MATHM252 EXAM — SUMMER 2003

Solutions

1) a) $P(A \cap B) = P(B|A)P(A) = \underline{1/12}$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 1/4 + P(B) - 1/12$$
$$= \underline{1/6 + P(B)} \text{ as req}^{\text{d}}$$

b) Must have $0 \le P(A \cup B) \le 1$ since it's a probability — hence $\underline{P(A \cup B) \le 1}$.

Also, since $A \subseteq A \cup B$ we must have $P(A) \le P(A \cup B)$ so that $\underline{P(A \cup B) \ge P(A) = 1/4}$.

$$1/4 \le P(A \cup B) \le 1 \Rightarrow 1/4 \le 1/6 + P(B) \le 1$$
$$\text{so } \underline{1/12 \le P(B) \le 5/6}$$

c) i) If disjoint, $P(A \cup B) = P(A) + P(B)$, which doesn't hold if $P(A \cup B) = 1/2$
$\Rightarrow$ not disjoint.

ii) For independence, $P(A \cap B) = P(A)P(B)$.
Now $P(A \cup B) = 1/6 + P(B) = 1/2$, so $P(B) = 1/3$.

Then $P(A)P(B) = 1/4 \times 1/3 = 1/12 = P(A \cap B)$
so $\underline{A \text{ \& } B \text{ are independent}}$.

2) a) (i) $\underline{\binom{13}{3}\binom{7}{3} / \binom{20}{6}}$ $(= 0.258)$

(ii) P(all the same colour) = P(all red) + P(all black)

$$= \frac{\binom{13}{6}\binom{7}{0} + \binom{13}{0}\binom{7}{6}}{\binom{20}{6}} = \underline{0.044}$$

(iii) P(5^{th} is third red) = P(2 reds in first 4, then red)

$$= \frac{\binom{13}{2}\binom{7}{2}}{\binom{20}{4}} \times \frac{5}{16} = \underline{0.106}$$

b) Let Y be # of repetitions; then $Y \sim \text{Geo}(0.258)$.

i) $P(Y=3) = (1-0.258)^2 \times 0.258 = \underline{0.142}$

ii) $P(Y \leq 5) = 1 - P(Y > 5)$
$= 1 - P(\text{no "successes" in 5 repetitions})$
$= 1 - (1-0.258)^5 = \underline{0.775}$

iii) $P(10 \leq Y \leq 20) = P(Y \leq 20) - P(Y \leq 9)$
$= P(Y > 9) - P(Y > 20)$
$= 0.742^9 - 0.742^{20}$
$= \underline{0.066}$

Median, m say, satisfies $P(Y > m) = 1/2$. Now $P(Y > m) = 0.742^m$. Putting $0.742^m = 1/2$ yields $m = \ln 1/2 \,/\, \ln 0.742 = 2.32$. But this is not a value that Y can take. It shows, however, that $P(Y \geq 3) = P(Y > 2) > 1/2$, and that $P(Y > 3) < 1/2$. Hence the median is 3.

3) a) i) Let X be # of wet days; then $X \sim Bin(31, 0$

$$P(X=10) = \binom{31}{10} 0.4^{10} 0.6^{21} = \underline{0.102}$$

ii) $P(X=10$, with no wet days in first week$)$
$= P(0$ wet days out of $7) \times P(10$ out of $24)$
(by independence)

$$= 0.6^{7} \times \binom{24}{10} 0.4^{10} 0.6^{14} = \underline{0.0045}$$

b) The distribution is $\Gamma(2, 1/5)$; hence mean is $2/{}^{1}/_{5} = \underline{10\text{mm}}$

(or derive from pdf)

Total rainfall in month is $\sum_{i=1}^{31} I_i Y_i$, where

I_i is an indicator for rain on day i & Y_i is drawn from the given distribution.
Then expected rainfall is $E\left(\sum_{i=1}^{31} I_i Y_i\right)$

$= \sum_{i=1}^{31} E(I_i Y_i) = \sum_{i=1}^{31} E(I_i)E(Y_i)$ [ASSUMING INDEPENDENCE OF I_i & Y_i] $= 31 \times 0.4 \times 10$

$= \underline{124\text{mm as required.}}$

(2 for "$0.4 \times 31 \times 10$")

c) $$P(Y_i > 20) = \int_{20}^{\infty} \frac{x}{25} e^{-x/5}\, dx$$

u ↑ ↖ du/dx

$$= \left[-\frac{x}{5} e^{-x/5}\right]_{20}^{\infty} + \frac{1}{5}\int_{20}^{\infty} e^{-x/5}\, dx$$

$$= 4e^{-4} + \left[-e^{-x/5}\right]_{20}^{\infty} = 5e^{-4} = 0.0$$

Counting a "success" as a day with more than 20mm of rain, and assuming that wet day rainfall amounts are mutually independent, the number of days with >20mm of rain follows a Bin(31, 0.092) distribution.

4) a) $P(B|A) = \dfrac{P(A|B)P(B)}{P(A)}$ providing $P(A) > 0$

Not required.

b) Let X_i be stem length of a type i plant.

