MAV24 Conference

Session B26 - Probability enriched through simulation in Mathematical Methods

Presenter: Frank Moya (frank.moya@bigpond.com)

Description

Simulation is an invaluable pedagogical tool for the teaching and learning of probability and statistical inference, and for carrying out mathematical investigations involving random events. In this session, participants will use **TI-Nspire CAS** technology to explore various techniques to set up and run simulations that are useful in teaching various topics from the probability area of study, as well as in investigation tasks. These simulation techniques can be adapted to other technology platforms. The session will aim to provide some innovative teaching ideas, as well as tips on effective use of technology, including ways of dynamically displaying simulation results to visualise key concepts and gain a deeper understanding of the topic.

Activity 1 – Bertrand's Box Problem – a conditional probability simulation

Connection to 2023-2027 VCE Study Design

VCE MM Unit 1 Area of Study 4

simulation using simple random generators such as coins, dice, spinners and pseudo-random generators using technology, and the display and interpretation of results …

VCE MM Unit 2 Area of Study 4

conditional probability in terms of reduced sample space, the relations $\Pr(A \mid B) = \dfrac{\Pr(A \cap B)}{\Pr(B)}$ …

Understanding Bertand's Box Problem

This problem was first posed by Joseph Bertrand in 1889, as follows.
Suppose that there 3 closed boxes.

- Box 1 contains two gold coins
- Box 2 contains two silver coins
- Box 3 contains one gold and one silver coin.

A box is chosen at random.
The coins in that box are chosen one at a time.
If the first coin chosen is gold, what is the probability that the second coin is also gold?

1.a. Exploring the problem: Intuition

What do you think the answer is, and can you justify your answer?
(Example. Student *A* argues that, since the box chosen must be 1 or 3, then the next coin is either gold or silver with probability 1/2 for each. Is Student *A* correct? Is there a flaw in his/her reasoning?)

1.b. Exploring the problem: Simulation

This document presents one possible approach to simulating this problem.
A *Lists & Spreadsheet* application will be set to simulate conducting 100 trials at a time.

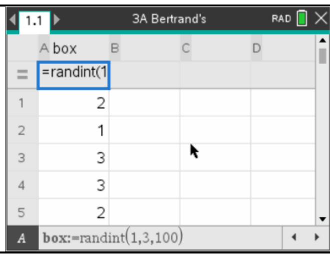
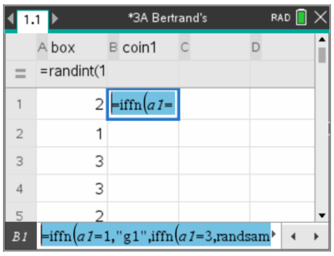
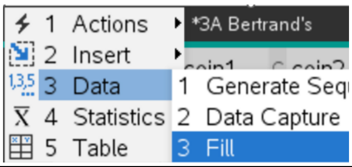
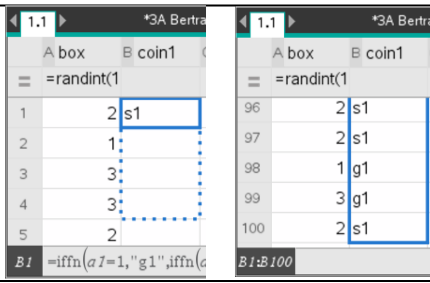
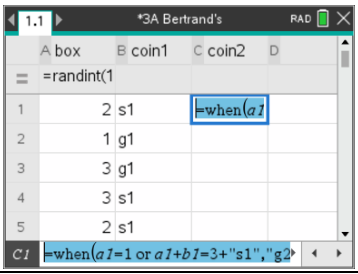
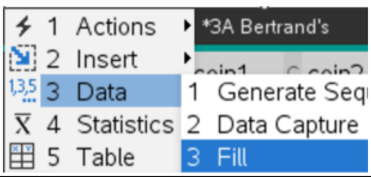
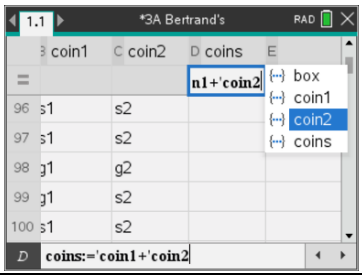
TI-Nspire Commands Used include

