Prelims Statistics and Data Analysis – Sheet 2 TT 2020

1. If $X_1, \ldots, X_n$ is a random sample from a geometric distribution with parameter p, find the maximum likelihood estimator $\widehat{p}$ of p.

Let $\theta = 1/p$. Find the likelihood as a function of θ, the maximum likelihood estimator $\widehat{\theta}$, and verify that $\widehat{\theta} = 1/\widehat{p}$.

Show that $\widehat{\theta}$ is unbiased. In the case $n = 1$ show that $E(\widehat{p}) > p$. [*In the $n = 1$ case, having first shown $E(\widehat{p}) > p$, can you find the value of $E(\widehat{p})$?*]

2. Suppose $X_1, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ distribution, where $\mu = \sigma^2 = \theta$. Show that the maximum likelihood estimator of θ is

$$\widehat{\theta} = \frac{1}{2}\left\{\left(1 + \frac{4}{n}\sum_{j=1}^{n} X_j^2\right)^{1/2} - 1\right\}.$$

3. A researcher wishes to estimate the density ρ of organisms per unit volume in a liquid. She conducts n independent experiments: in experiment $i = 1, \ldots, n$, she takes a fixed volume v_i of liquid and measures the number of organisms X_i in this volume – she assumes X_i has a Poisson distribution with mean ρv_i. Find the maximum likelihood estimator $\widehat{\rho}$ and find the bias of $\widehat{\rho}$.

If the total volume taken is fixed, $\sum_{i=1}^{n} v_i = a$ say, show that the variance of $\widehat{\rho}$ only depends on $v_1, \ldots, v_n$ via their sum a.

4. Suppose $X_1, \ldots, X_n$ is a random sample from a distribution with probability density function

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

Find the maximum likelihood estimator of θ.

5. The following data (from Dyer (1981)) are annual wages (in multiples of 100 US dollars) of a random sample of 30 production line workers in a large American industrial firm.

Annual wages (hundreds of US $)

112	154	119	108	112	156	123	103	115	107
125	119	128	132	107	151	103	104	116	140
108	105	158	104	119	111	101	157	112	115

(a) A standard probability model used for data on wages is the Pareto distribution, which has probability density function

$$f(x;\theta) = \theta\alpha^{\theta}x^{-(\theta+1)} \quad \text{for } x \geqslant \alpha,$$

where $\theta > 0$ and the constant α represents a statutory minimum wage. Find the maximum likelihood estimator of θ from a random sample $X_1, X_2, \ldots, X_n$, and, assuming $\alpha = 100$, the maximum likelihood estimate for the above dataset (for which $\sum \log x_i = 143.5$).

(b) Now suppose there is no statutory minimum wage, so that α is also an unknown parameter.

(i) Show that the MLE for α is $\widehat{\alpha} = \min_i X_i$. What is the MLE for θ?

(ii) Show that

$$P(\widehat{\alpha} > y) = \left(\frac{\alpha}{y}\right)^{n\theta} \quad \text{for } y \geqslant \alpha.$$

[*Use the fact that* $\min_i X_i > y \iff \{X_i > y \text{ for } 1 \leqslant i \leqslant n\}$.]

(iii) Deduce that, for each $\epsilon > 0$,

$$P(|\widehat{\alpha} - \alpha| > \epsilon) \to 0 \quad \text{as } n \to \infty.$$

6. (Using R or Matlab)

R: work through Rdemo-1 on the course website. At this stage the idea is to get some experience with R and to look at simple plots of the `trees` data. In a few lectures' time we will fit linear regression models to data of this type.

Matlab: work through the Matlab section on the final page of Rdemo-1.