MATH11400 Statistics 1 2010–11

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

Solution Sheet 7

1. (a) Let us interpret the question as meaning 'from which point on the ruler is the *average* distance to 1, 2 and 11 minimised?'. That is, find θ such that $d_1(\theta) = (|1-\theta| + |2-\theta| + |11-\theta|)/3$ is as small as possible. You cannot use calculus to solve this, since this function of θ is not always differentiable. But note that for $\theta \in (1,2)$, $d_1(\theta)$ is decreasing, since $|1-\theta|$ is getting bigger at rate 1, while $|2-\theta|$ and $|11-\theta|$ are both getting smaller at rate 1. Similarly on $(2,11)$, $d_1(\theta)$ is increasing. In fact $d_1(\theta)$ is decreasing for all $\theta < 2$ and increasing for all $\theta > 2$, so the minimum is at $\theta = 2$. It is not accident that 2 is the *median* of 1,2 and 11. In fact, it is a general result that for any odd n, the quantity $\{|x_i - \theta|\}$ is minimised when θ is the median of $\{x_i\}$. For even n, the function is constant between the middle two data values. These can be readily proved by generalising the argument for $n = 3$ given above.

   (b) Now we want to minimise $(|1-\theta|^2+|2-\theta|^2+|11-\theta|^2)/3$, or in general $(1/n)\sum_{i=1}^{n}(x_i-\theta)^2$, which we denote $(d_2(\theta))^2$. This is a *least squares* estimate. You can use calculus, and it is easy to see that $d_2(\theta)$ is minimised when θ is the mean $\overline{x}$ of the data values, or $14/3$ for the flea's $n = 3$ example.

   (c) In this 3rd version, we want to minimise $d_\infty(\theta) = \max\{|1-\theta|, |2-\theta|, |11-\theta|\}$, or in general $\max\{|x_i - \theta|, i = 1, 2, \ldots, n\}$. It is easy to see that the value of θ making this as small as possible is the one in the middle of the interval spanned by the data, i.e. $(x_{(1)}, x_{(n)})$, i.e. the *mid-range*, $(x_{(1)} + x_{(n)})/2$, or 6 for the flea's problem.

   (d) All of these three versions of the question ask us to define the 'centre' of the set of numbers, using different criteria. They are all examples of a general family of location estimators, those minimising $d_p(\theta) = ((1/n)\sum_{i=1}^{n}|x_i - \theta|^p)^{(1/p)}$. You can check that this definition agrees with those above when $p = 1, 2$ and $\to \infty$. All of these have some use in statistics – which is best to use depends on the statistical properties of the population from which the data can be assumed to be drawn.

2. Let $X_1, \ldots, X_n$ be a random sample of size n from a distribution with population mean denoted by $\mu = \mathrm{E}(X)$ and population variance denoted by $\sigma^2 = \mathrm{Var}(X)$, and let $\overline{X} = (X_1 + \cdots + X_n)/n$ denote the sample mean.

   (a) From your notes, the bias of $\overline{X}$ as an estimator of μ is defined as $\mathrm{E}(\overline{X} - \mu)$, and the mean square error of $\overline{X}$ as an estimator of μ is defined as $\mathrm{E}[(\overline{X} - \mu)^2]$.

$$
\begin{aligned}
\text{Now} \quad \mathrm{E}(\overline{X}) &= \mathrm{E}((X_1 + \cdots + X_n)/n) = \mathrm{E}(X_1/n) + \cdots + \mathrm{E}(X_n/n) \\
&= \mathrm{E}(X_1)/n + \cdots + E(X_n)/n = \mu/n + \cdots + \mu/n = n\mu/n \\
&= \mu.
\end{aligned}
$$

Thus, whatever the distribution of X, $\mathrm{E}(\overline{X} - \mu) = \mathrm{E}(\overline{X}) - \mu = \mu - \mu = 0$, so $\overline{X}$ has zero bias as an estimator of μ. We say $\overline{X}$ is unbiased as an estimator for the population mean.

(b) Also

$$
\begin{aligned}
\mathrm{Var}(\overline{X}) &= \mathrm{Var}((X_1 + \cdots + X_n)/n) \\
&= \mathrm{Var}(X_1/n) + \cdots + \mathrm{Var}(X_n/n) \text{ as the } X_i \text{ are independent} \\
&= \mathrm{Var}(X_1)/n^2 + \cdots + \mathrm{Var}(X_n)/n^2 \\
&= \sigma^2/n^2 + \cdots + \sigma^2/n^2 = n\sigma^2/n^2 \\
&= \sigma^2/n.
\end{aligned}
$$

Since $\mathrm{E}(\overline{X}) = \mu$, from above, $\mathrm{E}[(\overline{X} - \mu)^2] = \mathrm{Var}(\overline{X}) = \sigma^2/n$.

(c) Let $X_1, \ldots, X_n$ be a random sample of size n from a distribution with population mean denoted by the $\mu = \mathrm{E}(X)$ and population variance denoted by the $\sigma^2 = \mathrm{Var}(X)$. From notes, the condition for the sample variance $S^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$ to be an unbiased estimator of the population variance σ^2 is that $\mathrm{E}(\sum_{i=1}^n (X_i - \overline{X})^2/(n-1) - \sigma^2) = 0$.

From the handout $\sum_{i=1}^n (X_i - \mathrm{E}(X))^2 = \sum_{i=1}^n X_i^2 - n\overline{X}^2$.

