



Probability Distribution II

Ma. Louise De Las Penas, Phd

Ateneo de Manila University
Philippines

LEVEL

High School and University

OBJECTIVES

The aim of this note is to study probability distributions of discrete random variables with the aid of the graphics calculator

Corresponding eActivity

probdist.g1e

OVERVIEW

The **probability distribution of a discrete random variable** (random variable that assumes countable values) lists all the possible values that the random variable can assume and their corresponding probabilities.

In this paper, we will illustrate how the graphics calculator is a helpful tool in the numerical and graphical analysis of probability distributions of discrete random variables.

EXPLORATORY ACTIVITY

A consumer agency surveyed all 2750 families living in a small town to collect data on the number of television sets owned by them. The following table lists the frequency distribution of the data collected by the agency.

NUMBER OF TV SETS OWNED	0	1	2	3	4	5	6	7
Number of Families	120	970	730	410	270	150	75	25

- Construct a probability distribution table for the numbers of television sets owned by these families.
- Give a graphical presentation of the probability distribution.
- Let x denote the number of television sets owned by a randomly selected family from this town. Find the following probabilities:

Probability Distribution II

1. $P(x=1)$
2. $P(x>2)$
3. $P(x\leq 1)$
4. $P(1\leq x\leq 3)$

d. Calculate the mean and standard deviation for the probability distribution

Solution:

a. We access the Spreadsheet Editor.

Let x denote the number of television sets owned by a randomly selected family from this town. We enter the values of x in column A and the number of families (which correspond to the frequencies of each of the television sets) in column B as follows:

SHEET	A	B	C	D
1	0	120		
2	1	970		
3	2	730		
4	3	410		
5	4	270		

SHEET	A	B	C	D
6	5	150		
7	6	75		
8	7	25		
9				
10				

We first convert the frequencies in column B to relative frequencies, and store these in column C . We divide every entry of column B by 2750, the total number of families participating in the survey.

We enter [= \$B1÷2750] and position the cursor in the third column row 1, or C1(left screen below). The corresponding result will appear in the position specified and the formula entered also appears(right screen below).

SHEET	A	B	C	D
1	0	120		
2	1	970		
3	2	730		
4	3	410		
5	4	270		

= \$B1÷2750

SHEET	A	B	C	D
1	0	120	0.0436	
2	1	970		
3	2	730		
4	3	410		
5	4	270		

= \$B1÷2750

We copy the formula and paste this to the succeeding rows in the third column. Doing this will obtain the relative frequencies corresponding to all the other entries in column B :

SHEET	A	B	C	D
1	0	120	0.0436	
2	1	970		
3	2	730		
4	3	410		
5	4	270		

= \$B1÷2750

SHEET	A	B	C	D
1	0	120	0.0436	
2	1	970		
3	2	730		
4	3	410		
5	4	270		

= \$B1÷2750

SHEET	A	B	C	D
1	0	120	0.0436	
2	1	970	0.3527	
3	2	730		
4	3	410		
5	4	270		

= \$B2÷2750

SHEET	A	B	C	D
1	0	120	0.0436	
2	1	970	0.3527	
3	2	730	0.2654	
4	3	410		
5	4	270		

= \$B3÷2750

Probability Distribution II

Constructing the probability distribution using the Spreadsheet Editor allows the learner to explore and verify easily the characteristics of the probability distribution of a discrete random variable.

For instance, the relative frequencies, or the approximate probabilities obtained from the sample lies in the range 0 to 1. Moreover, the sum of the probabilities assigned to all possible values of x is equal to 1.

SHEE	A	B	C	D
6	5	150	0.0545	
7	6	75	0.0272	
8	7	25	9E-3	
9				
10				
=CellSum(C1:C8)				

SHEE	A	B	C	D
6	5	150	0.0545	
7	6	75	0.0272	
8	7	25	9E-3	
9				
10				
=CellSum(C1:C8)				

The sum of cells in the spreadsheet can be obtained using the command [= Cel Sum]. The cell identifiers, (consisting of the column letter and row number) of the cells to be added is specified.

As seen in the right screen above the sum of all probabilities of the probability distribution is equal to 1.

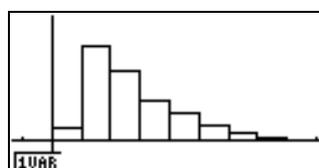
b. The probability distribution of a discrete random variable can be represented graphically using a histogram.

