MATH11400 Statistics 1 2010–11

Homepage http://www.stats.bris.ac.uk/%7Emapjg/Teach/Stats1/

Problem Sheet 7

*1. (a) (From a recent Guardian puzzle) A lazy flea is wandering along a ruler. He knows that at a certain time, he will receive an instruction to move to the 1 inch mark on the ruler, the 2 inch mark or the 11 inch mark. Which of these it will be is uncertain, and he can assume there is 1/3 probability of each of these possibilities. Where should he position himself to minimise the distance he has to move when instructed? (Like all puzzles, this is rather imprecisely stated: make reasonable assumptions and then solve it).

(b) How does your answer change if instead he want to minimise the mean *squared* distance he has to move?

(c) What if he wants to minimise the *maximum* distance he might have to move?

(d) What does this question have to do with the issue of giving a numerical summary of the centre of a sample of data $x_1, x_2, \ldots, x_n$?

*2. Let $X_1, \ldots, X_n$ be a random sample of size n from a general distribution with population mean denoted by $\mu = \mathrm{E}(X)$ and population variance denoted by $\sigma^2 = \mathrm{Var}(X)$. (Note: we are not assuming here that the population has a Normal distribution).

(a) Show that the sample mean $\overline{X} = (X_1 + \cdots + X_n)/n$ has expected value μ (and so $\overline{X}$ is always an unbiased estimator of the population mean).

(b) Show also that $\overline{X}$ has variance σ^2/n, (and so $\overline{X}$ always has mean square error equal to σ^2/n as an estimator of μ, where μ denotes the population mean and σ^2 denotes the population variance).

(c) As part of the proof of Theorem 6.9(b) on the handout (you can look ahead to this and understand this part without us getting to this section in the lectures), we showed that for any n random variables $X_1, \ldots, X_n$ we have $\sum_{j=1}^n X_j^2 = \sum_{j=1}^n (X_j - \overline{X})^2 + n\overline{X}^2$. Starting from this result, show that, whatever the distribution of X, the sample variance $S^2 = \sum_{j=1}^n (X_j - \overline{X})^2/(n-1)$ has expected value σ^2 (and so S^2 is always an unbiased estimator of the population variance σ^2).

3. Let Y have a Gamma(α, λ) distribution. Show that $\mathrm{E}(Y) = \alpha/\lambda$, and show that, for $\alpha > 1$, $\mathrm{E}(1/Y) = \lambda/(\alpha - 1)$. [Hint: Recall from your Probability 1 notes that $\int_0^\infty x^{a-1} e^{-bx} dx = \Gamma(a)/b^a$ for all $a > 0$ and $b > 0$.]

*4. Let $X_1, \ldots, X_n$ be a random sample of size n from the Exponential(θ) distribution. We found earlier that the maximum likelihood estimator of θ was $\hat{\theta}_{mle} = n / \sum_{i=1}^{n} X_i$.

(a) Find the moment generating function of $\sum_{i=1}^{n} X_i$. Hence show $\sum_{i=1}^{n} X_i$ has the Gamma(n, θ) distribution and state its mean.

(b) The population mean of the Exponential(θ) distribution is $1/\theta$ and the maximum likelihood estimator of this population mean is $1/\hat{\theta}$. Show that the maximum likelihood estimator has expected value $1/\theta$ (and so it is unbiased as as estimator of the population mean).

(c) Use the results of question 3 above to show that $\mathrm{E}(\hat{\theta}_{mle}) = \theta n/(n-1)$. Hence find the average error (i.e. the bias) of $\hat{\theta}_{mle}$ as an estimator of θ and show it is not an unbiased estimator of θ.