Section 1 of Unit 07 (Statistics 2)

Poisson distribution, linear combinations of two variables and continuous random variables

Recommended Prior Knowledge
Students must have studied S1 before starting this Section.

Context
Sections 1 and 2 must be studied in order, since the ideas in Section 1 are required throughout Section 2.

Outline
This Section introduces the Poisson distribution, considers its relation to the binomial distribution and looks at how it can be approximated by the normal distribution. The Section continues with a study of the mean and variance of linear combinations of two variables. It concludes with an in-depth study of continuous random variables, including the calculation of probabilities, mean, variance and median.

Topic	Learning Outcomes	Suggested Teaching Activities	Resources	On-Line Resources
1	**The Poisson Distribution** • Understand the relevance of the Poisson distribution to the distribution of random events and use the Poisson distribution as a model.	Time should be spent in looking at the general properties of a "Poisson" distribution. Students should realise that a "Poisson" distribution can be recognised by the fact that only the "mean" for a particular event is given. In a Poisson distribution, neither the probability of a single success nor the number of events is given. Introduce the students to different examples of Poisson distributions: • number of road deaths in a city centre in a month • number of flaws in a 100m length of cloth • number of breakdowns of a machine in a year • number of phone-calls coming into an exchange in a given time interval.	Have ready on OHP several examples of different types of Poisson distributions.	www.mathsrevision.net →A-Level Section →Statistics This is an excellent revision site containing sections appropriate to most of the topics in this syllabus.

	• Calculate probabilities for the distribution $P_o(\mu)$. • Use the Poisson distribution as an approximation to the binomial distribution where appropriate (np<5, approximately). • Use the fact that if $X \sim P_o(\mu)$, then the mean and variance of X are each equal to μ.	Students should be introduced to the general formula for the probability of r successes for the situation in which μ, the mean number of events is given. Students should be familiar with "e" from work covered in P3. They should then familiarise themselves with different examples using these formulae. Students should realise that if $P_o(5)$ is the distribution for the number of calls coming in to a telephone exchange in a 10-minute interval, then $P_o(1)$ is the number of calls coming into the exchange in a 2-minute interval. This will be referred to in more detail in the next section on linear combinations of independent variables. Students should realise that the Poisson distribution is derived from a binomial in which n is very large and p is very small. It is a worthwhile exercise taking a particular example, say, Po(5) representing 5 telephone calls coming into an office in a 10 minute interval. This can be thought of, approximately, as a binomial distribution with n = 600 (number of seconds in 10 minutes) and $p = \frac{1}{120}$ (mean number of calls per second). Candidates should be able to calculate the probabilities of 0, 1, 2, 3 phone calls in the 10 minute interval using either the Poisson formula or the binomial formula for n successes. Taking n = 6000 and $p = \frac{1}{1200}$, will lead to even closer agreement between the Poisson and binomial values. This is an appropriate time to show that since the variance of a binomial is $np(1-p)$ and $(1-p)$ is approximately 1, then the variance and the mean take the same value. This is required for the next syllabus item.	Have ready on OHP the different formulae for $P(X)$ for X = 0, 1, 2, 3 etc for $P_o(\mu)$. Have ready on OHP, the results for a comparison of the probabilities calculated for $P_o(5)$ and for $B(600, \frac{1}{120})$, $B(60, \frac{1}{12})$ and for $B(6000, \frac{1}{1200})$.	www.mathsrevision.net →A-Level Section →Statistics → The Poisson distribution.

	• Use the normal distribution, with continuity correction, as an approximation to the Poisson distribution where appropriate (μ>15, approximately).	If μ is large (>15), the Poisson distribution can be closely approximated by a normal distribution. If time allows, the students can calculate the probabilities (with each student taking a few) for the distribution $P_o(16)$. On plotting these, the students will appreciate that the resulting graph is approximately normal in shape. Using the normal distribution with mean 16 and variance 16, candidates should be able to obtain a value for the probability of, say 10 successes, by considering the interval ($9.5<X<10.5$). They should realise, from work covered in S1, the need for a continuity correction when a continuous distribution is used to approximate a discrete distribution. Lots of practice incorporating all these syllabus items is now needed. In particular, it is worth discussing the different ways in which different approximations are appropriate to questions.	Have ready on OHP, values for the probabilities of, say, 10,11,12, 13, 14, 15 etc. successes for $P_o(16)$ and the calculated values for these probabilities using N(16,16).	www.mathsrevision.net →A-Level Section →Statistics →Normal Approximations
2	**Linear combinations of random variables.** • Use, in the course of solving problems, the results that $\mathrm{E}(aX+b)=a\mathrm{E}(X)+b$ and $\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X)$, $\mathrm{E}(aX+bY)=a\mathrm{E}(X)+b\mathrm{E}(Y)$ and $\mathrm{Var}(aX+bY)=a^2\mathrm{Var}(X)+b^2\mathrm{Var}(Y)$.	The proofs in this section are difficult and are not required. For good students however, the proofs are available in text-books and can be covered if time allows. In other cases, it is a worthwhile exercise verifying the results by looking at a simple example which uses the distributions of two variables, X and Y. Starting with a set X ={1,2,3} and a set Y=(2,6), it is relatively east to verify such results as : $\mathrm{E}(X) = 2$, $\mathrm{Var}(X) = ⅔$, $\mathrm{E}(Y) = 4$, $\mathrm{Var}(Y) = 4$, $\mathrm{E}(2X) = 4$, $\mathrm{Var}(2X) = 2⅔$ (ie 2^2×⅔), $\mathrm{E}(3Y+2) = 14$, $\mathrm{Var}(3Y+2) = 36$ (ie 3^2×4), $\mathrm{E}(X+Y) = 6$, Var $(X+Y)$ = 4⅔ (ie 4+⅔), $\mathrm{E}(X-Y) = -2$, Var $(X-Y)$ = 4⅔. (ie 4+⅔), **not (4−⅔),** $\mathrm{E}(X_1+X_2) = 4$, Var (X_1+X_2) = Var (X) + Var(X) = 1⅓.	Have ready on OHP a table showing given values for the mean and variance of variables X and Y and then for such variables as $2X$, X_1+X_2, X_1-X_2, $2X+3Y$, etc.	

