

STK4130 / STK9130 Fall 2012 – Exam project

Deadline Monday December 3rd, before midnight (24:00)

You are not allowed to collaborate with other students on the project exam.

The exam project consists of 4 Problems over 3 pages. Make sure you have the complete exam set.

The answer to the exam project can be handed in on email to osamuels@math.uio.no or on paper to my mailbox on the 7th floor in NHA. You may deliver a handwritten or Latex/Word-processed answer to the project in English, Norwegian or any other Scandinavian language

Problem 1

In this problem we assume that $X_i, i = 1, \dots, n$ are iid and exponentially distributed, i.e. with density $f(x; \lambda) = \lambda \exp(-\lambda x)$ for $x > 0$. Then $\mu = E[X_i] = 1/\lambda$ and $\text{Var}[X_i] = 1/\lambda^2 = \mu^2$.

- Let $\hat{\lambda} = 1/\bar{X}$. Argue that $\hat{\lambda} \rightarrow_p \lambda$.
- Find the limiting distribution of $\sqrt{n}(\bar{X} - \mu)$ and from this derive the limiting distribution of $\sqrt{n}(\hat{\lambda} - \lambda)$ (without using that $\hat{\lambda}$ is the MLE of λ).
- Show that $\hat{\lambda}$ is MLE and use this to find the limit distribution of $\sqrt{n}(\hat{\lambda} - \lambda)$.
- Find the influence function of the functional $\lambda = \lambda(F) = 1/E[X_i]$ and based on this suggest the limiting distribution of $\sqrt{n}(\hat{\lambda} - \lambda)$. Compare the results in questions b), c) and d).
- Suppose that we erroneously assumed that $X_i \sim \exp(\lambda)$ and that in fact the X_i has density $g(x; \theta, k) = \frac{\theta k}{(\theta x + 1)^{k+1}}$ for $x > 0$ and $k > 2$. Characterize the parameter λ estimated by $\hat{\lambda} = 1/\bar{X}$ in terms of k and θ . Also derive the limiting distribution of $\sqrt{n}(\hat{\lambda} - \lambda)$ under the misspecified model.
- Twist the problem around. Suppose we erroneously assumed that the X_i were iid from $g(x; \theta, k)$ for a fixed k and unknown θ , but that the X_i in fact were exponentially distributed. Again characterize the θ being estimated by the misspecified MLE $\hat{\theta}$ and give an expression for large sample distribution of $\sqrt{n}(\hat{\theta} - \theta)$. Use numerical integration to investigate the effect of the misspecification.

Problem 2

As in Problem 1 assume that $X_i, i = 1, \dots, n$ are iid and exponentially distributed, i.e. with density $f(x; \lambda) = \lambda \exp(-\lambda x)$. Now the X_i 's are not observed directly. Rather one observes for given z_i the indicator $Y_i = I(X_i \leq z_i)$. This can be described as having only left-censored or right-censored observation where X_i is left-censored at z_i if $Y_i = 1$ and right-censored at z_i if $Y_i = 0$. The data are summarized as $(z_1, Y_1), \dots, (z_n, Y_n)$. Such data are sometimes referred to as "current status data".

- a) Show that the likelihood for the data $(z_1, Y_1), \dots, (z_n, Y_n)$ can be written

$$L(\lambda) = \prod_{i=1}^n [(1 - \exp(-\lambda z_i))^{Y_i} \exp(-\lambda z_i)^{(1-Y_i)}]$$

and derive the Newton-Raphson and the Fisher-scoring algorithms for calculating the MLE λ^* based on the left- and right-censored data.

- b) Since the current status data are incomplete compared to observing all the X_i one can also obtain λ^* based on the EM-algorithm. Derive this algorithm and in particular find the conditional expectations $E[X_i | X_i \leq z_i]$ and $E[X_i | X_i > z_i]$.
- c) If all $z_i = z$ are equal then the MLE $\lambda^* = -\frac{1}{z} \ln(1 - \frac{Y_\bullet}{n})$ where $Y_\bullet = \sum_{i=1}^n Y_i$. Show that this holds. Furthermore show that the EM-algorithm has a fixed point at λ^* , i.e. if the estimate in iteration k equals $\lambda^{(k)} = \lambda^*$ then also the estimate in the next iteration $\lambda^{(k+1)} = \lambda^*$.
- d) Generate a small data set for instance with $n = 10, \lambda = 1$ and $z_i = \frac{i}{4}, i = 1, \dots, 10$ and demonstrate that the MLE λ^* is a fixed point for the EM-algorithm also in this situation.

Problem 3

Assume again that $X_i | \lambda$ for $i = 1, \dots, n$ are iid and $\exp(\lambda)$. Furthermore assume that λ has a prior from the gamma(α, β) distribution.

- a) Find the posterior distribution of $\lambda | X_1, \dots, X_n$. Furthermore derive expressions for the posterior mean and posterior mode. Compare with the posterior median for a few choices of $\sum_{i=1}^n X_i, \alpha$ and β .

- b) Assume a quadratic loss function $L(\hat{\lambda}, \lambda) = (\hat{\lambda} - \lambda)^2$. Find the risk function and the Bayes risk of both the posterior mean and the posterior mode.
- c) Argue that the posterior mean is admissible. Is the posterior mode admissible?

Problem 4

Assume that $X_i, i = 1, \dots, n$ are iid $U[0, \theta]$ where θ is an unknown parameter.

- a) Show that both $\hat{\theta} = 2\bar{X}$ and $\theta^* = M \frac{n+1}{n}$ are unbiased estimators of θ where $M = \max(X_1, \dots, X_n)$. Find the variances of $\hat{\theta}$ and θ^* .
- b) Show that M is sufficient for θ . Argue that $\tilde{\theta} = E[\hat{\theta}|M]$ is unbiased with $\text{Var}(\tilde{\theta}) \leq \text{Var}(\hat{\theta})$.
- c) Derive an explicit expression for $\tilde{\theta}$. Discuss whether it is UMVU.