But $\mathrm{E}(X_i^2) = \mathrm{Var}(X_i) + [\mathrm{E}(X_i)]^2 = \sigma^2 + \mu^2$

and $\mathrm{E}(\overline{X}^2) = \mathrm{Var}(\overline{X}) + [\mathrm{E}(\overline{X})]^2 = \sigma^2/n + \mu^2$ from question 3 above, so

$$
\begin{aligned}
\mathrm{E}(\textstyle\sum_{i=1}^n (X_i - \overline{X})^2) &= \mathrm{E}(\textstyle\sum_{i=1}^n X_i^2) - \mathrm{E}(n\overline{X}^2) = \textstyle\sum_{i=1}^n \mathrm{E}(X_i^2) - n\,\mathrm{E}(\overline{X}^2) \\
&= \textstyle\sum_{i=1}^n (\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2) \\
&= n\sigma^2 + n\mu^2 - n\sigma^2/n - n\mu^2 = n\sigma^2 - \sigma^2 \\
&= (n-1)\sigma^2
\end{aligned}
$$

so $\mathrm{E}(\sum_{i=1}^n (X_i - \overline{X})^2/(n-1)) = \sigma^2$ and $\sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$ is unbiased as an estimator for σ^2.

3. The hint in the question reminds us that $\int_{x=0}^{\infty} x^{a-1} e^{-bx} dx = \Gamma(a)/b^a$ for $a > 0$ and $b > 0$.

$$
\begin{aligned}
\text{Now}\,\mathrm{E}(Y) &= \textstyle\int_{-\infty}^{\infty} x f(x) dx \\
&= \textstyle\int_{x=0}^{\infty} x \lambda^\alpha x^{\alpha-1} e^{-\lambda x} (\Gamma(\alpha))^{-1} dx = \lambda^\alpha (\Gamma(\alpha))^{-1} \int_{x=0}^{\infty} x^\alpha e^{-\lambda x} dx \\
&= \lambda^\alpha (\Gamma(\alpha))^{-1} \Gamma(\alpha+1)/\lambda^{\alpha+1} \text{from above with } a = \alpha + 1 > 0 \text{ and } b = \lambda > 0 \\
&= \lambda^\alpha/\lambda^{\alpha+1} \times \Gamma(\alpha+1)/\Gamma(\alpha) = \alpha/\lambda \text{ as } \Gamma(\alpha+1) = \alpha\Gamma(\alpha).
\end{aligned}
$$

$$
\begin{aligned}
\text{and}\,\mathrm{E}(1/Y) &= \textstyle\int_{-\infty}^{\infty} x^{-1} f(x) dx = \int_{x=0}^{\infty} x^{-1} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} (\Gamma(\alpha))^{-1} dx \\
&= \lambda^\alpha (\Gamma(\alpha))^{-1} \textstyle\int_{x=0}^{\infty} x^{\alpha-2} e^{-\lambda x} dx \\
&= \lambda^\alpha (\Gamma(\alpha))^{-1} \Gamma(\alpha-1)/\lambda^{\alpha-1} \text{from above with } a = \alpha - 1 > 0 \text{ and } b = \lambda > 0 \\
&= \lambda^\alpha/\lambda^{\alpha-1} \times \Gamma(\alpha-1)/\Gamma(\alpha) = \lambda/(\alpha-1) \text{ as } \Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1).
\end{aligned}
$$

4. (a) From your notes §6.1(ii) the Exponential(θ) distribution has moment generating function $\mathcal{M}_X(t) = \theta/(\theta - t)$ (defined for $t < \theta$). If $X_1, \ldots, X_n$ is a random sample from this distribution then, from your notes §6.1(v), it has moment generating function $[\mathcal{M}_X(t)]^n = \theta^n/(\theta - t)^n$, and from your notes §6.1(ii) this is the moment generating function of the Gamma(n, θ) distribution. Thus, if $X_1, \ldots, X_n$ is a random sample from the Exponential(θ) distribution, then $\sum_{i=1}^n X_i \sim$ Gamma(n, θ). Let $Y = \sum_{i=1}^n X_i$, then, from question 6 above, $\mathrm{E}(Y) = n/\theta$ and $\mathrm{E}(1/Y) = \theta/(n-1)$. Note that for a random variable Y it is generally not true that $\mathrm{E}(1/Y) = 1/\mathrm{E}(Y)$.

   (b) Let $\tau = 1/\theta$. Then $\hat{\theta}_{mle} = n/Y$ and $\hat{\tau}_{mle} = \tau(\hat{\theta}_{mle}) = 1/\hat{\theta} = Y/n$. Thus $\mathrm{E}(\hat{\tau}_{mle}) = \mathrm{E}(Y/n) = \mathrm{E}(Y)/n = n/\theta n = 1/\theta = \tau$, so $\hat{\tau}_{mle}$ is unbiased as as estimator of the population mean.

   (c) Now $\hat{\theta}_{mle} = n/Y$ so $\mathrm{E}(\hat{\theta}_{mle}) = \mathrm{E}(n/Y) = n\,\mathrm{E}(1/Y) = n\theta/(n-1) = \theta + [\theta/(n-1)]$. Thus $\mathrm{E}(\hat{\theta}_{mle}) - \theta = \theta/(n-1)$ so $\hat{\theta}_{mle}$ has bias $\theta/(n-1)$ as an estimator of θ, i.e. $\hat{\theta}_{mle}$ on average systematically overestimates θ by an amount $\theta/(n-1)$.