STATISTICS II

Problem sheet 2 - Solutions

(1) i) For any I_j we have $E(I_j) = p$, $E(I_j^2) = p$ so that $Var(I_j) = p - p^2 = p(1-p)$.

Now since $X = I_1 + I_2 + \cdots + I_n$ and the I_j are independent we have

$$E(X) = \sum_{j=1}^{n} E(I_j) \quad , \quad Var(X) = \sum_{j=1}^{n} Var(I_j)$$

$$= np \qquad\qquad = np(1-p)$$

ii) To calculate $E(X)$ & $Var(X)$ from the definition:

$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} x = \sum_{x=1}^{n} np \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} = np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y}$$

where $y = x - 1$.

Now $\sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} = (p + (1-p))^n = 1^n = 1$

$\Rightarrow E(X) = np$.

This calculation is tricky algebraically. To get $E(X^2)$ and hence $Var(X)$ by this method is even worse.

$$E(X^2) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} x^2 = \sum_{x=1}^{n} \binom{n}{x} p^x (1-p)^{n-x} x^2$$

$$= \sum_{x=1}^{n} \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x} x^2 = np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,(n-x)!} p^{(x-1)} (1-p)^{n-x} x$$

$$= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{(n-1)-y} (y+1) \qquad \text{where } y = x-1$$

$$= np \left\{ \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{(n-1)-y} \cdot y + \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{(n-1)-y} \right\}$$

Definition of mean of Bin $(n-1, p)$	Sum of probability function for Bin$(n-1, p)$

$$= np \left\{ (n-1)p + 1 \right\} = n^2 p^2 + np(1-p)$$

$$\Rightarrow Var(X) = E(X^2) - [E(X)]^2 = n^2p^2 + np(1-p) - n^2p^2$$

$$= np(1-p)$$

The moral is that it's easier to calculate the mean and variance by method i) or ii).

(2)

$X \sim Bin(10, 0.4)$.

i) $P(X<7) = P(X \le 6) = F_X(6) = 0.9452$

ii) $P(X \ge 3) = 1 - P(X \le 2) = 1 - F_X(2) = 1 - 0.1673$

$= 0.8327$

iii) $P(2 < X \le 5) = P(X \le 5) - P(X \le 2)$

$= F_X(5) - F_X(2)$

$= 0.8338 - 0.1673$

$= 0.6665$

(3)

Now let $X \sim Bin(10, 0.7)$. This value of p is greater than 0.5 so we can't read F_X directly from the statistical tables. Instead consider $Y = 10 - X$ (the number of failures).

Now $Y \sim Bin(10, 0.3)$ and we can express the probabilities i) – iii) in terms of events involving Y, and use tables.

i) $P(X<7) = P(Y>3) = 1 - P(Y \le 3) = 1 - F_Y(3)$

$= 1 - 0.6496 = 0.3504$

ii) $P(X \ge 3) = P(Y \le 7) = F_Y(7) = 0.9984$

iii) $P(2<X\le 5) = P(5 \le Y < 8) = P(4 < Y \le 7)$

$= F_Y(7) - F_Y(4) = 0.9984 - 0.8497 = 0.1487$

4

i) Assuming independence, the number of correct answes X should follow a $Bin(100, \frac{1}{6})$ distribution.

$$\Rightarrow \quad E(X) = 100 \times \frac{1}{6} = 16\tfrac{2}{3} \doteq 16.67$$

$$Var(X) = 100 \times \frac{1}{6} \times \frac{5}{6} = \frac{125}{9} \doteq 13.9$$

ii) To get the mean and variance of the score we can proceed in different ways.

a) Total score $Y = 2 \times X - 3(100 - X)$

$$= 5 \times X - 300$$

$$\Rightarrow \quad E(Y) = 5 \times E(X) - 300 = -216.67$$

$$Var(Y) = 25 \times Var(X) = 347.22$$

b) Alternatively we can work out the mean and variance of the student's score on a single question. (call this S).

$$E(S) = \frac{1}{6} \times 2 + \frac{5}{6} \times (-3) = -\frac{13}{6} = -2.167$$

$$E(S^2) = \frac{1}{6} \times 4 + \frac{5}{6} \times 9 = \frac{49}{6}$$

$$\Rightarrow \quad Var(S) = \frac{49}{6} - \left(\frac{13}{6}\right)^2 = 3.472$$

