Statistics - Exercises (Solutions) (9 pages; 7/4/20)

Binomial distribution

(1***) If $X \sim B(n, p)$, prove that:

(i) $E(X) = np$ (ii) $Var(X) = np(1-p)$

Solution

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

$= \sum_{x=1}^{n} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} x$ (as the 1st term vanishes)

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

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

Let $u = x - 1$ and $N = n - 1$

Then $E(X) = np \sum_{u=0}^{N} \frac{N!}{u!(N-u)!} p^u (1-p)^{N-u}$

$= np \sum_u P(X = u) = np$

(ii) Considering how x was made to cancel in (i), it will be easier to deal with $E[X(X-1)]$ than with $E(X^2)$.

We want to prove that $Var(X) = np(1-p)$.

LHS $= E(X^2) - \mu^2 = \mathrm{E[X(X-1) + X]} - \mu^2$,

so we need to prove that

$\mathrm{E[X(X-1) + X]} - \mu^2 = np(1-p)$;

ie that $\mathrm{E[X(X-1)]} = np(1-p) - E(X) + \mu^2$

$= np(1-p) - np + (np)^2$

$= (np)^2 - np^2 = np^2(n-1)$

Bearing this in mind,

$Var(X) = E(X^2) - \mu^2 = \mathrm{E}[\mathrm{X}(\mathrm{X}-1) + \mathrm{X}] - \mu^2 = \mathrm{E}[\mathrm{X}(\mathrm{X}-1)] + \mu - \mu^2$

$= [\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} x(x-1)] + np - (np)^2$

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

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

Let $u = x - 2$ and $N = n - 2$

Then Var(X)

$= n(n-1)p^2 \sum_{u=0}^{N} \frac{N!}{u!(N-u)!} p^u (1-p)^{N-u}] + np - (np)^2$

$= \{n(n-1)p^2 \sum_u P(X = u)\} + np - (np)^2$

$= n(n-1)p^2 + np - (np)^2$

$= np - np^2 = np(1-p)$

Normal distribution

(2***) Suppose that the heights (in cm) of adult males in the UK are distributed $N(174,49)$.

(i) Assuming that there are 2.5 cm to an inch, what proportion of adult males in the UK are over 6 ft? Give your answer to 1dp.

(ii) In another country, the heights of adult males are distributed Normally, such that 10% are over 6 ft and 5% are under 5ft. Find the mean and variance of the distribution. Give your answers to 1dp.

Solution

(i) 6 ft $= 6 \times 12 \times \frac{5}{2} = 180$ cm

If $X \sim N(174, 7^2)$, then $\text{Prob}(X > 180) = \text{Prob}(\frac{X-174}{7} > \frac{180-174}{7})$

$= \text{Prob}(Z > 0.857) = 1 - 0.8042 = 0.1958$, from tables.

So 19.6% of adult males in the UK are over 6 ft (1dp).

(ii) 5 ft $= 5 \times 12 \times \frac{5}{2} = 150$ cm

Let height, $Y \sim N(\mu, \sigma^2)$.

Then $\text{Prob}\left(Z > \frac{180-\mu}{\sigma}\right) = 0.1$ and $\text{Prob}\left(Z < \frac{150-\mu}{\sigma}\right) = 0.05$,

so that , from tables, $\frac{180-\mu}{\sigma} = 1.282$ and $\frac{150-\mu}{\sigma} = -1.645$

Hence $\frac{180-\mu}{1.282} = \frac{150-\mu}{-1.645} \Rightarrow -296.1 + 1.645\mu = 192.3 - 1.282\mu$

$\Rightarrow \mu = \frac{192.3+296.1}{1.645+1.282} = 166.86$

and $\sigma = \frac{180-166.86}{1.282} = 10.2496$ and $\sigma^2 = 105.054$

Thus the mean is 166.9 cm and the variance is 105.1 cm^2 (1 dp).

(3***) Show that 1 standard deviation to either side of the mean of the Normal distribution occurs at the point of inflexion of the Normal curve.

Solution

Considering $N(0,1)$, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

$\phi'(x) = \frac{-x}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

and $\phi''(x) = \frac{-1}{\sqrt{2\pi}} \{e^{-\frac{1}{2}x^2} + x(-x)e^{-\frac{1}{2}x^2}\}$

A point of inflexion is a turning point of the gradient, for which a necessary condition is that the gradient is stationary; ie $\frac{d}{dx}\phi'(x) = 0$ or $\phi''(x) = 0$

[Technically, to confirm that it is a turning point of the gradient, we should check that $\frac{d^2}{dx^2}\phi'(x) \neq 0$; ie $\phi'''(x) \neq 0$ (this is a sufficient condition; a necessary condition is that the first non-zero derivative of $\phi'(x)$ is an even derivative). However, we can see from the curve that there is a point of inflexion.]

