DATA

Q1

7SO1 (2007)

a) i) $E(X) = \sum_i x_i p(x_i)$, which is well defined providing the sum converges (this needs checking if X takes values in an infinite set)

ii) $E[\phi(X)] = \sum_i \phi(x_i) p(x_i)$

iii) $E(aX+b) = \sum_i (ax_i + b) p(x_i)$

$$= a\underbrace{\sum_i x_i p(x_i)}_{= E(X)} + b\underbrace{\sum_i p(x_i)}_{=1 \text{ since } p(\cdot) \text{ is a pmf}}$$

$$= \underline{\underline{aE(X) + b}} \text{ as required}$$

b) If X is continuous, replace $\sum$ by $\int$ & $p(x)$ by $f(x)dx$: $E(X) = \int_{\mathbb{R}} x f(x) dx$

c) i) $Var(X) = E(X-\mu)^2 = E(X^2 - 2\mu X + \mu^2)$

$$= E(X^2) - 2\mu \underbrace{E(X)}_{=\mu} + \mu^2$$

$$= \underline{\underline{E(X^2) - \mu^2}}, \text{ as required}$$

ii) $Var(aX+b) = E(aX + b - E(aX+b))^2$

and $E(aX+b) = a\mu + b$

$\Rightarrow Var(aX+b) = E(aX - a\mu)^2$

$$= a^2 E(X-\mu)^2$$

$$= \underline{\underline{a^2 Var(X)}} \text{ as required}$$

2/

a) Let $\{E_i\}$ be a partition of a sample space Ω (i.e. a collection of mutually exclusive subsets with $\bigcup_i E_i = \Omega$), and let F be any event. Then

$$P(F) = \sum_i P(F|E_i)P(E_i)$$

Proof: $F = F \cap \Omega = F \cap (\bigcup_i E_i)$
$= \bigcup_i (F \cap E_i)$ (distributive laws)

and, since the $\{E_i\}$ are mutually exclusive, so are the sets in the collection $\{F \cap E_i\}$. Hence, by Kolmogorov's third axiom,

$$P(F) = \sum_i P(F \cap E_i)$$

But from the definition of conditional probability,

$P(F|E_i) = P(F \cap E_i)/P(E_i)$

$\Rightarrow$ $P(F \cap E_i) = P(F|E_i)P(E_i)$. Result follows.

b) i) Let E be the event "mine is inert", and F_n the event "unexploded after being hit n times". Then $P(F_n) = P(F_n|E)P(E) + P(F_n|E^c)P(E^c)$ (L of T P)

$= (1 \times p) + ((1-\alpha)^n(1-p))$

$= p + (1-\alpha)^n(1-p)$

ii) $P(F_{n+1}|F_n) = P(F_{n+1} \cap F_n)/P(F_n) = P(F_{n+1})/P(F_n)$

$= [p + (1-\alpha)^{n+1}(1-p)] / [p + (1-\alpha)^n(1-p)]$

If α is 0 or 1, this is always equal to 1; otherwise, both $(1-\alpha)^n$ & $(1-\alpha)^{n+1}$ tend to zero as $n \to \infty$, and the limiting probability is 1.

T.O.

DATA

3/

a) For a pdf we require (i) $f(x) \geq 0$ everywhere (trivial) (ii) $\int_{\mathbb{R}} f(x)\,dx = 1$.

Here, $\int_{\mathbb{R}} f(x)\,dx = \int_0^\infty \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}\,dx$

$= (u = \alpha x^\beta,\ du = \alpha\beta x^{\beta-1}dx)\ \int_0^\infty e^{-u}\,du = 1$

$\Rightarrow$ $f(x)$ is a pdf as required

The distribution function is

$F(x) = \int_0^x f(t)\,dt = $ (same transformation as before)

$[-e^{-u}]_0^{\alpha x^\beta}$

$= 1 - e^{-\alpha x^\beta}$ $(x > 0)$

$F(x) = 0$ $(x \leq 0)$

b) If $Y = X^\gamma$ & $\gamma > 0$, then Y takes values in $(0, \infty)$. For y in this range,

$P(Y \leq y) = P(X \leq y^{1/\gamma}) = F_X(y^{1/\alpha})$

$= 1 - e^{-\alpha y^{\beta/\alpha}}$, using the result from part (a).

But this just the distribution function of a Weibull distribution with parameters α & β/γ

(or use transformation of variables formula:

$f_Y(y) = f_X(x)\,|dx/dy|$

but this is messier here)

ATA

4/

a) Let X be DOC concentration; then $X \sim N(6, 1.5^2)$.