Now $Y = S_1 + S_2 \cdots + S_{100}$ is the sum of 100 identically independently distributed r.v's with this mean and variance

$$\left.\begin{aligned} \Rightarrow \quad Var\,Y &= \sum_{i=1}^{100} Var(S_i) = 100 \times 3.472 = 347.2 \\ E(Y) &= \sum^{100} E(S_i) = 100 \times (-2.167) = -216.7 \end{aligned}\right\} \text{as above.}$$

(5) (i) Recall from lectures that if X_i, $i=1,\dots,r$ are i.i.d Geometric(p) random variables, then

$$Y = X_1 + X_2 + \cdots + X_r \sim NBin(r, p)$$

Also, recall (stats I), that $E(X_i) = \frac{1}{p}$, $var(X_i) = \frac{1-p}{p^2}$, for all $i = 1, 2, \dots, r$.

It follows that

$$E(Y) = E\left(\sum_{i=1}^{r} X_i\right) = \sum_i E(X_i) = \sum_i \frac{1}{p} = \frac{r}{p}$$

$$var(Y) = Var\left(\sum_{i=1}^{r} X_i\right) = \sum_i var(X_i) \quad \left[\text{Since } X_i \text{ are independent}\right]$$

$$= \frac{r(1-p)}{p^2}$$

(ii) (a) We want the dist^n of no. of "trials" until two "successes" (in a sequence of independent trials).

Therefore,

$$X \sim NBin\left(2, \frac{1}{6}\right)$$

(b) $P(X>2) = 1 - P(X \leq 2) = 1 - P(X=2)$

[as X can only take values $x = 2, 3, 4, \dots$]

$$= 1 - \frac{1}{6}\frac{1}{6} = \frac{35}{36}$$

$$= 0.972$$

⑥ (i) X represents the no. of trials required to achieve the first success in a sequence of independent trials, each with probability p of success.

$$P(X>k) = P(\text{first } k \text{ trials result in "F"})$$
$$= (1-p)^k, \quad k=1,2,3,\ldots \quad \left[\text{because of } \underline{\text{independence}} \text{ of trials}\right]$$

(ii) Now, if X_i, $i=1,\ldots,n$ are i.i.d with $X_i \sim \text{Geometric}(p)$ and $Y = \min\{X_1,\ldots,X_n\}$, then

$$P(Y>k) = P\{(X_1>k)\cap(X_2>k)\cap\cdots\cap(X_n>k)\}$$
$$= P(X_1>k)\times P(X_2>k)\times\cdots\times P(X_n>k)$$
[because of independence].
$$= P(X>k)^n \quad \text{[since distributions are identical]}$$
$$= \{(1-p)^k\}^n = \underline{\underline{(1-p)^{nk}}}.$$

(iii) From (ii) we can write

$$P(Y>k) = q^k, \quad \text{where } q=(1-p)^n$$

and comparing this with the probability in (i), we can conclude that

$$Y \sim \underline{\underline{\text{Geometric }(1-q)}}, \quad \text{with } q=(1-p)^n$$

[Recall that 2 r.v.'s with the same CDF – i.e. $P(X\leq k)$ or equivalently $P(X>k)$ – have the same distribution.]

7

The p.m.f. is given by

$$P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \quad , \quad x = 0, 1, 2, \dots \quad , \quad \lambda > 0 .$$

Then

$$E(X) = \sum_{x=0}^{\infty} x f_X(x) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{e^{-\lambda} \cdot \lambda^x}{(x-1)!}$$

$$= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!} \qquad \left[\text{and setting } x-1 = y\right]$$

$$= \lambda \sum_{y=0}^{\infty} \frac{e^{-\lambda} \cdot \lambda^y}{y!} = \lambda \cdot 1 = \underline{\underline{\lambda}}$$

Also,

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda} \lambda^x}{x!} = \sum_{x=1}^{\infty} \lambda . x . \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!}$$

[and setting again $x-1=y$]

$$= \lambda \sum_{y=0}^{\infty} (y+1) \frac{e^{-\lambda} \lambda^y}{y!} = \lambda \sum_{y=0} y \frac{e^{-\lambda} \lambda^y}{y!} + \lambda \sum_{y=0} \frac{e^{-\lambda} \cdot \lambda^y}{y!}$$

$$= \lambda . \lambda + \lambda = \lambda^2 + \lambda$$

Then:

$$\operatorname{var}(X) = E(X^2) - E^2(X) = \lambda^2 + \lambda - \lambda^2$$

$$= \underline{\underline{\lambda}} .$$