P(type I plant correctly classified)

$= P(X_I \leq 55) = P\left(Z \leq \dfrac{55-50}{3}\right)$

where $Z \sim N(0,1)$

$= P(Z \leq 1.667) = 0.95$ (tables)

P(type II plant correctly classified)

$= P\left(Z \geq \dfrac{55-60}{5}\right) = P(Z \geq -1)$

$= P(Z \leq 1)$ (symmetry) $= 0.84$

ii) $P(\text{misclassified}) = P(\text{misclass} \mid \text{Type I})P(\text{Type I}) + P(\text{misclass} \mid \text{Type II})P(\text{Type II})$

$= [(1-0.95) \times 0.25] + [(1-0.84) \times 0.75]$

$= 0.1325$

iii) P(Type I | classified as Type I)

$= \dfrac{P(\text{classified as I} \mid \text{Type I})P(\text{Type I})}{P(\text{classified as I})}$ (Bayes)

$= \dfrac{P(\text{classified as I} \mid \text{Type I})P(\text{Type I})}{P(\quad " \quad) + P(\text{classified as I} \mid \text{Type II})P(\text{Type II})}$

$= 0.95 \times 0.25 / (0.95 \times 0.25 + 0.16 \times 0.75) = 0.664$

c) Misclassification prob for Type I is now $P\left(Z > \frac{\alpha - 50}{3}\right)$

" " " " II " " $P\left(Z < \frac{\alpha - 60}{5}\right)$

$= P\left(Z > \frac{60-\alpha}{5}\right)$ by symmetry.

For equality, require $\frac{\alpha - 50}{3} = \frac{60 - \alpha}{5}$

$5\alpha - 250 = 180 - 3\alpha$

$8\alpha = 430$, $\underline{\alpha = 53.75 \text{ mm}}$

5) a) $\hat{\mu}_1 = \bar{x} = \frac{1}{11}\sum_{i=1}^{11} x_i = \underline{0.455}$

$\hat{\mu}_2 = \bar{y} = \frac{1}{14}\sum_{i=1}^{14} y_i = \underline{-21.79}$

$\hat{\sigma}_1^2 = s_1^2 = \frac{1}{10}\left(\sum_{i=1}^{11} x_i^2 - 11\bar{x}^2\right) = \underline{1367.3}$ (calculator)

$\hat{\sigma}_2^2 = s_2^2 = \underline{2106.2}$

b) Under $H_0 : \sigma_1^2 = \sigma_2^2$ (and assuming normal distributions), the test statistic

$$F = S_1^2 / S_2^2$$

follows an $F_{10,13}$ distribution. The upper 2.5% point of this is 3.250 (tables). The upper 2.5% point of $F_{13,10}$ is approximately $3.621 + (3.621 - 3.365)/12 = 3.642$ (interpolation from tables), so the lower 2.5% point of $F_{10,13}$ is $1/3.642 = 0.275$. 0.277

3.600

The observed value of F is $1367.3/2106.2 = 0.649$ which lies within the acceptance region $\Longrightarrow$ no evidence that variances differ.

c) CI is

$$\bar{x} - \bar{y} \pm \left(\underbrace{t_{14+11-2}(0.025)}_{2.5\% \text{ point of } t_{23} = 2.069} \times \hat{s.e.}(\bar{x} - \bar{y})\right)$$

Now $\hat{s.e.}(\bar{x} - \bar{y}) = S_p\sqrt{\frac{1}{11} + \frac{1}{14}}$, where $s_p^2 = \dfrac{10s_1^2 + 13s_2^2}{23} = 1784.9$.

Hence $\hat{s.e.}(\bar{x}-\bar{y}) = 17.02$, and the CI is

$$0.455 + 21.79 \pm (2.069 \times 17.02)$$
$$= \underline{(-12.97,\ 57.46)}$$

d) The CI above includes zero, & hence there is no evidence at the 95% level that the means of the groups are different. Combined with the result from (b), there is therefore little evidence that the treatment has any effect on weight increase.

The additional assumptions here are that observations within groups are independent (probably reasonable providing, e.g., no siblings within groups) and that both groups' observations are drawn from normal distributions. The latter is questionable, particularly since the observations have clearly been rounded to the nearest 5 (& sometimes to the nearest 10).

6) a) $E[e^{aX}] = \sum_{k=0}^{\infty} e^{ak} P(X=k)$

$= \sum_{k=0}^{\infty} e^{ak} \frac{m^k e^{-m}}{k!}$

$= e^{-m} \sum_{k=0}^{\infty} (me^a)^k / k!$

$= e^{-m} \cdot e^{me^a} = \underline{\underline{e^{m(e^a-1)}}}$ as req^d.

6) (i) $S_n \sim Poi(n\mu)$

Hence $E(e^{-n^{-1}S_n}) = \exp[n\mu(e^{-n^{-1}}-1)]$

(result from part (a) with $m = n\mu$, $a = -n^{-1}$)

and $E(T) = 1 - \exp[n\mu(e^{-n^{-1}}-1)] \neq p$.

<u>So T is biased for p, as required.</u>

Since $n\mu(e^{-n^{-1}}-1) = n\mu(-n^{-1} + \frac{n^{-2}}{2!} - \cdots)$

$= -\mu + \frac{\mu}{2n} - \cdots$,

then as $n \to \infty$ we must have $n\mu(e^{-n^{-1}}-1) \to -\mu$; hence $E(T) \to 1 - e^{-\mu} = p$, <u>as required</u>

(ii) $Y \sim Bin(n,p)$.

$E(Y) = np$ and $Var(Y) = np(1-p)$.

Hence $E(Y/n) = np/n = p$, and

<u>Y is unbiased for p, as required.</u>

$S.e.(Y/n) = \sqrt{Var(Y/n)}$

$= \underline{\underline{\sqrt{\frac{p(1-p)}{n}}}}$

(iii) A 'good' estimator is one that has both a small bias and a small standard error. Therefore, although T is biased & Y/n is not, T may be preferable if its bias is small and its standard error is less than that of Y/n. To choose between the estimators we would therefore need to calculate the standard error of T i.e. $\sqrt{Var(T)}$. A convenient way to combine the bias and variance of an estimator is via the criterion of mean squared error:

$$MSE = bias^2 + variance.$$

In general, given the choice of two estimators we will prefer the one with the smaller MSE.