13.9: Discrete Distributions- hypergeometric, binomial, and poisson

Authors: Tommy DiRaimondo, Rob Carr, Marc Palmer, Matt Pickvet
Stewards: Kaveh Saba, Arthur Edge, Kathryn Meintel, Renu Rao

9.1 What are Discrete Distributions?

In order to explain the meaning and utility of discrete distributions, it is necessary to first introduce the topic of Random Variables. The term *random* in Random Variable refers to the idea that each individual outcome has an equal chance of occurring. Therefore, each outcome is randomly determined. For example, a fair coin has an equal probability of flipping up heads as it does flipping up tails. Hence, the random variable is determined by the outcome of flipping a coin. However, it is frequently the case in chemical engineering that when a process is performed we are only interested in some function of the outcome rather than the outcome itself. For example, in a chemical reactor we might want to know the probability that two molecules collide and react to form a product at a certain temperature, however it is more useful to know the sum of the these collisions that form product rather than the individual occurrences. This is analogous to rolling two dice and wanting to know the sum of the dice rather than the individual numbers rolled. These discrete values defined on some sample space compose what are called Random Variables.

9.1.1. Random Variable Example

Suppose that our experiment consists of tossing 4 fair coins. If we define a variable $X$ and let it denote the number of heads flipped, then $X$ is a random variable with possible discrete values of 0,1,2,3,4 with respective probabilities:

$$P\{X = 0\} = P\{T,T,T,T\} = \frac{1}{16}$$
We can do the same for the probability of rolling sums of two dice. If we let $Y$ denote the sum, then it is a random variable that takes on the values of 2, 3, 4, 5, 6, 7, 8, 9, 10, & 12. Rather than writing out the probability statements we can represent $Y$ graphically:

The graph plots the probability of the $Y$ for each possible random value in the sample space (y-axis) versus the random value in the sample space (x-axis). From the graph one can infer that the sum with the greatest probability is $Y = 7$.

These are just two ways one can describe a random variable. What this leads into is representing these random variables as functions of probabilities. These are called the discrete distributions or probability mass functions. Furthermore, independent random events with known probabilities can be lumped into a discrete Random Variable. The Random Variable is defined by certain criteria, such as flipping up heads a certain number of times using a fair coin. The probability of a certain random variable equaling a discrete value can then be described by a discrete distribution. Therefore, a discrete distribution is useful in determining the probability of an outcome value without having to perform the actual trials. For example, if we wanted to know the probability of rolling a six 100 times out of 1000 rolls a distribution can be used rather than actually rolling the dice 1000 times.

Note: Here is a good way to think of the difference between discrete and continuous values or distributions. If there are two continuous values in a set, then there exists an infinite number of other continuous values between them. However, discrete values can only take on specific values rather than infinite divisions between them. For example, a valve that can only be completely open or completely closed is analogous to a discrete distribution, while a valve that can change the degree of openness is analogous to a continuous distribution.
The three discrete distributions we discuss in this article are the binomial distribution, hypergeometric distribution, and poisson distribution.

9.2 Binomial Distribution

This type of discrete distribution is used only when both of the following conditions are met:

1. The test has only two possible outcomes
2. The sample must be random

If both of the above conditions are met, then one is able to use this distribution function to predict the probability of a desired result. Common applications of this distribution range from scientific and engineering applications to military and medical ones. For example, a binomial distribution may be used in determining whether a new drug being tested has or has not contributed (“yes” or “no” labels) to alleviating symptoms of a disease.

