Probability Calculation Tools Tutorial

Purpose

The purpose of this document is to present a tutorial on the methods and tools used to find or calculate probabilities for common probability distributions. The primary focus of this document is on the binomial distribution but its methods will be extended to the hypergeometric, Poisson, and normal distributions.

Methods

There are four primary methods for finding probabilities for probability distributions:

- Formulas
- Approximations (see Appendix B.8)
- Tables
- Software, e.g. MINITAB
  - **Graph**> **Probability Distribution Plot**> **View Probability**
  - **Calc**> **Probability Distribution**

The formulas are fundamental but can be tedious to use. Tables can provide quick solutions but cover a limited range of the inputs and interpolation might be required. For the problems that tables don't cover, approximations are often available that put the problem solution back into the scope of another table. Software provides accuracy, convenience, and wide scope; however, the other methods will be necessary when software isn't available or accessible.

Binomial Distribution

Formulas

The probability of finding x events in n trials when the probability of an event occurring on any trial is p is given by the binomial probability distribution

$$b(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

where $\binom{n}{x}$ is the combination operation

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Example Find the exact binomial probability $b(x=2;n=10,p=0.05)$ using the formula.

Solution The exact solution to the problem can be obtained using the binomial distribution formula:

$$\begin{aligned} b(x=2;n=10,p=0.05) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \binom{10}{2} 0.05^2 (1-0.05)^8 \\ &= 0.075 \end{aligned}$$

The probability of x taking on any of the values from $x=0$ to $x=c$ is given by the sum of the individual probabilities, called the cumulative probability distribution

$$b(c;n,p) = \sum_{x=0}^{c} \binom{n}{x} p^x (1-p)^{n-x}$$

Cumulative probabilities are painful to solve using the formulas so tables or software are used when they are available.

Cumulative probability tables are also useful when calculating the probability for a range of x values by using two values from a $b(c;n,p)$ table or calculator. Be careful to choose the correct value of the lower limit of x when using this method.

Example Find the exact binomial probability $b(2 \le x \le 6; n = 20, p = 0.20)$.

Solution The requested probability is

$$b(2 \le x \le 6; n = 20, p = 0.20) = \sum_{x=2}^{6} b(x; n = 20, p = 0.20)$$

however, this involves calculating or looking up five $b(x;n,p)$ values. If a cumulative binomial probability table or calculator is available then it is easier to do the calculation in two parts:

$$b(2 \le x \le 6; n = 20, p = 0.20) = b(c = 6; n = 20, p = 0.20) - b(c = 1; n = 20, p = 0.20)$$

Approximations

Under certain conditions one probability distribution can be approximated by another. The benefit of using an approximation is that it may be difficult to solve the original exact problem but simple to solve the problem using another distribution.

Approximations between distributions are available when the two distributions' parameters are equated and their shapes are very similar. Approximating one distribution with another should only be done when the necessary conditions are satisfied. Some of the more commonly used approximations are presented in Appendix B8 with examples.

The binomial distribution plays a role in the following approximations:

- Approximate the hypergeometric distribution with the binomial distribution when the sample size n is small compared to the lot size N (typically less than 10%) using

$$h(x; D, N, n) \simeq b(x; n, p = D/N)$$

- Approximate the binomial distribution with the Poisson distribution when the sample size is large and the proportion is small (typically $n > 100$ and $p < 0.10$) using

$$b(x; n, p) \simeq Poisson(x; \lambda = np)$$

- Approximate the binomial distribution with the normal distribution when the sample size is large and the proportion is moderate (typically $n > 100$ and $0.10 < p < 0.90$) using

$$b(x; n, p) \simeq \Phi\left(x - \frac{1}{2} < x' < x + \frac{1}{2}; \mu = np, \sigma = \sqrt{np(1-p)}\right)$$

where the continuity correction is used to map the discrete x value to the interval $\left(x - \frac{1}{2}, x + \frac{1}{2}\right)$.

Tables

Cumulative Probability Tables

When appropriate software is not available, probability calculation problem can often be solved using values from published tables. The tables can only provide discrete values so it may be necessary to interpolate between values in a table or a problem might fall outside the scope of a table and another calculation method will be required.

