Avd. Matematisk statistik

KTH Matematik

EXAM IN SF2955 COMPUTER INTENSIVE METHODS
WEDNESDAY 19th of August 2009 14.00 – 19.00.

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 September 1st. You will be informed by email if you specify your email address.

Good luck!

- -

Uppgift 1

Space Shuttle Challenger (NASA=(National Aeroscpace Administration in USA) Orbiter Vehicle Designation: OV-099) was NASA's second Space Shuttle orbiter with maiden flight on April 4, 1983, and it completed nine missions before distintegrating 73 seconds after the launch of its tenth mission, STS-51-L on January 28, 1986.
The disaster commission concluded that the accident was caused by a combustion gas leak through a joint in one of the solid rocket boosters (SRB), which was sealed by a rubber device called *O-ring*. The rings never functioned according to design.
Weather forecasts for January 28 had predicted an unusually cold morning (for Kennedy Space Center in Florida), with temperatures close to 31 ^oF ($-$ 1 ^oC). Prompted by this, several engineers of the contractor responsible for the construction and maintenance of the SRBs pointed out that if the O-rings were colder than 53 ^oF, they did not have enough data to determine whether the rings would seal properly. Following decisions at the managerial level[1] the launch was made at 31 ^oF.
A statistical analysis, done independently of the disaster commission, aimed at predicting the performance of the O-rings as a function of the ambient (=omgivande) temperature was based on the following model. The pertinent random variable is

$$X = \text{ number of incidents of thermally distressed O-rings.} \tag{1}$$

(There are two kinds of distress: erosion and blowby. Erosion is caused by excessive heat burning up the O-ring, blowby can occur when gases rush through the O-ring.) There are six primary O-rings per vehicle, and it is assumed that

$$X \sim \mathrm{Bin}(6, p(t))$$

[1]This has been described as 'unethical decision-making resulting from intense customer (=NASA) intimidation'.

where t is the ambient temperature in Fahrenheit degrees, ${}^oF$. The data consists of observations of temperature t_i and the number of incidents x_i

$$(t_i, x_i), i = 1, 2, \ldots, 23$$

(a slightly simplified set of data is displayed in Uppgift 2 below) during 23 shuttle missions prior to the disastrous launch.

a) Explain how to construct a non-parametric 90 % bootstrap confidence interval for the expected number of thermally distressed O-rings for a fixed temperature in $[{}^oF]$. Here you are asked to write down a guideline of the computation. You do not in principle need to do any calculations. (4 p)

b) We can use a logistic regression model for $p(t)$, i.e.,

$$\ln\left[\frac{p(t)}{1-p(t)}\right] = \alpha + \beta t. \tag{2}$$

Explain how to construct a parametric 90 % bootstrap confidence interval for the expected number of thermally distressed O-rings as a function of the temperature $[{}^oF]$ using this model. Here you are asked to write down a guideline of the computation. *Hint:* Find the expression for $p(t)$. (4 p)

c) In the Figure (see next page) we see a 90 % bootstrap confidence interval for the expected number of thermally distressed O-rings based on the model (2). The asterisks in the middle denote the estimated expected numbers. What is Your analysis of launches at temperatures lower than 53 ${}^oF$?
When β in (2) was estimated from data, it turned out to have a negative sign. What does this imply for the expected number of incidents with thermally distressed O-rings ? (2 p)

