Assignment sheet 4

October 11, 2017

Assignment 1. Let $X_1, \dots, X_n$ be an i.i.d. sample from a probability distribution F. Suppose that $\mathbb{E}[X_1^2] < \infty$. Define $\mu = \mathbb{E}[X_1]$ and $\sigma^2 = \text{Var}(X_1)$. Let $\overline{X}$ be the sample mean and $S^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \overline{X})^2$ be the sample variance.
(a) Show that $S^2 \xrightarrow{p} \sigma^2$ as $n \to \infty$.
(*Hint : Write* $S^2 = (n-1)^{-1} \sum_{i=1}^n X_i^2 - [n/(n-1)]\overline{X}^2$ *and apply the large of large numbers to the two terms followed by continuous mapping theorem and Slutsky's theorem.*)
(b) Using continuous mapping theorem and Slutsky's theorem, show that

$$T_n := \frac{\sqrt{n}(\overline{X} - \mu)}{S} \xrightarrow{d} N(0, 1)$$

as $n \to \infty$.
(Note : The statistic T_n is called a standardized or a Studentized version of $\sqrt{n}(\overline{X} - \mu)$.)
(c) Suppose that F is the $N(\mu, \sigma^2)$ distribution. What do you know about the exact sampling distribution of T_n for any fixed $n \geq 2$?
(d) Use part (b) to determine the behaviour as $n \to \infty$ of the exact distribution obtained in part (c). Compare your answer with that obtained for Exercise 7 in Week 3.

Assignment 2. Let $X_1, \dots, X_n \overset{iid}{\sim} Ber(p)$ for some $p \in (0, 1)$. Let $U_n = \overline{X}(1 - \overline{X})$, where $\overline{X}$ is the sample mean.
(a) What is U_n estimating ? Why ?
(b) Is U_n an unbiased estimator of $p(1-p)$? Justify.
(c) Is U_n a consistent estimator of $p(1-p)$? Justify.
(d) Find out the asymptotic distribution of $\sqrt{n}[U_n - p(1-p)]$ as $n \to \infty$.
(*Hint : Use the central limit theorem and the delta method.*)

Assignment 3. Let $X_1, \dots, X_n \overset{iid}{\sim} Ber(p)$ for some $p \in (0, 1)$. Let $V_n = \sum_{i=1}^n X_i$.
(a) Is X_1 unbiased for p ? Why ?
(b) What is a minimal sufficient statistic for p ?
(c) Use the theorem in Slide 91 to arrive at the answer in part (b).
(*Hint : Proceed along the steps in the proof in Slide 93.*)
(d) Find $W_n = \mathbb{E}[X_1 \mid V_n]$.
(e) Verify that $\mathbb{E}[W_n] = p$. Can you say this without doing any calculation ?
(*Hint : Use the properties of conditional expectation.*)
(f) Show directly that $\text{Var}(W_n) \leq \text{Var}(X_1)$. Is equality attained ?
(Note : This is a verification of the Rao-Blackwell theorem and W_n is the "Rao-Blackwellised" version of X_1. Also, this is the best possible "Rao-Blackwellisation".)
(g) Find the Cramer-Rao lower bound for the variance of an unbiased estimator of p. Is this lower bound attained by any estimator of p ?

Assignment 4. In a casino one plays the following game. You pay 1 franc. With probability $p = 0.49$, you win 2 francs, whereas with probability $1 - p = 0.51$ you do not win anything.
(a) Which random variable X will you use in order to describe this game ?
(b) Suppose you start with 1000 francs and play this game 1000 times. Write a formula for the probability that you have at least 1000 francs at the end.
(c) Use the central limit theorem and the R function `pnorm` to approximate the probability in (b).

(d) Use the R commands `pnorm` and `pbinom` to visualise the approximation in (c).
(e) Why do we use the normal approximation and not the Poisson approximation for the law of rare events (assignment 3(f,g), week 2) ?

Assignment 5. The exponential family structure can sometimes help in calculating integrals.
(a) Let $X \sim Gamma(5, \lambda)$ with density $x^4 e^{-\lambda x}\lambda^5/\Gamma(5)$ on $[0, \infty)$ and zero otherwise. Using results on exponential families, find $\mathbb{E}\,X$ and $\operatorname{Var} X$. Compare with the result in slide 64.
(b) Let $\theta > 0$ be a parameter and X be a random variable with density

$$f(x;\theta) = \begin{cases} \theta x^{-\theta-1}, & 1 \le x \\ 0, & \text{otherwise.} \end{cases}$$

Find $\mathbb{E}[\log X]$ et $\mathbb{E}[(\log X)^2]$.
Hint : instead of calculating painful integrals, notice that this is an exponential family with sufficient statistic related to $\log X$.
Remark. This is known as a *Pareto* distribution and it is used for modelling income.