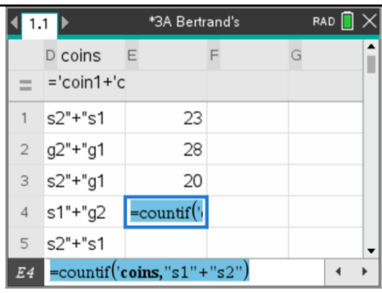
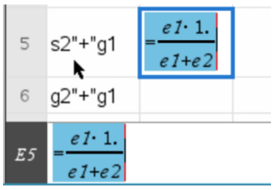
Accessing the Catalogue of all commands, by pressing 📖 1
📖 RandSeed *Number* (*Number* could be, say, the last 4 digits of your phone number)
📖 randInt(lowerbound, upperbound[, number of trials])
📖 randSamp(list, sample_size[, no repetition =1])
📖 ifFn(BooleanExpr, Value_If_true [,Value_If_false [,Value_If_unknown]])

when(Condition, trueResult [,falseResult] [,unknownResult])
count(list)
countIf(list, condition)
List $\{a1, b1, c1, \ldots\}[n]$ returns the n^{th} element in the list. E.g. $\{a1, b1, c1, \ldots\}[2]$ is $b1$ (2^{nd} element)

Setting up a simple simulation in Lists & Spreadsheet for 100 trials

- Before setting up simulations, make sure RandSeed *Number* has been performed on your TI technology.
- Save document ([ctrl] then [S]) with a name such as *Bertrand* before the end of the session.

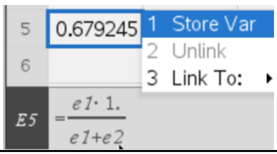
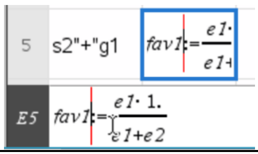
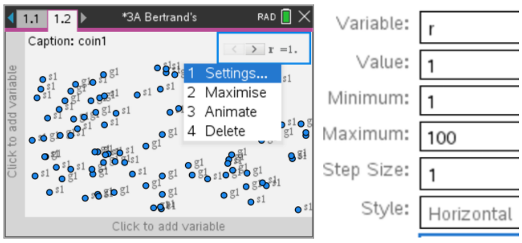
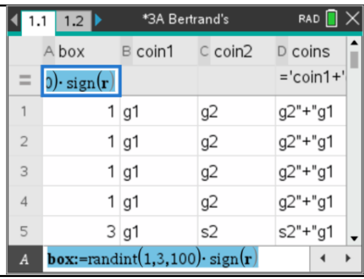
Instructions and keypad help	Screenshot
Selecting Box 1, 2 or 3 at random * Open a Lists & Spreadsheet page in a new document. - [on] > Click spreadsheet icon. OR [ctrl] [N] > [4] Add Lists & S… * Column A – name: **Box** * Column A – formula: **=randint(1,3,100)**	
Selecting **first** coin from the box Let "g1" and "s1" denote "gold first coin" and "silver first coin" N.B. " " (quotation marks) press [?!▸] * Column B – name: **coin1** * Cell B1 – =ifFn(a1=1,"g1",ifFn(a1=3,randsamp({"g1","s1"},1,1)[1],"s1")) * Press [enter] to lock in the formula.	
Fill the formula down to cell B100 - Click cell B1. - Then [menu] > Data > Fill. Then down arrow ▼ to B100 and [enter].  Top of list: [ctrl] [7] Bottom of list: [ctrl] [1]	
Selecting **second** coin from the box Let "g2" and "s2" denote "gold 2nd coin" and "silver 2nd coin" N.B. " " (quotation marks) press [?!▸] * Column C – name: **coin2** * Cell C1 – =when(a1=1 or a1+b1=3+"s1","g2","s2","s2") * Press [enter] to lock in the formula.	
Fill the formula down to cell C100 - Click cell C1. - Then [menu] > Data > Fill. Then down arrow ▼ to C100 and [enter].	
Combine the results for coin 1 and coin 2. * Column D – name: **coins** * Column D – formula: N.B. choose lists 'coin1 , 'coin2 from [var] **coins:=** [var] select **coin1** [+] [var] select **coin2** then [enter] Formula: **coins:='coin1+'coin2**	
Count the number of each possible result	

