Statistical Inference, Master Wiskunde (Prof. C. Ley)
Exam May 2017

Name : **Surname :**

This written exam lasts for three hours and counts for 75% of the final mark. It is an open book exam. We require you to write out properly and all in detail your answers to the questions. Readability is important. Best of luck !

Question 1 : Understanding concepts

a) Explain in very clear and simple words the statement "The bootstrap is to the sample what the sample is to the population."

b) At the very beginning of the course, we have seen how to model the number of scored goals in a football match with Poisson distributions. Supposing that all played matches (say, N) are independent and hence the exact scores are our observations, what would the likelihood function look like ? How would you explain to a layman the benefits of estimating the strength parameters by means of maximum likelihood ?

c) What is the reason that power calculations are often considered at so-called local alternatives ? Explain with an example of a test statistic.

d) It is a well-known fact that the sum of two independent Poisson distributions with parameters λ_1 and λ_2 is a Poisson distribution with parameter $\lambda_1 + \lambda_2$, and vice-versa. Bearing this in mind, establish the asymptotic distribution of

$$\frac{X_n - n}{\sqrt{n}}$$

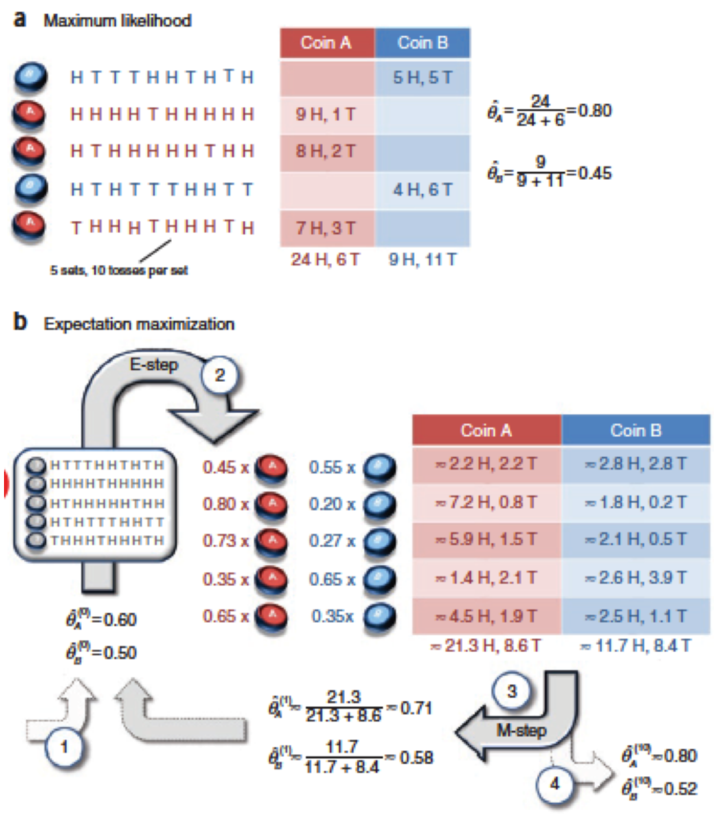
as $n \longrightarrow \infty$, where X_n is a Poisson distribution with parameter n.

e) Show that the Inverse Gamma distribution with density

$$\sigma^2 \mapsto \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma^2}\right)$$

with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$ is the conjugate prior distribution for a $\mathcal{N}(\mu, \sigma^2)$ with fixed location parameter μ.

f) Explain to a layman, on basis of the schematic representation below, how the EM algorithm works in comparison to the usual Maximum Likelihood estimation.



Question 2 : In all likelihood

Consider the following CDF :

$$F(x, \theta_1, \theta_2) = \begin{cases} 1 - \left(\dfrac{\theta_1}{x}\right)^{\theta_2} & \text{if } \theta_1 \leq x \\ 0 & \text{otherwise,} \end{cases}$$

with parameters $\theta_1, \theta_2 > 0$.

1. Find the density function $f(x, \theta_1, \theta_2)$. Provide the likelihood and log-likelihood functions.
2. Find the MLEs of θ_1 and θ_2.

Question 3 : On Fisher's Information

$X_1, ..., X_n$ are drawn from the distribution with density function $f(x, \theta) = \dfrac{3\theta^3}{(x+\theta)^4}$ for $0 < x < \infty$ and $0 < \theta < \infty$.

1. Find the (expected) Fisher Information matrix for one subject.
2. Provide the total Fisher Information matrix and interpret it.

Question 4 : The Gamma distribution

$X_1, ..., X_n$ are drawn from the Gamma distribution $\Gamma(2, \theta)$ with density function $f(x, \theta) = \dfrac{x}{\theta^2} e^{-\frac{x}{\theta}}$, with $\theta > 0$, $0 < x < \infty$, $E(X) = 2\theta$ and $Var(X) = 2\theta^2$.

1. Show that the MLE of θ is $\dfrac{\bar{X}}{2}$.

2. Find the MLE of $Var(X)$.
3. Show analytically that the MLE is consistent for θ and asymptotically normal.
4. Provide the (expected and observed) total Fisher Information matrix.
5. We want to test $H_0 : \theta = 0.5$ vs $H_1 : \theta \neq 0.5$. Provide the Wald statistic and describe the rejection region. ***In order to avoid confusion, consider the one with total expected Fisher information matrix with plug-in estimator***.
6. Find the Score test and Likelihood Ratio test.