Since only two outcomes are possible, they can be denoted as \( M_S \) and \( M_F \) for the number of successes and number of failures respectively. The term \( p \) is the frequency with which the desired number, \( M_S \) or \( M_F \) will occur. The probability of a collection of the two outcomes is determined by the following equation.

\[
P(M_S, M_F) = k p^{M_S}(1-p)^{M_F}
\]

\( k \) in the above equation is simply a proportionality constant. For the binomial distribution it can be defined as the number of different combinations possible

\[
k = C(M_S + M_F) = \frac{(M_S + M_F)!}{M_S!M_F!}
\]

! is the factorial operator. For example, 4! = 4 * 3 * 2 * 1 and x! = x * (x - 1) * (x - 2) * ... * 2 * 1. In the above equation, the denominator, \( M_S!M_F! \), represents the number of ways one could arrange results containing \( M_S \) successes and \( M_F \) failures. Therefore, the total probability of a collection of the two outcomes can be described by combining the two above equations to produce the binomial distribution function.

\[
P(M_S, M_F) = \frac{(M_S + M_F)!}{M_S!M_F!} p^{M_S}(1-p)^{M_F}
\]

Which can be simplified as the equation

\[
\text{} \Gamma(X-k) - \frac{n!}{(n-k)!k!} p^k(1-p)^n
\]

for \( k = 0, 1, 2, 3 \ldots \).

In the above equation, \( n \) represents the total number of possibilities, or \( M_S + M_F \), and \( k \) represents the number of desired outcomes, or \( M_S \). These equations are valid for all non-negative integers of \( M_S \), \( M_F \), \( n \), and \( k \) and also for \( p \) values between 0 and 1.

Below is a sample binomial distribution for 30 random samples with a frequency of occurrence being 0.5 for either result.

https://eng.libretexts.org/Bookshelves/Industrial_and_Systems_Engineering/Book%3A_Chemical_Process_Dynamics_and_C…

Updated: Tue, 29 Dec 2020 13:33:12 GMT
Powered by
This example is synonymous to flipping a coin 30 times with $k$ being equal to the number of flips resulting in heads.

An important note about these distributions is that the area under the curve, or the probability of each integer added together, always sums up to 1.

One is able to determine the mean and standard deviation, which is described in the Basic Statistics article, as shown below.

\[
\mu = np \tag{5}
\]

\[
\sigma = \sqrt{np(1 - p)} \tag{6}
\]

An example of a common question that may arise when dealing with binomial distributions is “What is the probability of a certain result occurring 33 – 66% of the time?” In the above case, it would be synonymous to the likelihood of heads turning up between 10 and 20 of the trials. This can be determined by summing the individual probabilities of each integer from 10 to 20.

\[
P(33 \% > X > 66\%) = \sum_{\text{probabilities}} P(X=k) = P(10) + P(11) + ... + P(19) + P(20)
\]

\[
= 0.0280 + 0.0509 + ... + 0.0509 + 0.0280 = 0.9572
\]

The probability of heads resulting in 33 – 66% of the trials when a coin is flipped 30 times is 95.72%.

The probability for binomial distribution can also be calculated using Mathematica. This will simplify the calculations, save time, and reduce the errors associated with calculating the probability using other methods such as a calculator. The syntax needed to calculate the probability in Mathematica is shown below in the screen shot. An example is also given for various probabilities based on 10 coin tosses.
where \( k \) is the number of events, \( n \) is the number of independent samples, and \( p \) is the known probability. The example above also generates a set of values for the probabilities of getting exactly 0 to 10 heads out of 10 total tosses. This will be useful because it simplifies calculating probabilities such as getting 0-6 heads, 0-7 heads, etc. out of 10 tosses because the probabilities just need to be summed according to the exact probabilities generated by the table.

In addition to calling the binomial function generated above, Mathematica also has a built-in Binomial Distribution function that is displayed below:

\[
\text{PDF}[\text{BinomialDistribution}[n,p],k]
\]

where \( n, p, \) and \( k \) still represent the same variables as before.

This built-in function can be applied to the same coin toss example.

As expected, the probabilities calculated using the built-in binomial function matches the probabilities derived from before. Both methods can be used to ensure accuracy of results.

### 9.3 Poisson Distribution

**Note: The variables used in this section are defined the same way as seen above in the "Binomial Distribution" section.**

A Poisson distribution has several applications, and is essentially a derived limiting case of the binomial distribution. It is most applicable to a situation in which the total number of successes is known, but the number of trials is not. An example of such a situation would be if you know the mean expected number of cancer cells present per sample and wanted to determine the probability of finding 1.5 times that amount of cells in any given sample, you would use a Poisson distribution.

