MATH11712 Statistics I
Semester 2, 2022 / 2023
Example Sheet 8

Please submit your solutions to question 3 through blackboard by 12noon Monday 20 March.

1. Let X_1, X_2, X_3 be a random sample from a $N(\mu, \sigma^2)$ distribution, where μ is unknown. Consider the two estimators of μ given by:

$$\hat{\mu}_1 = \frac{X_1 + X_2 + X_3}{3}$$

$$\hat{\mu}_2 = \frac{X_1 + 2X_2 + X_3}{4}$$

(i) Show that both of these estimators are unbiased for μ.

(ii) Find $\text{Var}(\hat{\mu}_1)$ and $\text{Var}(\hat{\mu}_2)$. Which estimator would you prefer to use in practice?

(iii) What are the sampling distributions of each of these estimators?

2. Let $\bar{X}_1$ and $\bar{X}_2$ be the sample means of two independent random samples of sizes n_1 and n_2, respectively from the $N(\mu, \sigma^2)$ distribution, where μ is unknown but σ is known. Consider the estimator

$$\hat{\mu} = a\bar{X}_1 + (1-a)\bar{X}_2 \text{ for } 0 \leq a \leq 1.$$

Show that the $\text{bias}(\hat{\mu}) = 0$, $\forall a$ and that $\text{Var}(\hat{\mu})$ is minimized when $a = \frac{n_1}{n_1+n_2}$.

3. $*$ Let $X_1, \ldots, X_n$ be a random sample from the distribution with p.d.f.

$$f(x) = \begin{cases} e^{-(x-\delta)}, & x > \delta \\ 0, & \text{otherwise}. \end{cases}$$

where the value of δ is unknown.

(i)* Show that $\bar{X}_n$ is a biased estimator of δ. [2 marks]

(ii)* Derive an expression for $\text{bias}(\bar{X}_n)$ and discuss what happens to its value as $n \to \infty$. [2 marks]

(iii)* Can you propose an alternative, unbiased estimator of δ? [1 mark]

4. Let $x_1, \ldots, x_5$ be a random sample of values, which we assume were generated from a $Bi(3, p)$ distribution, where p is unknown.

(i) Write down the likelihood function, $L(p)$, for these data. [Recall that each x_i is a count of the number of successes in $m = 3$ Bernoulli trials.]

The following observations were actually obtained by random sampling from the $\text{Bi}(3, p)$ distribution, where p is unknown:

$$1, 3, 2, 2, 3$$

Suppose that, for the purposes of this question, p can only take one of four different values which are given by the set $\{0.5, 0.6, 0.7, 0.8\}$.

(ii) Use these data to calculate the values of the likelihood function $L(p)$ for each of the four different possible values of p. Which of the possible values of p would you choose to be the maximum likelihood estimate based on the above data?

5. Let $x_1, \ldots, x_n$ be a random sample of n observations from a Geometric distribution with unknown parameter p.

(i) Write down the likelihood and log-likelihood functions for these data.

[Look up the appropriate probability mass function in the notes 'Some Common Probability Distributions' on Blackboard, or in a text book.]

(ii) Use calculus on the log-likelihood function to show that the maximum likelihood estimate of p is given by

$$\hat{p} = \frac{1}{\bar{x}}.$$

6. The polychlorinated biphenyl (PCB) concentration in a fish caught in Lake Michigan was measured by a technique known to give a measurement error that is normally distributed with a standard deviation of 0.8 parts per million. The results of $n = 10$ independent measurements of the concentration are

$$11.2, 12.4, 10.8, 11.6, 12.5, 10.1, 11.0, 12.2, 12.4, 10.6$$

(each in parts per million). Calculate a (i) 95% and a (ii) 99% confidence interval estimate of the mean PCB level of this fish.

7. An engineering firm manufactures a component that has a lifetime that is normally distributed with a known standard deviation of 3.4 hours. A random sample of $n = 9$ components is obtained, in which the average lifetime is 100.8 hours. Find a (i) 95 percent and a (ii) 98 percent confidence interval estimate of the mean component lifetime.