Statistical Inference (Level M) 2020-21
Tutorial Sheet 3

*These questions relate to Chapter 3 (part 1), lectures 7-9. Please attempt at least the * questions before the tutorial.*

1. Suppose that $X_1, X_2, \ldots, X_n$ are independent observations from a $\text{Poi}(\lambda)$ distribution (where $\lambda \geq 0$ is unknown). It is proposed to estimate λ using the sample mean,

$$t(\mathbf{X}) = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

   Show that $t(\mathbf{X})$ has the same range as λ, and that it is an unbiased and consistent estimator of λ.

2. * Suppose that $X_1, X_2, \ldots, X_n$ are independent observations from a $\text{N}(\mu, \sigma^2)$ distribution (where μ is unknown and σ^2 is known). It is proposed to estimate μ using the sample mean,

$$t(\mathbf{X}) = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

   Show that $t(\mathbf{X})$ has the same range as μ, and that it is an unbiased and consistent estimator of μ.

3. Suppose that $X_1, X_2, \ldots, X_n$ are independent $\text{U}(0, \theta)$ random variables, for unknown $\theta > 0$. It is proposed to estimate θ by $t(\mathbf{X}) = 2\bar{X}$.

   By using the fact that a $\text{U}(0, \theta)$ random variable has $\text{E}(X_i) = \frac{1}{2}\theta$ and $\text{var}(X_i) = \frac{1}{12}\theta^2$. Show that $t(\mathbf{X})$ is an unbiased and consistent point estimator of θ.

4. Suppose that $x_1, x_2 \ldots, x_n$ are observations of the independent random variables $X_1, X_2, \ldots, X_n$ respectively. Obtain the maximum likelihood estimator of the unknown parameter θ in each of the following models for $X_1, X_2, \ldots, X_n$.

   (a) $X_i \sim \text{Bi}(m_i, \theta)$ where $0 < \theta < 1$ and $m_1, m_2, \ldots, m_n$ are known positive integers.

   (b) $X_i \sim \text{Geo}(\theta)$ where $0 < \theta < 1$

   (c) * $X_i \sim \text{Ga}(\alpha, \theta)$ where $0 < \theta$ and α is a known positive constant

   (d) $X_i \sim \text{U}(-\theta, \theta)$ where $0 < \theta$

5. * As part of a project on the occurence and distribution of a species of hoverfly in Trinidad[1] data were collected on the numbers of larvae found on 20 randomly sampled flowering bracts of *Heliconia* plants. The observed numbers were

$$9,\ 14,\ 3,\ 3,\ 8,\ 7,\ 7,\ 6,\ 7,\ 0,\ 6,\ 0,\ 5,\ 1,\ 3,\ 12,\ 2,\ 4,\ 0,\ 11$$

One possible model for the number of larvae (y_i) is that they are observations of independent random variables with probability function

$$f(y_i) = \theta^{y_i}(1-\theta) \quad 0 < \theta < 1$$

where the y_i are non-negative integers.

(a) Write down the log likelihood function for θ.

(b) Find the maximum likelihood estimate of θ, and check that it is a *maximum* likelihood estimate.

6. In a study looking at the effect of rising average temperatures on insect pest populations, the time from egg hatch to adulthood was recorded for 20 fruitflies of the species *Drosophila Simulans*. The recorded times (T_i) in days are:

```
11.7 15.0 17.5 14.1 12.6 13.2 11.4 14.5 11.4 13.8
13.0 14.9 12.0 13.5 13.4 13.0 12.6 10.3 11.6 11.4
```

The minimum development time implied by the physiological maximum development rate is about 10 days. A reasonable model for these data is that they are observations of independent r.v.s. $T_i = 10 + X_i$ where X_i has a distribution with p.d.f.

$$f(x_i) = \frac{1}{2\beta^3} x_i^2 e^{-x_i/\beta} \quad x_i > 0, \quad \beta > 0$$

Find the maximum likelihood estimate of β.

7. * Suppose that we have data $x_1, \ldots, x_n$ which we model as observations of independent random variables X_i with p.d.f.

$$f(x_i) = \begin{cases} (\beta+1)x_i^{\beta} & 0 < x_i < 1 \\ 0 & \text{otherwise} \end{cases}$$

where β is an unknown parameter. Find the maximum likelihood estimator of β, and check that it is a *maximum* likelihood estimate.

[1]Data collected by Sharon Kennedy

8. The leaves of a certain species of plant are examined for insect infestation. When at least one insect is present on a leaf, the total number of insects on the leaf is recorded. Otherwise, no record is kept of that particular leaf. The table below shows the results from examining 100 leaves.

no. of insects, k	1	2	3	4	5	6	7
frequency, n_k	26	27	18	18	8	2	1

It is believed that the number of insects infesting a leaf follows a Poisson distribution. Since the zero category is missing (no recorded data), the recorded data follow what is known as a Truncated Poisson distribution.

Model: $X_1, X_2, \ldots, X_n$ independent, with each X_i following a Truncated Poisson distribution

Data: $x_1, x_2, \ldots, x_n$, where n_k observed values are equal to k $(k = 1, 2, \ldots)$

The probability mass function for each random variable is:

$$\frac{e^{-\lambda}\lambda^x}{(1-e^{-\lambda})x!}$$

It can be shown that the MLE for λ satisfies the equation:

$$\frac{\lambda}{1-e^{-\lambda}} - \bar{x}$$

If the data were not truncated, then the MLE would occur at the sample mean, $\bar{x}$. This is a sensible first approximation to the MLE for the truncated case.

By setting $g(\lambda) = \frac{\lambda}{1-e^{-\lambda}} - \bar{x}$, with $\bar{x} = 2.65$, use the Newton-Raphson method of optimisation to find $\hat{\lambda}_{MLE}$. For a first approximation $\lambda^{(0)}$ take $\bar{x} = \lambda^{(0)} = 2.65$. For $g'(\lambda)$ use $g'(\lambda) = 1 - \bar{x}e^{-\lambda}$.

Complete the following table to help with this:

Iteration	$\lambda^{(j)}$	$g(\lambda^{(j)})$	$g'(\lambda^{(j)})$	$g(\lambda^{(j)})/g'(\lambda^{(j)})$
0	2.65			
1				
2				
3				