Tables of $b(x;n,p)$ are available; however, tables of $b(c;n,p)$ are more useful and so they are more common. The reason is that 1) most practical problems involve a range of x values and 2) two adjacent entries in a $b(c;n,p)$ table can be used to calculate $b(x;n,p)$:

$$b(x;n,p) = b(c = x;n,p) - b(c = x-1;n,p)$$

When the problem to be solved involves a range of x values two table entries must be used:

$$b(x_1 \le x \le x_2;n,p) = b(c = x_2;n,p) - b(c = x_1 - 1;n,p)$$

Be careful to look up the correct value for the lower bound.

A table of the cumulative binomial probability distribution $b(c;n,p)$ is presented in Appendix B.1. There are subtables for each n and the tables are indexed by c values in rows and p values in columns. The values in the body of the table are the $b(c;n,p)$. Table B.1 covers $2 \le n \le 20$, $0 \le c \le n$, and $p = 0.01, 0.02, 0.05, 0.10, 0.20, \cdots, 0.80, 0.90, 0.95,$ and 0.99. As an example, Figure 1 shows the $n = 10$ table excerpt from Appendix B.1.

Interpolation is necessary between between adjacent p values. The differences between adjacent p values are small enough that simple linear interpolations are possible - usually by inspection.

CUMULATIVE BINOMIAL PROBABILITIES
b(c; n, p)

n	c	0.01	0.02	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
10	0	0.904	0.817	0.599	0.349	0.107	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000	0.000
10	1	0.996	0.984	0.914	0.736	0.376	0.149	0.046	0.011	0.002	0.000	0.000	0.000	0.000	0.000
10	2	1.000	0.999	0.988	0.930	0.678	0.383	0.167	0.055	0.012	0.002	0.000	0.000	0.000	0.000
10	3	1.000	1.000	0.999	0.987	0.879	0.650	0.382	0.172	0.055	0.011	0.001	0.000	0.000	0.000
10	4	1.000	1.000	1.000	0.998	0.967	0.850	0.633	0.377	0.166	0.047	0.006	0.000	0.000	0.000
10	5	1.000	1.000	1.000	1.000	0.994	0.953	0.834	0.623	0.367	0.150	0.033	0.002	0.000	0.000
10	6	1.000	1.000	1.000	1.000	0.999	0.989	0.945	0.828	0.618	0.350	0.121	0.013	0.001	0.000
10	7	1.000	1.000	1.000	1.000	1.000	0.998	0.988	0.945	0.833	0.617	0.322	0.070	0.012	0.000
10	8	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.989	0.954	0.851	0.624	0.264	0.086	0.004
10	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.972	0.893	0.651	0.401	0.096
10	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 1. Appendix B.1 excerpt for $b(c;n = 10,p)$.

Example: Find the cumulative binomial probability $b(c = 2;n = 10,p = 0.05)$.

Solution: The $n = 10$ table is shown in Figure 1. The answer to the problem is found at the intersection of the $c = 2$ row and the $p = 0.05$ column as shown in Figure 2. The answer is $b(c = 2;n = 10,p = 0.05) = 0.988$.

n	c	0.01	0.02	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
10	0	0.904	0.817	0.599	0.349	0.107	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000	0.000
10	1	0.996	0.984	0.914	0.736	0.376	0.149	0.046	0.011	0.002	0.000	0.000	0.000	0.000	0.000
10	2	1.000	0.999	0.988	0.930	0.678	0.383	0.167	0.055	0.012	0.002	0.000	0.000	0.000	0.000
10	3	1.000	1.000	0.999	0.987	0.879	0.650	0.382	0.172	0.055	0.011	0.001	0.000	0.000	0.000
10	4	1.000	1.000	1.000	0.998	0.967	0.850	0.633	0.377	0.166	0.047	0.006	0.000	0.000	0.000
10	5	1.000	1.000	1.000	1.000	0.994	0.953	0.834	0.623	0.367	0.150	0.033	0.002	0.000	0.000
10	6	1.000	1.000	1.000	1.000	0.999	0.989	0.945	0.828	0.618	0.350	0.121	0.013	0.001	0.000
10	7	1.000	1.000	1.000	1.000	1.000	0.998	0.988	0.945	0.833	0.617	0.322	0.070	0.012	0.000
10	8	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.989	0.954	0.851	0.624	0.264	0.086	0.004
10	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.972	0.893	0.651	0.401	0.096
10	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 2. Solution to $b(c = 2;n = 10,p = 0.05)$.

Example: Use table B.1 to find the binomial probability $b(x = 2; n = 10, p = 0.05)$.

