

UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in: STK4051/STK9051 - Computational Statistics —

Day of examination: 02.06.25

Examination hours: 15.00–19.00

This problem set consists of 5 pages.

Appendices: None.

Permitted aids: Approved calculator.

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1 EM Algorithm

Consider the following data, reported in Lange (2003, *The EM Algorithm*),

c	0	1	2	3	4	5	6
y_c	3062	587	284	103	33	4	2

The table summarises the number of widows (y_c) and related number of dependent children (c) from a Swedish pension fund. For this exercise, it is important to note that a simple Poisson model gives poor fit, clearly because most of the widows have no dependent children. We can obtain better model if we suppose that the widows come from two populations: one group (group A) consists of widows who did not want to have children, and the other group (group B) consists of those who wanted to have children (but might have not managed). For the widows in group B, we assume a Poisson distribution with mean μ for the number of children. Let p denote the probability that a widow belongs to group A (thus the probability that she belongs to group B is $1 - p$).

a

Show that the log-likelihood for the observed data is

$$\ell(p, \mu) = \text{const} + y_0 \log[p + (1 - p)e^{-\mu}] + \sum_{c=1}^6 y_c [\log(1 - p) + c \log \mu - \mu - \log c!]$$

[note: if you need, the last page contains the densities of some distributions].

b

Derive the log-likelihood for the complete data, denoting with x_A the number of widows without children from group A, and with x_B the number of widows without children from group B.

(Continued on page 2.)

c

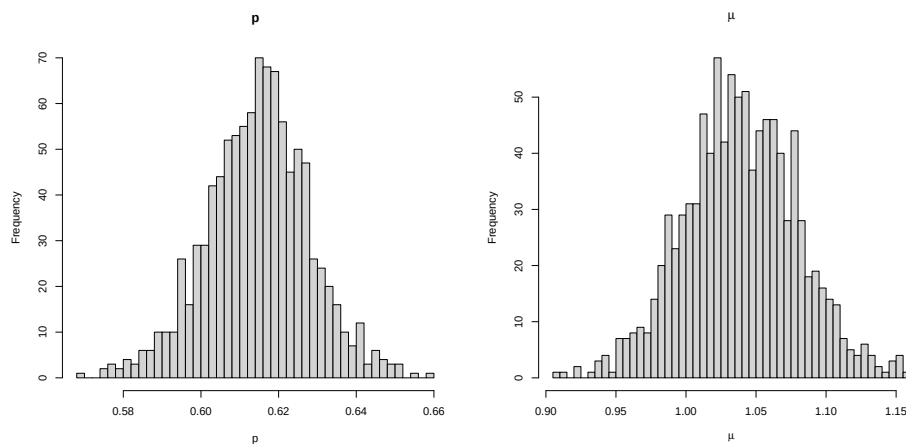
Describe briefly what the $Q(p, \mu | p^{(k-1)}, \mu^{(k-1)})$ function is and derive it in this case.

d

Derive the updating equations for the M-step and report the numerical results for the first three iterations, using the point $(0.75, 0.40)$ as the initial guess.

e

In order to give an estimate of the uncertainty, the algorithm has been applied to 1000 bootstrap samples (re-sampled with replacement from the original sample). The results are summarised in the following figures,



Based on these figures:

- Write down a bootstrap confidence interval for both p and μ (it is not important to provide very precise numbers, but rather that you explain clearly how you found them).
- How could you have obtained a more precise confidence interval?
- Also provide an example of bootstrap sample that could have likely been generated by the bootstrap procedure, in the form of a table like the one at the beginning of this exercise.

Problem 2 Bootstrap

a

Consider a simple regression model, $Y_i = x_i^T \beta + \epsilon_i$, $i = 1, \dots, n$, where:

- ϵ_i are i.i.d. with 0 mean and constant variance;
- x_i is a p -dimensional vector of predictors (covariates);
- β a p -dimensional vector of parameters.

(Continued on page 3.)

In order to make inference on β , we proceed as follows:

- we resample from the collection of response values a new pseudo-response, say Y_i , for each observed x_i , generating a bootstrap dataset;
- we estimate $\hat{\beta}^*$ from this bootstrap dataset;
- we repeat this procedure B times and use the empirical distribution of $\hat{\beta}^*$ as an approximation of the distribution of β .

Explain why the procedure above is wrong, and suggest two alternative approaches to approximate the distribution of β . Describe briefly the two alternatives.

b

Consider the estimation of θ for the $\text{Unif}(0, \theta)$ distribution. Explain the reason for which the bootstrap fails when one uses the maximum likelihood estimator, and provide an alternative for which we can successfully apply the bootstrap procedure.

c

Another situation in which a naive application of bootstrapping does not work is the presence of dependent data:

- Explain why, in this case, applying a simple bootstrap procedure does not work.
- To solve this issue, many approaches have been proposed. Describe the moving block bootstrapping approach, including the advantage(s) with respect to the non-moving block approach and the issues related to the edge effect.
- Comment on the choice of the block length l , explaining in particular what happens if l is too small for the problem at hand, and what occurs when it is too large.