To display the histogram for the given data, we access the GRPH menu. We assign the first graph, GPH1 to the given data. We specify the columns representing the x values and the relative frequencies, then select " histogram"[Scatter] as the graph type as follows:

StatGraph1	
Graph Type:	Hist
XCellRange:	A1:A8
Frequency :	C1:C8
GPH1 GPH2 GPH3	

The histogram is given below. We use the width 1 for the each bar in the graph.

SH	
Histogram Settings	
Start:	0
Width:	1
Draw:	[EXEC]
GPH1 GPH2 GPH3 SEL	
SET	



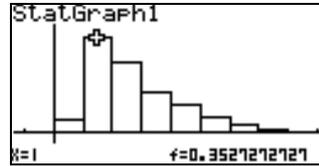
c.

1. $P(x=1)$

The probability of a family having exactly one television set can be obtained directly from the spreadsheet. This can also be obtained when we trace the histogram. From the histogram, we can visualize that more families have 1 television set.

Probability Distribution II

SHEET	A	B	C	D
1	0	120	0.0436	
2	1	970	0.3521	
3	2	730	0.2654	
4	3	410	0.149	
5	4	270	0.0981	
			=B2÷2750	



2. $P(x > 2)$

The probability that a family owns more than two television sets is given by the sum of the probabilities of owning three, four, five, six and seven television sets. Thus,

$$P(x > 2) = P(x=3) + P(x=4) + P(x=5) + P(x=6) + P(x=7)$$

Using the calculator, we add rows 4 to 8 of column C of the spreadsheet, corresponding to the formula: $P(x=3) + P(x=4) + P(x=5) + P(x=6) + P(x=7)$. We store the sum in a location of one's choice in the spreadsheet:

SHEET	A	B	C	D
7	6	75	0.0272	
8	7	25	9E-3	
9				
10				
11				
			=CellSum(C4:C8)	

SHEET	A	B	C	D
7	6	75	0.0272	
8	7	25	9E-3	
9				
10	0.3381			
11				
			=CellSum(C4:C8)	

The result is given by $P(x > 2) = 0.3381$.

3. $P(x \leq 1)$

The probability of a family having at most one television set is given by

$$P(x \leq 1) = P(x=0) + P(x=1)$$

Using the calculator, we add rows 1 and 2 of column C of the spreadsheet, corresponding to the formula $P(x=0) + P(x=1)$:

SHEET	A	B	C	D
7	6	75	0.0272	
8	7	25	9E-3	
9				
10				
11				
			=CellSum(C1:C2)	

SHEET	A	B	C	D
7	6	75	0.0272	
8	7	25	9E-3	
9				
10	0.3963			
11				
			=CellSum(C1:C2)	

The answer is $P(x \leq 1) = 0.3963$.

4. $P(1 \leq x \leq 3)$

The probability of a family having one to three television sets is given by the sum of the probabilities of one, two and three television sets.

$$P(1 \leq x \leq 3) = P(x=1) + P(x=2) + P(x=3)$$

We add rows 2, 3 and 4 of column C of the spreadsheet, corresponding to the formula

Probability Distribution II

$$P(x=1) + P(x=2) + P(x=3):$$

SHEE	A	B	C	D
7	6	75	0.0272	
8	7	25	9E-3	
9				
10				
11				
=CellSum(C2:C4)				

SHEE	A	B	C	D
7	6	75	0.0272	
8	7	25	9E-3	
9				
10	0.7672			
11				
=CellSum(C2:C4)				

The answer we obtain is $P(1 \leq x \leq 3) = 0.7672$.

d. The 1Var calculations can be used to determine some information about the probability distribution. We enter [CALC 1Var] after adjusting the calculation settings(left screen below). The mean number of televisions per family is about 2.2. This is also the expected value of x , that is $E(x) = 2.2$. The standard deviation is 1.47.

1Var XCell:A1:A8
1Var Freq :C1:C8
2Var XCell:A1:A5
2Var YCell:B1:B5
2Var Freq :1
ICELL

1-Variable
\bar{x} =2.22363636
Σx =2.22363636
Σx^2 =7.11818181
$x\sigma n$ =1.47432124
$x\sigma n-1$ =
n =1

Remarks: In the study of frequency distributions, it is at times helpful to refer to the cumulative probability distribution.

We illustrate how a cumulative probability distribution is obtained in the Spreadsheet Editor. We refer to the previous example given.

We first copy the first entry in the relative frequencies column, in this case C1, to the topmost row of the fourth column, D1, which will store our cumulative relative frequencies.