In order to derive the Poisson distribution, you must first start off with the binomial distribution function.

\[
P(X = k) = \frac{n^k}{(n-k)!} p^k (1-p)^{n-k} \quad \text{for} \quad k = 0, 1, 2, 3, \ldots
\]
Due to the fact that the Poisson distribution does not require an explicit statement of the total number of trials, you must eliminate \( n \) out of the binomial distribution function. This is done by first introducing a new variable (\( \mu \)), which is defined as the expected number of successes during the given interval and can be mathematically described as:

\[
\mu = np \quad (5)
\]

We can then solve this equation for \( p \), substitute into equation 5, and obtain equation 7, which is a modified version of the binomial distribution function.

\[
P \{ X = k \} = \frac{n!}{(n-k)!k!} \left( \frac{\mu}{n} \right)^k \left( 1 - \frac{\mu}{n} \right)^{n-k} \quad (7)
\]

If we keep \( \mu \) finite and allow the sample size to approach infinity we obtain equation 8, which is equal to the Poisson distribution function. (Notice to solve the Poisson distribution, you do not need to know the total number of trials)

\[
P \{ X = k \} = \frac{\mu^k e^{-\mu}}{k!} \quad (8)
\]

In the above equation:

- \( e \) is the base of the natural logarithm (\( e = 2.71828... \))
- \( k \) is the number of occurrences of an event
- \( k! \) is the factorial of \( k \)
- \( \mu \) is a positive real number, equal to the expected number of occurrences that occur during the given interval.

Using the Poisson distribution a probability versus number of successes plot can be made for several different numbers of successes.
This graph represents several key characteristics of a Poisson process.

1. The number of successes is independent of the time since the last event.
2. Probabilities, or outcomes, are discrete.
3. The probability of two or more successes in a short interval is extremely small (approximately zero).

These rules hint that the Poisson processes only model rare events, yet this is not true. What these characteristics say is that if you are limited to a small interval window, you will observe less than two events. It is important to know that the Poisson distribution is actually an approximation of the binomial distribution. As \( n \) increases and \( p \) decreases, the Poisson distribution becomes equal to the binomial distribution.

### 9.4 Hypergeometric Distribution

Note: The definitions of the variables in this section are different than the previous sections.

A hypergeometric distribution function is used only if the following three conditions can be met:

1. Only two outcomes are possible
2. The sample must be random
3. Selections are not replaced

Hypergeometric distributions are used to describe samples where the selections from a binary set of items are not replaced. This distribution applies in situations with a discrete number of elements in a group of \( N \) items where there are \( K \) items that are different. As a simple example imagine we are removing 10 balls from a jar of mixed red and green balls without replacing it. What is the chance of selecting 2 green balls? 5 red balls? In general terms, if there are \( K \) green balls and \( N \) total balls, the chance of selecting \( k \) green balls by selecting \( n \) balls is given by

\[
P(X = k) = \frac{\text{(# of ways to select k green balls)} \cdot \text{(# of ways to select n-k red balls)}}{\text{(total number of ways to select)}}
\]

(9)

As another example, say there are two reactants sitting in a CSTR totaling to \( N \) molecules. The third reactant can either react with A or B to make one of two products. Say now that there are \( K \) molecules of reactant A, and \( N-K \) molecules of reactant B. If we let \( n \) denote the number of molecules consumed, then the probability that \( k \) molecules of reactant of A were consumed can be described by the hypergeometric distribution. Note: This assumes that there is no reverse or side reactions with the products.