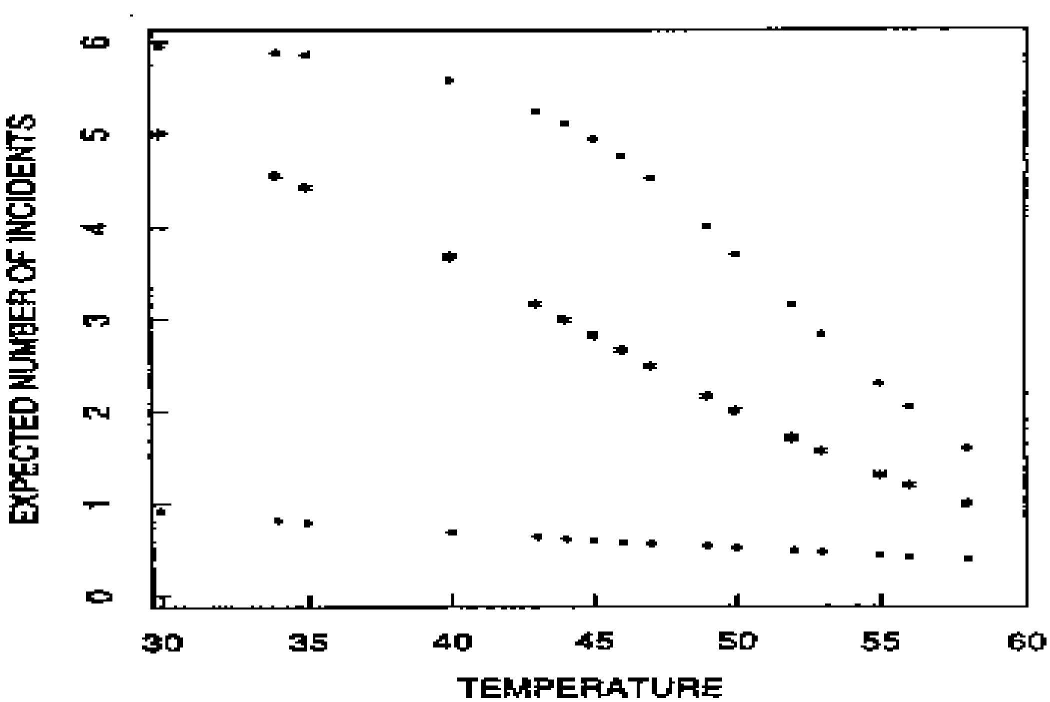

Figure 5. 90% Bootstrap Confidence Limits, Field-Joint Distress: Primary O-Rings and Binomial-Logit Model: N = 6.

Uppgift 2

We continue with Space Shuttle incident data. This assignment is, however, independent of the preceding, and can be solved even if You have not solved Uppgift 1 or any part of it. We consider the random variable Y defined by using X in (1) as

$$Y = \begin{cases} 1 & \text{if } X > 0 \\ 0 & \text{if } X = 0. \end{cases}$$

We want to estimate the parameter $\theta = P(Y \geq 1)$ and have the 23 (independent) observations of X. We disregard here the temperature data.

allenger Field Joint
/ O-Ring Failure Data

Flight Date	X Number of Field Joint Primary O-Ring Incidents	t Temperature at Launch (°F)
4/12/81	0	66
11/12/81	1	70
3/22/82	0	69
11/11/82	0	68
4/4/83	0	67
6/18/83	0	72
8/30/83	0	73
11/28/83	0	70
2/3/84	1	57
4/6/84	1	63
8/30/84	1	70
10/5/84	0	78
11/8/84	0	67
1/24/85	2	53
4/12/85	0	67
4/29/85	0	75
6/17/85	0	70
7/29/85	0	81
8/27/85	0	76
10/3/85	0	79
10/30/85	2	75
11/26/85	0	76
1/12/86	1	58

a) Give the plug-in estimate of θ numerically. Give the distribution of this estimate (in terms of θ). (5 p)

b) This estimate is bootstrapped. Give the distribution of this bootstrap estimate. (5 p)

Uppgift 3

Let X_i be I.I.D. and $E[X_i] = \mu$ and $V(X_i) = \sigma^2$. We define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

and for $i = 1, 2, \ldots, n$

$$\bar{X}_{(i)} = \frac{1}{n-1}\sum_{l \neq i}^{n} X_l. \tag{3}$$

a) Check that the following three equalities hold:

$$\bar{X}_{(i)} = \frac{n\bar{X} - X_i}{n-1}, \tag{4}$$

$$\bar{X}_{(i)} - \bar{X} = \frac{\bar{X} - X_i}{n-1}, \tag{5}$$

and

$$\bar{X}_{(i)} + \bar{X} = \frac{2(n-1)\bar{X} + \bar{X} - X_i}{n-1}. \tag{6}$$

(2 p)

b) As the estimator of the parameter $\theta = \mu^2$ we use $\widehat{\theta} = \bar{X}^2$, and then define using (3)

$$\widehat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} \bar{X}_{(i)}^2.$$

The jackknife bias estimator is defined by

$$\widehat{bias}_{jack} = (n-1)\left(\widehat{\theta}_{(\cdot)} - \widehat{\theta}\right).$$

Show that

$$\widehat{bias}_{jack} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2. \tag{7}$$