Solution: The answer can be obtained from two adjacent entries from the table as shown in Figure 3:

$$
\begin{aligned}
b(x = 2; n = 10, p = 0.05) &= b(c = 2; n = 10, p = 0.05) - b(c = 1; n = 10, p = 0.05) \\
&= 0.988 - 0.914 \\
&= 0.074
\end{aligned}
$$

The exact answer, calculated directly from the formula is

$$
\begin{aligned}
b(x = 2; n = 10, p = 0.05) &= \binom{10}{2} 0.05^2 (1 - 0.05)^8 \\
&= 0.0746
\end{aligned}
$$

The discrepancy between the exact calculation and the value obtained using the table is due to round-off error in the table and is negligible.

n	c	0.01	0.02	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
10	0	0.904	0.817	0.599	0.349	0.107	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000	0.000
10	1	0.996	0.984	0.914	0.736	0.376	0.149	0.046	0.011	0.002	0.000	0.000	0.000	0.000	0.000
10	2	1.000	0.999	0.988	0.930	0.678	0.383	0.167	0.055	0.012	0.002	0.000	0.000	0.000	0.000
10	3	1.000	1.000	0.999	0.987	0.879	0.650	0.382	0.172	0.055	0.011	0.001	0.000	0.000	0.000
10	4	1.000	1.000	1.000	0.998	0.967	0.850	0.633	0.377	0.166	0.047	0.006	0.000	0.000	0.000
10	5	1.000	1.000	1.000	1.000	0.994	0.953	0.834	0.623	0.367	0.150	0.033	0.002	0.000	0.000
10	6	1.000	1.000	1.000	1.000	0.999	0.989	0.945	0.828	0.618	0.350	0.121	0.013	0.001	0.000
10	7	1.000	1.000	1.000	1.000	1.000	0.998	0.988	0.945	0.833	0.617	0.322	0.070	0.012	0.000
10	8	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.989	0.954	0.851	0.624	0.264	0.086	0.004
10	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.972	0.893	0.651	0.401	0.096
10	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 3. Entries of $b(c; n = 10, p = 0.05)$ to find $b(x = 2; n = 10, p = 0.05)$.

Example: Find the binomial probability $b(1 \le x \le 4; n = 10, p = 0.20)$ using table B.1.

Solution: The solution to the problem can be obtained from two entries from the table:

$$
\begin{aligned}
b(1 \le x \le 4; n = 10, p = 0.20) &= b(c = 4; n = 10, p = 0.20) - b(c = 0; n = 10, p = 0.20) \\
&= 0.967 - 0.107 \\
&= 0.860
\end{aligned}
$$

n	c	0.01	0.02	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
10	0	0.904	0.817	0.599	0.349	0.107	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000	0.000
10	1	0.996	0.984	0.914	0.736	0.376	0.149	0.046	0.011	0.002	0.000	0.000	0.000	0.000	0.000
10	2	1.000	0.999	0.988	0.930	0.678	0.383	0.167	0.055	0.012	0.002	0.000	0.000	0.000	0.000
10	3	1.000	1.000	0.999	0.987	0.879	0.650	0.382	0.172	0.055	0.011	0.001	0.000	0.000	0.000
10	4	1.000	1.000	1.000	0.998	0.967	0.850	0.633	0.377	0.166	0.047	0.006	0.000	0.000	0.000
10	5	1.000	1.000	1.000	1.000	0.994	0.953	0.834	0.623	0.367	0.150	0.033	0.002	0.000	0.000
10	6	1.000	1.000	1.000	1.000	0.999	0.989	0.945	0.828	0.618	0.350	0.121	0.013	0.001	0.000
10	7	1.000	1.000	1.000	1.000	1.000	0.998	0.988	0.945	0.833	0.617	0.322	0.070	0.012	0.000
10	8	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.989	0.954	0.851	0.624	0.264	0.086	0.004
10	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.972	0.893	0.651	0.401	0.096
10	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 4. Entries of $b(c; n = 10, p = 0.05)$ to find $b(1 \le x \le 4; n = 10, p = 0.20)$.

Inverse Cumulative Probability

In some cases it is necessary to find the c value for given n, p, and $b(c; n, p)$. This requires finding the specified n subtable and its p column and then searching down the column of $b(c; n, p)$ values in the body of the table to find that value that meets the specified $b(c; n, p)$ condition. This problem is called the *inverse cumulative probability* problem. The output of the calculation is the c value. It's unlikely that the problem will be solved exactly for the specified c value so you will have to round up or down as required for the best compromise.