In mathematical terms this becomes

\[
\hat{P}\{X = k\} = \frac{(N - K)! \cdot K! \cdot (N - n)!}{k!(K-k)! \cdot (N-K)! \cdot (N - n)! \cdot (n-k)! \cdot N!}
\]

(10)

where,
• \( N \) = total number of items
• \( K \) = total number of items with desired trait
• \( n \) = number of items in the sample
• \( k \) = number of items with desired trait in the sample

This can be written in shorthand as

\[
P \{ X = k \} = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}
\]  
(11)

Where

\[
\binom{A}{B} = \frac{A!}{(A-B)!B!}
\]  
(12)

The formula can be simplified as follows: There are

\[
\binom{N}{n}
\]
possible samples (without replacement). There are

\[
\binom{K}{k}
\]
ways to obtain \( k \) green balls and there are

\[
\binom{N-k}{n-k}
\]
ways to fill out the rest of the sample with red balls.

If the probabilities \( P \) are plotted versus \( k \), then a distribution plot similar to the other types of distributions is seen.
Question:

Suppose that you have a bag filled with 50 marbles, 15 of which are green. What is the probability of choosing exactly 3 green marbles if a total of 10 marbles are selected?

Answer:

\[ P \{ X = 3 \} = \frac{\binom{15}{3} \binom{50 - 15}{10 - 3}}{\binom{50}{10}} \]

\[ P \{ X = 3 \} = \frac{(50 - 15)!15!(50 - 10)!}{3!(15 - 3)!(50 - 15 + 3 - 10)!(10 - 3)!50!} \]

\[ P \{ X = 3 \} = 0.2979 \]

9.4.1 Fisher's exact

A special case of the hypergeometric distribution is the Fisher's exact method. Fisher's exact is the probability of sampling a specific configuration of a 2 by 2 table with constrained marginals. Marginals in this case refer to the sums for each row and each column. Therefore, every Fisher's exact will have 4 marginals, those 4 being the sum of the first column, the sum of the second column, the sum of the first row, and the sum of the second row. Since these values are constant, that also means that the sum of all the elements in the table will always equal the same thing. This is clarified with the image below:
In the image above, constant marginals would mean that E, F, G, and H would be held constant. Since those values would be constant, I would also be constant as I can be thought of as the sum of E and F or G and H are of which are constants.

In theory this test can be used for any 2 by 2 table, but most commonly, it is used when necessary sample size conditions for using the z-test or chi-square test are violated. The table below shows such a situation:

From this $p_{\text{fisher}}$ can be calculated:

$$p_{\text{fisher}} = \frac{(a + b)! (c + d)! (a + c)! (b + d)!}{(a + b + c + d)! a! b! c! d!}$$

where the numerator is the number of ways the marginals can be arranged, the first term in the denominator is the number of ways the total can be arranged, and the remaining terms in the denominator are the number of ways each observation can be arranged.

As stated before, this calculated value is only the probability of creating the specific 2x2 from which the $p_{\text{fisher}}$ value was calculated.

Another way to calculate $p_{\text{fisher}}$ is to use Mathematica. The syntax as well as an example with numbers can be found in the screen shot below of the Mathematica notebook. For further clarification, the screen shot below also shows the calculated value of $p_{\text{fisher}}$ with numbers. This is useful to know because it reduces the chances of making a simple algebra error.

Another useful tool to calculate $p_{\text{fisher}}$ is using available online tools. The following link provides a helpful online calculator to quickly calculate $p_{\text{fisher}}$. [1]

Using Mathematica or a similar calculating tool will greatly simplify the process of solving for $p_{\text{fisher}}$.

After the value of $p_{\text{fisher}}$ is found, the p-value is the summation of the Fisher exact value(s) for the more extreme case(s), if applicable. The p-value determines whether the null hypothesis is true or false. An example of this can be found in the worked out hypergeometric distribution example below.

Finding the p-value
As elaborated further here: [2], the p-value allows one to either reject the null hypothesis or not reject the null hypothesis. Just because the null hypothesis isn't rejected doesn't mean that it is advocated, it just means that there isn't currently enough evidence to reject the null hypothesis.

In order to fine the p-value, it is necessary to find the probabilities of not only a specific table of results, but of results considered even more "extreme" and then sum those calculated probabilities. An example of this is shown below in example 2.

What is considered "more extreme" depends on the situation. In the most general sense, more extreme means further from the expected or random result.