Hint: You may find it useful at some stage to apply suitably the algebraic identity $a^2 - b^2 = (a+b)(a-b)$ and then to take advantage of (5) -(6). (4 p)

c) Show that the jackknife estimator $\bar{\theta}$(=jackknife bias corrected $\widehat{\theta}$) is unbiased. *Hint:* You may find it useful to note the following identity in section 13.3 of *Formel- och tabellsamling*

$$\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2,$$

(You need not prove this) or, in particular, that this gives

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

You are also allowed to use without proof the following well known result in mathematical statistics:

$$E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right]=\sigma^2.$$

You may solve this part of Uppgift, even if you have not solved part **b)**. (4 p)

Uppgift 4

Let the target distribution $\mathbf{f}=(f_i)_{i=-\infty}^{\infty}$ on the integers be

$$f_i = C\cdot e^{-i^4}, i=\ldots,-2,-1,0,1,2,\ldots,.$$

Give the Metropolis-Hastings algorithm for this target distribution, when the proposal distribution is given by a simple random walk , i.e, the matrix $\mathbf{Q}$ is given by

$$q_{i|j}=\begin{cases} p & \text{if } j=i+1 \\ 1-p & \text{if } j=i-1 \\ 0 & \text{otherwise.}\end{cases}$$

How does the value of p influence the algorithm ? (10 p)

Uppgift 5

Consider the m-file `pivot.m`:

function [val]=pivot(x,a);

n=length(x);

val=(mean(x)-a)/(std(x)/sqrt(n));

You obtain the following (slightly edited) print from a session with Matlab^R

n=length(x)

n=10

a=mean(x)

a =0.3363

std(x)

ans =0.8779

boot=bootstrp(1000,'pivot',x,a);

low=prctile(boot,2.5)

low =-2.7395

upper=prctile(boot,97.5)

upper =2.1042

Compute using these results a 95% confidence interval for θ, where the interval is based on the pivotal statistic

$$\frac{\bar{X}-\theta}{s/\sqrt{n}}.$$

(10 p)

Uppgift 6

Two models $\mathcal{M}_p$, $p = 1, 2$, for $Y_1, Y_2, \ldots, Y_n$ 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 **BIC (= Bayesian Information Criterion) principle of data driven model choice** selects the model that minimizes the BIC criterion

$$Q_{\mathrm{BIC}}\left(\mathcal{M}_p\right) = -\frac{1}{n} \ln f_{Y_1,Y_2,\ldots,Y_n}\left(y_1, y_2, \ldots, y_n; \widehat{\theta}^{(p)}\right) + \frac{p \ln(n)}{2n}, \tag{8}$$

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

$$-\ln f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n; \widehat{\theta}^{(2)}) = \frac{n}{2} \ln 2\pi + n \ln \sigma + \min_{\alpha,\beta} \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - (\alpha + \beta x_i)\right)^2. \tag{9}$$

(1 p)

Commment: By (9), 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**)

$$\widehat{\alpha} = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i. \tag{10}$$

b) Next show that

$$Q_{\text{BIC}}\left(\mathcal{M}_2\right) = \frac{1}{2}\ln 2\pi + \ln\sigma + \frac{1}{2n\sigma^2}\left(S_{yy} - \widehat{\beta}S_{xy}\right) + \frac{\ln(n)}{n}, \tag{11}$$

and that

$$Q_{\text{BIC}}\left(\mathcal{M}_1\right) = \frac{1}{2}\ln 2\pi + \ln\sigma + \frac{1}{2\sigma^2 n}S_{yy} + \frac{\ln(n)}{2n}. \tag{12}$$

In these formulas, following section 13.3 on p. 6 of the *Formel- och tabellsamling*, we have

$$S_{yy} = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2, \tag{13}$$

and

$$S_{xy} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right), \tag{14}$$

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

$$\widehat{\beta} = \frac{S_{xy}}{S_{xx}}, \tag{15}$$

where

$$S_{xx} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2. \tag{16}$$

Hint: Note that β^*_{obs} in the *Formel- och tabellsamling* is $= \widehat{\beta}$ in our notation. You are allowed to use (9) from **a)** even if you have not done **a)**. (2 p)

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

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

What is a statistician′s interpretation of the method of BIC model choice in view of this result ? You are allowed to use **b)** even if you have not done it. (3 p)

d) Show that the BIC is asymptotically consistent. You may solve this even if you have not done **c)**. (4 p)

Hint: Study the probability

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

You need to regard $\widehat{\beta}$ in (15) as a random variable, i.e., you must take

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

In addition, 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 19th of MAY 2009 02.00 p.m. – 07.00 p.m..

