Actuarial Mathematics and Statistics
Statistics 5 Part 2: Statistical Inference
Tutorial Problems

1. Which of the following statements relate to probabilities that can be interpreted as frequencies? For those that can, describe the corresponding sequence of experiments.

    (a) The probability that a '6' is scored with a given die is $\frac{1}{6}$.

    (b) The probability that a randomly chosen 20-year old male survives until age 60 is 0.8.

    (c) The probability that Wayne Rooney (famous footballer) survives until age 60 is 0.8.

    (d) The probability that Jack the Ripper was a member of the Royal Family is 0.05.

    (e) The probability that the next pint of milk you buy goes off within 1 week of purchase is 0.6.

2. Which standard distributions might be appropriate to model the following experimental outcomes?

    (a) Select 10 individuals randomly from the student population of Heriot-Watt and count the number of 1st-year students in your sample.

    (b) Throw a dart at a dartboard repeatedly until you score a treble 20. Count the number of throws required.

    (c) Monitor a stretch of road continuously for 1 year and count the number of accidents that occur.

    (d) Select a male student at random from the HW population and measure their height.

    (e) Put a new battery in a watch and measure the time elapsed before the watch stops.

3. Let $\underline{X}$ denote a random sample of size n from a distibution with parameter θ and let $g(\underline{X})$ denote an estimator for θ.

    (a) Show that $\text{MSE}(g(\underline{X})) = \text{Var}(g(\underline{X})) + (\text{bias}(g(\underline{X})))^2$.

(b) Suppose the mean and variance of X are μ and σ^2 respectively and that $\bar{X}$ and S^2 denote the sample mean and variance. Show that (i) $E(\bar{X}) = \mu$, (ii) $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$, (iii) $E(S^2) = \sigma^2$.

4. Let $\underline{X}$ denote a random sample of size n from a $Poisson(\lambda)$ distribution. Consider the estimator $\hat{\lambda} = \bar{X}$.

   (a) Is $\hat{\lambda}$ an unbiased estimator for λ? Give reasons.

   (b) Calculate the MSE of $\hat{\lambda}$.

   (c) Calculate the Cramer-Rao Lower Bound and determine whether $\hat{\lambda}$ is the most efficient unbiased estimator.

   (d) Consider the 'family' of estimators $\hat{\lambda} = a\bar{X}$ where $a > 0$. Calculate the MSE of such an estimator as a function of a. Which value of a gives the estimator with the minimum MSE?

5. Consider the example 2.2.1. from the notes.

   (a) Prove carefully that a^* and b^* are given by the expressions (i.e. minimise the expression for the MSE with respect to a and b.)

   (b) For the case where $\theta = 100$, $\sigma_1^2 = \sigma_2^2 = 1$ calculate the MSE for the optimal estimator (given by a^* and b^*) and the MSE for the optimal unbiased estimator.

6. Let $\underline{X}$ be a random sample from $\exp(1/\theta)$ (i.e. with mean θ).

   (a) Write down $E(\bar{X}), \text{Var}(\bar{X})$ and $\text{MSE}(\bar{X})$.

   (b) Now consider the estimator $Y = a\bar{X}$. Write down $\text{MSE}(Y)$ and determine the value of a which minimizes it. Compare $\bar{X}$ and the optimum Y for both small and large values of n. Comment.

   (c) For the case where $n = 200$ use Chebyshev's inequality to obtain an upper bound for the frequency with which the estimator $\bar{X}$ will not fall within 10% of the true value of θ (i.e. between 0.9θ and 1.1θ.)

7. In the situation described in question 6, for a sample of size n, determine the Cramer-Rao lower bound for the variance of an unbiased estimator of θ. Comment.

8. Let $X \sim Bin(n, p)$. Show that $\frac{X}{n}$ is a consistent estimator for p.

9. Let $\underline{X}$ be a random sample from $N(0, \sigma^2)$.

(a)Show that the Cramer-Rao lower bound for the unbiased estimators of σ^2 is $\frac{2\sigma^4}{n}$.

(b) Investigate the following estimators for σ^2 (i.e. comment on bias and efficiency).

