F71SB2 Statistics 2 Assignment 1

NAME (BLOCK LETTERS)..
I affirm that this assignment is my sole and original work.
SIGNATURE..
TUTORIAL CLASS: Day............Time............Room............

Due: 5 February 2007 at 4.30pm (use AMS course-work box, outside EM 1.25).

INSTRUCTIONS: Attempt ALL FOUR questions. To receive full credit you must **show your work** and **explain your answers**. There are 40 marks available.

1. Consider the following game: a fair die is rolled repeatedly and you win a prize if a '5' or a '6' is achieved. The outcomes of the rolls are independent from each other.

   (a) You decide to play this game until you win the prize for the first time.

   i. Identify fully the distribution of the number, X, of rolls until you win the prize. [1 *mark*]

   Probability of "success" is $\frac{1}{3}$. Therefore $X \sim \text{Geometric}(\frac{1}{3})$

   ii. Calculate the probability that more than 4 rolls will be required until you win the prize, given that you didn't win it with the first roll. [4 *marks*]

   $P(X>4/X>1) = P(X>3)$ [using lack-of-memory]

   $= P(\text{first 3 rolls "failures"})$

   $= (2/3)^3$ [by independence]

   $= 8/27$

   (b) Now consider that you decide to play this game until you win the prize twice.

   i. Identify fully the distribution of the number, Y, of rolls until you win twice. [1 *mark*]

   $Y \sim \text{Neg. Binomial}(2, \frac{1}{3})$.

   ii. Find the probability that the number of rolls required to win twice is less than 3. [4 *marks*]

   $P(Y<3) = P(Y\leq 2)$

   $= P(Y=2)$ [y can only take values 2,3,4,...]

   $= \frac{1}{3}\cdot\frac{1}{3}$ [independence]

   $= 1/9$

2. (a) Let X be a random variable with range $S_x = \{1, 2, 3, 4\}$. For each of the following determine whether the given values can serve as the values of the cumulative distribution function (*cdf*) of X, giving reasons for your answers:

i) $F_X(1) = 0.3, F_X(2) = 0.5, F_X(3) = 0.8, F_X(4) = 1.2$; [2 *marks*]

No. $F_X(4)$ must be equal to 1.

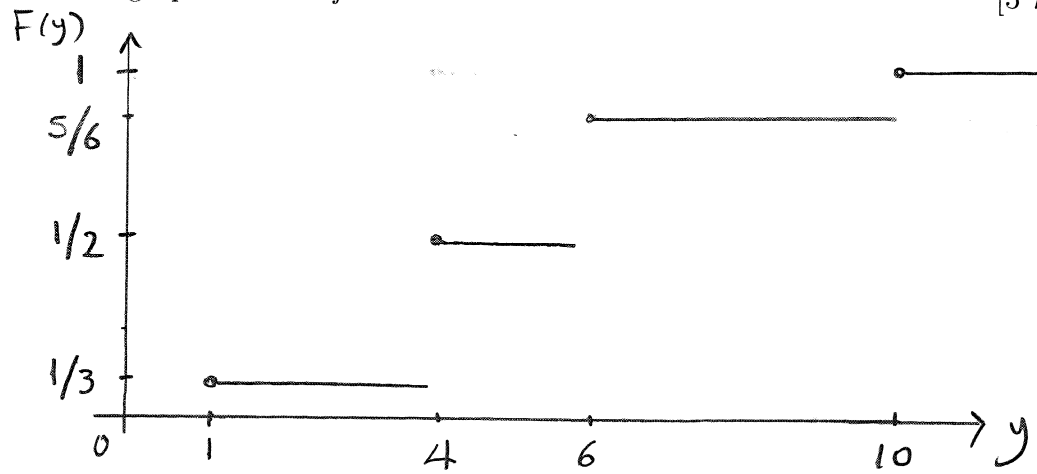
ii) $F_X(1) = 0.5, F_X(2) = 0.4, F_X(3) = 0.7, F_X(4) = 1.0$. [2 *marks*]

No. $F(x)$ must be non-decreasing.

(b) Now consider a new discrete random variable Y, with the following *cdf*:

$$F_Y(y) = \begin{cases} 0, & \text{for } y < 1 \\ \frac{1}{3}, & \text{for } 1 \leq y < 4 \\ \frac{1}{2}, & \text{for } 4 \leq y < 6 \\ \frac{5}{6}, & \text{for } 6 \leq y < 10 \\ 1, & \text{for } y \geq 10. \end{cases}$$

i) Draw the graph of the *cdf* of Y. [3 *marks*]



ii) Calculate $P(2 < Y \leq 6)$. [3 *marks*]

$$P(2 < Y \leq 6) = P(Y \leq 6) - P(Y \leq 1)$$
$$= F(6) - F(1) = \frac{5}{6} - \frac{1}{3}$$
$$= 1/2$$

iii) Calculate $P(Y = 4)$. [3 *marks*]

$$P(Y = 4) = P(Y \leq 4) - P(Y \leq 1)$$
$$= F(4) - F(1) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

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. [3 *marks*]

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

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. [4 *marks*]

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. [2 *marks*]

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

The assumed Poisson model might not be appropriate.

4. Suppose that the time, T (measured in hours), until you receive the next call on your mobile phone, follows an Exp(0.8) distribution.

   (a) Calculate the probability that the time until you receive a phone call is more than three hours, assuming that you have not received a call in the last one hour.

   [4 *marks*]

$P(T>3 / T>1) = P(T>2)$ [using lack-of-memory]

$= 1 - (1 - e^{-2\times 0.8})$

$= e^{-1.6} = \underline{0.202}$

   (b) As above, asssume that the times (measured in hours) between successive calls on your mobile phone are independently distributed as Exp(0.8) random variables. Consider a new random variable, Z, giving the number of calls on your mobile phone during a lecture that lasts 45 minutes. Identify fully the distribution of Z. (*Hint: first determine the distribution of the time between successive calls measured in 45-minute periods.*)

   [4 *marks*]

If Y is the time between calls in 45-minute periods then

$T = \frac{45}{60} Y \Rightarrow Y = \frac{4}{3} T$, with $T \sim \text{Exp}(0.8)$ (in hours).

Then, using the rescaling property

$Y \sim \text{Exp}\left(\frac{0.8}{4/3}\right)$ i.e. $\text{Exp}(0.6)$

It follows (using connection between exponential and Poisson) that

$\underline{Z \sim \text{Poisson}(0.6)}$

End of assignment