$P(X > 8) = P\left(Z > \frac{8-6}{1.5}\right) = P(Z > 1.333)$ where $Z \sim N(0,1)$

$= 1 - P(Z \le 1.333) = 1 - 0.909$ (Table 4)

$= \underline{\underline{0.091}}$

b) Let τ be required concentration; then $P(X > \tau) = 0.01$.

$P\left(Z > \frac{\tau - 6}{1.5}\right) = 0.01 \Rightarrow \frac{\tau - 6}{1.5} = 2.3263$ (Table 5)

Hence $\underline{\underline{\tau = 9.49 \text{ ppm}}}$

c) Let Y be # of samples with DOC concentration > 8 ppm.

Then $Y \sim Bin(5, 0.091)$.

$P(Y \ge 2) = 1 - P(Y=0) - P(Y=1)$

$= 1 - (0.909)^5 - 5(0.091)(0.909)^4$

$= \underline{\underline{0.0687}}$

d) In this case we're sampling without replacement, hence Y now has a hypergeometric distribution.

$P(Y \ge 2) = 1 - P(Y=0) - P(Y=1)$

$= 1 - \frac{\binom{4}{0}\binom{16}{5}}{\binom{20}{5}} - \frac{\binom{4}{1}\binom{16}{4}}{\binom{20}{5}}$

$= 1 - \frac{4368 + 7280}{15504} = \underline{0.2487}$

URfol

DATA

5/

a) The differences (Test 1 - Test 2) are

Student	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Difference	-21	-1	-8	16	7	-7	9	14	-13	38	12	-2	16	2	4

Stem & Leaf plot:

-2	1
-1	3
-0	1 2 7 8
0	2 4 7 9
1	2 4 6 6
2	
3	8

Leaf unit = 1%

(for full marks, require statement of units, and also vertical alignment of columns)

b) Since the marks are obtained from the same set of students each time, a paired t-test is appropriate. Let μ_1 & μ_2 be the means of the distributions from which the two sets of marks are drawn. Then, under the null hypothesis $H_0: \mu_1 = \mu_2$, the test statistic

$$t = \frac{\bar{d}}{S_d/\sqrt{n}}$$

has a t distribution with $(n-1)$ degrees of freedom, providing the observed differences are drawn independently and the underlying distribution of those differences is normal.

In the formula above, $\bar{d}$ and S_d are the sample mean & standard deviation of the differences:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \quad \text{and} \quad S_d = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} d_i^2 - n\bar{d}^2\right)}$$

5 (contd)

and n is the sample size (15 here).
The upper and lower 2.5% points of the t_{14} distribution are ± 2.145 (Table 10); hence we will reject H_0 if the observed value of $|t|$ is greater than 2.145, and accept otherwise.

For the data given, $\sum_{i=1}^{15} d_i = 66$ and $\sum_{i=1}^{15} d_i^2 = 3174$.

Hence $\bar{d} = 4.4$ and $s_d^2 = \frac{1}{14}(3174 - 290.4) = 206.0$

So observed value of t is $\frac{4.4}{\sqrt{206/15}} = 1.187$

Since $|1.187| < 2.145$, we accept H_0 and conclude there is no evidence (at the 5% level) for a difference in the underlying means.

Regarding the assumptions: the assumption of independence should be OK if the students didn't copy from each other, and the stem-&-leaf plot shows that normality probably is OK, although the 38 may be a bit of an outlier.

(+ any other intelligent / aware comments)

DATA

61

a) i) Unbiased if $\underline{E(T_n) = \theta}$

ii) Consistent: $\underline{\text{EITHER}}$ if $\lim_{n\to\infty} P(|T_n - \theta| < \varepsilon) = 1$ for any $\varepsilon > 0$, $\underline{\text{OR}}$ if $E(T_n) = \theta$ and $\lim_{n\to\infty} Var(T_n) = 0$.

(definitions aren't exactly equivalent, but either is acceptable).

b) i) $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n\mu = \mu$

$\Rightarrow$ $\underline{\bar{X} \text{ is unbiased for } \mu, \text{ as required}}$

$Var(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \underline{\sigma^2/n}$ as req^d.

ii) $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[(X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}-\mu) + (\bar{X}-\mu)^2\right]$$

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

But $\sum_{i=1}^{n}(X_i-\mu) = n(\bar{X}-\mu)$, hence

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

Taking expectations, we have

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

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