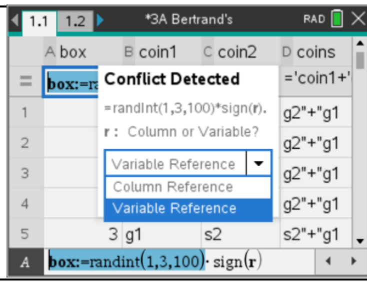
* Cell E1 – count result **coins** is "g1"+"g2" Formula: =countif('**coins**,"g1"+"g2") (chooses '**coins** from var) * Fill formula down to Cell E4 enter. Then EDIT cells E2 to E4 * Cell E2 – Edit Formula to: =countif('**coins**, "g1"+"s2") * Cell E3 – Edit Formula to: =countif('**coins**, "s1"+"g2") * Cell E4 – Edit Formula to: =countif('**coins**, "s1"+"s2")	
Calculate proportion of 'favourable' outcomes Let a 'favourable' outcome be "g1"+"g2" given "g1". That is, 'favourable' proportion is $\frac{count("g1"+"g2")}{count("g1"+"g2")+count("g1"+"s2")}$ * Cell E5 – calculate the favourable proportion Formula: =E1/(E1+E2)×1.0 enter or =approx(E1/(E1+E2)) enter	

Repeating the simulation – Press ctrl then R.

Long-run Proportion: Capturing and displaying the results for repeated runs of the simulation

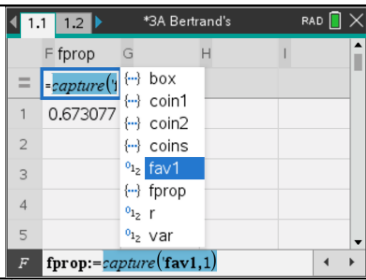
You may observe that for repeated runs of the simulation of 100 trials, there is variation in the proportion of 'favourable' outcomes between different runs. It is therefore useful to observe the long-run proportion if the simulation is repeated many times.

Instructions and keypad help	Screenshot
Store Cell E5 value ('favourable' proportion) as a variable - Select Cell E5 then press var. Select '1 Store Var' - Name the variable: *fav1*, then enter.	
Getting ready to display results in Data & Statistics page Set up a Data & Statistics page where the captured results from repeated runs will be displayed. - Add a Data & Statistics page. - Add a slider to the page: menu > Actions > Insert Slider - Slider Settings: variable: r, Value: 1, Min: 1, Max: 100, Minimise:Yes	
Tweaking the simulation Return to the spreadsheet. Edit Column 1 formula so that a new run of the simulation will occur when the slider value changes. - Column A formula cell – edit **box**:=randint(1,3,100) to **box**:=randint(1,3,100)×sign(r)	

Instructions and keypad help	Screenshot
Tweaking the simulation Return to the spreadsheet. Edit Column 1 formula so that a new run of the simulation will occur when the slider value changes. - Column A formula – edit **box**:=randint(1,3,100) to **box**:=randint(1,3,100)×sign(r) enter - 'Conflict detected' dialog box – select 'Variable Reference' OK	

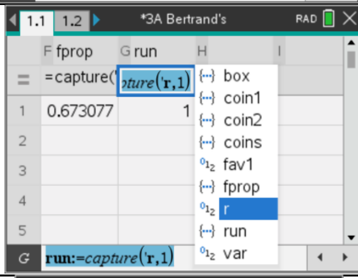
Capturing the values of variable ***fav1***

- Column F. Name: fprop then [enter].
- Then [menu] > Data > Data Capture > Automatic
- Then [var] > *fav1* then [enter]
- Column F formula should now be: **fprop**:=capture('**fav1**,1)



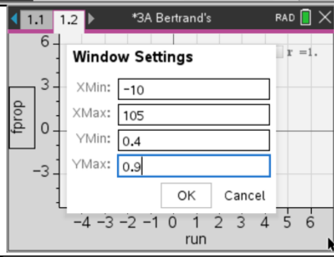
Capturing the values of variable ***r***

