STEP Support Programme

STEP 3 Specification Probability and Statistics Notes

These notes are designed to help students in preparing for STEP 3. They cover the "bold and italic" sections of the STEP 3 specification which are not covered in the A-level single Mathematics specifications, or A-level Further Maths Common Core. Many of these topics will be covered in the optional A Level Further Mathematics modules.

There are more notes on the various sections of the specification in the STEP 3 modules.

Independent Random Variables

Two events are *independent* if the occurrence (or not) of one event, A, has no affect on the probability that another event, B, will occur. For example, getting a head on a coin will have no affect on the probability that you get a six on a dice.

Getting an even number on a dice and getting a prime number on the same roll of that dice are not independent as if you get an even number the probability that it is prime is $\frac{1}{3}$ but if you get an odd number the probability that it is prime is $\frac{2}{3}$. Another example of two events which are not independent is whether or not it rains in Cambridge on two consecutive days.

Events A and B are independent iff:

$$\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B)$$

Two random variables are independent if the outcome of one has no affect on the outcome of the other. More formally:

- Two *discrete* random variables X and Y are independent if the events $X = x$ and $Y = y$ are independent for all possible x and y.
- Two *continuous* random variables X and Y are independent if the events $X \leqslant x$ and $Y \leqslant y$ are independent for all possible x and y.[1]

You do not need to know these formal conditions, an informal understanding is enough.

Algebra of Expectation

You are expected to know, and use, the following:

- $\mathrm{E}(aX + bY + c) = a\mathrm{E}(X) + b\mathrm{E}(Y) + c$
- $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$
- and for **independent random variables** we have $\mathrm{Var}(aX \pm bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$

See the Appendix for some derivations of these.

[1]For a continuous distribution the probability that X is exactly equal to x is zero. Hence $\mathrm{P}(X = x)$ and $\mathrm{P}(Y = y)$ are both equal to 0 and so are (trivially and fairly uselessly) independent.

Example: 2005 STEP 3 Q12

12 Five independent timers time a runner as she runs four laps of a track. Four of the timers measure the individual lap times, the results of the measurements being the random variables T_1 to T_4, each of which has variance σ^2 and expectation equal to the true time for the lap. The fifth timer measures the total time for the race, the result of the measurement being the random variable T which has variance σ^2 and expectation equal to the true race time (which is equal to the sum of the four true lap times).

Find a random variable X of the form $aT + b(T_1 + T_2 + T_3 + T_4)$, where a and b are constants independent of the true lap times, with the two properties:

(1) whatever the true lap times, the expectation of X is equal to the true race time;

(2) the variance of X is as small as possible.

Find also a random variable Y of the form $cT + d(T_1 + T_2 + T_3 + T_4)$, where c and d are constants independent of the true lap times, with the property that, whatever the true lap times, the expectation of Y^2 is equal to σ^2.

In one particular race, T takes the value 220 seconds and $(T_1 + T_2 + T_3 + T_4)$ takes the value 220.5 seconds. Use the random variables X and Y to estimate an interval in which the true race time lies.

Solution:

Let the true times of the laps be t_1, t_2, t_3, t_4 and the true race time be t, so that we have $\mathrm{E}(T_1) = t_1$ etc. We also have:

$$t_1 + t_2 + t_3 + t_4 = t \qquad (*)$$

This is because t_1 is the true time of lap 1 etc.

Let $X = aT + b(T_1 + T_2 + T_3 + T_4)$, so we have:

$$\begin{aligned} \mathrm{E}(X) &= at + b(t_1 + t_2 + t_3 + t_4) \\ \mathrm{Var}(X) &= a^2\sigma^2 + b^2 \times 4\sigma^2 \end{aligned}$$

The given conditions tell us that we want $\mathrm{E}(X) = t$, and that we want to minimise $\mathrm{Var}(X) = \sigma^2(a^2 + 4b^2)$.

The first condition gives us $at + b(t_1 + t_2 + t_3 + t_4) = t \implies at + bt = t$, using (*), and so we have $a + b = 1$. Substituting this into the variance we have:

$$\begin{aligned} \mathrm{Var}(X) &= \sigma^2\left(a^2 + 4(1-a)^2\right) \\ &= \sigma^2\left(5a^2 - 8a + 4\right) \\ &= \sigma^2\left(5\left(a - \tfrac{4}{5}\right)^2 - 5 \times \tfrac{16}{25} + 4\right) \\ &= \sigma^2\left(5\left(a - \tfrac{4}{5}\right)^2 + \tfrac{4}{5}\right) \end{aligned}$$

Therefore to minimise the variance take $a = \frac{4}{5}$ which means that $b = \frac{1}{5}$.

