) = P(X = x \mid \theta) = \begin{pmatrix} n \\ k \end{pmatrix} \theta^x (1-\theta)^{n-x}.$$

If we take $p(\theta)$ as the uniform density on $0 < \theta < 1$, then we know (see the *Formulas on Bayesian Inference*) that the posterior density is $\mathcal{B}\mathrm{eta}(x+1, n+1-x)$. From *Formulas on Bayesian Inference* we obtain also that

$$\widehat{\theta}_{\mathrm{ML}} = \frac{x}{n}.$$

Hence the sought-after Bayesian sample-resample algorithm is

1. Generate a random number Θ which is uniformly distributed on $[0,1]$.
2. Generate a random number U, independently of Θ, which is uniformly distributed on $[0,1]$ and accept the proposal $\Theta = \theta$, if

$$U \leq \frac{\theta^x (1-\theta)^{n-x}}{\left(\frac{x}{n}\right)^x (1-\frac{x}{n})^{n-x}}.$$

3. Otherwise generate a new proposal according to 1. and then repeat 2. and keep doing that until we finally accept a proposal.

Uppgift 5

a) The Metropolis-Hastings algorithm (with symmetric proposal distribution $\mathbf{Q}$) has the acceptance probability

$$\alpha_{i,j} = \min\left(1, \frac{f_j}{f_i}\right)$$

Hence, if $Y_{n+1} = j$ is proposed, when $X_n = i$, and $f_j = f_i$, the proposal is always accepted. In Barker's algorithm we get

$$\alpha^{\mathrm{B}}_{i,j} = \frac{f_j}{f_i + f_j} = \frac{1}{2}.$$

Hence Barker's algorithm tosses a fair coin about accepting the proposal $Y_{n+1} = j$ or staying at the current value $X_n = i$. Metropolis-Hastings algorithm seems therefore preferable, as it leads to a more exhaustive sampling of the set $0, 1, \ldots, J \leq \infty$.

b) When $i \neq j$ we get by definition

$$p_{i|j} = P(X_{n+1} = j \mid X_n = i) = P(Y_{n+1} = j \quad \& \quad \text{accepted} \mid X_n = i)$$

$$= P(Y_{n+1} = j \mid \text{accepted}, X_n = i)\, P(\text{accepted} \mid X_n = i)$$

$$= P(Y_{n+1} = j \mid X_n = i)\, P(\text{accepted} \mid X_n = i)$$

since Y_{n+1} is generated independently of acceptance. By construction of Barker's algorithm for MCMC we thus get

$$p_{i|j} = q_{i|j}\alpha^{\mathrm{B}}_{i,j}.$$

$$\text{ANSWER } \mathbf{b)}\text{: } \underline{p_{i|j} = q_{i|j}\alpha^{\mathrm{B}}_{i,j}.}$$

c) The chain $\{X_n\}_{n\geq 0}$ has clearly the Markov property, as the probability of the next value X_{n+1} by construction does not depend on other preceding values than X_n. In order to show that Barker's algorithm generates a Markov chain $\{X_n\}_{n\geq 0}$ with the target distribution $\mathbf{f}$ as the stationary distribution we check the reversibility condition, i.e.,

$$f_i p_{i|j} = f_j p_{j|i}.$$

Now we have from the answer in **b)** and from (4)

$$f_i p_{i|j} = f_i q_{i|j}\alpha^{\mathrm{B}}_{i,j} = f_i q_{i|j}\frac{f_j}{f_i + f_j} =$$

$$= f_j q_{i|j}\frac{f_i}{f_i + f_j} = f_j q_{j|i}\frac{f_i}{f_j + f_i}$$

as $\mathbf{Q}$ is assumed to be symmetric. Then the answer in **b)** and (4) give that

$$= f_j q_{j|i}\alpha^{\mathrm{B}}_{j,i} = f_j p_{j|i}.$$

In other words we have shown that the Markov chain $\{X_n\}_{n\geq 0}$ is reversible for $\mathbf{f}$, and therefore $\mathbf{f}$ is a stationary distribution for the chain.