Once the more extreme tables have been created and the probability for each value obtained, they can be added together to find the p-value corresponding to the least extreme table.

It is important to note that if the probabilities for every possible table were summed, the result would invariably be one. This should be expected as the probability of a table being included in the entire population of tables is 1.

- **Example 1**

The example table below correlates the the amount of time that one spent studying for an exam with how well one did on an exam.

<table>
<thead>
<tr>
<th>Score Poor on Exam</th>
<th>Does not study</th>
<th>Does Study</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

There are 5 possible configurations for the table above, which is listed in the Mathematica code below. All the \( p_{\text{fisher}} \) for each configuration is shown in the Mathematica code below.

- **Example 2**

What are the odds of choosing the samples described in the table below with these marginals in this configuration or a more extreme one?
First calculate the probability for the initial configuration.

\[ P_{fisher} = \frac{6!19!6!19!}{25!5!11!11!18!} = 0.000643704 \]

Then create a new table that shows the most extreme case that also maintains the same marginals that were in the original table.

<table>
<thead>
<tr>
<th></th>
<th>catalyst A</th>
<th>catalyst B</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad product</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Good product</td>
<td>0</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Totals</td>
<td>6</td>
<td>19</td>
<td>25</td>
</tr>
</tbody>
</table>

Then calculate the probability for the most extreme case.

\[ P_{fisher} = \frac{6!19!6!19!}{25!6!0!0!19!} = 0.00000564653 \]

Finally add the two probabilities together.

\[ p_{fisher} = 0.000643704 + 0.00000564653 = 0.00064 \]

### 9.5 Maximum Entropy Function

The maximum entropy principle uses all the testable information about multiple given probability distributions to describe the true probability distribution by maximizing the information entropy. To derive a distribution with the maximum entropy principle, you obtain some testable information \( I \) about some quantity \( x \). This information is expressed as \( m \) constraints on the expectation values of the functions \( f_k \). This makes the Bayesian probability distribution to satisfy:

\[
F_k = \sum_{i=1}^{n} Pr(x_i|I) f_k(x_i) \quad k = 1, \ldots, m
\]

We also know that all of these probabilities must sum to 1, so the following constraint is introduced:

\[
\sum_{i=1}^{n} Pr(x_i|I) = 1
\]
Then the probability distribution with maximum information entropy that satisfies all these constraints is:

$$P(x_i | I) = \frac{1}{Z(\lambda_1, \ldots, \lambda_m)} \exp [\lambda_1 f_1(x_i) + \cdots + \lambda_m f_m(x_i)]$$

The normalization constant is determined by the classical partition function:

$$Z(\lambda_1, \ldots, \lambda_m) = \sum_{i=1}^{n} \exp [\lambda_1 f_1(x_i) + \cdots + \lambda_m f_m(x_i)]$$

The $\lambda_k$ parameters are Lagrange multipliers with values determined by:

$$F_k = \frac{\partial}{\partial \lambda_k} \log Z(\lambda_1, \ldots, \lambda_m)$$

All of the well-known distributions in statistics are maximum entropy distributions given appropriate moment constraints. For example, if we assume that the above was constrained by the second moment, and then one would derive the Gaussian distribution with a mean of 0 and a variance of the second moment.

The use of the maximum entropy distribution function will soon become more predominant. In 2007 it had been shown that Bayes’ Rule and the Principle of Maximum Entropy were compatible. It was also shown that maximum entropy reproduces every aspect of orthodox Bayesian inference methods. This enables engineers to tackle problems that could not be simply addressed by the principle of maximum entropy or Bayesian methods, individually, in the past.

### 9.6 Summary

The three discrete distributions that are discussed in this article include the Binomial, Hypergeometric, and Poisson distributions. These distributions are useful in finding the chances that a certain random variable will produce a desired outcome.

