MA20226 - Question Sheet Seven

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

2011/12 Semester I

Task: These are questions on the material covered in the final three lectures of the course and some of them may be used for tutorial discussion. If, at any point, you want any of these questions marking then just hand them in to me in my office, 4W4.10, and I'll mark them for you and return the sheet to you.

1. Two voltmeters, A and B, were used to measure the voltage of a standard cell. Ten independent observations were taken using A and 8 using B. The observations were:

Meter A	1.11	1.15	1.14	1.10	1.09	1.11	1.12	1.15	1.13	1.14
Meter B	1.21	1.15	1.11	1.17	1.20	1.14	1.15	1.12		

   (a) Assuming that readings from A $\sim N(\mu_1, \sigma_1^2)$, and those from B $\sim N(\mu_2, \sigma_2^2)$, and that all readings are independent, test the hypothesis that the variances of observations from the two meters are equal against the hypothesis that they are different.

   (b) If you conclude that there is insufficient evidence to suggest the variances are not equal, form a pooled estimate of the common σ^2 and test the hypothesis that the means from the two meters are equal against the hypothesis that they are different.

2. Suppose that $X_1, \ldots, X_n$ are iid $N(\mu_X, \sigma^2)$ random quantities independent of $Y_1, \ldots,$ Y_m which are iid $N(\mu_Y, \sigma^2)$ random quantities. What are the distributions of the following:

   (a) $\overline{X}$ and $\overline{Y}$

   (b) S_X^2 and S_Y^2

   (c) $\frac{\overline{X}-\mu_X}{\sigma/\sqrt{n}}$

   (d) $\frac{\overline{Y}-\mu_Y}{S_Y/\sqrt{m}}$

   (e) $\overline{X} - \overline{Y}$

   (f) S_X^2/S_Y^2

   (g) S_Y^2/S_X^2

   (h) $S_p^2 = \frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}$

   (i) $\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)}}$

(j) $\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}$

3. The breaking strengths of fourteen fibres (seven chosen randomly from a large batch of type A, and seven chosen at random from a batch of type B) is recorded:

	1	2	3	4	5	6	7
Batch A	15.7	16.1	16.0	14.8	14.1	15.2	16.3
Batch B	15.0	14.6	14.3	15.2	12.9	15.1	14.6

Test whether the two fibre types have the same mean breaking strength.

4. Let $x_1, \ldots, x_n$ be a random sample from a geometric distribution

$$P(X = x \,|\, p) \quad = \quad (1-p)p^{x-1} \quad x = 1, 2, \ldots$$

where $p \in (0, 1)$. Hence, this is the distribution for the number of trials until the first failure in a sequence of independent trials with each trial having a probability p of being a success.

(a) Explain why the sample average $\overline{x}$ is a sufficient statistic for calculating the likelihood function when the sample size n is known.

(b) Find the maximum likelihood estimate for p in terms of $\overline{x}$.

In an ecological study of the feeding behaviour of birds, the number of hops between flights was counted for several birds. For the following data, fit a geometric distribution to these data using the maximum likelihood estimate of p and test for goodness of fit using Pearson's chi-square statistic, remembering the "rule of thumb" to "pool" counts in adjacent cells so that all the resulting cell expectations are at least 5.

Number of hops	1	2	3	4	5	6	7	8	9	10	11	12
Count	48	31	20	9	6	5	4	2	1	1	2	1

5. An offspring in a breeding experiment can be of three types with probabilities, independently of other offspring,

$$\tfrac{1}{4}(2+p) \quad \tfrac{1}{2}(1-p) \quad \tfrac{1}{4}p$$

(a) Argue that for n offspring the probability that there are a, b and c of the three types, respectively, is of the form

$$C(2+p)^a(1-p)^b p^c$$

where C does not depend upon p.

(b) Show that the maximum likelihood estimate $\hat{p}$ of p is a solution of $np^2 + (2b+c-a)p - 2c = 0$.

(c) Suppose that an experiment gives $a = 58$, $b = 33$ and $c = 9$. Find the maximum likelihood estimate $\hat{p}$.

(d) Use $\hat{p}$ to calculate expected frequencies of the three types of offspring, and test the adequacy of the genetic model using Pearson's chi-square statistic.