Uppgift 6

a) As we have in the model $\mathcal{M}_2$ independent $Y_i \in N(\alpha + \beta x_i, \sigma)$ for $i = 1, 2, \ldots, n$, we get

$$f_{Y_1,Y_2,\ldots,Y_n}\left(y_1, y_2, \ldots, y_n; \theta^{(2)}\right) = \frac{1}{\sqrt{2\pi}^n \sigma^n} e^{-\frac{\sum_{i=1}^n (y_i - (\alpha - \beta x_i))^2}{2\sigma^2}}.$$

Then

$$-\ln f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n; \widehat{\theta}^{(2)}) = \min_{\alpha,\beta}\left(\frac{n}{2}\ln 2\pi + n\ln\sigma + \frac{\sum_{i=1}^n (y_i - (\alpha - \beta x_i))^2}{2\sigma^2}\right) =$$

$$= \frac{n}{2}\ln 2\pi + n\ln\sigma + \min_{\alpha,\beta}\frac{1}{2\sigma^2}\sum_{i=1}^n \left(y_i - (\alpha + \beta x_i)\right)^2.$$

which verifies (6).

b) We have from **a)**

$$Q_{\mathrm{AIC}}(\mathcal{M}_2) = -\frac{1}{n}\ln f_{Y_1,Y_2,\ldots,Y_n}\left(y_1, y_2, \ldots, y_n; \widehat{\theta}^{(2)}\right) + \frac{2}{n} =$$

$$\frac{1}{2}\ln 2\pi + \ln\sigma + \min_{\alpha,\beta}\frac{1}{2n\sigma^2}\sum_{i=1}^n \left(y_i - (\alpha + \beta x_i)\right)^2 + \frac{2}{n}.$$

A formula on p. 6 of the *Formel- och tabellsamling* gives directly that

$$\min_{\alpha,\beta}\sum_{i=1}^{n}\left(y_i-(\alpha+\beta x_i)\right)^2=\left(S_{yy}-\widehat{\beta}S_{xy}\right).$$

Thus we get $Q_{\mathrm{AIC}}\left(\mathcal{M}_2\right)$ as asserted.

c) As we have in the model $\mathcal{M}_1$ independent $Y_i\in N(\alpha,\sigma)$ for $i=1,2,\ldots,n$, we get

$$f_{Y_1,Y_2,\ldots,Y_n}\left(y_1,y_2,\ldots,y_n;\theta^{(1)}\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\dfrac{(y_1-\alpha)^2}{2\sigma^2}}\cdots\frac{1}{\sqrt{2\pi}\sigma}e^{-\dfrac{(y_n-\alpha)^2}{2\sigma^2}}$$

$$=\frac{1}{\sqrt{2\pi}^n\sigma^n}e^{-\dfrac{\sum_{i=1}^{n}(y_i-\alpha)^2}{2\sigma^2}}.$$

Then

$$-\ln f_{Y_1,Y_2,\ldots,Y_n}(y_1,y_2,\ldots,y_n;\widehat{\theta}^{(1)})=\frac{n}{2}\ln 2\pi+n\ln\sigma+\min_{\alpha}\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha)^2$$

and as the maximum likelihood estimate is $\widehat{\alpha}=\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$,

$$=\frac{n}{2}\ln 2\pi+n\ln\sigma+\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2,$$

and thus

$$Q_{\mathrm{AIC}}\left(\mathcal{M}_1\right)=\frac{1}{2}\ln 2\pi+\ln\sigma+\frac{1}{2\sigma^2 n}S_{yy}+\frac{1}{n},$$

by definition of S_{yy} in (9).