Uppgift 1

a) As X = number of incidents of thermally distressed O-rings $\sim \text{Bin}(6, p(t))$, we have that

$$m = E[X] = 6 \cdot p(t).$$

For fixed t_i the plug-in (or maximum likelihood) estimate of $p(t)$ is

$$\widehat{p} = \widehat{p}(t_i) = \frac{x_i}{6}$$

Then we know that the bootstrap distribution of $6\widehat{p}$ is

$$6\widehat{p} \sim \text{Bin}\left(6, \frac{x_i}{6}\right).$$

Then we can use the *Percentile method*, and obtain the bootstrap condidence interval for m as

$$\left[\widehat{m}^{*(0.05}, \widehat{m}^{*(0.95}\right]$$

where $G_{Bin}\left(\widehat{m}^{*(0.05}\right) = 0.05$ and $G_{bin}\left(\widehat{m}^{*(0.95}\right) = 0.95$ and $G_{Bin}(z)$ is the distribution function of $\text{Bin}_{0.05}\left(6, \frac{x_i}{6}\right)$. This can be repeated for all t_i.

b) We can estimate α and β by least squares with (2). Let us suppose that $\widehat{\alpha}$ and $\widehat{\beta}$ are the least squares estimates. From (2)

$$\frac{p(t)}{1-p(t)} = e^{\alpha+\beta t} \Rightarrow p(t) = \frac{1}{1+e^{-(\alpha+\beta t)}}.$$

and thereby we estimate

$$\widehat{p}(t) = \frac{1}{1+e^{-(\widehat{\alpha}+\widehat{\beta} t)}}.$$

Here one may in fact continue as in **a)** using the bootstrap distribution

$$6\widehat{p}(t) \sim \text{Bin}\left(6, \widehat{p}(t)\right).$$

With a parametric bootstrap more in line with the standard notion we draw a sample of six outcomes from the Bernoulli distribution $\text{Ber}(\widehat{p}(t))$ for $i = 1, 2, \ldots, B$, compute the estimates $\widehat{p}_i^{*(0.05}(t)$, find the percentiles of these samples, which gives us the confidence interval for $E[X]$ as

$$\left[6\widehat{p}_B^{*(0.05}(t), 6\widehat{p}_B^{*(0.95}(t)\right].$$

Comment: Note that when we use (2) we can compute a bootstrap confidence intervals for every t, not only for those t_i, where we have data. This is an important difference to the method in a) .

c) It is clear from the Figure that the confidence intervals become wider with decreasing t, indicating that our data gives less and less precise idea about the expected number of incidents under colder temperatures, statistically confirming the engineering insight that there was not enough data to determine whether the rings seal properly.
When $\widehat{\beta} < 0$ we set $a = -\widehat{\beta}$, and then put

$$\widehat{p}(t) = \frac{1}{1 + e^{-\widehat{\alpha}} e^{at}}.$$

Since now $a > 0$ we have for $t_1 < t_2$ that $1 + e^{-\widehat{\alpha}} e^{at_1} < 1 + e^{-\widehat{\alpha}} e^{at_2}$. This means that

$$\frac{1}{1 + e^{-\widehat{\alpha}} e^{at_2}} < \frac{1}{1 + e^{-\widehat{\alpha}} e^{at_1}},$$

i.e.,

$$\widehat{p}(t_2) < \widehat{p}(t_1).$$

In words, we predict that the expected number of incidents with O-rings will increase for lower temperatures.

Uppgift 2

a) The plug-in estimate of $\theta = P(Y \geq 1)$ is $\widehat{\theta} =$ the relative frequency of the missions with at least one incident for joint O-rings. Hence we get from the given data that $\widehat{\theta} = 7/23 = 0.304$. The corresponding statistic $23\widehat{\theta}(X_1, X_2, \cdots, X_{23})$ is distributed as $\text{Bin}(23, \theta)$.

b) For a non-parametric bootstrap one samples the 23 observations with replacement. This is the defining situation for Binomial distribution and gives that $23\widehat{\theta}^*$ has the distribtion $\text{Bin}(23, 0.3)$.