SHEE	A	B	C	D
1	0	120	0.0436	0.0436
2	1	970	0.3527	
3	2	730	0.2654	
4	3	410	0.149	
5	4	270	0.0981	
=B1+2750				

Using [= GRAB] we enter the sum formula for the first two relative frequencies, D1 + C2 and store in D2:

SHEE	A	B	C	D
1	0	120	0.0436	0.0436
2	1	970	0.3527	
3	2	730	0.2654	
4	3	410	0.149	
5	4	270	0.0981	
=				

SHEE	A	B	C	D
1	0	120	0.0436	0.0436
2	1	970	0.3527	0.3963
3	2	730	0.2654	
4	3	410	0.149	
5	4	270	0.0981	
=D1+C2				

We repeat this formula throughout the fourth column, by the copy and paste command. In the 3rd row of column 4, we have the sum of the first three relative frequencies, D2+C3; in the 4th row, D3+C4 and so on. Finally in the last entry the sum of all relative

Probability Distribution II

frequencies must be 1.

SHEET	A	B	C	D
1	0	120	0.0436	0.0436
2	1	970	0.3527	0.3963
3	2	730	0.2654	0.6618
4	3	410	0.149	0.8109
5	4	270	0.0981	0.909

PASTE

SHEET	A	B	C	D
1	0	120	0.0436	0.0436
2	1	970	0.3527	0.3963
3	2	730	0.2654	0.6618
4	3	410	0.149	0.8109
5	4	270	0.0981	0.909

PASTE =D2+C3

SHEET	A	B	C	D
1	0	120	0.0436	0.0436
2	1	970	0.3527	0.3963
3	2	730	0.2654	0.6618
4	3	410	0.149	0.8109
5	4	270	0.0981	0.909

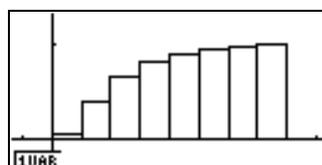
PASTE =D3+C4

SHEET	A	B	C	D
4	3	410	0.149	0.8109
5	4	270	0.0981	0.909
6	5	150	0.0545	0.9636
7	6	75	0.0272	0.9909
8	7	25	9E-3	1

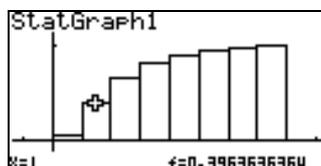
PASTE =D7+C8

The graph of the cumulative probability distribution is given as follows:

```
StatGraph1
Graph Type:Hist
XCellRange:A1:A8
Frequency :D1:D8
```



We can answer the question given earlier in Item c of number 3, $P(x \leq 1) = 0.3963636364$ by tracing the cumulative probability distribution:



Remarks: After entering the data and other relevant settings during the activity, we save all information in the spreadsheet menu. We call the file name "TVSETS" and may be referred to at a later time. Similarly, for the exercise that follows, we call the file name "Machines".

```
Spread Sheet Name
[TVSETS ]
```

```
Spread Sheet Name
[MACHINES]
```

Exercise: Elmo's Sporting Goods sells exercise machines as well as other sporting goods. On different days, it sells different numbers of these machines. The table that follows is constructed using past data and lists the probability distribution of the number of exercise machines sold per day at Elmo's.

MACHINES SOLD PER DAY	4	5	6	7	8	9	10
Probability	0.08	0.11	0.14	0.19	0.20	0.16	0.12

Probability Distribution II

- Give a graphical presentation of the probability distribution.
- Determine the probability that the number of exercise machines sold at Elmo's on a given day is
 - exactly 6
 - less than 7
 - 5 to 8
 - at most 6
- Calculate the mean and standard deviation for the probability distribution

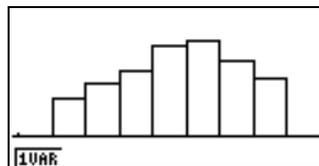
Solution:

- Let x be the number of machines sold per day. We access the Spreadsheet Editor, and enter the x values in the first column and the probabilities in the second column.

