LAB 2 INSTRUCTIONS

PROBABILITY DISTRIBUTIONS

There is a wide range of discrete probability distributions available in Excel. They can be accessed through the **Insert Function** feature. The most common applications for any discrete probability distribution are those that return an exact probability (probability mass function) or return the value for a cumulative probability (cumulative distribution function).

In this lab, we will discuss some of the above applications for binomial, negative binomial, and Poisson distributions. Examples are provided to illustrate how to use the tools in simple problems. For **Activating the Data Analysis Add-In** or **Inserting Excel Output into a Word Document**, see the **Lab 1 instructions**.

1. Binomial Distribution

The distribution of the number of successes (X) in n independent trials, each with the same probability of success (p), is the binomial distribution with parameters n and p. The binomial probabilities in Excel can be obtained by the BINOM.DIST or BINOMDIST functions. Either version takes four arguments as described below or via Microsoft.

BINOM.DIST(number_s,trials,probability_s,cumulative)

The BINOM.DIST function syntax has the following arguments:

number_s	The number of successes in trials.
trials	The number of independent trials.
probability_s	The probability of success on each trial.
cumulative	A logical value that determines the form of the function. If cumulative is TRUE, then BINOM.DIST returns the cumulative distribution function, which is the probability that there are at most **number_s** successes; if FALSE, it returns the probability mass function, which is the probability that there are exactly **number_s** successes.

The arguments in the BINOM.DIST function must satisfy the following conditions: **number_s** is a non-negative integer, **trials** is a positive integer (greater or equal to **number_s**), **probability_s** is between 0 and 1, and **cumulative** is either TRUE or FALSE.

Examples:

(a) Find the probability of getting exactly 10 correct answers in a 20 question multiple-choice exam, each with 5 options such that a student guesses randomly (indicating $p = 1/5 = 0.2$).

If $X \sim B(n = 10, p = 0.2)$, then $P(X = 10) = $ BINOM.DIST(10, 20, 0.2, FALSE) = 0.002031

(b) Find the probability of getting at least 11 correct answers in the same exam.

$P(X \geq 11) = 1 - P(X \leq 10) = 1 - $ BINOM.DIST(10, 20, 0.2, TRUE) = $1 - 0.999437 = 0.000563$

The interactive template *Binomial* available in the *lab2.xlsx* Excel file that can be downloaded from eClass allows you to calculate the binomial probabilities without using the function directly; rather, one enters the parameters of the binomial distribution. The exact probability (point probability) and cumulative probabilities (left-interval and right-interval) will be calculated automatically and displayed in your worksheet.

2. Negative Binomial (and Geometric) Distribution

The distribution of the number of trials (X) to obtain r successes, each with the same probability of success (p), is the negative binomial distribution with parameters r and p. The negative binomial probabilities in Excel can be obtained by the NEGBINOM.DIST function. The function takes four arguments as described below or via Microsoft. Note that NEGBINOMDIST only has three arguments, so cannot be used for cumulative probability.

NEGBINOM.DIST(number_f,number_s,probability_s,cumulative)

The NEGBINOM.DIST function syntax has the following arguments:

number_f   The number of failures.
number_s   The threshold number of successes.
probability_s   The probability of success.
cumulative   A logical value that determines the form of the function. If cumulative is TRUE, then NEGBINOM.DIST returns the cumulative distribution function; if FALSE, it returns the probability mass function.

The arguments in the NEGBINOM.DIST function must satisfy the following conditions: **number_f** and **number_s** are non-negative integers, **probability_s** is between 0 and 1, and **cumulative** is either TRUE or FALSE. Note that the geometric distribution is a special case of the negative binomial where **number_s** = 1.

Examples:

(a) Suppose a multiple-choice exam, each with 5 options such that a student guesses randomly (indicating $p = 1/5 = 0.2$). Find the probability of completing exactly 20 questions to obtain 10 correct answers. Note that **number_f** requires calculation from $r = 10$ and $X = 20$.

If $X \sim NB(r = 10, p = 0.2)$, then $P(X = 20) = $ NEGBINOM.DIST(20 − 10, 10, 0.2, FALSE) = 0.001016

(b) Find the probability of completing at least 21 questions to obtain 10 correct answers.

$P(X \geq 21) = 1 − P(X \leq 20) = 1 − $ NEGBINOM.DIST(20 − 10, 20, 0.2, TRUE) = 1 − 0.004072 = 0.995928

(c) Find the probability of completing exactly 20 questions to obtain the first correct answer.

If $X \sim NB(r = 1, p = 0.2)$, then $P(X = 20) = $ NEGBINOM.DIST(20 − 1, 1, 0.2, FALSE) = 0.002882

Using the geometric distribution formula, $P(X = 20) = (1 − p)^{x − 1}p = (1 − 0.2)^{20 − 1}(0.2) = 0.002882$, which is the same answer, but there is no such function in Excel, so NEGBINOM.DIST should be used as seen above.

(d) Find the probability of completing at least 21 questions to obtain the first correct answer.

