F71SB2 STATISTICS II

HWU MARCH 2008 - SOLUTIONS

① (a) Number of selections containing no set of twins:

$$|A| = \binom{5}{3}\binom{2}{1}\binom{2}{1}\binom{2}{1} = 80$$

Therefore,

$$\Pr\{\text{no pairs of twins}\} = \frac{|A|}{|S|} = \frac{\binom{5}{3}\binom{2}{1}\binom{2}{1}\binom{2}{1}}{\binom{10}{3}}$$

$$= \frac{80}{120} = \frac{2}{3}$$

(b) Event here is the complement of event in (a).

$$\Pr\{\text{one pair}\} = 1 - \Pr\{\text{no pairs}\} = 1 - \frac{2}{3} = \frac{1}{3}$$

② X ~~takes~~ values 1, 2, ..., 6 with probability 1/6.

$$E\{g(X)\} = \sum_{x=1}^{6} (2x^2+1)\frac{1}{6}$$

$$= \frac{1}{6}\Big\{(2\times1+1) + (2\times4+1) + (2\times9+1) + (2\times16+1) + (2\times25+1) + (2\times36+1)\Big\}$$

$$= \frac{1}{6}(3 + 9 + 19 + 33 + 51 + 73)$$

$$= \frac{188}{6} = 31.333$$

3

A study on the number of accidents per day, occurring on a busy street is conducted. A random sample of the number of accidents for 100 days is collected, and is given in the following table.

Number of accidents per day (x) :	0	1	2	3	4	≥ 5
Number of days (f_x) :	23	38	27	11	1	0

It is believed that the number of accidents per day, X, follows a Poisson(λ) distribution.(*Recall that for this distribution* $\mathrm{E}(X) = \lambda$.)

(a) Apply the method of moments to estimate the parameter of the Poisson distribution λ.

$$\bar{x} = \frac{1}{100}\left\{38 + 27\times 2 + 11\times 3 + 4\right\} = \frac{129}{100} = 1.29$$

For MME we need $E(X) = \bar{x}$

$$\Rightarrow \quad \hat{\lambda} = 1.29$$

(b) Calculate the expected frequencies for the ~~number of accidents~~ per day under the assumed distribution and using your estimate of λ from above.

Use $P(X=k) = \dfrac{e^{-\hat{\lambda}}\hat{\lambda}^k}{k!}$ $\quad \left[\text{and/or } P(X=k+1) = P(X=k)\dfrac{\hat{\lambda}}{k+1}\right]$

$P(X=0) = e^{-1.29} = 0.275 \Rightarrow E_0 = 100\times 0.275 = 27.5$

$P(X=1) = 0.275 \times 1.29 = 0.355 \Rightarrow E_1 = 35.5$

$P(X=2) = 0.355\,\dfrac{1.29}{2} = 0.229 \Rightarrow E_2 = 22.9$

$P(X=3) = 0.229\,\dfrac{1.29}{3} = 0.098 \Rightarrow E_3 = 9.8$

$P(X=4) = 0.098\,\dfrac{1.29}{4} = 0.032 \Rightarrow E_4 = 3.2$

$P(X\geq 5) = 1 - P(X\leq 4) = 1 - \sum_{k=0}^{4} P(X=k) = 0.011$

$$\Rightarrow E_{\geq 5} = 1.1$$

(c) Compare the expected numbers in each category, with the observed given in the table. Do you think that the observed data come from the assumed distribution? Give brief reasons for your answer.

Observed data look under-dispersed (observed frequencies lower at the tails and higher in the middle).

4 (a) We need

$$\int_{-1}^{5} f(x)\,dx = 1 \Rightarrow \int_{-1}^{5} \frac{1}{c}(x+1)\,dx = 1$$

$$\Rightarrow \frac{1}{c}\left[\frac{x^2}{2}+x\right]_{-1}^{5} = 1$$

$$\Rightarrow \frac{1}{c}\left(\frac{25}{2}+5-\frac{1}{2}+1\right) = 1 \Rightarrow \frac{18}{c} = 1$$

$$\Rightarrow \underline{c = 18}$$

(b) $$E(X) = \int_{-1}^{5} x f(x)\,dx = \int_{-1}^{5} x\frac{1}{18}(x+1)\,dx$$

$$= \int_{-1}^{5} \frac{1}{18}(x^2+x)\,dx = \frac{1}{18}\left[\frac{x^3}{3}+\frac{x^2}{2}\right]_{-1}^{5}$$

$$= \frac{1}{18}\left(\frac{125}{3}+\frac{25}{2}+\frac{1}{3}-\frac{1}{2}\right) = \frac{54}{18}$$