	• Use the result that if X has a normal distribution, then so does *aX+b.* • Use the result that if X and Y have independent normal distributions, then $aX+bY$ has a normal distribution. • Use the result that if X and Y have independent Poisson distributions, then $X+Y$ has a Poisson distribution.	It is important that students realise the difference between this last result (for (X_1+X_2)) and the result for $(2X)$. The variable (X_1+X_2) refers to two values of X selected from the distribution (ie 1+1, 1+2, 1+3, 2+1 …etc.) whereas 2X refers to one reading doubled (i.e. 2, 4, 6). Students need to realise that the above generalisations apply to normal distributions and need plenty of practice on different types of examples. The question below illustrates the difference in the distributions $(2X)$ and (X_1+X_2) and it is worthwhile working this through with the students. **Question**: Students in a college have masses which are normally distributed with mean 70kg and standard deviation 10 kg. Four students get a in a lift which has a safety limit of 300kg. Find the probability that the safety limit is exceeded. Repeat the calculation with a single student who is accompanied by luggage equal to three times his mass. The students should be able to appreciate that the limit is more likely to be exceeded in the second case. Introduce students to the result that if X and Y have independent Poisson distributions, then $X+Y$ has a Poisson distribution. This is a suitable opportunity to revise the work done earlier on the Poisson distribution and at the same time justify the previously used fact that if X represents a certain Poisson distribution, then so does the variable Y where $Y=kX$ where k is a constant.	It is worthwhile having this, or a similar question, ready and fully worked on an OHP.	

3	**Continuous random variables** • Understand the concept of a continuous random variable, and recall and use properties of a probability density function (restricted to functions defined over a single interval).	Students should have met the ideas of discrete and continuous distributions, both at IGCSE and in S1. Discuss generally with the students how probability might be related to the shape of the curve. Students should be able to appreciate that probability is related to “area under the curve”. Introduce the idea of probability density function by presenting the students with a particular case, say f(x) = k(4−x) for 2≤x≤4, = 0 elsewhere. Students should appreciate that because of the shape of the function, there is greater likelihood of a reading selected at random lying in an interval containing x = 2.2 (say 2.1<x<2.3) than in an interval of the same length containing x = 3.5 (say 3.4<x<3.6). Students should also appreciate that: • readings must lie in the interval 2 to 4 • the mean and the median is each less than the mid-point of the interval (less than 3) • $\int_2^4 k(4-x)\mathrm{d}x = 1$ leading to the value of k (=0.5) • the probability of a reading lying in the interval 2 to 3 is $\int_2^3 k(4-x)\mathrm{d}x$.	Have ready on an OHP a fully worked question for a given f(x), including a constant k, in which all possible questions are given. This should include finding k, finding probabilities and finding the mean, variance, median and quartiles.	www.mathsrevision.net →A-Level Section →Statistics →Continuous Random Variables

	• Use a probability density function to solve problems involving probabilities, and to calculate the mean and variance of a distribution (explicit knowledge of the cumulative distribution function is not included, but location of the median, for example, in simple cases by direct consideration of an area may be required).	Comparing probability density functions with discrete probability distributions should enable students to appreciate the formulae for the mean and variance of such distributions. The various formulae are identical except that "∫" replaces "Σ". Although explicit knowledge of the cumulative distribution function is not included, candidates should be able to calculate the median, *M,* from $\int_2^M k(4-x)\mathrm{d}x = 0.5$ and the lower quartile, *L*, from $\int_2^L k(4-x)\mathrm{d}x = 0.25$. Candidates need a lot of practice with different probability density functions, including such questions with the above distribution as: "If 3 readings are taken from the distribution, find the probability that exactly two of the three are less than 3".		