For Y we have $\mathrm{E}(Y) = ct + dt$ and $\mathrm{Var}(Y) = \sigma^2(c^2 + 4d^2)$. We also know that $\mathrm{Var}(Y) = \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2$. If $\mathrm{E}(Y^2) = \sigma^2$ then we have:

$$\sigma^2(c^2 + 4d^2) = \sigma^2 - [(c+d)t]^2$$

Since we want this to hold whatever the true lap times are, we want it to be true for all values of t. Hence we need $c + d = 0$ and $c^2 + 4d^2 = 1$. This means that $5c^2 = 1$ and so $c = \pm\frac{1}{\sqrt{5}}$ and $d = \mp\frac{1}{\sqrt{5}}$.

For the particular race we have:

$$\begin{aligned} x &= \frac{4}{5} \times 220 + \frac{1}{5} \times 220.5 \\ &= 220.1 \\ y &= \pm\left(\frac{1}{\sqrt{5}} \times 220 - \frac{1}{\sqrt{5}} \times 220.5\right) \\ &= \mp\frac{0.5}{\sqrt{5}} \end{aligned}$$

So our estimate of the race time is 220.1 seconds and the estimate of the standard deviation of one of the times is $\sigma = \frac{1}{2\sqrt{5}}$.

The variance of X is equal to $\sigma^2 \times \frac{4}{5}$, and so we estimate this as $\frac{1}{4\times 5} \times \frac{4}{5} = \frac{1}{25}$. Therefore our estimate for the standard deviation of X is $\frac{1}{5}$.

One way estimate for an interval in which the mean lies is to take an interval two standard deviations either side of the mean (for an approximately normal distribution this would give a confidence interval of 95%). Therefore our interval would be:

$$220.1 \pm 2 \times \frac{1}{5} = 220.1 \pm 0.4$$

It wouldn't matter much if you used 3 standard deviations, or even one.

Cumulative distribution functions

The *cumulative distribution function* $\mathrm{F}(x)$ is defined by:

$$\mathrm{F}(x) = \mathrm{P}(X \leqslant x) = \int_{-\infty}^{x} \mathrm{f}(t)\,\mathrm{d}t$$

$\mathrm{f}(x)$ is the *probability density function* of X.

To find the probability density function of $Y = X^2$, start by considering the cumulative distribution function.

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

To find the p.d.f. of Y, differentiate the c.d.f. of Y.

Example 2014 STEP 3 Question 12

The random variable X has probability density function $\mathrm{f}(x)$ (which you may assume is differentiable) and cumulative distribution function $\mathrm{F}(x)$ where $-\infty < x < \infty$. The random variable Y is defined by $Y = \mathrm{e}^X$. You may assume throughout this question that X and Y have unique modes.

(i) Find the median value y_m of Y in terms of the median value x_m of X.

(ii) Show that the probability density function of Y is $\mathrm{f}(\ln y)/y$, and deduce that the mode λ of Y satisfies $\mathrm{f}'(\ln\lambda) = \mathrm{f}(\ln\lambda)$.

Note that there are two more parts of this question which I have not reproduced here.

(i)

The median of X satisfies $\mathrm{P}(X \leqslant x_m) = \displaystyle\int_{-\infty}^{x_m} \mathrm{f}(x)\,\mathrm{d}x = \tfrac{1}{2}$. We want to find y_m such that $\mathrm{P}(Y \leqslant y_m) = \tfrac{1}{2}$.

$$\begin{aligned}
\mathrm{P}(Y \leqslant y_m) &= \tfrac{1}{2}\\
\mathrm{P}(\mathrm{e}^X \leqslant y_m) &= \tfrac{1}{2}\\
\mathrm{P}(X \leqslant \ln y_m) &= \tfrac{1}{2}
\end{aligned}$$

note that e^X *is a strictly increasing function*

Hence we have $x_m = \ln y_m \implies y_m = \mathrm{e}^{x_m}$.

Since e^X is an increasing function we can just take logarithms of both sides and preserve the inequality. If instead we were considering $Y = X^2$, or $Y = \sin X$ etc. things would be slightly more complicated.

(ii)

We have:

$$\begin{aligned} \mathrm{P}(Y \leqslant y) &= \mathrm{P}(X \leqslant \ln y) \\ &= \int_{-\infty}^{\ln y} \mathrm{f}(t)\,\mathrm{d}t \\ &= \mathrm{F}(\ln y) \\ &\textit{(Where } \mathrm{F}(X) \textit{ is the c.d.f. of } X\textit{)} \end{aligned}$$

and so the p.d.f. of Y is given by:

$$\begin{aligned} \mathrm{f}_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\mathrm{F}(\ln y) \\ &= \frac{1}{y} \times \mathrm{f}(\ln y) \end{aligned}$$

The mode of Y is where the p.d.f. has a maximum. Differentiating $\mathrm{f}_Y(y)$ with respect to y gives:

$$\begin{aligned} \mathrm{f}'_Y(y) &= \frac{1}{y} \times \frac{1}{y}\mathrm{f}'(\ln y) - \frac{1}{y^2}\mathrm{f}(\ln y) \\ &= \frac{1}{y^2}\left(\mathrm{f}'(\ln y) - \mathrm{f}(\ln y)\right) \end{aligned}$$

At the mode we have $\mathrm{f}'_Y(\lambda) = 0$, and so we have $\mathrm{f}'(\ln \lambda) = \mathrm{f}(\ln \lambda)$.

