F71SB2 STATISTICS II

MARCH 2006 — SOLUTIONS

① (a) $P(\text{no pair of twins}) = \dfrac{\binom{2}{1}\binom{2}{1}\binom{2}{1}}{\binom{8}{3}} \binom{4}{3}$

$= \dfrac{2^3 \times 4}{\frac{8!}{3!\,5!}} = \dfrac{4}{7}$

(b) $P(\text{one pair of twins}) = 1 - P(\text{no pair}) = \dfrac{3}{7}$

② (a) $F_X(x) = P(X \le x)$ for any possible value of r.v. X.

<u>Properties</u>: (i) $\lim_{x \to -\infty} F_X(x) = 0$

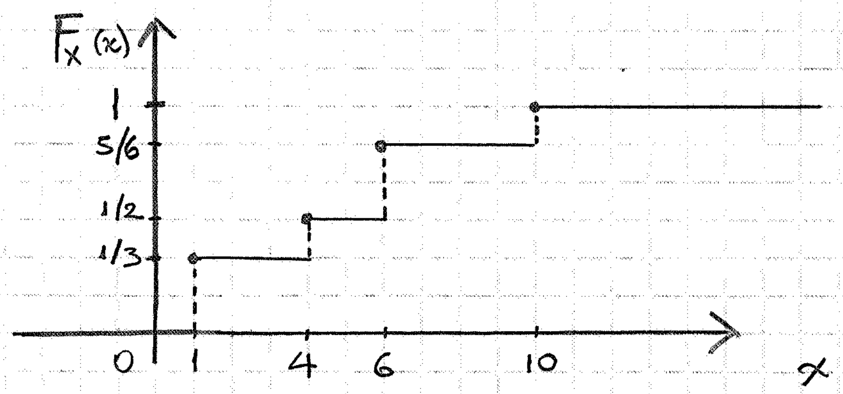
(ii) $\lim_{x \to \infty} F_X(x) = 1$

(iii) $F_X(x)$ is a non-decreasing function of x.

(b) (i) No, because $F_X(4) \neq 1$.

(ii) No, because $F_X(2) < F_X(1)$.

(c) (i)



(ii) $P(2 < X \le 6) = P(X \le 6) - P(X \le 1)$

$= 5/6 - 1/3 = \frac{1}{2}$

(iii) $P(X = 4) = P(X \le 4) - P(X \le 1)$

$= 1/2 - 1/3 = \frac{1}{6}$

3 (a) $\bar{x} = \frac{1}{100}\left(42 + 27\times2 + 20\times3 + 11\times4\right) = 2$

For M.M.E. we need $E(X) = \bar{x}$, i.e.

$$\frac{1}{\hat{p}} = \bar{x} \Rightarrow \hat{p} = \frac{1}{\bar{x}} \Rightarrow \underline{\underline{\hat{p} = 0.5}}$$

(b)

x:	1	2	3	4	≥ 5
$P(X=x)$:	0.50	0.25	0.125	0.0625	0.0625
$100 \times P(X=x)$:	50	25	12.5	6.25	6.25

(c) The model <u>does not</u> fit the data well.
Data appear to be <u>under-dispersed</u> compared to what would be expected under the geometric model.

4 (a) $E(X) = \int_0^1 x f(x)\,dx = \int_0^1 \theta x^{\theta}\,dx = \theta\left[\frac{x^{\theta+1}}{\theta+1}\right]_0^1$

$$= \underline{\underline{\frac{\theta}{\theta+1}}}$$

(b) For these data we have

$$\bar{x} = \frac{1}{5}\,4.06 = 0.812$$

For MME we need

$$E(X) = \bar{x} \Rightarrow \frac{\theta}{\theta+1} = \bar{x} \Rightarrow \hat{\theta} = \frac{\bar{x}}{1-\bar{x}}$$

$$\Rightarrow \hat{\theta} = \frac{0.812}{1-0.812}$$

$$= \underline{\underline{4.319}}$$

(5) (a) $Z \in (-\infty, \infty)$ and is an increasing function of X.

Also, $z = \frac{x-\mu}{\sigma} \Rightarrow x = \mu + \sigma z$, $\frac{dx}{dz} = \sigma$.

Then:

$$f_Z(z) = f_X(x) \frac{dx}{dz} = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(\mu + \sigma z - \mu)^2\right\} \cdot \sigma$$

$$= \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

This implies that $Z \sim N(0,1)$.

(b) i) $P(X > 85) = P\left(Z > \frac{85-80}{4}\right)$, with $Z \sim N(0,1)$

$$= P(Z > 1.25) = 1 - \Phi(1.25)$$
$$= 1 - 0.8944$$
$$= 0.1056$$

ii) $P(77 < X < 81) = P\left(\frac{77-80}{4} < Z < \frac{81-80}{4}\right)$, $Z \sim N(0,1)$

$$= P(Z < 0.25) - P(Z < -0.75)$$
$$= \Phi(0.25) - (1 - \Phi(0.75))$$
$$= \Phi(0.25) + \Phi(0.75) - 1$$
$$= 0.5987 + 0.7734 - 1$$
$$= 0.3721.$$

(6) (a) $x_{(1)} = 25.1$, $x_{(30)} = 36.3$

Median $= x_{\left(\frac{30+1}{2}\right)} = x_{(15.5)} = 31.6$

$Q_1 = x_{\left(\frac{30+1}{4}\right)} = x_{(7.75)} = 29.3 + \frac{3}{4}0.2 = 29.45$

$Q_3 = x_{\left(\frac{3(30+1)}{4}\right)} = x_{(23.25)} = 33.2 + \frac{1}{4}0.4 = 33.3$

(b)

25	1 5
26	
27	2 9
28	8 9
29	3 5 9
30	1 3 6 6
31	1 5 7 7 9
32	2 4 6 9
33	2 6 8
34	2 5 9
35	3
36	3

The distribution looks fairly symmetric and "bell" shaped.

Plot suggests data are Normal.

(7) (a) $P(T>3 \mid T>1) = P(T>2)$ from memoryless property

$= e^{-0.5\times 2} = e^{-1}$

$= 0.3679$

(b) $X \sim \text{Poisson}(0.5)$

$P(X=0) = \dfrac{e^{-0.5} \times 0.5^0}{0!} = e^{-0.5} = 0.6065$

(or, using $T \sim \text{Exp}(0.5)$).

(c) From CLT $\bar{T} \sim N\left(\frac{1}{0.5}, \frac{1}{25\times 0.5^2}\right)$ approximately,

i.e. $\bar{T} \sim N(2, 0.16)$ approx.

Then, $P(\bar{T}>1) = P\left(Z > \frac{1-2}{0.4}\right)$, $Z \sim N(0,1)$.

$= P(Z > -2.5) = \Phi(2.5) = 0.99379$

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

$$s^2 = \frac{1}{n-1}\left\{\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right\} = 0.298$$

(b) $s = \sqrt{0.298} = 0.546$

A 95% C.I. for μ_A is given by

$$\bar{x} \pm t_{7,0.025}\frac{s}{\sqrt{n}} = 7.088 \pm 2.365\,\frac{0.546}{\sqrt{8}}$$

i.e. $(6.631, 7.545)$

(c) If we construct a large number of such intervals (with different samples), we expect 95% of them to contain the true value μ.

(d) Since the interval does not contain the value 0, there is evidence that the mean length is different between the two populations.
Bacteria A appear to have larger mean length.