Example Find the c value that meets the condition $b(c;n = 10,p = 0.40) \leq 0.50$.

Solution Parsing down the $p = 0.40$ column of the $n = 10$ table the closest value of c that meets the condition is

$$b(c = 3;n = 10,p = 0.40) = 0.382$$

Software (MINITAB)

MINITAB has two methods that can be used to solve probability distribution problems:

- **Graph**> **Probability Distribution Plot**> **View Probability**
- **Calc**> **Probability Distributions**

The **Graph**> **Probability Distribution Plot**> **View Probability** method is more versatile than the **Calc**> **Probability Distributions** method. The **Graph**> **Probability Distribution Plot**> **View Probability** method is easier to set up, more flexible, its output is graphical so it is easier to interpret, and many problems can be solved using a single call to this method. It's only limitation is when problem solutions fall far out in the tails of a distribution and probabilities get rounded to zero or 1, i.e. the answer may not provide enough significant digits. If the output from the **Graph**> **Probability Distribution Plot**> **View Probability** method doesn't include enough significant digits then you will have to use the **Calc**> **Probability Distributions** method instead.

The **Calc**> **Probability Distributions** method is limited to solving left tail area problems which is fine for one-tailed problems but two-tailed problems require two calls to the method and then a manual subtraction to find the answer. The method's menu is a bit more confusing and its output is tabular, not graphical, so its results are a bit harder to interpret. The primary benefit of this method is that it delivers as many significant digits as are required. If the default output from the method doesn't show enough significant digits, you can just right click on the output and ask for more.

An excellent resource for more details about these calculation methods are the MINITAB Help web pages that can be accessed from the **Help** command button in the bottom left hand corner of every MINITAB menu.

Graph> Probability Distribution Plot> View Probability

After opening the **Graph**> **Probability Distribution Plot**> **View Probability** menu you will presented with a window with two tabs. You must populate fields in both tabs to configure a calculation. The first tab (**Distribution**) allows you to choose the distribution (hypergeometric, binomial, Poisson, normal, ...) and set its parameter values. The second tab (**Shaded Area**) allows you to specify the method and region to be shaded under the distribution function.

Figure 5 shows the **Distribution** tab of the menu configured for the binomial distribution with $n = 10$ and $p = 0.05$ and Figure 6 shows the **Shaded Area** tab. This tab allows the shaded area to be specified in terms of a probability (**Probability**) or in terms of an x axis coordinate value (**X Value**). In **Probability** mode you specify the input probability and

MINITAB will output the x values as in Figure 6. In **X Value** mode you specify the x value (or values) and MINITAB will output the probabilities as in Figure 7.

Both methods (**Probability** and **X Values**) support calculations for the **Right Tail**, **Left Tail**, **Both Tails**, or **Middle**. The stylized graphs in the menu (they are customized for the **Probability** and **X Values** modes) help identify the inputs required for the chosen method.

