# STK4011 / STK9011 Fall 2013 – Exam project

**Deadline Friday November 15th**

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

*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

Assume that $Y_i \sim n(\mu_i, 1), i = 1, \ldots, n$, are independent with $\mu_i = \alpha + \beta x_i$. The $x_i$ are given numbers, i.e. we have a simple linear regression model.

a) Write down the joint distribution of $(Y_1, \ldots, Y_n)$ on the form of an exponential family. In particular find the minimal sufficient statistics for this family.

b) The least squares estimators for $(\alpha, \beta)$ can be written as

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\widehat{\alpha} = \bar{Y} - \widehat{\beta}\bar{x}$$

where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (You are not asked to show this).

Explain why $\widehat{\alpha}$ and $\widehat{\beta}$ are UMVUE (Uniformly minimum variance among unbiased estimators).

c) Assume now that $Y_i \sim n(\mu_i, \sigma^2)$ are independent and $\mu_i = \alpha + \beta x_i$ as above, but where $\sigma$ is a third unknown parameter. Discuss the representation of the joint distribution of $(Y_1, \ldots, Y_n)$ as an exponential family and find minimal sufficient statistics.

Are $\widehat{\alpha}$ and $\widehat{\beta}$ still UMVUE?

Can you find an UMVUE for $\sigma^2$?

d) Now assume that $Y_i$ are independent binary random variables with

$$\mathrm{P}(Y_i = y_i) = \frac{\exp(y_i(\alpha + \beta x_i))}{1 + \exp(\alpha + \beta x_i)}$$

for $y_i = 0$ and $y_i = 1$. Also here the $x_i$ are given numbers and so we have a logistic regression model.

Show that also here the joint distribution of the $(Y_1, \ldots, Y_n)$ is an exponential family and show that minimal sufficient statistics are given by $T_1 = \sum_{i=1}^n Y_i$ and $T_2 = \sum_{i=1}^n x_i Y_i$.

e) Find $\mathrm{E}[T_1]$ and $\mathrm{E}[T_2]$ using a general result for exponential families.

Verify the results using standard methods for expectations.

f) Argue that the joint distribution of $(T_1, T_2)$ can be written

$$\mathrm{P}(T_1 = t_1, T_2 = t_2) = \frac{\exp(\alpha t_1 + \beta t_2)}{\prod_{i=1}^n [1 + \exp(\alpha + \beta x_i)]} K(t_1, t_2)$$

where $K(t_1, t_2) = \#\{(y_1, \ldots, y_n) : \sum_{i=1}^n y_i = t_1, \sum_{i=1}^n x_i y_i = t_2\}$. (This is a number that may require a fair deal of combinatorics to calculate).

Furthermore argue that the marginal distribution of $T_1$ can be written

$$\mathrm{P}(T_1 = t_1) = \frac{\exp(\alpha t_1)}{\prod_{i=1}^n [1 + \exp(\alpha + \beta x_i)]} H(\beta)$$

where $H(\beta) = \sum \exp(\beta \sum_{i=1}^n y_i x_i)$ and the sum is taken over all $(y_1, \ldots, y_n)$ such that $\sum_{i=1}^n y_i = t_1$ (the explicit expression for $H(\beta)$ is usually not nice).

Show that the conditional distribution of $T_2$ given $T_1 = t_1$ (i) depends on $\beta$, but not on $\alpha$ and (ii) is a one-parameter exponential family.

## Problem 2

Assume that $X_i, i = 1, \ldots, n$, are independent with density

$$f(x|\mu) = \exp(x - \mu)/(1 + \exp(x - \mu))^2, \quad x \in \Re$$

and $\mu$ is an unknown parameter in the parameter space $\Re$.

a) Find the derivative of the log-likelihood for $X_1, \ldots, X_n$.

Show explicitly that this derivative has expectation zero (i.e. not as a consequence of a general result).

Hint: You can use $\int_{-\infty}^{\infty} F(x) f(x) dx = \int_0^1 u du = \frac{1}{2}$ when $F(x)$ is a continuous cumulative distribution function and $f(x) = F'(x)$.

b) Show that the Cramér-Rao lower bound in this situation equals $\frac{3}{n}$.

c) Explain why $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimator for $\mu$ and show that its variance is larger than the Cramér-Rao lower bound. You can use that $\mathrm{Var}(X_i) = \frac{\pi^2}{3}$.

Another unbiased estimator of $\mu$ is the sample median. Explain why.

In a large sample the variance of the sample median is approximately equal to $1/(4nf(m|\mu)^2)$ where $m$ is the theoretical median. Compare this variance to the Cramér-Rao lower bound.

**Problem 3**

Assume that $X_i|\lambda, i = 1, \ldots, n$ are iid and exponentially distributed with mean $1/\lambda$ given $\lambda$. Assume furthermore that $\lambda$ has prior distribution $gamma(\alpha, \beta)$.

a) Find the posterior distribution of $\lambda$ given $X_1, \ldots, X_n$.

b) Show that the posterior mean is given by

$$\lambda^\star = \mathrm{E}[\lambda|X_1, \ldots, X_n] = \frac{\alpha + n}{1/\beta + \sum_{i=1}^{n} X_i}$$

(using the parametrization in Casella & Bergers book).

c) Find the limit distribution of $\sqrt{n}(\lambda^\star - \lambda)$.

Compare with the limit distribution of $\sqrt{n}(\widehat{\lambda} - \lambda)$ where $\widehat{\lambda} = 1/\bar{X}$ is the MLE of the $X_i$.

**Problem 4**

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

a) Show that both $\widehat{\theta} = 2\bar{X}$ and $\theta^\star = M\frac{n+1}{n}$ are unbiased estimators of $\theta$ where $M = \max(X_1, \ldots, X_n)$. Find the variances of $\widehat{\theta}$ and $\theta^\star$.

b) Show that $M$ is sufficient for $\theta$. Argue that $\tilde{\theta} = \mathrm{E}[\widehat{\theta}|M]$ is an unbiased estimator with $\mathrm{Var}(\tilde{\theta}) \leq \mathrm{Var}(\widehat{\theta})$.

c) Derive an explicit expression for $\tilde{\theta}$. Discuss whether it is UMVUE.