MA20226 - Question Sheet One

Simon Shaw, s.shaw@bath.ac.uk
http://people.bath.ac.uk/masss/ma20226.html

2011/12 Semester I

Set: Tutorials of Week 2, commencing Monday 10th October 2011.
Due in: Practicals of Week 3, commencing Monday 17th October 2011. You should write both your name and tutorial group on your work. If you are unable to make your practical then you should ensure that you hand the work to me personally *before* the practical is held; my office is 4W4.10.
Task: Attempt questions 1-3; questions 4-6 are additional questions which may be used for tutorial discussion.

1. State the sample space Ω and the parameter space Θ for the following random quantities. In each case, what is the interpretation of the parameters? Based on this, try to suggest an intuitive estimator for each parameter.

   (a) $X \sim Bernoulli(p)$

   (b) $X \sim N(\mu, \sigma^2)$

   (c) $X \sim Bin(n, p)$ (you may assume that n is known)

   (d) $X \sim Exp(\lambda)$ (the exponential distribution, parameterised so that $E(X \mid \lambda) = 1/\lambda$)

   (e) $X \sim Geo(p)$ (the geometric distribution, counting the number of trials up to and including the first success)

   (f) $X \sim U(a, b)$ (the uniform distribution)

2. Suppose $X_1, \ldots, X_n$ are independent and identically distributed (iid) $N(\mu, \sigma^2)$ random quantities. Using the properties of independent normals and expectation and variance operators, explain why the sampling distribution of $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is $N(\mu, \sigma^2/n)$.

3. Let $x_1, \ldots, x_n$ be a random sample (so that the corresponding X_i are independent) from an exponential distribution with probability density function

$$f(x|\tau) \quad = \quad \frac{1}{\tau}\exp(-x/\tau) \quad 0 \leq x < \infty$$

   and zero otherwise, where $\tau > 0$.

   (a) Show that if the model is correct it is sufficient to know the sample size n and the sample mean $\overline{x}$ to evaluate the likelihood function for any value of τ.

   (b) Find the maximum likelihood estimator $\tilde{\tau}$, a function of $X_1, \ldots, X_n$, of τ.

   (c) Show that $\tilde{\tau}$ is an unbiased estimator of τ; that is, $E(\tilde{\tau}|\tau) = \tau$ for all $\tau > 0$.

4. *Capture-recapture.* How does one estimate N, the total number of fish in a lake? The following technique is actually used for this and other problems concerning sizes of wildlife populations. A net is set up to catch a total of M fish. These are marked and returned to the lake. At a later date, a second batch of n fish are caught. The size of the fish population can then be estimated from seeing how many of the marked fish have been caught in the second sample. We denote this number by Y.

   (a) Explain why $E(Y \mid M, n, N) = nM/N$ and hence suggest an estimator for N. Evaluate your estimator for the possible values of y, the observed number of marked fish caught in the second stage, when six fish are marked in the first stage and ten fish are to be caught in the second stage.

   (b) Explain why

$$P(Y = y \mid M, n, N) \quad = \quad \frac{\binom{M}{y}\binom{N-M}{n-y}}{\binom{N}{n}}.$$

   For an observed value y of Y, discuss how you might use $P(Y = y \mid M, n, N)$ (when considered as a function of N) as providing an alternative route to estimating N. What is your estimate when $M = 6$, $n = 10$ and $y = 3$ and how does this compare to your estimate in (a)?

5. Let $X_1, \ldots, X_n$ be iid random quantities, each having probability density function

$$f(x \mid \theta) \quad = \quad \begin{cases} \frac{x}{\theta^2} \exp\left(-\frac{x}{\theta}\right) & 0 < x < \infty, \\ 0 & \text{otherwise} \end{cases}$$

   for $0 < \theta < \infty$. Find the maximum likelihood estimator, $T = T(X_1, \ldots, X_n)$, of θ.

6. Suppose that a parameter θ can assume one of three possible values $\theta_1 = 1$, $\theta_2 = 10$ and $\theta_3 = 20$. The distribution of a discrete random quantity Y, with possible values y_1, y_2, y_3, y_4, depends on θ as shown in the table below.

	θ_1	θ_2	θ_3
y_1	0.1	0.2	0.4
y_2	0.1	0.2	0.3
y_3	0.2	0.3	0.1
y_4	0.6	0.3	0.2

   Thus, each column gives the distribution of Y given the value of θ at the head of the column. For example, $P(Y = y_4 \mid \theta = \theta_1) = 0.6$.

   (a) Write down the parameter space Θ.

   (b) A single observation of Y is made. Evaluate the maximum likelihood estimator $\tilde{\theta}$ of θ (so $\tilde{\theta}$ is a function of the four possible outcomes of Y).

   (c) Evaluate the sampling distribution of $\tilde{\theta}$; that is, for each θ compute the probability distribution of $\tilde{\theta}$, based on a single observation of Y, i.e. $P(\tilde{\theta} = \theta_i \mid \theta = \theta_j)$ for each $i, j = 1, \ldots, 3$. Display your answer in tabular form.

   (d) Find $E(\tilde{\theta} \mid \theta = \theta_1)$. Is $\tilde{\theta}$ an unbiased estimator of θ?