Laboratory Assignment 1. Probability Distributions

The Objective: to study discrete and continuous probability distributions and explore their characteristics using Mathcad.

Description.

This laboratory work consists of three parts: 1) Discrete Probability Distributions; 2) Continuous Probability Distributions; 3) Characteristics of Random Variables.

During the first part, you will learn about ***pmf*** and ***cdf*** of discrete probability distributions and find out the influence of their parameters on the form of ***pmf***. The distributions under consideration are:

- ***binomial;***
- ***Poisson;***
- ***geometric;***
- ***negative binomial;***
- ***hypergeometric;***
- ***discrete uniform.***

Some of these distributions were covered in Lecture 4, others were not - you should explore them by yourself. Also, there are built-in Mathcad functions for all distributions mentioned above except for the ***discrete uniform***. You should write your own functions for it in Mathcad.

You can consult recommended textbooks, slides, and Internet sources (Wikipedia included) to learn more on ***negative binomial*** and ***hypergeometric*** distributions. You are advised to thoroughly examine the sections of *Mathcad Help* concerning probability distributions and *Quick Sheets* on *Statistics*.

Please, pay attention to some discrepancies between the notations in textbooks and in *Mathcad Help*: some parameters can be denoted by different letters.

The part of your report concerning discrete probability distributions must include Mathcad file (or files) and the textual part presenting brief descriptions of the distributions under study, formulae for their ***pmf*** and ***cdf***, and your observations of the influence of distributions' parameters on ***pmf***s.

The ***binomial*** distribution is considered below as an example. The other distributions should be dealt with by an analogy with few adjustments.

Example. Binomial Distribution

Preliminary Notes

Mathcad has built-in functions for several probability distributions; full list is available in *Mathcad Help → Functions → Statistics → Density and Distribution Functions*. All Mathcad functions for ***cdf*** start with letter "*p*" followed by abbreviation for the name of the distribution. The functions for ***pdf*** (and ***pmf***) start with "*d*" followed by the same abbreviation. Therefore, in case of ***binomial*** distribution, ***pbinom*** returns its ***cdf*** and ***dbinom*** returns its ***pmf***.

The variable and distribution parameters must be specified in brackets.

Brief Description

In probability theory and statistics, the binomial distribution with parameters $\boldsymbol{n} \geq 0$ and $\boldsymbol{p} \in [\boldsymbol{0}, \boldsymbol{1}]$ is the discrete probability distribution of the number of successes in a sequence of ***n*** Bernoulli trials. The binomial distribution is frequently used to model the number of successes in a sample of size ***n*** drawn with replacement from a population of size ***N***. [Wikipedia]

The domain for a binomial random variable with parameter ***n*** is a set {0,1,2,…, *n*}.

The probability mass function:

$$p_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} \;\; for \;\; 0 \leq k \leq n; \\ 0 \quad otherwise. \end{cases}$$

The cumulative distribution function:

$$F_X(x) = \begin{cases} 0, \quad if\ x < 0; \\ \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}, \qquad if\ \ 0 \leq x < n; \\ 1, \quad if\ x \geq n. \end{cases}$$

Remark

Henceforward, we assume the argument of **pmf** *is always an integer number, whereas the argument of* **cdf** *is real.*

Start with defining the distribution parameters. You can set the values arbitrarily.

$$n := 15 \qquad\qquad p := 0.3$$

Next, you should define ranges of variables for plotting the graphs of ***pmf*** and ***cdf***.

$$k := 0..n \qquad x := -2, -1.999..17$$

To demonstrate the behavior of the ***cdf*** outside the support of the distribution you should extend the range slightly for variable ***x***. Also, for distributions with infinite support, the bounds for the variables ***k*** and ***x*** should be set such that no significant changes in values of functions were visible.

The next step is defining functions of ***pmf*** and ***cdf***:

$$p_X(x) := \mathrm{dbinom}(x, n, p) \qquad F_X(x) := \mathrm{pbinom}(x, n, p)$$

The use of subscript is not required. However, it is useful to avoid confusion between the notation of ***pmf*** and parameter ***p***. You can add the subscript by pressing "dot" key on your keyboard after the name of the function.

After plotting the ***pmf***, you should see the graph similar to Figure 1.