- ColumnG. Name: run then [enter].
- Then [menu] > Data > Data Capture > Automatic
- Then [var] > *r* then [enter]
- Column G formula should now be: **run**:=capture('**r**,1)



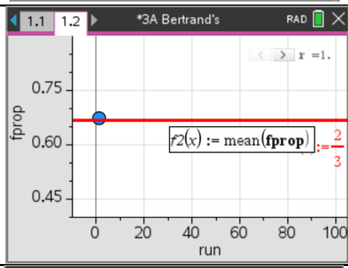
Setting up data display window for repeated runs

- Go to Data & Statistics page.
- Horizontal variable: select ***run***
- Vertical variable: select ***fprop***
- Adjust window settings:
  [menu] > Window/Zoom > Window Settings then adjust
  XMin: -10, XMax: 105, YMin: 0.4, YMax: 0.9



Setting up data display functions for repeated runs

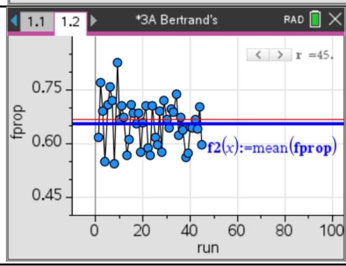
- Go to Data & Statistics page.
- Plot the function $f1(x) = 2/3$ and change line colour to red
  [menu] > Analyse > Plot function. Edit equation to $f1(x) := 2/3$
- Plot the function $f1(x) = mean(fprop)$
  [menu] > Analyse > Plot function. Edit to $f1(x) := mean(\boldsymbol{fprop})$
- In Cell E6 of spreadsheet, enter formula $= mean(\boldsymbol{fprop})$



Display results for repeated runs of the simulation

- Click or animate the slider to produce repeated runs of the simulation.
- To animate slider: Hover over slider, then [ctrl][menu] > Animate
- To stop animation:
  Hover over slider, then [ctrl][menu] > Stop Animate

Optional. Hover over a data point [ctrl][menu] > Connect Data Values



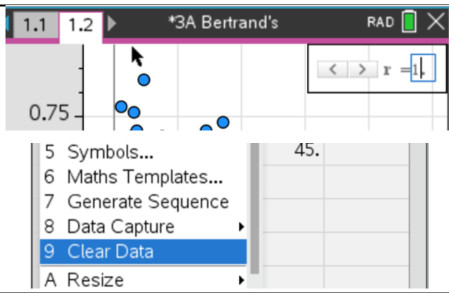
Comment on the mean value of the list *fprop* as the number of runs of the simulation increases.

Resetting the run of simulations and clearing the data points

* Step 1. Slider value set to 1.
- Click on the slider **value** until it becomes editable. Edit it to 1.
- Alternatively, click slider, [ctrl][menu] > Settings > Value: 1 [enter]

* Step 2. Clear Data in Column F of the spreadsheet
- Click Column **F formula** cell, [ctrl][menu] > [9] Clear Data

* Step 3. Clear Data in Column G of the spreadsheet
- Click Column **G formula** cell, [ctrl][menu] > [9] Clear Data



Calculation

1. Intuitive reasoning might be something like the following.
   Box 2 could not have been selected because it contains no gold coins. It is more likely that the first gold coin comes from box 1 than box 3, leading to the probability 2/3.

2. In setting up the spreadsheet simulation, it is clear that the 'success fraction' of "Gold second coin" given "Gold first coin" is given $\dfrac{count(\text{"g1"+"g2"})}{count(\text{"g1"+"g2"})+count(\text{"g1"+"s2"})}$

$$= \frac{Pr(Box1) \times Pr(Gold\ first\ having\ chosen\ Box1)}{\left(Pr(Box1) \times Pr(Gold\ first\ having\ chosen\ Box1)\right) + \left(Pr(Box3) \times Pr(Gold\ first\ having\ chosen\ Box3)\right)}$$

$$= \frac{\frac{1}{3} \times 1}{\left(\frac{1}{3} \times 1\right) + \left(\frac{1}{3} \times \frac{1}{2}\right)} = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{6}} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

This is an intuitive way of understanding **Bayes' Theorem**.

$$\Pr("g1 + g2"|g1) = \frac{\Pr("g1 + g2" \cap "g1")}{\Pr("g1")}$$

Investigation Suggestions: Related Problems

Investigate similar problems, such as the ***Monty Hall*** problem.