#### 9.6.1 Binomial Distribution Function

The Binomial distribution function is used when there are only two possible outcomes, a \textit{success} or a \textit{failure}. A success occurs with the probability $p$ and a failure with the probability $1-p$. Suppose now that in $n$ independent trials the \textit{binomial} random variable $X$ represents the number of successes. The following equation applies to binomial random variables:

$$P \{ X = k \} = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \quad \text{for} \quad k = 0, 1, 2, \ldots, n$$

$$P \{ X \leq i \} = \sum_{k=0}^{i} \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$
9.6.2 Poisson Distribution Function

The Poisson distribution can be used in a large variety of situations. It can also be used to approximate the Binomial Distribution when \( n \) is large and \( p \) is small yielding a moderate \( np \).

The following example is a situation in which the Poisson Distribution applies: Suppose a CSTR is full of molecules and the probability of any one molecule reacting to form product is small (say due to a low temperature) while the number of molecules is large, then the probability distribution function would fit the Poisson Distribution well (Note: all the molecules are the same here, unlike in the hypergeometric example to follow). Other examples include, the number of wrong telephone numbers that are dialed each day or the number of people in a community that live to 100 years of age.

The following equation is the Poisson Distribution function:

\[
P \{ X = k \} = \frac{\mu^k e^{-\mu}}{k!} \quad \text{for} \quad k = 0, 1, 2, \ldots, n
\]

\[
P \{ X \leq i \} = \sum_{k=0}^{i} \frac{\mu^k e^{-\mu}}{k!}
\]

9.6.3 Summary of Key Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Requirements</th>
<th>Distribution Function</th>
</tr>
</thead>
</table>
| Binomial     | • Only two outcomes are possible.  
• Range is random.  
• Success is independent of the trials.  
• The probability of non-consecutive successes is small. | \( P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \) |
| Poisson      | • Only two outcomes are possible.  
• Range is random.  
• Success is independent of the trials. | \( P(X = k) = \frac{\mu^k e^{-\mu}}{k!} \) |
| Hypergeometric | • Selections are not replaced. | \( P(X = k) = \binom{N-k}{n-k} \binom{K}{k} \binom{N-K}{n-k} \binom{N}{n} \) |

9.6.4 Summary of Distribution Approximations

There are many useful approximations for discrete random variables. An important approximation is that which yields a normal distribution because it allows for confidence intervals and probabilities to be continuous. Additionally, since the normal distribution is so well understood it is used for convenience when speaking to those who have yet to study probability theory. Below is a table of useful approximations. The majority of the theory behind these approximations is in the application of the Central Limit Theorem and Large Number Law but these are outside the scope of this text.
The above chart can be read as "A random variable under the following distribution can is approximately a <blank> random variable when the following requirements are met." For example, the final line of the table can be read: "A random variable following a Poisson distribution with parameter λ is approximately a continuous Normal random variable where the mean is λ and the variance is λ assuming that there are at least 100 samples that have been taken."

9.7 Worked out Binomial Distribution Example

In order for a vaccination of Polio to be efficient, the shot must contain at least 67% of the appropriate molecule, VPOLIO. To ensure efficacy, a large pharmaceutical company manufactured a batch of vaccines with each syringe containing 75% VPOLIO. Your doctor draws a syringe from this batch which should contain 75% VPOLIO. What is the probability that your shot will successfully prevent you from acquiring Polio? Assume the syringe contains 100 molecules and that all molecules are able to be delivered from the syringe to your bloodstream.

9.7.1 Solution

This can be done by first setting up the binomial distribution function. In order to do this, it is best to set up an Excel spreadsheet with values of k from 0 to 100, including each integer. The frequency of pulling a VPOLIO molecule is 0.75. Randomly drawing any molecule, the probability that this molecule will never be VPOLIO, or in other words, the probability of your shot containing 0 VPOLIO molecules is

\[ P = \frac{100!}{(100 - 0)!0!} (0.75)^0 (1 - 0.75)^{100-0} \]

Note that 0!=1

Your spreadsheet should contain probabilities for all off the possible values of VPOLIO in the shot. A link to our spreadsheet can be seen at the end of this article. A graph of our distribution is shown below.
The next step is to sum the probabilities from 67 to 100 of the molecules being VPOLIO. This is calculation is shown in the sample spreadsheet. The total probability of at least 67 of the molecules being VPOLIO is 0.9724. Thus, there is a 97.24% chance that you will be protected from Polio.