Note that $y \neq 0$ as we have $Y = \mathrm{e}^X$.

Appendix

A key observation for these derivations is that a double sum (or integral) can be done in either order (so you can sum over the x values and then the y values or vice versa). **You are not expected to know these derivations**, but they are included for completeness.

- $\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$

  If X and Y are discrete then we have:

$$\begin{aligned}
\mathrm{E}(aX+bY) &= \sum_x \sum_y (ax+by)\mathrm{P}(X=x\cap Y=y) \\
&= \sum_x \sum_y ax\mathrm{P}(X=x\cap Y=y) + \sum_x \sum_y by\mathrm{P}(X=x\cap Y=y) \\
&= \sum_x ax \left(\sum_y \mathrm{P}(X=x\cap Y=y)\right) + \sum_y by \left(\sum_x \mathrm{P}(X=x\cap Y=y)\right) \\
&= \sum_x ax\mathrm{P}(X=x) + \sum_y by\mathrm{P}(Y=y) \\
&= a\mathrm{E}(X) + b\mathrm{E}(Y)
\end{aligned}$$

  Note that $\sum_y \mathrm{P}(X = x \cap Y = y)$ is the sum of all the probabilities that $X = x$ for all the different values of Y, so is the sum of all the possible (disjoint) ways in which $X = x$ can happen. Therefore we have $\sum_y \mathrm{P}(X = x \cap Y = y) = \mathrm{P}(X = x)$.

  Note also that with a double sum over x and y we can pull out everything independent of y and then sum the y bits first — so for example $\sum_x \sum_y xy = \sum_x x\left(\sum_y y\right)$. We could alternatively pull out all the bits independent of x and the find the sum over x first.

  For continuous X and Y the argument is very similar, but we will need some definitions first:

  - The *joint distribution function* of X and Y is the function F given by

$$\mathrm{F}(x,y) = \mathrm{P}(X \leqslant x, Y \leqslant y)$$

  - $\mathrm{F}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} \mathrm{f}(u,v)\,\mathrm{d}u\,\mathrm{d}v$
  - The *marginal distribution function* of X is

$$\mathrm{F}_X(x) = \mathrm{P}(X \leqslant x) = \lim_{y\to\infty} \mathrm{F}(x,y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} \mathrm{f}(u,y)\,\mathrm{d}y\right)\,\mathrm{d}u$$

  - The *marginal density function* of X is:

$$\mathrm{f}_X(x) = \int_{-\infty}^{\infty} \mathrm{f}(x,y)\,\mathrm{d}y$$

Using these we have:

$$\begin{aligned}
\mathrm{E}(aX+bY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)\mathrm{f}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}ax\mathrm{f}(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}by\mathrm{f}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty}ax\left(\int_{-\infty}^{\infty}\mathrm{f}(x,y)\,\mathrm{d}y\right)\,\mathrm{d}x + \int_{-\infty}^{\infty}by\left(\int_{-\infty}^{\infty}\mathrm{f}(x,y)\,\mathrm{d}x\right)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}ax\mathrm{f}_X(x)\,\mathrm{d}x + \int_{-\infty}^{\infty}by\mathrm{f}_Y(y)\,\mathrm{d}y \\
&= a\mathrm{E}(X)+b\mathrm{E}(Y)
\end{aligned}$$

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

  Variance is the mean squared distance from the mean, so for a discrete random variable we have:

$$\begin{aligned}
\mathrm{Var}(X) &= \sum\left[\left(x-\mathrm{E}(X)\right)^2\times\mathrm{P}(X=x)\right] \\
&= \sum x^2\mathrm{P}(X=x) - 2\sum\mathrm{E}(X)\times x\mathrm{P}(X=x) + \sum[\mathrm{E}(X)]^2\mathrm{P}(X=x) \\
&= \mathrm{E}(X^2) - 2\mathrm{E}(X)\sum x\mathrm{P}(X=x) + [\mathrm{E}(X)]^2\sum\mathrm{P}(X=x) \\
&= \mathrm{E}(X^2) - 2[\mathrm{E}(X)]^2 + [\mathrm{E}(X)]^2 \\
&= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2
\end{aligned}$$

  Note that $\mathrm{E}(X)$ is a constant so can be taken outside the sum.