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2$$

$$S_{n+2}^2 = \frac{1}{n+1}\sum_{i=1}^{n}X_i^2.$$

Note: For a random sample from $N(\mu, \sigma^2)$, we have $Var(S^2) = \frac{2\sigma^4}{n-1}$ (we will prove this later). Also, for a $N(0, \sigma^2)$ random variable X, $Var(X^2) = 2\sigma^4$ (try to prove this using moment generating functions!).

10. A population consists of 100 individuals of whom some unknown proportion p carry a certain gene. Suppose you randomly select n of these *without replacement* and test whether they carry the gene, generating observations $Y_1, Y_2, ..., Y_n$, where $Y_i = 1$ if the ith individual carries the gene and is equal to zero otherwise.

(a) Prove that the marginal distribution of Y_i is $Bernoulli(p)$ for all $i = 1, \ldots, n$.

(b) Explain why the Y_i are not independent of each other.

(c) If $p = 0.4$ and $n = 50$ calculate the MSE of the estimator $\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n}$.

(d) Use Chebyshev's inequality to calculate an upper bound for the frequency with which $|\hat{p} - p|$ exceeds 0.05, if $p = 0.4$ and $n = 50$.

11. A health-and-safety officer believes that the number of industrial accidents occuring each month comes from a $Poisson(\lambda)$ distribution with the numbers being independent between months. In the 3 successive months they observe 3, 6, and 2 accidents, respectively.

(a) Construct the likelihood $L(\lambda)$ and the loglikelihood $l(\lambda)$ for this data set.

(b) Maximise the log-likelihood to obtain the MLE of λ for this data set.

(c) Find the method of moments estimator for λ. How does it compare with the MLE?

12. For random (i.i.d.) samples from the following distributions, find the MLE and the MME of the indicated parameters:

(a) p in $\text{Bin}(n, p)$ with n known and sample size 1.

(b) λ in $\exp(\lambda)$, sample size n.

(c) θ in $\text{Beta}(\theta, 1)$, sample size n. (Density is given by $f(x; \theta) = \theta x^{\theta-1}$ for $0 < x < 1$).

13. A bag contains 10 balls in total, r of which are black with the remainder white, where r is unknown. A random selection of 4 balls (without replacement) consists of 2 black and 2 white balls.

(a) Identify the range of values of r for which the likelihood $L(r)$ is non-zero.

(b) Show that the the likelihood $L(r)$ satisfies:

$$L(r) \propto r(r - 1)(10 - r)(10 - r - 1).$$

(c) Evaluate the r.h.s. of the expression in (b) for different values of r and hence find the MLE of r. Is it what you expected?

14. A random sample of size 2, i.e. $\underline{X} = (X_1, X_2)$, is taken from a population with density given by

$$f(x) = \frac{2}{\theta}(1 - \frac{x}{\theta}) \quad \text{for} \quad 0 < x < \theta$$

where $\theta > 0$ (note that the *range* of X depends on θ). Draw a rough graph of the likelihood function $L(\theta)$ and determine the MLE $\hat{\theta}$. Determine the MME of θ and compare this to $\hat{\theta}$.

15. Let $x_1, \ldots, x_n$ be the values in a random sample of size n from a $U(a, b)$ distribution (where $b > a$) and let $x_{(1)}$ and $x_{(n)}$ denote the smallest and largest values in the sample respectively.

(a) Sketch the region of parameter space where the likelihood is non-zero and for points (a, b) lying in this region determine the value of the likelihood $L(a, b)$.

(b) Hence show that the maximum likelihood estimate is given by $(\hat{a}, \hat{b}) = (x_{(1)}, x_{(n)})$.

(b) (Harder) Consider the sampling properties of $\hat{a}$ and $\hat{b}$. Do you think that they are independent of each other? Calculate $Cov(\hat{a}, \hat{b})$ for the case where $n = 2$, $a = 0$ and $b = 1$. (*Hint: Consider the joint density of (X_1, X_2) can you identify the joint density of $(X_{(1)}, X_{(2)})$?*)

16. The number of cars that exceed the speed limit on a stretch of road on any given day is believed to be $Poisson(\lambda)$ where λ is unknown. Furthermore it is believed that the numbers on different days are independent of each other. To estimate λ you install a speed camera to record the number of cars on 10 successive days. However, the camera is defective and only records the first speeding car each day. At the end of the trial you find that there were no speeding cars on 4 of the days with at least one culprit on the other 6 days.