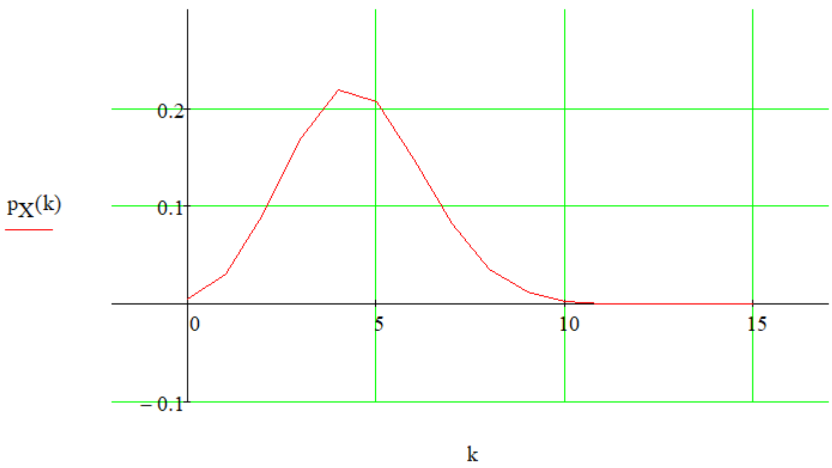

Figure 1

However, it is common practice to draw ***pmf*** as a stem plot. To change the view, you should open "*Formatting Currently Selecting X-Y Plot*" dialog and select *Type* of your graph to "stem", as shown in Figure 2.

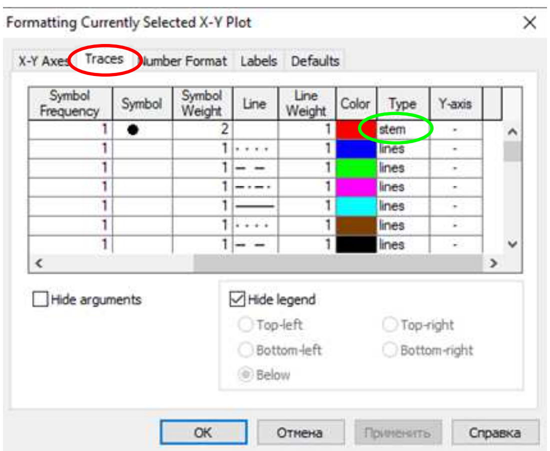

Figure 2

Figure 3 demonstrates the resulting ***pmf*** plot.

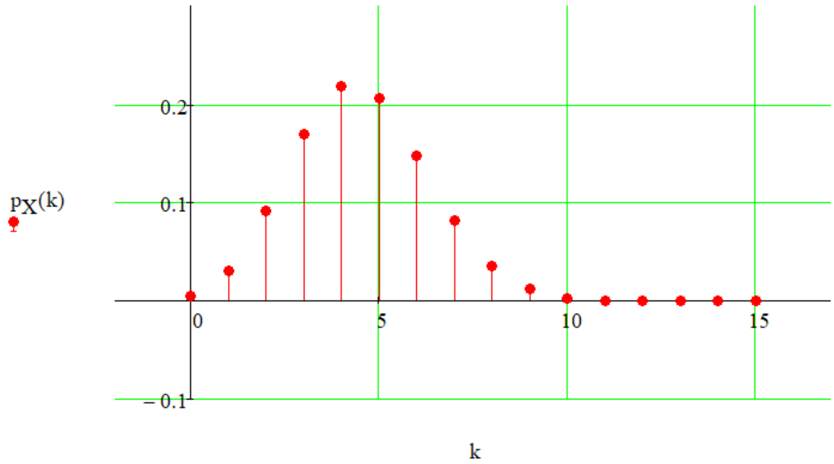

Figure 3

One of the major features of your graph is the most likely value of the random variable, known as the **mode**. As you can see, the mode for the ***Bin****(15, 0.3)* is 4. The value of probability for the mode obtained as

$$p_X(4) = 0.219$$

You should write down this result or put it in the table in order to trace the influence of distribution parameters on the mode.

Next, we should plot the cdf. The result is shown in Figure 4.

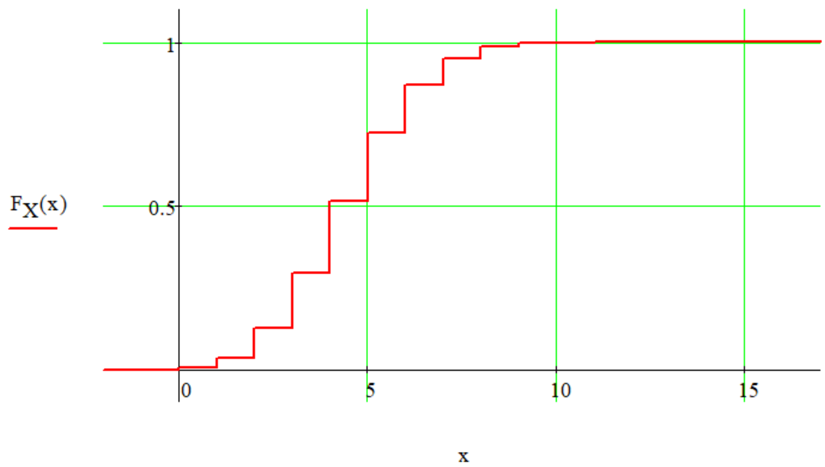

Figure 4

The graph of ***cdf*** is good as it is. Still, it is better to mark the values of ***cdf*** for all ***x = k***. We can add those values on the same plot (as points) as shown in Figure 5.