SHEE	A	B	C	D
1	4	0.08		
2	5	0.11		
3	6	0.14		
4	7	0.19		
5	8	0.2		

SHEE	A	B	C	D
6	9	0.16		
7	10	0.12		
8				
9				
10				

- The histogram is given as follows:



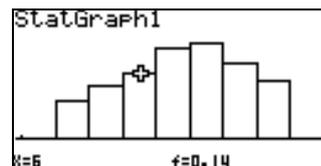
The following settings are used:

StatGraph1	
Graph Type:	Hist
XCellRange:	A1:A7
Frequency:	B1:B7

Histogram Settings	
Start:	4
Width:	1
Draw:	[EXE]

- The probability that the number of exercise machines sold at Elmo's on a given day is exactly 6 is given by 0.14. This value can be obtained directly from the spreadsheet. The value also appears when we trace the histogram.

SHEE	A	B	C	D
1	4	0.08		
2	5	0.11		
3	6	0.14		
4	7	0.19		
5	8	0.2		



- The probability that the number of exercise machines sold at Elmo's on a given day is

Probability Distribution II

less than 7 is given by

$$P(x < 7) = P(x=6) + P(x=5) + P(x=4) = 0.33,$$

since $P(x=3) = P(x=2) = P(x=1) = 0$.

SHEET	A	B	C	D
5	8	0.2		
6	9	0.16		
7	10	0.12		
8				
9				
		=CellSum(B1:B3)		

SHEET	A	B	C	D
6	9	0.16		
7	10	0.12		
8				
9		0.33		
10				

As seen from the screen dump above, we add the first to third rows in column B.

3. The probability that the number of exercise machines sold at Elmo's on a given day is 5 to 8 is given by

$$P(5 \leq x \leq 8) = P(x=5) + P(x=6) + P(x=7) + P(x=8)$$

The sum is obtained by adding the 2nd row to the 5th row of column B.

SHEET	A	B	C	D
6	9	0.16		
7	10	0.12		
8				
9				
10				
		=CellSum(B2:B5)		

SHEET	A	B	C	D
6	9	0.16		
7	10	0.12		
8				
9		0.64		
10				
		=CellSum(B2:B5)		

4. The probability that the number of exercise machines sold at Elmo's on a given day is at most 6 is obtained through the formula

$$P(x \leq 6) = P(x=4) + P(x=5) + P(x=6) = 0.33$$

which is calculated in #2 above.

c. The mean and the standard deviation is obtained from the 1VAR screen as follows:

1Var XCell: A1:A7
1Var Freq : B1:B7
2Var XCell: A1:A5
2Var YCell: B1:B5
2Var Freq : 1
CELL

1-Variable
\bar{x} = 7.28
S_x = 7.28
S_x^2 = 56.14
σ_n = 1.77245592
σ_{n-1} =
n = 1

The mean is given by 7.28 and the standard deviation is 1.77245592.

Remarks:

The statistical calculations and graphs relative to the probability distributions obtained in the Spreadsheet Editor can also be obtained in the Stat Editor using almost the same steps. Some slight changes in the commands are given below:

Probability Distribution II

The x values are stored in List1 and the frequencies in List 2. Consider for instance the exploratory activity given earlier. The data is given as follows:

	List 1	List 2	List 3	List 4
SUB				
2	1	970		
3	2	730		
4	3	410		
5	4	270		

GRPH CALC TEST INTR DIST

Frequencies are converted to relative frequencies and stored in List 3 using the command:

	List 1	List 2	List 3	List 4
SUB				
1	0	120		
2	1	970		
3	2	730		
4	3	410		

List 2=2750

	List 1	List 2	List 3	List 4
SUB				
1	0	120	0.0436	
2	1	970	0.3527	
3	2	730	0.2654	
4	3	410	0.149	

0.04363636364
List L→M Dim Fill Seq

The cumulative [CUM] command on the probability distribution produces the probability distribution in List 4. Observe that the sum of all relative frequencies is 1.

	List 1	List 2	List 3	List 4
SUB				
1	0	120	0.0436	
2	1	970	0.3527	
3	2	730	0.2654	
4	3	410	0.149	

Cum1 List 3
List L→M Dim Fill Seq

	List 1	List 2	List 3	List 4
SUB				
1	0	120	0.0436	0.0436
2	1	970	0.3527	0.3963
3	2	730	0.2654	0.6618
4	3	410	0.149	0.8109

0.04363636364
List L→M Dim Fill Seq

	List 1	List 2	List 3	List 4
SUB				
5	4	270	0.0981	0.909
6	5	150	0.0545	0.9636
7	6	75	0.0272	0.9909
8	7	25	9E-3	1

Sum Prod Cuml % 1
List L→M Dim Fill Seq

REFERENCE

[1] Mann, Prem. *Introductory Statistics*, 5th Edition, Wiley and Sons, Inc 2004.