VCE Specialist Mathematics
Units 3 and 4

Sample modelling or problem-solving task – sums of Bernoulli random variables

Introduction

A binomial distributed random variable with parameters *n* and *p* is the sum of *n* independent Bernoulli distributed random variables. Through this task students have the opportunity to see that sample proportions are a special case of sample means and connect work between the Statistics and probability areas of study in Mathematical Methods and Specialist Mathematics.

Part 1

Consider a question which has a yes – no response. Let the probability of a ‘yes’ be *p.* The random variable is Bernoulli distributed with the probability of a ‘yes’ being *p* and the probability of a ‘no’ being 1– *p*.

Let *X*­­­1 be the response by a person. The random variable is Bernoulli distributed with the probability of a ‘yes’ being *p* and the probability of a ‘no’ being 1– *p*. We write Pr(*X*1­ = 1) = *p* and Pr(*X*1 = 0) = 1 –*p*. Let *X*­­­2 be the response of a second person and *X*­­­3 a third (assume independence of the responses).

If *Z* = *X*­­­1 + *X*­­­2 + *X*3 then Pr(*Z* = *z*) = Pr(*X*­­­1 + *X*­­­2 + *X*3= *z*).

The values of *z* can be 0,1, 2 or 3. The numbers correspond to 0 yes’s, 1 yes, 2 yes’s and 3 yes’s. For example:

Pr(*Z* =2) = Pr(*X*­­­1= 1) Pr(*X*­­­2= 1) Pr(*X*­­­3= 0) + Pr(*X*­­­1= 1) Pr(*X*­­­2= 0) Pr(*X*­­­3= 1)

+ Pr(*X*­­­1= 1) Pr(*X*­­­2= 0) Pr(*X*­­­3= 1) + Pr(*X*­­­1= 0) Pr(*X*­­­2= 1) Pr(*X*­­­3= 1)

=$\left(\begin{matrix}3\\2\end{matrix}\right)$ $\left(\begin{matrix}3\\2\end{matrix}\right)$(*p*)2(1 – *p*)

Continue in this way to form the rule for a binomial distribution.

Part 2

Distribution of sample proportions.

1. Show that for taking samples of size *n* from a population that = ­­­­­­…­­­
2. Find the mean and variance of the distribution for by using these results where the *X*­­­*i* are independent Bernoulli random variables.

Part 3

Suppose that you toss 3 fair coins. For each Heads, you win $1, and for each Tails, you lose $1. Let *Y* be you net winning.

Then *Y* is a random variable that can take the values −3, −1, 1, 3 and:

Pr(*Y* = –3) = Pr(*Y* = 3) = , Pr(*Y* = –1) = Pr(*Y* = 1) = .

Suppose that the game is repeated a number of times and the winnings and losses ‘totalled’ and the mean

calculated from this. There is a new random variable *T* = ­­­­­­…­­­ defined in this way where the *Yi* are

independent and identical to the *Y*.

Investigate the distribution of *T*, starting with *n* = 2 and *n* = 3.

Areas of study

The following content from the areas of study is addressed through this task.

|  |  |
| --- | --- |
| **Area of study** | **Content dot point** |
| Data analysis, probability and statistics | 1, 2, 3, 4, 5, 6 |

Outcomes

The following outcomes, key knowledge and key skills are addressed through this task.

|  |  |  |
| --- | --- | --- |
| **Outcome** | **Key knowledge dot point** | **Key skill dot point** |
| **1** | 15 | 15 |
| **2** | 6 | 3, 4, 5 |
| **3** | 1, 2, 3, 4, 5, 6 | 1, 2, 3, 4, 5, 6, 7, 9, 10, 11 |

The following outcomes, key knowledge and key skills are addressed through this task.