Media: Binomial_Distribution_Example1.xls

9.8 Gaussian Approximation Of A Binomial Distribution Example

Calculation of the binomial function with \( n \) greater than 20 can be tedious, whereas calculation of the Gauss function is always simple. To illustrate this, consider the following example.

Suppose we want to know the probability of getting 23 heads in 36 tosses of a coin. This probability is given by the following binomial distribution:

\[
P = \frac{36!}{23!13!}(0.5)^{36} = 3.36%\]

To use a Gaussian Approximation, we must first calculate the mean and standard deviation.

\[
\mu = np = 18
\]

\[
\sigma = \sqrt{np(1 - p)} = 3
\]

Now we can approximate the probability by the Gauss function below.
This approximation is very close and requires much less calculation due to the lack of factorials. The usefulness of the Gaussian approximation is even more apparent when $n$ is very large and the factorials involve very intensive calculation.

### 9.9 Worked out Hypergeometric Distribution Example

A teacher has 12 students in her class. On her most recent exam, 7 students failed while the other 5 students passed. Curious as to why so many students failed the exam, she took a survey, asking the students whether or not they studied the night before. Of the students who failed, 4 students did study and 3 did not. Of the students who passed, 1 student did study and 4 did not. After seeing the results of this survey, the teacher concludes that those who study will almost always fail, and she proceeds to enforce a "no-studying" policy. Was this the correct conclusion and action?

#### 9.9.1 Solution

This is a perfect situation to apply the Fisher's exact test to determine if there is any association between studying and performance on the exam. First, create a 2 by 2 table describing this particular outcome for the class, and then calculate the probability of seeing this exact configuration. This is shown below.

<table>
<thead>
<tr>
<th></th>
<th>Did Study</th>
<th>Did Not Study</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed Exam</td>
<td>4</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Passed Exam</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Totals</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
</tbody>
</table>

$$p_{fisher,1} = \frac{(4 + 3)!(1 + 4)!(4 + 1)!(3 + 4)!}{(4 + 3 + 1 + 4)!3!1!4!} = 0.221$$

Next, create 2 by 2 tables describing any configurations with the exact same marginals that are more extreme than the one shown above, and then calculate the probability of seeing each configuration. Fortunately, for this example, there is only one configuration that is more extreme, which is shown below.

<table>
<thead>
<tr>
<th></th>
<th>Did Study</th>
<th>Did Not Study</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed Exam</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Passed Exam</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Totals</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
</tbody>
</table>

$$p_{fisher,2} = \frac{(5 + 2)!(0 + 5)!(5 + 0)!(2 + 5)!}{(5 + 2 + 0 + 5)!5!2!0!5!} = 0.0265$$

Finally, test the significance by calculating the p-value for the problem statement. This is done by adding up all the previously calculated probabilities.
Thus, the p-value is greater than 0.05, which is the standard accepted level of confidence. Therefore, the null hypothesis cannot be rejected, which means there is no significant association between studying and performance on the exam. Unfortunately, the teacher was wrong to enforce a "no-studying" policy.

9.10 Worked out Poisson Example

The hormone PREGO is only found in the female human body during the onset of pregnancy. There is only 1 hormone molecule per 10,000 found in the urine of a pregnant woman. If we are given a sample of 5,000 hormone molecules what is the probability of finding exactly 1 PREGO? If we need at least 10 PREGO molecules to be 95% positive that a woman is pregnant, how many total hormone molecules must we collect? If the concentration of hormone molecules in the urine is 100,000 molecules/mL of urine, what is the minimum amount of urine (in mL) necessary to insure an accurate test (95% positive for pregnancy)?

9.10.1 Solution

This satisfies the characteristics of a Poisson process because

1.) PREGO hormone molecules are randomly distributed in the urine
2.) Hormone molecule numbers are discrete
3.) If the interval size is made smaller (i.e. our sample size is reduced), the probability of finding a PREGO hormone molecule goes to zero

Therefore we will assume that the number of PREGO hormone molecules is distributed according to a Poisson distribution.