Uppgift 3

a) By definition we get

$$\bar{X}_{(i)} = \frac{1}{n-1} \sum_{l \neq i}^{n} X_l = \frac{1}{n-1} \left(\sum_{l=1}^{n} X_l - X_i \right)$$

$$= \frac{1}{n-1} \left(n\bar{X} - X_i \right).$$

which shows the first of the identities (4), which then gives

$$\bar{X}_{(i)} - \bar{X} = \frac{1}{n-1} \left(n\bar{X} - X_i \right) - \bar{X}$$

$$= \frac{1}{n-1} \left(n\bar{X} - X_i - (n-1)\bar{X} \right) = \frac{1}{n-1} \left(\bar{X} - X_i \right),$$

as was to be proved. Finally, invoking (4) again,

$$\bar{X}_{(i)} + \bar{X} = \frac{1}{n-1} \left(n\bar{X} - X_i \right) + \bar{X} =$$

$$= \frac{1}{n-1}\left(n\bar{X} - X_i + (n-1)\overline{X}\right)$$
$$= \frac{2(n-1)\bar{X} + \bar{X} - X_i}{n-1}.$$

Of course, this way of writing is artificial, but we happen to need the expression in this very form in the sequel.

b) With

$$\widehat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} \bar{X}^2_{(i)}.$$

we get

$$\widehat{bias}_{jack} = (n-1)\left(\widehat{\theta}_{(\cdot)} - \widehat{\theta}\right)$$
$$= (n-1)\left(\frac{1}{n}\sum_{i=1}^{n} \bar{X}^2_{(i)} - \overline{X}^2\right)$$
$$= \frac{(n-1)}{n}\sum_{i=1}^{n}\left(\bar{X}^2_{(i)} - \overline{X}^2\right)$$
$$= \frac{(n-1)}{n}\sum_{i=1}^{n}\left(\bar{X}_{(i)} - \overline{X}\right)\left(\bar{X}_{(i)} + \overline{X}\right),$$

where we used $a^2 - b^2 = (a+b)(a-b)$. Then we get from (5) -(6) that this equals

$$= \frac{(n-1)}{n}\frac{1}{n-1}\frac{1}{n-1}\sum_{i=1}^{n}\left(\overline{X} - \bar{X}_i\right)\left(2(n-1)\bar{X} + \bar{X} - X_i\right)$$

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

But it is clear that

$$\sum_{i=1}^{n}\left(\bar{X} - X_i\right) = n\bar{X} - \sum_{i=1}^{n} X_i = 0.$$

Hence

$$\widehat{bias}_{jack} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2,$$

as was to be shown.

c) The jackknife estimator $\bar{\theta}$ is

$$\bar{\theta} = \widehat{\theta} - \widehat{bias}_{jack}.$$

We are asked to show that

$$E\left[\bar{\theta}\right] = \mu^2.$$

We have

$$E\left[\bar{\theta}\right] = E\left[\widehat{\theta}\right] - E\left[\widehat{bias}_{jack}\right]. \tag{17}$$

If we use the proposed hint, we get

$$E\left[\widehat{\theta}\right] = E\left[\bar{X}^2\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i^2\right] - E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right].$$

We set

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

Then we get, recalling $V(X) = E(X^2) - (E(X))^2$, and since X_i be I.I.D. and $E\left[X_i\right] = \mu$ and $V(X_i) = \sigma^2$ that

$$E\left[\bar{X}^2\right] = \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2 + \mu^2\right) - \left(\frac{n-1}{n}\right)E\left[s^2\right] = \sigma^2 + \mu^2 - \left(\frac{n-1}{n}\right)\sigma^2, \tag{18}$$

where we used the second hint proposed, or a theorem in mathematical statistics.
Next we get by **b)** that

$$E\left[\widehat{bias}_{jack}\right] = E\left[\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \frac{1}{n}E\left[s^2\right] = \frac{1}{n}\sigma^2. \tag{19}$$

When we insert (18) and (19) in the right hand side of (17) we obtain

$$E\left[\bar{\theta}\right] = \sigma^2 + \mu^2 - \left(\frac{n-1}{n}\right)\sigma^2 - \frac{1}{n}\sigma^2 = \mu^2. \tag{20}$$

But this shows by definition that $\bar{\theta}$ is unbiased, as was asserted.