(a) Find the likelihood $L(\lambda)$ by writing the probability of the observations as a function of λ. Maximise it to obtain the MLE of λ.

(b) What is the distribution of the number of days out of the 10 on which there are no speeding cars. (*Hint: Think of a day on with no speeding cars as a success.*)

5

(c) Use the result from (ii) and properties of MLEs to give a 'quicker' derivation of the MLE of λ.

17. In a game involving a biased coin with $P(H) = p$ (where p is unknown) players toss the coin 3 times and win if they obtain at least one Head. Suppose that 20 players play the game, resulting in 10 winners. Find the MLE of p.

18. The diameters of tubers (measured in cms) of a certain variety of potato follow an $Exp(\lambda)$ distribution where λ is unknown. You are provided with a random sample of 100 tubers and two sorting grids. Any tuber less than 6 cms. in diameter will pass through the first grid while any tuber less than 3 cms. will pass though the second grid. Out of the 100 tubers, 20 pass through neither grid, and 50 pass through both grids.

(a) Show carefully that the likelihood $L(\lambda)$ is given by

$$L(\lambda) = (1 - e^{-3\lambda})^{80} e^{-210\lambda}$$

(b) Find the value of λ which maximises this expression. (Hint: Make the substitution $p = e^{-3\lambda}$ and find the MLE of p first of all.)

19. Let $\underline{X}$ be a random sample of size n from Poisson(μ). Of the n observations, n_0 equal 0, n_1 equal 1, and the remaining $n - n_0 - n_1$ observations are greater than 1.

(a) Obtain an equation satisfied by the MLE $\hat{\mu}$.

(b) In the case $n = 20, n_0 = 8, n_1 = 7$, use numerical methods to find $\hat{\mu}$ correct to 4 decimal places.

Note: Use Poisson tables to find a good starting value for your numerical calculations.

20. Let $\underline{X}$ be a random sample from $\Gamma(t, \lambda)$.

(a) Obtain a set of equations satisfied by the MLEs $\hat{t}$ and $\hat{\lambda}$. Suggest how these might be solved.

(b) Determine the MMEs for t and λ. Which method (MLE or MME) seems easier?

6

21. *Independent* samples of size n_1 and n_2 are taken from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ respectively. Determine the MLEs $\hat{\mu}_1, \hat{\mu}_2$, and $\hat{\sigma}^2$. (Hint: $L = L_1 \cdot L_2$).

22. Let $\underline{X}$ be a random sample from a truncated $N(\mu, 1)$ distribution with density given by

$$f(x; \mu) = \left(\frac{1}{1 - \Phi(-\mu)} \right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2} \right) \quad \text{for} \quad 0 < x < \infty.$$

(a) Show that both the MLE and MME are given by the solution to

$$\bar{x} - \mu - \frac{\phi(-\mu)}{1 - \Phi(-\mu)} = 0$$

where ϕ and Φ are the density and distribution functions, respectively, of $N(0, 1)$.

(b) If $\bar{x} = 1.42$, find $\hat{\mu}$ to 1 decimal place.

23. The time till relapse in months of patients treated with a certain drug is believed to follow a $\Gamma(2, \beta)$ distribution where β is unknown. To estimate β you take a random sample of 10 patients at time $t = 0.0$ and monitor them over the following 6 months. In this time 6 patients relapse at times $t_1, \ldots, t_6$ for which $\sum t_i = 13.2$. The remaining 4 patients do not relapse.

(a) Show that the probability that a given patient does not relapse (as a function of β)is given by

$$p(\beta) = (1 + 6\beta)e^{-6\beta}.$$

(b) Explain carefully why the likelihood satisfies

$$L(\beta) \propto \beta^{12}(1 + 6\beta)^4 e^{-37.2\beta}.$$

Hint. You need to include a factor corresponding to each of the 10 patients.

(c) Find a quadratic equation satisfied by the MLE of β and hence find the MLE. (Remember that β must be positive!)