Activity 2 – Simulation of the sampling distribution of sample proportions

Connection 2023-2027 VCE Study Design
VCE MM Unit 1 Area of Study 4
simulation using simple random generators …, including informal consideration of ***proportions in samples***
VCE MM Units 3&4 Area of Study 4
simulation of random sampling, for a variety of values of p and a range of sample sizes, to illustrate the distribution of $\hat{P}$ and variations in confidence intervals between samples.

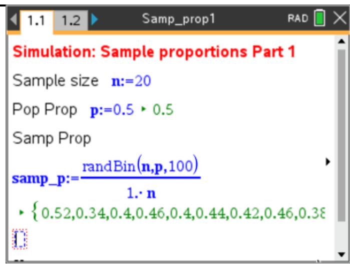
Sample proportion simulation. Assume that in a very large city 50% of people named on the electoral roll are female. If random samples of n names are drawn from the electoral roll, simulate the distribution of the proportion of females named in samples.

Part 1 of Activity 2: Express simulation

Select random samples of size $n = 20$ from population with a true population proportion, $p = 0.5$. Repeat the sampling 100 times. Use: **randBin(*n*, *p* [,#Trials])**.

To select 100 samples of size $n = 20$ from $\text{Bi}(n = 50, p = 0.5)$ and calculate the sample proportion, on a **Notes** page:

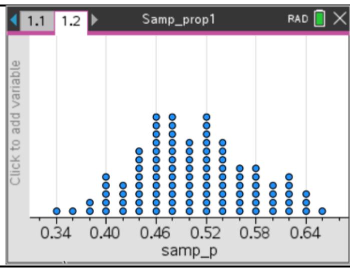
- In different rows - as shown - enter the words '**Sample Size**', '**Pop Prop**', '**Samp Prop'**.
- Press [ctrl] [M] to insert a **Maths Box** next to *each* of the words.
- In the Maths Box next to the words '**Sample Size**', enter $\boldsymbol{n} \coloneqq \mathbf{20}$ (for ':=' press [ctrl] [templates] (:=)).
- In the Maths Box next to the words '**Pop Prop**', enter $\boldsymbol{p} \coloneqq \mathbf{0.5}$.
- In the Maths Box next to the words '**Samp Prop'**, input ***s_prop*** :=. Then press [menu] > **Calculations** > **Probability** > **Random** > **Binomial** and enter $\frac{\boldsymbol{randBin}(\boldsymbol{n},\boldsymbol{p},\mathbf{100})}{\mathbf{1.0}\cdot\boldsymbol{n}}$.



To obtain plots of the sample data, add a **Data & Statistics** page to the document, then:

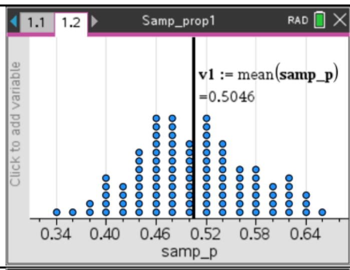
- Press [tab] and select **samp_p** on the horizontal axis.

To select and recalculate a new set of 100 samples of size 20, navigate to page 1.1, click the '**samp_p'** Maths Box and press enter.



To display a plot value of the mean of the 100 sample proportions, on page 1.2:

- Press [menu] > **Analyse > Plot Value**, then enter $\boldsymbol{var1 := mean(samp\_p)}$.
- Select new sets of 100 samples (as described above) and observe the recalculated plot value for the mean of the sample proportions.



To obtain a histogram of the sample data:

- On page 1.2, press [ctrl] [menu] > **Histogram**.

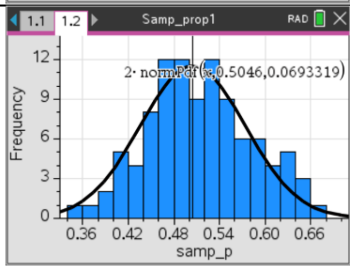
To compare the data to a normal distribution curve:

- Press [menu] > **Analyse > Show Normal PDF**.