$\phi''(x) = 0 \Rightarrow 1 - x^2 = 0 \Rightarrow x = \pm 1$, as required.

Rank Correlation

(4***) Show that the formula $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ can be written as $r_s = 1 - \frac{6\sum {d_i}^2}{n(n^2-1)}$

when the data items are ranks.

[In other words, instead of using the formula for Spearman's coefficient, it is theoretically possible to use the standard formula for r.]

Solution

$S_{xx} = \sum {x_i}^2 - \frac{(\sum x_i)^2}{n}$

As the x_i are just the numbers 1 to n in some order,

$\sum {x_i}^2 = \frac{n}{6}(n+1)(2n+1)$ and $\sum x_i = \frac{n}{2}(n+1)$

and so $S_{xx} = \frac{n}{6}(n+1)(2n+1) - \frac{n(n+1)^2}{4}$

By the same reasoning, S_{yy} will also have this value,

so the denominator of r is

$\frac{n}{6}(n+1)(2n+1) - \frac{n(n+1)^2}{4} = \frac{n}{12}(n+1)\{4n+2-(3n+3)\}$

$= \frac{n}{12}(n+1)(n-1)$

Then $S_{xy} = \sum x_i y_i - \frac{(\sum x_i)\,(\sum y_i)}{n}$

Now $\sum {d_i}^2 = \sum(x_i - y_i)^2 = (\sum {x_i}^2) + (\sum {y_i}^2) - 2\sum x_i y_i$,

so that $\sum x_i y_i = \frac{(\sum {x_i}^2)+(\sum {y_i}^2)-\sum {d_i}^2}{2}$

$= \frac{1}{2}\left\{2.\frac{n}{6}(n+1)(2n+1) - \sum {d_i}^2\right\}$

Hence $S_{xy} = \frac{n}{6}(n+1)(2n+1) - \frac{1}{2}\sum {d_i}^2 - \frac{\left(\frac{n}{2}(n+1)\right)^2}{n}$

$= \frac{n}{12}(n+1)\{4n+2-(3n+3)\} - \frac{1}{2}\sum {d_i}^2$

$= \frac{n}{12}(n+1)(n-1) - \frac{1}{2}\sum {d_i}^2$

and $r = \frac{\frac{n}{12}(n+1)(n-1) - \frac{1}{2}\sum {d_i}^2}{\frac{n}{12}(n+1)(n-1)} = 1 - \frac{6\sum {d_i}^2}{n(n^2-1)}$

Venn Diagrams

(5**) The events A and B are independent and are such that $P(A) = \frac{1}{5}$ and $P(A \cup B) = \frac{1}{2}$

Find:

(i) $P(B)$

(ii) $P(A' \cup B')$

(iii) $P(B' \cap A)$

(iv) $P(B'|A)$

Solution

(i) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

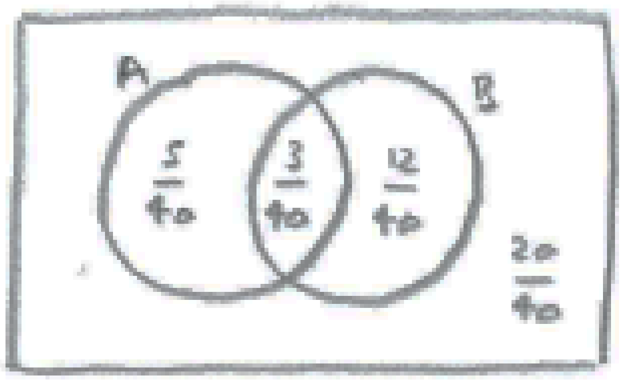
& $P(A \cap B) = P(A)P(B)$, from independence

Hence $\frac{1}{2} = \frac{1}{5} + P(B) - \frac{1}{5}P(B) \Rightarrow \left(\frac{1}{2} - \frac{1}{5}\right) = \frac{4}{5}P(B)$

$\Rightarrow P(B) = \frac{3}{10}\left(\frac{5}{4}\right) = \frac{3}{8}$

[At this point, a Venn diagram could be filled in,

using $P(A \cap B) = \frac{1}{5}P(B) = \frac{3}{40}$]