Assignment 6. Maximising the likelihood is a way to obtain parameter estimators. In this exercise you are asked to compute the m.l.e for the following distributions :
(i) The Bernoulli distribution.
(ii) The Exponential distribution.
(iii) The Normal distribution (for both μ and σ^2)
(iv) The uniform distribution.

Assignment 7. Maximum likelihood estimation is a recipe to construct estimators. In this assignment we introduce another such recipe, *the method of moments*. Let $X \sim f(x;\theta)$ be a random variable whose distribution depends on a parameter θ. The expectation $\mathbb{E}\,X$ will therefore also depend on θ. (We assume that it is defined for all θ.) Call this function $m(\theta)$.
(a) Let $X_1, \ldots, X_n$ be an independent sample from X. What can you say about $\overline{X}_n = \sum_{i=1}^n X_i/n$ and $m(\theta)$ when n is large ?
(b) Assume that m is continuously invertible. Explain why $\widetilde{\theta} = m^{-1}(\overline{X}_n)$ is a sensible estimator of θ. It is called the *method of moments* estimator of θ.
(c) Suppose that $X \sim Exp(\lambda)$. Find the method of moments estimator $\widetilde{\lambda}$ of λ and compare with the maximum likelihood estimator $\widehat{\lambda}$.
(d) Suppose that $X \sim Unif(0, \kappa)$. Find the method of moments estimator $\widetilde{\kappa}$ of κ and compare with the maximum likelihood estimator $\widehat{\kappa}$.
(e) Suppose that $f(x;\theta) = \theta x^{-\theta-1}$ for $x \ge 1$ and zero otherwise (as in a previous assignment), with $\theta > 1$. Find the method of moments estimator $\widetilde{\theta}$ of θ and compare with the maximum likelihood estimator $\widehat{\theta}$.
(f) Compare the mean squared errors for the two types of estimators in parts (c) and (d). *Hint : some of the required calculations have been already carried out in the course.*
(g) Assuming $\theta > 2$, compare the *asymptotic* variance of the two estimators in part (e). *Hint : for* $\widetilde{\theta}$ *use the central limit theorem and the delta method ; for* $\widehat{\theta}$ *use calculations from a previous assignment.*

Assignment 8. The goal of this exercise is to see the theorem on Slide 114 (CLT for 1-parameter exponential families) applied in practice.
Let $X_1, \ldots, X_n$ be a sample from a Poisson(λ) distribution.

(i). (a) Write the minimal sufficient statistics T and call $\overline{T} = T/n$. Use theorem in Slide 100 to compute the mean and the variance of both T and $\overline{T}$.

(b) Use the theorem on Slide 114 to find the approximate sampling distribution for T.

(ii). Check the previous result graphically.
Ideally you could use the true distribution of $\overline{T}$. Alternatively, you could write a sample from $\overline{T}$ in this way

```
library(Matrix)
library(stats)

N=10
m=1000
lambda=5
sample <- matrix(rpois(N*m, lambda), N, m)
sample
suff_stat <- colMeans(sample)
suff_stat
```

Standardize by the true mean and the true variance as in theorem on Slide 114. Use `ecdf` and `pnorm` to plot the empirical cdf of $\overline{T}$ against the cdf of a standard normal distribution.

(iii). Find the exact distribution of T. Repeat the computation of part (ii) sampling from the true distribution of T.
(*Use the moment generating fuction*)

Assignment 9. Last week we have found a sufficient statistics for some member of the exponential family. This week we focus on minimally sufficient.

(i) Prove that $T(z) = \sum_{i=1}^{n} z_i$ is a minimal sufficient statistics for the Binomial distribution.
(i) Prove that $T(y) = \sum_{i=1}^{n} y_i$ is a minimal sufficient statistics for the Poisson distribution.
(iii) Prove that $T_1(y) = \sum_{i=1}^{n} y_i$ is a minimal sufficient statistics for the mean of a Normal $\mathcal{N}(\mu, \sigma^2)$ distribution and $T = (T_1, T_2)$ with $T_2(y) = \sum_{i=1}^{n} y_i^2$ is a minimum sufficient statistics for σ (and thus for both parameters).