Problem 3 MCMC

To generate data from the following distribution,

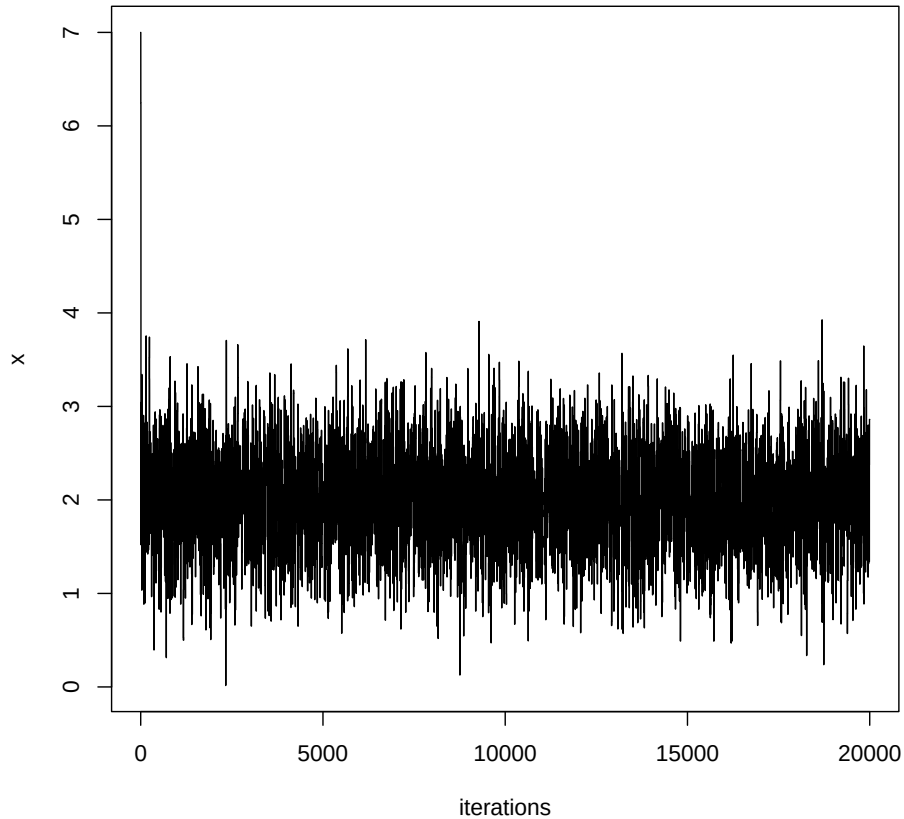
$$0.5\mathcal{N}(2, 0.5) + 0.5\mathcal{N}(16, 0.5)$$

we performed 20,000 iterations of a Markov Chain using a Metropolis-Hastings algorithm.

(Continued on page 4.)

a

When plotting the sample path, we obtained the following figure,



- Describe how this plot should appear in the case of a chain with good mixing properties and relate your description to this plot.
- After removing the burn-in, can we use the realization of the chain as a representative sample for the distribution above? Why or why not?
- The acceptance rate in this case is 0.29. Is it a good value? Explain why this value should neither be too small nor too large.

b

Finally, consider the Gelman & Rubin statistic,

$$R = \frac{\frac{L-1}{L}W + \frac{1}{L}B}{W},$$

where B is the between-chain variance, W the within-chain variance, and L the chains' length after the burn-in.

- Describe briefly the idea behind the R statistic and what it is used for.
- Provide a plausible value for the experiment above (it is not important the value itself, but the reasoning behind your choice, that you should explain well).

(Continued on page 5.)

SOME DISTRIBUTIONS AND THEIR DENSITIES*(not all of them are relevant for this exam)*

- $X \sim \text{Bin}(n, p) \rightarrow f(x) = \binom{n}{x} p^x (1-p)^{n-x};$
- $X \sim \text{Poisson}(\lambda) \rightarrow f(x) = \frac{\lambda^x}{x!} e^{-\lambda};$
- $X \sim \mathcal{N}(\mu, \sigma^2) \rightarrow f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$
- $X \sim \text{Exp}(\lambda) \rightarrow f(x) = \lambda e^{-\lambda x};$
- $X \sim \text{Gamma}(\alpha, \beta) \rightarrow f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x};$
- $X \sim \text{Unif}(a, b) \rightarrow f(x) = \frac{1}{b-a};$
- $X \sim \text{Beta}(\alpha, \beta) \rightarrow f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \text{ with } B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)};$
- $X \sim \text{Multinom}(n, p_1, \dots, p_k) \rightarrow f(x) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \times \dots \times p_k^{x_k}.$

THE END