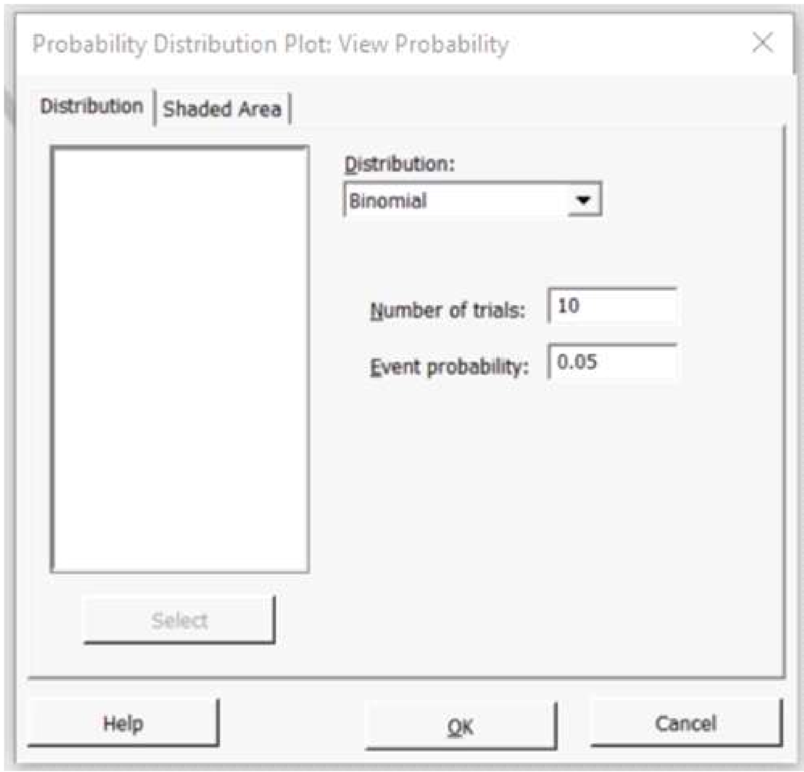

Figure 5. **Graph**> **Probability Distribution Plot**> **View**

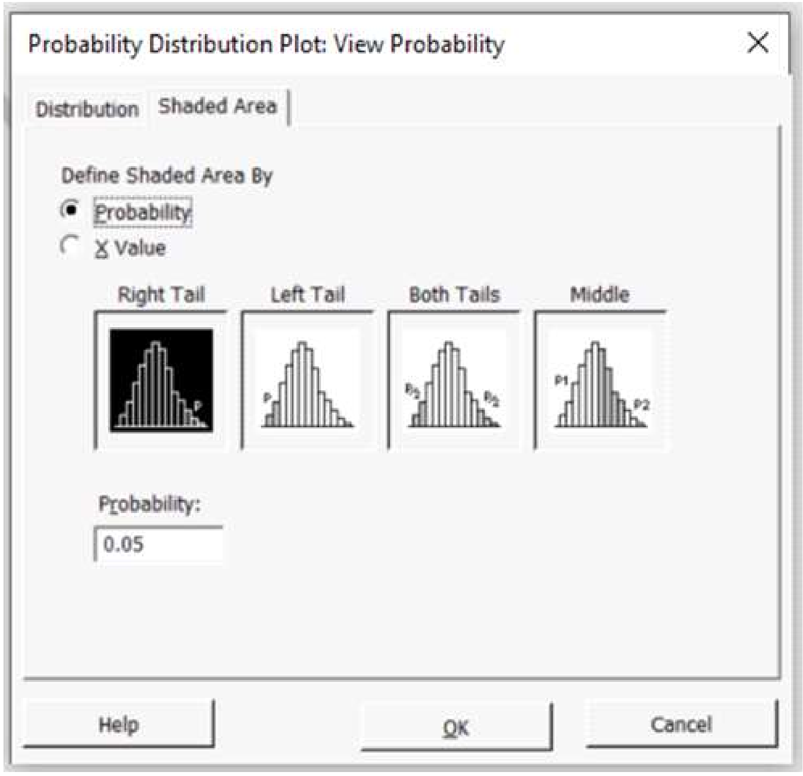

Figure 6. **Graph**> **Probability Distribution Plot**> **View**

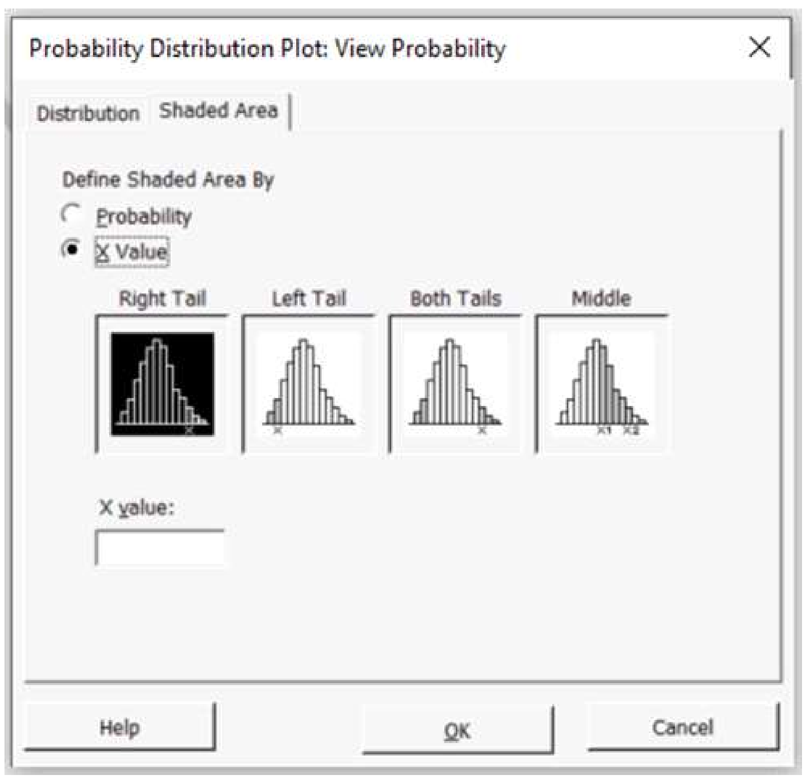

Figure 6. **Graph**> **Probability Distribution Plot**> **View**

Calc> **Probability Distributions**

Like the **Graph**> **Probability Distribution Plot**> **View Probability** method, the **Calc**> **Probability Distributions** method also provides probability calculation functions; however, it is limited to operations in the left tail so you will have to cast any right tail or two tail problems into the correct form to use it. (The **Calc**> **Probability Distributions** method is a legacy method that has been in MINITAB forever. The **Graph**> **Probability Distribution Plot**> **View Probability** menu is relatively new and was added for its additional features.)

To use the **Calc**> **Probability Distributions** menu open the menu and choose the desired distribution, such as **Calc**> **Probability Distributions**> **Binomial** as shown in Figure 8. The three radio buttons at the top allow you to specify three types of outputs: **Probability**, **Cumulative Probability**, and **Inverse Cumulative Probability**. Here is the purpose of each method:

- The **Probability** method takes as an input a single x value or a column of x values and reports their corresponding probabilities.