  For a continuous random variable the argument is exactly the same apart from you start with $\mathrm{Var}(X) = \displaystyle\int (x-\mathrm{E}(X))^2\mathrm{f}(x)\,\mathrm{d}x$.

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

  Using $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$ we have:

$$\begin{aligned}
\mathrm{Var}(aX+b) &= \mathrm{E}\left((aX+b)^2\right) - [\mathrm{E}(aX+b)]^2 \\
&= \sum\left[(ax+b)^2\times\mathrm{P}(X=x)\right] - [a\mathrm{E}(X)+b]^2 \\
&= \sum a^2x^2\mathrm{P}(X=x) + 2ab\sum x\mathrm{P}(X=x) + \sum b^2\mathrm{P}(X=x) - [a\mathrm{E}(X)+b]^2 \\
&= a^2\mathrm{E}(X^2) + \cancel{2ab\mathrm{E}(X)} + \cancel{b^2} - a^2[\mathrm{E}(X)]^2 - \cancel{2ab\mathrm{E}(X)} - \cancel{b^2} \\
&= a^2\left[\mathrm{E}(X^2) - [\mathrm{E}(X)]^2\right] \\
&= a^2\,\mathrm{Var}(X)
\end{aligned}$$

  The derivation for continuous random variables is very similar, starting with $\mathrm{Var}(aX+b) = \displaystyle\int (ax+b)^2\mathrm{f}(x)\,\mathrm{d}x - [\mathrm{E}(aX+b)]^2$.

- If X and Y are independent $\mathrm{Var}(aX \pm bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)$

  Note that if X and Y are independent then $\mathrm{P}(X = x \cap Y = y) = \mathrm{P}(X = x) \times \mathrm{P}(Y = y)$. We have:

$$
\begin{aligned}
\mathrm{Var}(aX \pm bY) &= \mathrm{E}[(aX \pm bY)^2] - [\mathrm{E}(aX \pm bY)]^2 \\
&= \sum_x \sum_y (a^2x^2 \pm 2abxy + b^2y^2)\mathrm{P}(X = x \cap Y = y) - [a\mathrm{E}(X) \pm b\mathrm{E}(Y)]^2 \\
&= \sum_x a^2x^2\mathrm{P}(X = x) \sum_y \mathrm{P}(Y = y) \pm \sum_x 2abx\mathrm{P}(X = x) \sum_y y\mathrm{P}(Y = y) \\
&\quad + \sum_x b^2\mathrm{P}(X = x) \sum_y y^2\mathrm{P}(Y = y) - a^2[\mathrm{E}(X)]^2 \mp 2ab\mathrm{E}(X)\mathrm{E}(Y) - b^2[\mathrm{E}(Y)]^2 \\
&= \sum_x a^2x^2\mathrm{P}(X = x) \times 1 \pm \sum_x 2abx\mathrm{P}(X = x) \times \mathrm{E}(Y) \\
&\quad + \sum_x b^2\mathrm{P}(X = x) \times \mathrm{E}(Y^2) - a^2[\mathrm{E}(X)]^2 \mp 2ab\mathrm{E}(X)\mathrm{E}(Y) - b^2[\mathrm{E}(Y)]^2 \\
&= a^2\mathrm{E}(X^2) \pm \cancel{2ab\mathrm{E}(X)\mathrm{E}(Y)} + b^2\mathrm{E}(Y^2) - a^2[\mathrm{E}(X)]^2 \mp \cancel{2ab\mathrm{E}(X)\mathrm{E}(Y)} - b^2[\mathrm{E}(Y)]^2 \\
&= a^2\Big(\mathrm{E}(X^2) - [\mathrm{E}(X)]^2\Big) + b^2\Big(\mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2\Big) \\
&= a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)
\end{aligned}
$$

- A similar result (but not one that you are expect to know) is that if X and Y are independent then $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$.

$$
\begin{aligned}
\mathrm{E}(XY) &= \sum_x \sum_y xy\mathrm{P}(X = x \cap Y = y) \\
&= \sum_x \sum_y xy\mathrm{P}(X = x)\mathrm{P}(Y = y) \\
&\qquad\qquad \text{since } X \text{ and } Y \text{ are independent} \\
&= \sum_x x\mathrm{P}(X = x)\left(\sum_y y\mathrm{P}(Y = y)\right) \\
&\qquad\qquad \text{taking the bits independent of } y \text{ out of the } y \text{ sum} \\
&= \sum_x x\mathrm{P}(X = x) \times \mathrm{E}(Y) \\
&= \mathrm{E}(Y) \times \left(\sum_x x\mathrm{P}(X = x)\right) \\
&= \mathrm{E}(X)\mathrm{E}(Y)
\end{aligned}
$$