Compare the parameters of the normal pdf (and the 'One Var' statistics of the sample proportions) with theoretical values:

$$\mathrm{E}(\hat{P}) = p = 0.5 \text{ and } \mathrm{sd}(\hat{P}) = \sqrt{\frac{p(1-p)}{n}} \approx 0.0707.$$

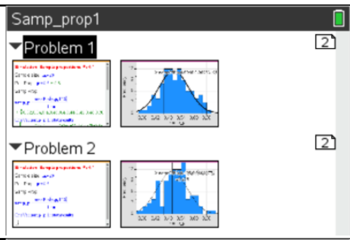
- Observe the effects of changing the values of p or n by editing the value in the relevant Maths Box on page 1.1.



Part 2 of Activity 2: Preferred simulation – illustrating variation between samples and building the distribution one sample at a time

To build on the previous simulation in a new problem:

- Press [ctrl] ▲ to obtain a thumbnail view.
- Press ▲ again to select the title 'Problem 1'.
- Press [ctrl] [C] then [ctrl] [V] to copy and paste, creating a clone.
- Click the **second** page of **Problem 2** to open it.

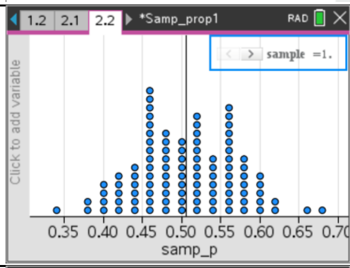


On page 2.2, add a slider titles 'sample', as follows.

- Press [menu] > **Actions > Insert Slider**.
- Input the following slider settings:

Variable: **Sample**; Value: **1**; Min.: **1**; Max.: **200**; Step Size: **1**.

The slider will be used to draw samples, one at a time.

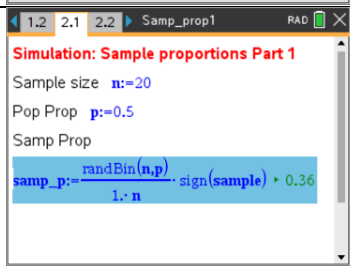


Navigate to page 2.1 and edit to generate one sample at a time.

- Edit the bottom Maths Box to

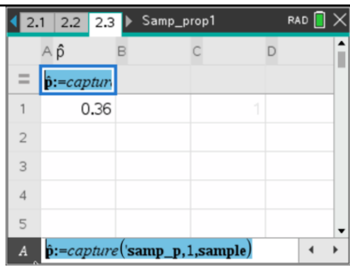
$$\boldsymbol{samp\_p := \frac{randBin(n,p)}{1.0 \cdot n}} \times \mathrm{sign}(sample)$$

Note. $\mathrm{sign}(x) = 1$ *for* $x > 0$. *When the slider value 'sample' changes, the Maths Box containing '*$\ldots \times \boldsymbol{sign(sample)}$*' detects a change and draws a new sample. However,* $sign(x) = 1$ *and doesn't affect the value of the result. It serves only to trigger selection of a new sample.*



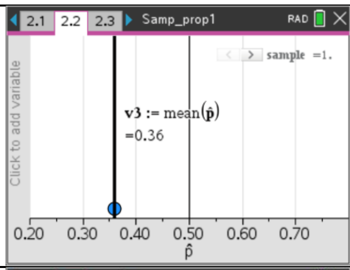
To capture the value of the sample proportion:

- From p. 2.2, add a **Lists & Spreadsheet** page. This becomes p. 2.3.
- In the column A title cell, input $\hat{\boldsymbol{p}}$ and press [enter] (*Note. select* $\hat{p}$ *from the* [ctrl] [∞β°] *list*).
- With the column A formula cell selected, press [menu] > **Data** > **Data Capture** > **Automatic**.
- Enter $\hat{\boldsymbol{p}} \coloneqq \boldsymbol{capture}('\boldsymbol{samp\_p}, \mathbf{1}, \boldsymbol{sample})$. *Note. A little-known trick is used here. The* ***capture*** *command will not normally capture consecutive identical values. By adding '**…, sample**' to the* ***capture*** *command, it detects a change and captures* ***all*** *values, including consecutive values that are identical.)*