- The **Cumulative Probability** method takes as an input a single x value or a column of x values and reports their corresponding left tail cumulative probabilities.
- The **Inverse Cumulative Probability** method takes as an input a single or a column of left tail cumulative probability values and reports their corresponding left tail x values.

The input boxes immediately below the radio buttons will vary depending on the distribution chosen but will always ask for the distribution's parameter values. For the binomial distribution the parameters are the **Number of trials**: n and the **Event probability**: p.

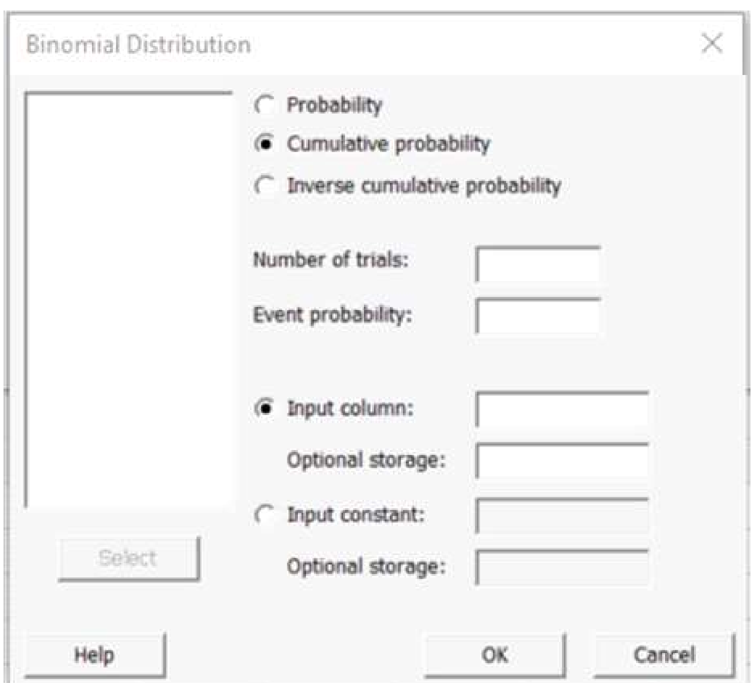

Figure 8. **Calc**> **Probability Distributions**> **Binomial**

Underneath the fields for parameter entry are the **Input column**: and **Input constant**: input fields. **Input constant**: is used most often but occasionally there is the need to use **Input column**:. For the **Probability** and **Cumulative Probability** methods the **Input constant**: or **Input column**: values will be x values and the output will be a probability or cumulative probability, respectively. For the **Inverse Cumulative Probability** method the **Input constant**: or **Input column**: values will be a single or column of x values.

Example Use the **Calc**> **Probability Distributions**> **Binomial** method to calculate the binomial probability $b(x = 2; n = 10, p = 0.20)$.

Solution Figure 9 shows how to configure the menu to calculate the binomial probability and the output from MINITAB. The answer is $b(x = 2; n = 10, p = 0.20) = 0.3020$.