To answer the first question, we begin by finding the expected number of PREGO hormone molecules:

\[
\mu = np = (5,000) \left( \frac{1}{10,000} \right) = 0.5
\]

Next we use the Poisson distribution to calculate the probability of finding exactly one PREGO hormone molecule:

\[
P\{X = k\} = \frac{\mu^k e^{-\mu}}{k!} = 0.303 \quad \text{where} \quad k = 1 \quad \mu = 0.5
\]

The second problem is more difficult because we are calculating \( n \). We begin as before by calculating the expected number of PREGO hormone molecules:
\[ \mu = np = \frac{n}{10,000} \]

Now we apply the Poisson Distribution Function equation:

\[ P(X \geq 10) = 0.95 = \sum_{k=10}^{\infty} \frac{e^{-\mu} \mu^k}{k!} - 1.0 = \sum_{k=10}^{\infty} \left( \frac{n}{10,000} \right)^k e^{-\frac{n}{10,000}} \frac{1}{k!} \]

It is easiest to solve this in Mathematica:

The function `FindRoot[]` was used in Mathematica because the `Solve[]` function has problems solving polynomials of degree order greater than 4 as well as exponential functions. However, `FindRoot[]` requires an initial guess for the variable trying to be solved, in this case \( n \) was estimated to be around 100,000. As you can see from the Mathematica screen shot the total number of hormone molecules necessary to be 95% sure of pregnancy (or 95% chance of having at least 10 PREGO molecules) was 169,622 molecules.

For the last step we use the concentration of total hormone molecules found in the urine and calculate the volume of urine required to contain 169,622 total hormone molecules as this will yield a 95% chance of an accurate test:

\[
\text{Concentration} = \frac{100,000 \ \text{molecules}}{mL \ \text{urine}}
\]

\[
\text{Volume of urine} = \frac{169,622}{100,000} = 1.69622 \ mL
\]

### 9.11 Example: Gaussian Approximation to a Poisson Distribution

To illustrate the Gaussian approximation to the Poisson distribution, consider a distribution where the mean (\( \mu \)) is 64 and the number of observations in a definite interval (\( N \)) is 72. Determine the probability of these 72 observations?

Using the Poisson Distribution
\[ P(N) = e^{-\mu} \cdot \frac{\mu^N}{N!} \]

\[ P(72) = e^{-64} \cdot \frac{64^{72}}{72!} = 2.9\% \]

This can be difficult to solve when the parameters \( N \) and \( \mu \) are large. An easier approximation can be done with the Gauss function:

\[ P(72) = G_{64,8} = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 3.0\% \]

where, \( X=\mu \) and \( \sigma=\sqrt{\mu} \).

### 9.12 Multiple Choice Question 1

All of the following are characteristics of the Poisson distribution **EXCEPT**:

a) The number of successes is independent of the interval

b) There are imaginary numbers

c) Probabilities, or outcomes, are discrete

d) Two or more success in a short interval is extremely small

### 9.13 Multiple Choice Question 2

If there are \( K \) blue balls and \( N \) total balls, the chance of selecting \( k \) blue balls by selecting \( n \) balls in shorthand notation is given by:

a) 

\[ P \{ X = k \} = \frac{\binom{N}{n} \binom{K-k}{N-n}}{\binom{K}{n}} \]
\[ P \{ X = k \} = \frac{\binom{n}{k} \binom{n - K}{N - k}}{\binom{K}{n}} \]

c) \[ P \{ X = k \} = \frac{\binom{K}{k} \binom{N - K}{n - k}}{\binom{N}{n}} \]

d) \[ P \{ X = k \} = \frac{\binom{K}{n} \binom{N - n}{K - N}}{\binom{N}{n}} \]

9.14 Discrete Distribution Presentation: Clown Time

www.youtube.com/v/6WOieU664Ew

9.15 References