$P(X \geq 21) = 1 − P(X \leq 20) = 1 − $ NEGBINOM.DIST(20 − 1, 1, 0.2, TRUE) = 1 − 0.988471 = 0.011529

The interactive template *Neg. Bin.* available in the *lab2.xlsx* Excel file that can be downloaded from eClass allows you to calculate the negative binomial probabilities without using the function directly; rather, one enters the parameters of the negative binomial distribution. The exact probability (point probability) and cumulative probabilities (left-interval and right-interval) will be calculated automatically and displayed in your worksheet.

3. Poisson Distribution

The distribution of the number of successes (X) over some period of time or space, with an average rate of success (λ), is the Poisson distribution with parameter λ. The Poisson probabilities in Excel can be obtained by the POISSON.DIST or POISSON functions. Either version takes three arguments as described below or via Microsoft.

POISSON.DIST(x,mean,cumulative)

The POISSON.DIST function syntax has the following arguments:

x   The number of successes.
mean   The average rate of success.
cumulative   A logical value that determines the form of the function. If cumulative is TRUE, then BINOM.DIST returns the cumulative distribution function; if FALSE, it returns the probability mass function.

The arguments in the POISSON.DIST function must satisfy the following conditions: **x** is a non-negative integer, **mean** is a positive real number, and cumulative is either TRUE or FALSE.

Examples:

(a) Suppose vehicles arrive at an intersection at a rate of 10 per minute. What is the probability that 7 vehicles will arrive in one minute?

If $X \sim Poisson(\lambda = 10)$, then $P(X = 7) = $ POISSON.DIST(7, 10, FALSE) $= 0.090079$

(b) Find the probability that at least 8 vehicles will arrive in one minute?

$P(X \geq 8) = 1 - P(X \leq 7) = 1 - $ POISSON.DIST(7, 10, TRUE) $= 1 - 0.220221 = 0.779779$

(c) A traffic light cycle lasts 45 seconds. Find the probability that exactly 7 vehicles will arrive at the intersection at a randomly chosen cycle.

Since 45 seconds is 0.75 minutes, then the number of vehicles that arrive at the intersection follows a Poisson distribution with a mean of $\lambda = 10 \times 0.75 = 7.5$ to adjust for the time period.

If $Y \sim Poisson(\lambda = 7.5)$, then $P(Y = 7) = $ POISSON.DIST(7, 7.5, FALSE) $= 0.146484$

(d) Using the traffic light cycle of 45 seconds, find the probability that at least 8 vehicles will arrive at the intersection at a randomly chosen cycle.
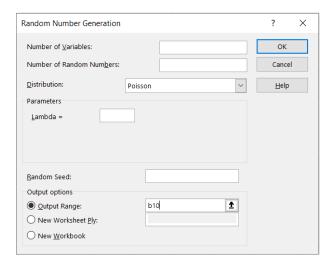
$P(Y \geq 8) = 1 - P(Y \leq 7) = 1 - $ POISSON.DIST(7, 7.5, TRUE) $= 1 - 0.524639 = 0.475361$

The interactive template *Poisson* available in the *lab2.xlsx* Excel file that can be downloaded from eClass allows you to calculate the Poisson probabilities without using the function directly; rather, one enters the parameters of the Poisson distribution. The exact probability (probability at x) and cumulative probability will be calculated automatically and displayed in your worksheet.

4. Using Excel to Generate Random Numbers

Excel includes the **Random Number Generation** feature that fills a range of a worksheet with random numbers from one of seven distributions: uniform, normal, Bernoulli, binomial, Poisson, patterned, and discrete.

In order to access the feature, choose the **Data** tab and click on the **Data Analysis** feature on the far right. (If the feature is not available, see the Lab 1 instructions to activate it.) Then, choose **Random Number Generation** from the available list and click **OK**. The following dialog box will appear.

For example, creating 50 samples of size $n = 5$ for a Poisson distribution with $\lambda = 1$ would require a value of 50 in the *Number of Variables* entry, a value of 5 in the *Number of Random Numbers* entry, *Poisson* to be selected from the drop-down menu for *Distribution*, and a value of 1 in the *Lambda* entry. For the same random numbers to be created for each student, a *Random Seed* is essential (see lab assignments for the number). For the **Output Options**, it is preferred to select *Output Range* and choose a single cell to present the upper-left corner of all the corresponding output.

5. COUNTIF Function

The COUNTIF function is used to count the number of cells in a given range that meet a single criterion (or several criteria contained in one overall expression). The function is accessible either from the **Insert Function** feature or by entering the following formula in a blank cell on the worksheet.

=COUNTIF(range, criteria)

The function takes two arguments as described below or via Microsoft.

The **range** argument is the cell addresses you want Excel to evaluate, and **criteria** is the value you want counted or the conditon to apply to the range.

Examples:

To count all cells that contain the label NO in the range A1:A100, use COUNTIF(A1:A100, "NO").

To count all cells in the range A1:A100 with entries exceeding 10, use COUNTIF(A1:A100,">10").

To count all cells in the range A1:A100 with entries identical to the contents of cell C1, use COUNTIF(A1:A100, C1).

To count all cells in the range A1:A100 with the entries from the interval [1,2], use the following.

COUNTIF(A1:A100,"<=2") – COUNTIF(A1:A100,"<=1")

To count all cells outside of the interval [1,2] in the same range, use the following.

COUNTIF(A1:A100,"<1") + COUNTIF(A1:A100,">2")