d) The AIC criterion is minimized by the model $\mathcal{M}_2$, as soon as

$$Q_{\mathrm{AIC}}\left(\mathcal{M}_1\right)>Q_{\mathrm{AIC}}\left(\mathcal{M}_2\right)\Leftrightarrow$$

$$Q_{\mathrm{AIC}}\left(\mathcal{M}_1\right)-Q_{\mathrm{AIC}}\left(\mathcal{M}_2\right)>0,$$

and from the above

$$Q_{\mathrm{AIC}}\left(\mathcal{M}_1\right)-Q_{\mathrm{AIC}}\left(\mathcal{M}_2\right)=\frac{1}{2\sigma^2 n}\widehat{\beta}S_{xy}-\frac{1}{n}.$$

Thus, when we apply (11),

$$Q_{\mathrm{AIC}}\left(\mathcal{M}_1\right)>Q_{\mathrm{AIC}}\left(\mathcal{M}_2\right)\Leftrightarrow\frac{1}{2\sigma^2 n}\widehat{\beta}S_{xy}>\frac{1}{n}\Leftrightarrow\frac{\widehat{\beta}S_{xy}}{\sigma^2}>2\Leftrightarrow\frac{\frac{S_{xy}}{S_{xx}}S_{xy}}{\sigma^2}>2$$

Now we use again (11) or $\widehat{\beta}=\frac{S_{xy}}{S_{xx}}$ to get

$$\Leftrightarrow\frac{\frac{S_{xy}^2}{S_{xx}^2}S_{xx}}{\sigma^2}>2\Leftrightarrow\frac{\widehat{\beta}^2S_{xx}}{\sigma^2}>2\Leftrightarrow\frac{\widehat{\beta}^2}{\frac{\sigma^2}{S_{xx}}}>2\Leftrightarrow\frac{\mid\widehat{\beta}\mid}{\frac{\sigma}{\sqrt{S_{xx}}}}>\sqrt{2},$$

as was to be shown.

e) By the result in **d)** the probability

$$P\left(\mathcal{M}_2 \text{ is selected by AIC } \mid \mathcal{M}_1 \text{ is true}\right).$$

is the same as the probability

$$P\left(\frac{\mid \widehat{\beta} \mid}{\frac{\sigma}{\sqrt{S_{xx}}}} > \sqrt{2} \mid \mathcal{M}_1 \text{ is true}\right).$$

Here we consider $\widehat{\beta}$ as the random variable

$$\widehat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(Y_i - \bar{Y}\right)}{S_{xx}}.$$

The formula a) in section 13.1 on p. 5 of *Formel- och tabellsamling* gives that

$$\widehat{\beta} \in N\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right).$$

Then, if $\mathcal{M}_1$ is true, and since the models are nested,

$$\widehat{\beta} \in N\left(0, \frac{\sigma}{\sqrt{S_{xx}}}\right)$$

and

$$\frac{\widehat{\beta}}{\frac{\sigma}{\sqrt{S_{xx}}}} \in N\left(0, 1\right).$$

Let now for simplicity of writing set

$$Z \equiv \frac{\widehat{\beta}}{\frac{\sigma}{\sqrt{S_{xx}}}}.$$

Then

$$P\left(\mid Z \mid > \sqrt{2}\right) = P\left(Z > \sqrt{2}\right) + P\left(Z < -\sqrt{2}\right)$$

$$= 1 - P\left(Z \leq \sqrt{2}\right) + P\left(Z < -\sqrt{2}\right) = 2\left(1 - \Phi\left(\sqrt{2}\right)\right),$$

where $\Phi\left(\sqrt{2}\right)$ is the cumulative distribution function of the standard normal distribution $N(0,1)$ evaluated at $\sqrt{2}$. Therefore

$$P\left(\mathcal{M}_2 \text{ is selected by AIC } \mid \mathcal{M}_1 \text{ is true}\right) = 2\left(1 - \Phi\left(\sqrt{2}\right)\right) \approx 0.1573.$$

Hence, the AIC model choice is not asymptotically consistent, since the probability of erroneous model choice does not turn to zero, as $n \to \infty$.

ANSWER **e)**: $\underline{P\left(\mathcal{M}_2 \text{ is selected by AIC } \mid \mathcal{M}_1 \text{ is true}\right) = 0.16.}$