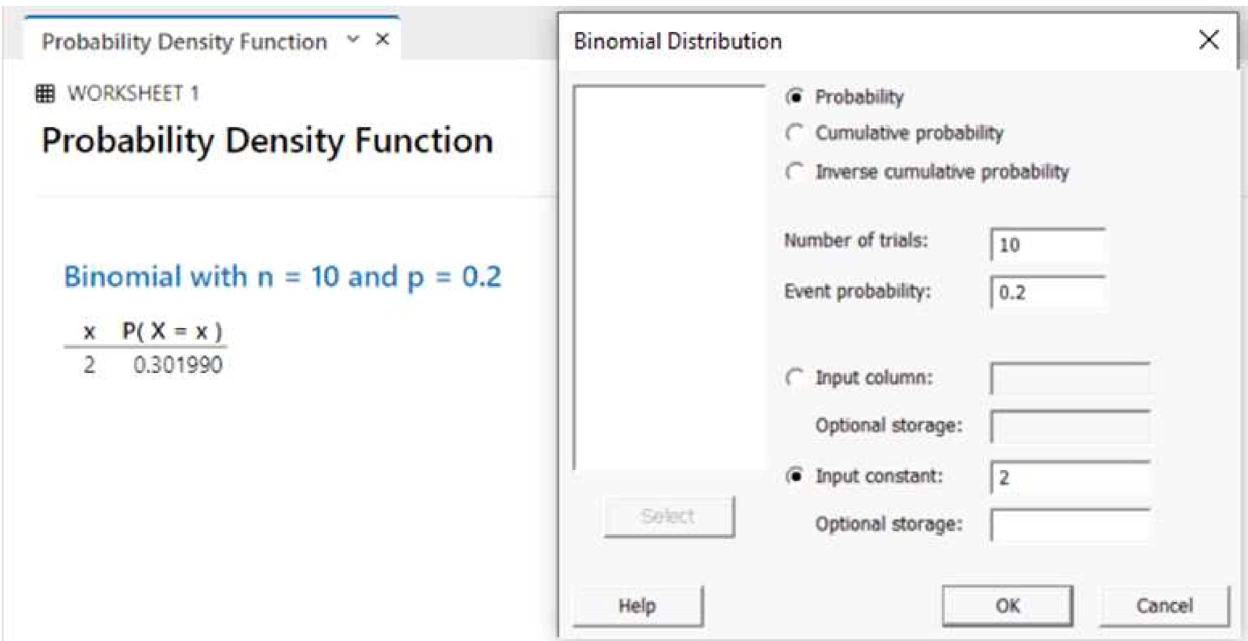

Figure 9. Calculation of $b(x = 2; n = 10, 0 = 0.20)$ using the **Calculate**> **Probability**

Example Use the **Calc**> **Probability Distributions**> **Binomial** method to calculate the binomial probability $b(c = 2; n = 10, p = 0.20)$.

Solution Figure 10 shows how to configure the menu to calculate the cumulative binomial probability and the output from MINITAB. The answer is $b(c = 2; n = 10, p = 0.20) = 0.6778$.

