STEP Support Programme

STEP 3 Statistics Topic Notes

Algebra of Expectations

- $\mathrm{E}(aX + bY + c) = a\mathrm{E}(X) + b\mathrm{E}(Y) + c$
- $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$
- $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$

If X and Y are independent random variables:

- $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$
- $\mathrm{Var}(aX + bY + c) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)$

Distribution functions

The **probability distribution function** is also called the **probability density function**.

If you are given the probability distribution function of X, then you can find the distribution function of related random variables, such as $Y = X^2$. You do this via the **cumulative distribution function**, e.g.:

$$\begin{aligned}\mathrm{F}_Y(y) &= \mathrm{P}(Y \leqslant y)\\ &= \mathrm{P}(X^2 \leqslant y)\\ &= \mathrm{P}(-\sqrt{y} \leqslant X \leqslant \sqrt{y})\\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \mathrm{f}(t)\,\mathrm{d}t\end{aligned}$$

Once you have the C.D.F. for Y you can differentiate (with respect to y) and hence find the P.D.F.

Geometric Distribution

The *Geometric* distribution models the number of trials needed up to and including the first "success". It has one parameter, p, the probability of success. It is not on all A-level specifications, and is not on the 2019 STEP specifications but does sometimes appear in older STEP questions. The notes here are intended for interest, and to help you if you are trying to attempt one of these types of question.

One example of a geometric distributions is X where X is the number of rolls of a dice until a six is rolled. Then the probability of success is $p = \frac{1}{6}$ and $X \sim \mathrm{Geo}\left(\frac{1}{6}\right)$.

The probability of a six occurring on the r^{th} roll (and not before) is $\mathrm{P}(X = r) = (1-p)^{r-1}p = \left(\frac{5}{6}\right)^{r-1}\left(\frac{1}{6}\right)$ (as the first $r-1$ rolls are all "not sixes" followed by a six on the r^{th} roll).

Writing $q = 1 - p$ we have:

$$\begin{aligned}
\mathrm{P}(X=1)+\mathrm{P}(X=2)+\mathrm{P}(X=3)+\mathrm{P}(X=4)+\ldots &= p+qp+q^2p+q^3p+\ldots \\
&= p\left(1+q+q^2+q^3+\ldots\right) \\
&= p \times \frac{1}{1-q} \qquad \text{Sum of a infinite GP} \\
&= p \times \frac{1}{p} = 1 \qquad \text{as expected!}
\end{aligned}$$

To find the expectation we want to find:

$$\begin{aligned}
\mathrm{E}(X) &= \mathrm{P}(X=1)+2\mathrm{P}(X=2)+3\mathrm{P}(X=3)+4\mathrm{P}(X=4)+\ldots \\
&= p+2qp+3q^2p+4q^3p+\ldots \\
&= p\left(1+2q+3q^2+4q^3+\ldots\right).
\end{aligned}$$

This looks a lot like a derivative! Starting with $1+q+q^2+q^3+q^4+\ldots = (1-q)^{-1}$ we can differentiate with respect to q to get:

$$0+1+2q+3q^2+4q^3+\ldots = (1-q)^{-2} \qquad (*)$$

We then have $\mathrm{E}(X) = p \times \frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$.

To find $\mathrm{Var}(X)$ start by differentiating $(*)$ to get:

$$2+3\times 2q+4\times 3q^2+5\times 4q^3+\ldots = 2(1-q)^{-3} \qquad (\dagger)$$

Then $(\dagger)-(*)$ gives:

$$1+2\times 2q+3\times 3q^2+4\times 4q^3+\ldots = \frac{2}{p^3}-\frac{1}{p^2} = \frac{2-p}{p^3}.$$

Hence $\mathrm{E}(X^2) = p+2^2qp+3^2q^2p+4^2q^3p+\cdots = \frac{2-p}{p^2}$ and

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}(X^2) - \left[\mathrm{E}(X)\right]^2 \\
&= \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 \\
&= \frac{1-p}{p^2}
\end{aligned}$$