(ii) $P(A' \cup B') = P([A \cap B]') = 1 - P(A \cap B) = 1 - \frac{1}{5}P(B) = 1 - \frac{3}{40} = \frac{37}{40}$ (or from Venn diagram)

(iii) As A and B are independent, so are A and B'

[Independence of A and B ⇒ knowledge that A has occurred doesn't affect the probability of B occurring, or the probability of B not occurring]

Hence $P(B' \cap A) = P(B')P(A) = (1 - P(B))P(A) = \frac{5}{8}\left(\frac{1}{5}\right) = \frac{1}{8}$

(or from Venn diagram)

(iv) As A and B' are independent, $P(B'|A) = P(B') = 1 - P(B) = \frac{5}{8}$

(or from Venn diagram; or $P(B'|A) = \frac{P(B' \cap A)}{P(A)} = \frac{\left(\frac{1}{8}\right)}{\left(\frac{1}{5}\right)} = \frac{5}{8}$)

Probability generating functions

(6***) Given that $X_1, X_2, \ldots, X_N$ & N are independent random variables, where the X_i are all distributed as X, and that

$S_N = X_1 + X_2 + \cdots + X_N$,

prove that $Var(S_N) = E(N)Var(X) + Var(N)[E(X)]^2$

The following results may be used:

(A) $E(X) = G_X'(1)$

(B) $VarX = G_X''(1) + G_X'(1) - [G_X'(1)]^2$

(C) $G_{S_N}(s) = G_N(G_X(s))$

(D) $E(S_N) = E(N)E(X)$

Solution

From (B), $Var(S_N) = G''_{S_N}(1) + G'_{S_N}(1) - \left[G'_{S_N}(1)\right]^2$

From (C), $G'_{S_N}(s) = G'_N\big(G_X(s)\big)G'_X(s)$

$= \{\frac{d}{du}G_N(u)\}\frac{du}{ds}$, where $u = G_X(s)$

Then $G''_{S_N}(s) = \frac{d}{ds}[\left\{\frac{d}{du}G_N(u)\right\}\frac{du}{ds}]$

$= \left\{\frac{d^2}{du^2}G_N(u)\right\}\frac{du}{ds}.\frac{du}{ds} + \left\{\frac{d}{du}G_N(u)\right\}\frac{d^2u}{ds^2}$

so that $G''_{S_N}(1) = G''_N(1)[G'_X(1)]^2 + G'_N\big(G_X(1)\big)G''_X(1)$

From (B), $VarN = G''_N(1) + G'_N(1) - [G'_N(1)]^2$,

so that $Var(S_N) = \{VarN - G'_N(1) + [G'_N(1)]^2\}[G'_X(1)]^2$

$+G'_N(1)G''_X(1) + G'_{S_N}(1) - \left[G'_{S_N}(1)\right]^2$, since $G_X(1) = 1$

$= \{VarN - E(N) + [E(N)]^2\}[E(X)]^2$

$+E(N)\{G''_X(1)\} + G'_{S_N}(1) - \left[G'_{S_N}(1)\right]^2$

Then, from (B), $VarX = G''_X(1) + G'_X(1) - [G'_X(1)]^2$

and, from (A), $E(X) = G'_X(1)$

Also, from (A)&(D), $G'_{S_N}(1) = E(N)E(X)$

Hence $Var(S_N) = \{VarN - E(N) + [E(N)]^2\}[E(X)]^2$

$+E(N)\{VarX - E(X) + [E(X)]^2\} + E(N)E(X) - [E(N)E(X)]^2$

$= VarN[E(X)]^2 + E(N)VarX$

$= E(N)Var(X) + Var(N)[E(X)]^2$, as required

(7***) ['Poisson hen'] A hen lays N eggs, where $N \sim P_o(\lambda)$, and each egg has probability p of hatching. Using any results about probability generating functions , show that the total number of eggs that hatch $\sim P_o(\lambda p)$.

Solution

Let the total number of eggs that hatch be $Z = X_1 + \cdots + X_N$,

where the $X_i \sim$Bernouilli(p).

Then $G_Z(s) = G_N(G_X(s))$

with $G_N(s) = e^{\lambda(s-1)}$ and $G_X(s) = (1-p) + ps$,

so that $G_Z(s) = e^{\lambda(-p+ps)} = e^{\lambda p(s-1)}$

and hence $Z \sim P_o(\lambda p)$, as required.