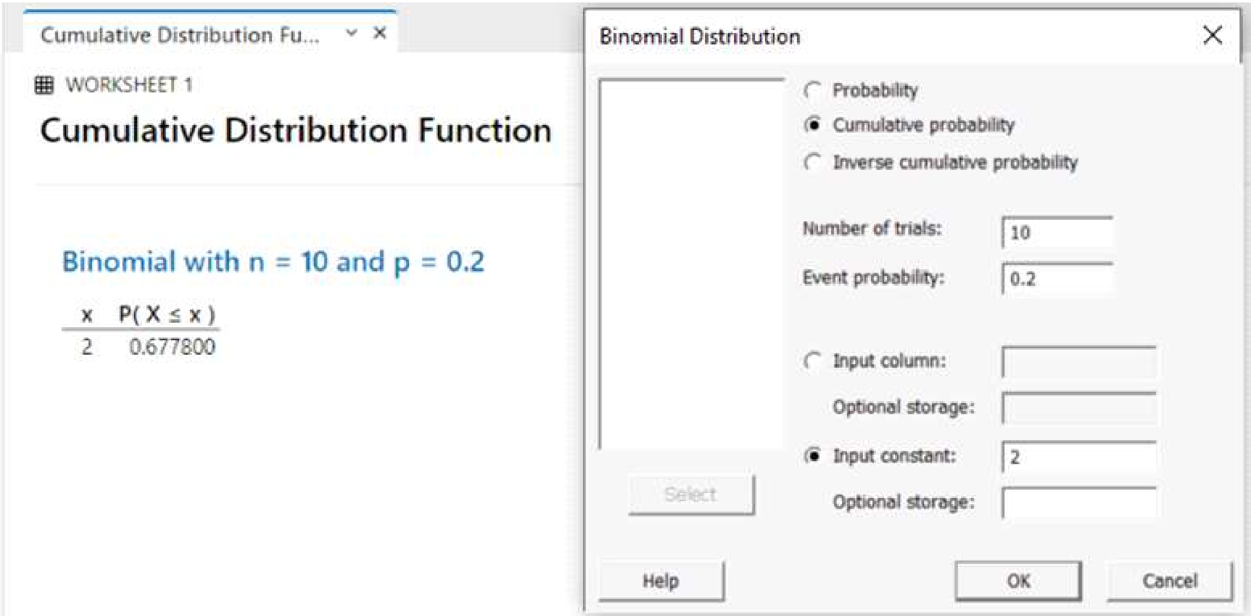

Figure 10. Calculation of $b(c = 2; n = 10, 0 = 0.20)$ using the **Calculate**> **Probability**

Example Use the **Calc**> **Probability Distributions**> **Binomial** method to calculate the inverse cumulative binomial probability (i.e. the c value) for $b(c; n = 10, p = 0.20) = 0.6778$.

Solution Figure 11 shows how to configure the menu to calculate the inverse cumulative binomial probability and the output from MINITAB. Note that MINITAB reports answers for $c = 2$ and $c = 3$ because the random variable x is discrete and the cumulative probability input 0.70 falls between two consecutive x values. You will have to compromise by choosing the c value that makes sense for your application.

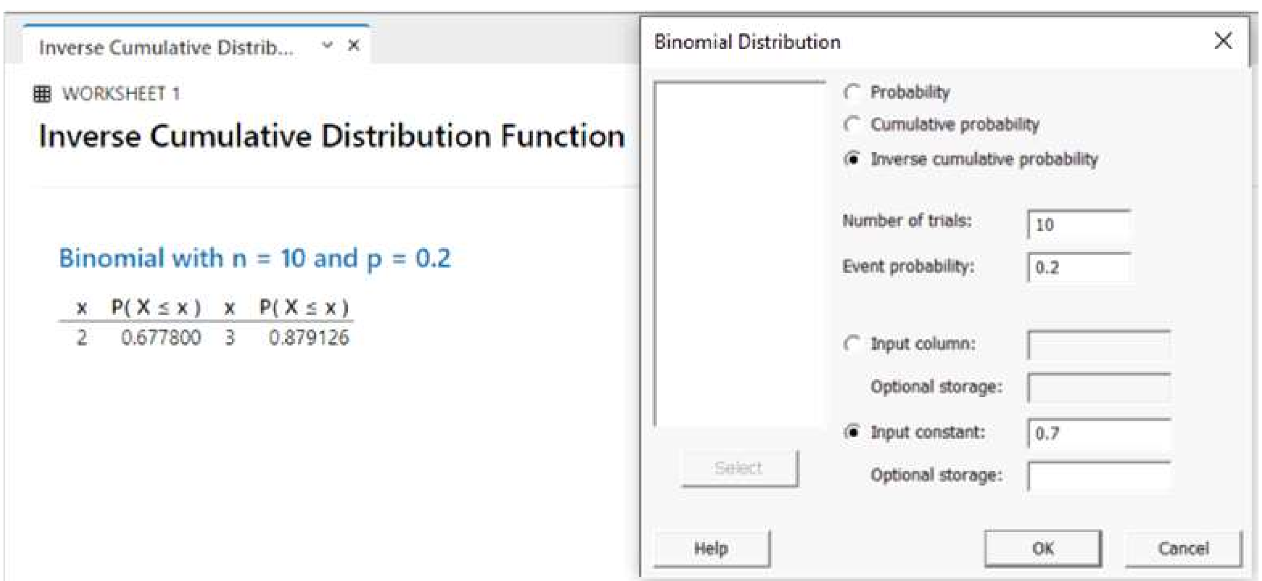

Figure 11. Calculation of c for $b(c; n = 10, 0 = 0.20) = 0.70$ using the **Calculate**> **Probability**