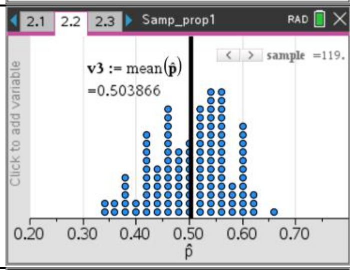
In preparation to plot sample proportions captured in column A:

- On page 2.2, press [menu] > **Window/Zoom** > **Window Settings** and set XMin: **0.2**, XMax: **0.8**.
- Using [menu] > **Analyse** > **Plot value**, set up plot values $\boldsymbol{v2} \coloneqq \boldsymbol{p}$ and $\boldsymbol{v3} \coloneqq \boldsymbol{mean}(\hat{\boldsymbol{p}})$



To populate column A and plot the captured values:

- On page 2.2, click or animate the slider to select new samples.
- As the number of samples taken increases, observe the behaviour of plot value $\boldsymbol{v3} \coloneqq \boldsymbol{mean}(\hat{\boldsymbol{p}})$**.**



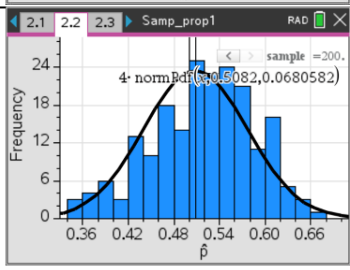
To obtain a histogram of the sample data:

- On page 1.2, press [ctrl] [menu] > **Histogram**.

To compare the data to a normal distribution curve:

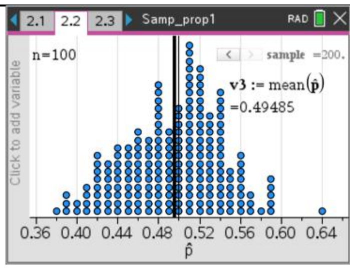
- Press [menu] > **Analyse** > **Show Normal PDF**.

Compare normal pdf parameters with $\mathrm{E}(\hat{P})$ and $\mathrm{sd}(\hat{P}) = \sqrt{\frac{p(1-p)}{n}}$



Explore the effect of changing the sample size, n, and the population proportion, p, by making one change at a time on page 2.1 and resetting the simulation each time.
To **reset the simulation**:

- On page 2.2, set the slider value to **1**.
- On page 2.3, click the column A formula cell,
- Press [ctrl] [menu] > **Clear Data**.



Activity 3 – Investigation: Number of Bernoulli trials needed for the first 'success'

Connection 2023-2027 VCE Study Design

MM Unit 1 Area of Study 4

simulation using simple random generators … and pseudo-random generators using technology, and the display and interpretation of results …

MM Unit 2 Area of Study 4

simulation to estimate probabilities involving selection with and without replacement

MM Units 3&4 Area of Study 4

Discrete random variables. Bernoulli trials …

The Investigation.

(a) Investigate the number of rolls of a fair six-sided die needed for the first 'six'.

This concept is about the probability of needing a certain number of trials before achieving the first success in a series of Bernoulli trials.

Extension

(a) Use simulation to explore the sampling distribution of the sample mean when samples of size n are taken from the asymmetric distribution generated in part (a) above.

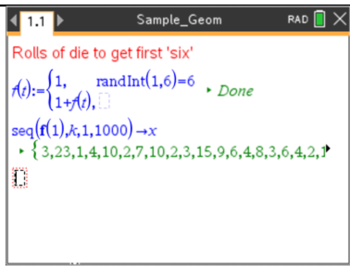
(a) To set up a simulation for the number of rolls of a die needed to get the first 'six', on a **Notes** page define a piecewise recursive function, $f(t)$, as follows.

- Press ctrl M to insert a **Maths Box**.
- In the **Maths Box**, enter $\boldsymbol{f(t) := \begin{cases} 1, & \text{randInt}(1,6) = 6 \\ 1 + f(t), & \text{else} \end{cases}}$.

Notes. (1) *Press* [template key] *to select the piecewise template*. (2) *The 't' in the recursive function is a 'dummy' input. The function adds 1 to the previous value until the random integer is a '6'.*

- Insert a new **Maths Box** and input $\boldsymbol{\text{seq}(f(1),k,1,100)}$
- Press ctrl sto→, input $\boldsymbol{x}$ then press enter.