$$= \underline{3}$$

$$E(X^2) = \int_{-1}^{5} x^2 f(x)\,dx = \int_{-1}^{5} \frac{1}{18}(x^3+x^2)\,dx$$

$$= \frac{1}{18}\left[\frac{x^4}{4}+\frac{x^3}{3}\right]_{-1}^{5} = \frac{1}{18}\left(\frac{625}{4}+\frac{125}{3}-\frac{1}{4}+\frac{1}{3}\right)$$

$$= \frac{198}{18} = \underline{11}$$

$$\mathrm{var}(X) = E(X^2) - E^2(X) = 11 - 9 = \underline{2}$$

④ (c) $P(X>2) = \int_2^5 f(x)\,dx = \int_2^5 \frac{1}{18}(x+1)\,dx$

$= \frac{1}{18}\left[\frac{x^2}{2} + x\right]_2^5$

$= \frac{1}{18}\left(\frac{25}{2} + 5 - 2 - 2\right)$

$= \underline{0.75}$.

⑤ (a) $Z \sim N(0,1)$

(b) $P(77 < X < 81) = P\left(\frac{77-80}{4} < \frac{X-80}{4} < \frac{81-80}{4}\right)$

$= P(-0.75 < Z < 0.25)$

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

$= \Phi(0.25) - \Phi(-0.75)$

$= \Phi(0.25) - \{1 - \Phi(0.75)\}$

$= \Phi(0.25) + \Phi(0.75) - 1$

$= 0.5987 + 0.7734 - 1$

$= \underline{0.3721}$.

(c) $P(1 \le T \le 2) = P(T \le 2) - P(T \le 1)$

$= F_T(2) - F_T(1) = 1 - e^{-0.5\times 2} - 1 + e^{-0.5\times 1}$

$= 0.6065 - 0.3679 = \underline{0.2386}$

(6) (a) Median $= x_{\left(\frac{30+1}{2}\right)} = x_{(15.5)} = \underline{29}$

$Q_1 = x_{\left(\frac{30+1}{4}\right)} = x_{(7.75)} = \underline{18}$

$Q_3 = x_{\left(\frac{3(30+1)}{4}\right)} = x_{(23.25)} = \underline{43.25}$

$IQR = Q_3 - Q_1 = \underline{25.25}$

(b)

1	0 3 4
1	5 6 7 8 8 9
2	1 2 4
2	6 7 9 9
3	
3	6
4	0 0 0 1 2 3 4 4
4	5 6 6 9
5	0
5	

Distribution <u>does not</u> look normal (bimodal).

(7) (a) $E(\sum X_i) = E(X_1) + \cdots E(X_n) = \mu + \mu + \cdots + \mu$

$= n\mu$

$\text{var}(\sum X_i) = \text{var}(X_1) + \cdots + \text{var}(X_n)$ (<u>by independence</u>)

$= n\sigma^2$

from CLT: $Y = \sum X_i \overset{\text{approx}}{\sim} N(n\mu, n\sigma^2)$

(b) Let Y be the no. of heads

Then $Y \sim \text{Bin}(100, 0.5)$
and from CLT $Y \overset{\text{approx}}{\sim} N\left(100 \times 0.5, 100 \times 0.5^2\right)$

i.e. $Y \overset{\text{approx.}}{\sim} N(50, 25)$.

$P(Y > 60) \simeq P(Y > 60.5)$ (continuity correction)

$= P\left(Z > \frac{60.5 - 50}{5}\right)$ where $Z \sim N(0,1)$

$= 1 - \Phi(2.1) = 1 - 0.98214$
$= 0.0179$.

(8) (a) $\bar{x} = \frac{\sum x_i}{n} = \frac{56.7}{8} = 7.0875$

A 90% CI for μ is given by

$$\left(\bar{x} - 1.6449 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.6449 \frac{\sigma}{\sqrt{n}}\right)$$

i.e. $\left(7.0875 - 1.6449 \frac{0.5}{\sqrt{8}}, 7.0875 + 1.6449 \frac{0.5}{\sqrt{8}}\right)$

i.e. $(6.797, 7.378)$

(b) $s^2 = \frac{1}{n-1}\left(\sum x_i^2 - \left(\sum x_i\right)^2/n\right) = 0.298$, $s = 0.546$.

(c) A 90% CI for μ is now:

$\bar{x} \pm t_{7, 0.05} \frac{s}{\sqrt{n}} = 7.088 \pm 1.895 \frac{0.546}{\sqrt{8}}$

i.e. $(6.722, 7.453)$