Uppgift 4

The Metropolis-Hastings algorithm generates a Markov chain $(X_n)_{n\geq 0}$ with the integers as state space such that $\mathbf{f} = (f_i)_{i=-\infty}^{\infty}$ is its stationary distribution.
Assume that $X_n = i$. Then the algorithm proposes a new value $Y_{n+1} \sim \{q_{i|j}\}_{j=0}^{J}$, i.e.,

$$P\left(Y_{n+1} = i+1 | X_n = i\right) = p, P\left(Y_{n+1} = i-1 | X_n = i\right) = 1-p$$

$$P\left(Y_{n+1} = k | X_n = i\right) = 0, k \neq i+1, i-1.$$

We have precomputed $\alpha_{i,j}$ for $j = i+1$

$$\alpha_{i,i+1} = \min\left\{1, \frac{e^{-(i+1)^4}(1-p)}{e^{-(i)^4}p}\right\}.$$

and $j = i-1$

$$\alpha_{i,i-1} = \min\left\{1, \frac{e^{-(i-1)^4}p}{e^{-(i)^4}(1-p)}\right\}.$$

Then we take

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

and continue by setting $n+1 \rightarrow n$.
The transition probability $p_{i|j}$ of the M.C. is

$$p_{i|j} = q_{i|j}\alpha_{i,i-1} = \begin{cases} (1-p)\min\left\{1, \frac{e^{-(i-1)^4}p}{e^{-(i)^4}(1-p)}\right\} & j = i-1 \\ p\min\left\{1, \frac{e^{-(i+1)^4}(1-p)}{e^{-(i)^4}p}\right\} & j = i+1 \\ 1 - p_{i|i-1} - p_{i|i+1} & i = j \\ 0 & \text{otherwise.} \end{cases}$$

If $p \approx 1$, then $\frac{p}{(1-p)}$ is large, thus $\alpha_{i,i-1} \approx 1$, and hence $p_{i|i-1} \approx 0$, and, since $(1-p)/p \approx$, we get $\alpha_{i,i+1} \approx 0$. If $p \approx 0$, then $\alpha_{i,i-1} \approx 0$, and since then $\frac{1-p}{p}$ is large $\alpha_{i,i+1} \approx p \times 1 \approx 0$. Thus for large and small values of p the Markov chain tends to refuse the proposals, and is a slow process.

Uppgift 5

To bootstrap statistic corresponding to $T = (\bar{x} - \theta)/(s/\sqrt{n})$ is $T^* = (\bar{x}^* - \bar{x})/(s^*/\sqrt{n})$. If we had access to the percentiles z_α of $G(z) = P(T \le z)$, the confidence interval for θ would be

$$(\widehat{\theta} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}, \widehat{\theta} - z_{\alpha/2}\frac{s}{\sqrt{n}}).$$

Now, we estimate the percentiles for the bootstrap statistic T^* by simulation of its distribution. With data from the Matlab print we see that the confidence interval becomes

$$(\bar{x} - z^*_{0.975}\frac{s}{\sqrt{n}}, \bar{x} - z^*_{0.025}\frac{s}{\sqrt{n}}) =$$

$$(0.3363 - 2.1042\frac{0.8779}{\sqrt{10}}, 0.3363 - (-2.7393)\frac{0.8779}{\sqrt{10}}) = \underline{(-0.2479, 1.0968).}$$

Uppgift 6

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

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

Then

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

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

which verifies (9).
b) We have from **a)**

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

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

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 $Q_{\text{BIC}}\left(\mathcal{M}_2\right)$ as asserted. 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

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

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

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

by definition of S_{yy} in (13).

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

$$Q_{\text{BIC}}\left(\mathcal{M}_1\right) > Q_{\text{BIC}}\left(\mathcal{M}_2\right) \Leftrightarrow$$

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

and from the above

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

Thus, when we apply (15),

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

Now we use again (15) or $\widehat{\beta}=\frac{S_{xy}}{S_{xx}}$ to get

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

as was to be shown. For a statistician this means that the BIC criterion has in this case the form of a statistical test.

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

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

is the same as the probability

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

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

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

Let now for simplicity of writing set

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

Then

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

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

$$P\left(\mathcal{M}_2 \text{ is selected by BIC } \mid \mathcal{M}_1 \text{ is true}\right) = 2\left(1 - \Phi\left(\sqrt{\ln(n)}\right)\right) \to 0.$$

Hence, the BIC model choice is asymptotically consistent, since the probability of erroneous model choice converges to zero, as $n \to \infty$.

ANSWER **d)**: $\underline{P\left(\mathcal{M}_2 \text{ is selected by BIC } \mid \mathcal{M}_1 \text{ is true}\right) \to 0, \text{as} \quad n \to \infty}$