STEP Support Programme

STEP III Statistics: Hints

1 **(i)** Since the pond is **large** the frog cannot jump over it!

Consider what $p_2(2)$ actually is, and then think about the different things that can happen with the first jump.

(ii) u_1, u_2 and u_3 are the expected number of jumps the frog takes to land in the pond if it starts $\frac{1}{2}$ m, $1\frac{1}{2}$ m and $2\frac{1}{2}$ m away from the edge. In each case start by finding the different possible number of jumps and use $q = 1 - p$ to write your final answers in terms of q.

(iii) There are three constants to find and three values of u_i so this looks like being a case of solving simultaneous equations. The constants will all be in terms of q. As n gets large certain parts of u_n can be ignored. For the last part, start by working out the expected distance covered by the frog in one jump.

2 To find $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$ use the given pgf for Y and the given results for probability generating functions on page 7 of the **formula book** which can be found here. You will need to do some differentiation using the product and chain rule, and initially things will look a bit messy when finding $\mathrm{Var}(Y)$ but then lots of terms cancel. Be careful not to confuse $\mathrm{G}(t)$ and $\mathrm{G}_Y(t)$.

The number of tosses until a head occurs has a *Geometric* distribution. On page 7 of the formula book you can find the mean, variance and pgf for a Geometric distribution (along with others).

X_i is the number of heads shown on the i^{th} toss, which can be thought of as a binomial distribution, but it is probably easier just to say that $\mathrm{P}(X_i = 0) = \mathrm{P}(X_i = 1) = \frac{1}{2}$. You can find the pgf of X by using $\mathrm{E}(t^X)$ (which will only have two terms).

To find $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$ you can use the results shown earlier. To find the pgf of Y use $\mathrm{G}(\mathrm{H}(t))$. Finally $\mathrm{P}(Y = r)$ is the coefficient of t^r in $\mathrm{G}_Y(t)$ — use $\mathrm{G}_Y(t) = \mathrm{G}(\mathrm{H}(t))$ and then you will need to use a binomial expansion.

3 The formula for covariance, $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$ is given in the formula book (although there it has μ_X for $\mathrm{E}(X)$ etc.)

You will also need the formulae for *independent* random variables X and Y:

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y) \qquad \text{and}$$
$$\mathrm{Var}(aX \pm bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)$$

These are also given in the formula book!

As long as you have these formulae the rest of the question is basically applying these, expanding some brackets and solving some simultaneous equations. It gets a little messy in places.

For the last part, take a guess as to what X_i will be (hint, it involves μ_i, σ_i and Y_i) and then show that this satisfies the required conditions.

4 This question might look a bit off-putting, but it is basically an integration question and you are given the key result needed.

Remember that $\displaystyle\int_{\text{all } x} \mathrm{f}(x)\,\mathrm{d}x = 1$, and using the given result will enable you to find C.

Have a look at the limits of the integrals to help you find “a suitable substitution”.

For the median we need $\displaystyle\int_0^m \mathrm{f}(x)\,\mathrm{d}x = \int_m^\infty \mathrm{f}(x)\,\mathrm{d}x$. “Hence” is an instruction which must be obeyed!

The expectation is $\displaystyle\int_{\text{all } x} x \times \mathrm{f}(x)\,\mathrm{d}x$ and the result given at the start of the question will be useful again.

The question asks you to find the *density* function of T — this is another name for the probability distribution function, i.e. you want to find $\mathrm{f}(x)$ such that $\mathrm{P}(T < t) = \displaystyle\int_0^t \mathrm{f}(x)\,\mathrm{d}x$. It is possible to write the density function of T in the same form as the density function of V.