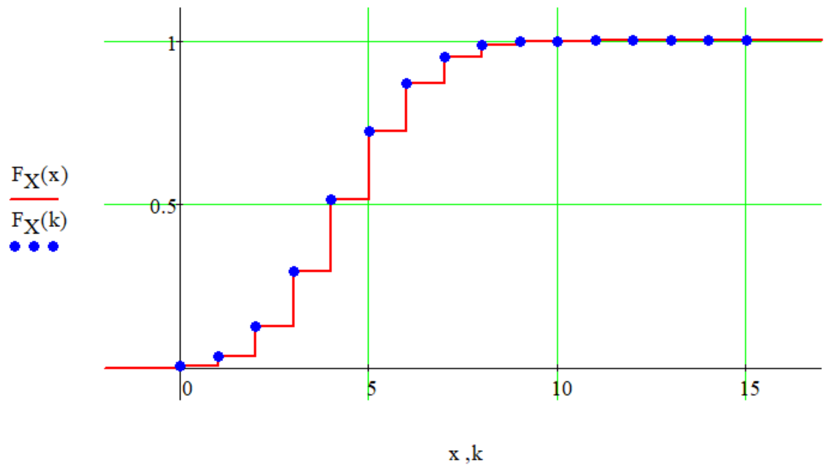

Figure 5

Now, we need to explore how the distribution parameters influence the ***pmf***. The binomial distribution has two parameters, so, it is necessary to find out the impact of each parameter separately.

First, we increase parameter ***n*** while keeping ***p*** as initially defined (Figure 6).

$$n := 16 \qquad p := 0.3$$

$$p_X(x) := \mathrm{dbinom}(x, n, p)$$

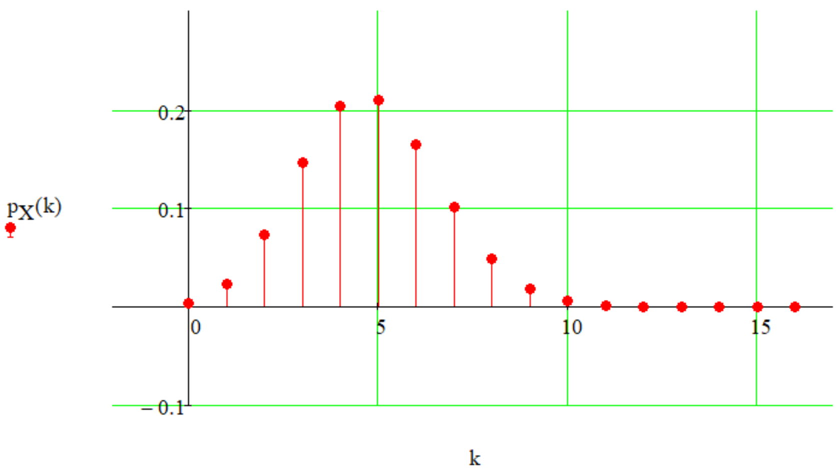

Figure 6

We can observe the value of the mode has increased (x = 5), and the value of probability for the mode has decreased

$$p_X(5) = 0.21$$

Now, we can reason that mode increases with the increase of ***n*** and its probability decreases. However, we need to try out several different values of ***n***, both greater and less than our initial value, in order to make sure that this relation is valid.

Next, we increase parameter ***p*** maintaining ***n*** at its initial value (Figure 7).

$$n := 15 \qquad p := 0.4$$

$$p_X(x) := \text{dbinom}(x, n, p)$$

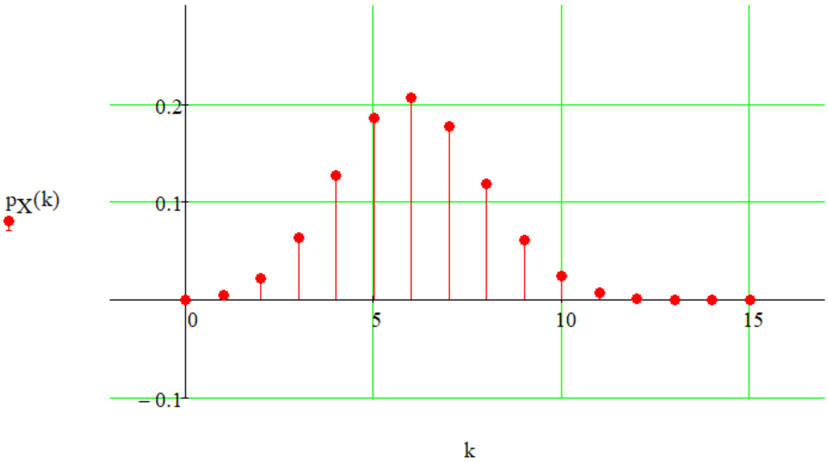

Figure 7

Then, we should repeat the chain of actions to find out the impact parameter ***p*** has on the mode and its probability.

Be sure to explore thoroughly the behavior of the ***pmf*** with regard to distribution parameters. In your report, you should present your findings backing them up with tables and/or plots.

You should adhere to this plan of action for each probability distribution.