The list of 'number of rolls for first 'six'' is stored as list $\boldsymbol{x}$.

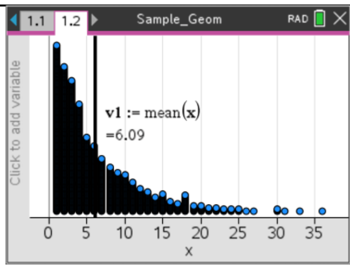


To obtain a plot of a model of the distribution of the number of rolls for the first 'six', X, add a **Data & Statistics** page, then:

- Press tab and select $\boldsymbol{x}$ on the horizontal axis.
- Press menu > **Analyse** > **Plot Value** and enter $\boldsymbol{v1 := \text{mean}(x)}$

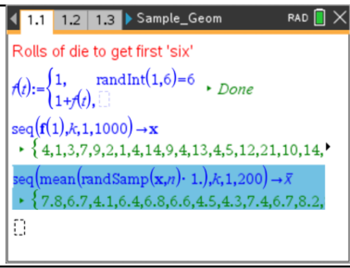
Note. This simulation models the Geometric distribution for the number of Bernoulli trials until the first 'success', with parameter $p = \frac{1}{6}$. *The Geometric distribution is the discrete analogue of the Exponential distribution, which is studied in Topic 2.*

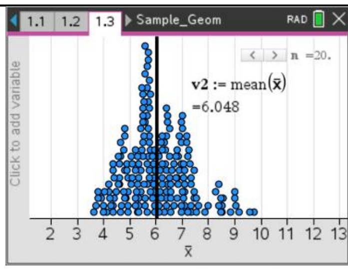
$$\text{E}(X) = \tfrac{1}{p} = 6 \text{ and } \sigma = \text{sd}(X) = \frac{\sqrt{1-p}}{p} = \frac{\sqrt{\frac{5}{6}}}{\frac{1}{6}} \approx 5.46$$



(b) To explore the sampling distribution of the sample mean when samples of size n are taken from the distribution generated in part (a) above, add a new **Data & Statistics** page, then:

- Press menu > **Actions** > **Insert Slider**.
- Enter slider settings: Variable: $\boldsymbol{n}$, Value: **20**, Min.: **5**, Max.: **100**, Step Size: **10**, Minimise ☑



• Navigate to the **Notes** page 1.1 and in a new **Maths Box** input **seq(mean(randSamp(*x*,*n*)·1.0),*k*,1,200)** (don't press [enter] yet.) • Press [ctrl] [sto→] then press [ctrl] [∞β°], select $\bar{x}$ then press [enter]. *Note. This will select 200 samples of size* n *(slider value) from list* x*, calculate the sample means and store the means as list* $\bar{x}$.	
To plot a model of the sampling distribution of sample means taken from the asymmetric distribution, navigate to **Data & Statistics** page 1.3, then: • Click below the horizontal axis and select the variable $\bar{x}$. • Press [menu] > **Analyse** > **Plot Value** and enter $\boldsymbol{v2} := \textbf{mean}(\bar{x})$ (select $\bar{x}$ from [var] menu). • Observe the effect of on the dot plot of increasing the sample size, n, using the slider.	
To observe the approximate normality of the model of the sampling distribution of sample means using a Normal PDF curve, on the **Data & Statistics** page 1.3: • Set the slider value to $n = 20$. • Press [menu] > **Plot Type** > **Histogram**. • Press [menu] > **Window/Zoom** > **Zoom – Data** • Press [menu] > **Analyse** > **Show Normal PDF**. • Edit the value of ***n*** to larger sample sizes and observe the effect on the shape and spread of the histogram and the Normal PDF parameters. • Adjust the bin width and zoom to the data as necessary. *Note. This simulation should demonstrate that, despite the samples being drawn from such a skewed distribution, the distribution of sample means is approximately Normal. The Normal PDF parameters for* $n = 20$ *should be consistent with* $\mathrm{E}(\bar{X}) = \mu = 6$ *and* $\mathrm{sd}(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{5.46}{\sqrt{20}} \approx 1.22$.