Journal of Cost Analysis and Parametrics, 4:106–118, 2011

ISSN: 1941-658X print
DOI: 10.1080/1941658X.2011.585331

A Probabilistic Approach to Determining the Number of Units to Build in a Yield-Constrained Process

TIMOTHY P. ANDERSON

The Aerospace Corporation, Arlington, Virginia

Many cost estimating problems involve determining the number of units to build in a yield-constrained manufacturing process, when it takes, on average, n attempts to produce m successes (m ≤ n). Examples include computer chips, focal plane arrays, circuit boards, field programmable gate arrays, etc. The simplistic approach to this problem is to multiply the number of units needed, m, by the expected number of attempts needed to produce a single success, n. For example, if a contractor reports that it takes, on average, 10 attempts to build one working unit, and if four such units are needed for a space-borne application, then the simplistic approach would be to plan for 4 × 10 = 40 units, and estimate the cost accordingly. However, if the cost analyst uses the simplistic approach, he or she is likely to be disappointed, as the probability that 40 attempts will actually produce four working units is only about 57%. Consequently, there is a 43% probability that 40 attempts will be insufficient. In fact, if the analyst wants to have, say, 80% confidence that four working units will be available, then he/she should plan for 54 attempts! Obviously, this could have a huge impact on the cost estimate. The purpose of this research is to describe the nature of the problem, to justify modeling the problem in terms of a negative binomial random variable, and to develop the necessary thought process that one must go through in order to adequately determine the number of units to build given a desired level of confidence. This understanding will be of great benefit to cost analysts who are in the position of estimating costs when certain hardware elements behave as described previously. The technique will also be very useful in cost uncertainty analysis, enabling the cost analyst to determine the appropriate probability distribution for the number of units needed to achieve success in their programs.

Introduction

Consider an example that was recently experienced by the author in which an independent cost estimate was being performed for a space-borne system that required four identical units of a particular item per platform. The unit cost of the units was estimated using an engineering buildup method. For the purposes of this article, we will assume the unit cost is \$500K per unit whether it meets specifications or not, and with little or no cost improvement for subsequent units (i.e., 100% learning curve). That is, the \$500K cost includes fabrication of the unit as well as testing to ensure that the unit meets specifications (i.e., it works).

It was known that there were yield issues in this manufacturing process such that, on average, 10 units had to be manufactured for every deliverable, fully functioning unit. The question posed to the author at the time was, "how many unit builds should be included in the independent cost estimate to ensure that a total of four units will be delivered?" At first glance, it seemed that we should use the "expected value" approach, estimating the total number of units by multiplying the number needed by the number required to achieve a

Address correspondence to Timothy P. Anderson, The Aerospace Corporation, 15049 Conference Center Drive, Suite 600, Chantilly, VA 20151. E-mail: timothy.p.anderson@aero.org

success. That is, if four are needed, and the expected yield is 1 out of 10, then one would expect to have to build $4 \times 10 = 40$ individual units in order to achieve the desired outcome, with a resulting cost estimate of $40 \times \$500K = \$20M$.

However, upon further investigation, looking at the estimate from a cost-risk perspective, it became apparent that the "expected-value" approach carried a substantial amount of uncertainty. We could, for example, get lucky and meet our goal with only half that number. On the other hand, we could fall upon bad times, in which 40 units still would not be enough. So, the author looked at how to model this manufacturing situation from a probabilistic perspective. Considering the number of units to build as a random variable, with a rate of one "success" for every 10 "trials," the obvious choice was to model the number of successes as a negative binomial random variable. As a result, it will be shown below that the probability of obtaining four suitable units from a batch of 40 attempts is about 57%. Moreover, if we want to have, say, 80% confidence that we could achieve four successes, then we need to plan to build 54 units!

The remainder of this article discusses the methodology used to arrive at this conclusion, describes the negative binomial distribution and how it applies to this problem, and uses the method to (1) determine an appropriate number of units to plan for, (2) model the number of units as a probability distribution for use in a cost-risk analysis, and (3) describe the use of the approach in calculating the Government's most probable cost (MPC) in a source selection activity.

The Negative Binomial Distribution

The negative binomial probability distribution is a discrete probability distribution that models the number of trials, n, needed to achieve a specified number of successes, m, ($m \leq n$), under a key set of assumptions, and is built up from first principles as described in the following sections. The discrete probability distributions described below can be found in most probability and statistics texts, such as Devore (1995), Mendenhall et al. (1990), and Rice (1995).

First, the Bernoulli Random Variable

Suppose a unit manufacturing process is such that it takes, on average, n attempts ("trials") in order to produce one satisfactory unit (a "success"). One can then view each attempt as a Bernoulli random variable with probability of success, p:

$$p = 1/n,$$

provided the trials have these critical properties:

1. The result of each trial is classified as either a success or a failure;
2. The probability, p, of success is the same in every trial; and
3. The trials are independent—the outcome of one trial has no influence on later outcomes.

Next, the Binomial Random Variable

Discrete probability theory enables us to extend the concept of Bernoulli random variables to count the number of successes, m, in n repeated Bernoulli trials, each with probability of success, p. Since it is a function of Bernoulli trials, it follows that a binomial experiment is one that possesses the following properties:

1. The experiment consists of n identical trials;
2. Each trial results in one of two outcomes—success or failure;
3. The probability of success on a single trial is p, and remains the same from trial to trial—consequently, the probability of failure is $q = 1 - p$;
4. The trials are independent; and
5. The random variable of interest is Y, the number of successes achieved during the n trials.

The probability mass function (PMF) of the binomial distribution is given as follows. A random variable Y is said to have a *binomial distribution* based on n trials with success probability p if and only if:

$$P(Y = y) = \binom{n}{y} p^y (1 - p)^{n-y}, y = 0, 1, 2, \ldots, n, \quad \text{and} \quad 0 \le p \le 1,$$

where

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

The mean and variance of the binomial distribution are given as follows:

$$\mu = E(Y) = np,$$

$$\sigma^2 = \mathrm{Var}(Y) = np(1 - p).$$

Unfortunately, the binomial random variable does not address the problem at hand, since it counts the number of successes in n trials, when what we desire is the number of trials needed to achieve m successes. But it represents the next step in the buildup to the negative binomial distribution.

Then, the Geometric Random Variable

The geometric distribution is defined for an experiment that is very similar to the binomial experiment. It is also concerned with identical and independent trials, each of which can result in either success or failure. The probability of success is p, and is constant from trial to trial. However, instead of counting the number of successes that occur in n trials, the geometric random variable Y represents the number of the individual trial on which the first success occurs. Thus, the experiment consists of a series of trials, and concludes with the first success.

The PMF of the geometric distribution is given as follows. A random variable Y is said to have a *geometric probability distribution* if and only if:

$$P(Y = y) = (1 - p)^{y-1} p, \quad y = 1, 2, 3, \ldots, 0 \le p \le 1.$$

The mean and variance of the geometric distribution are given as follows:

$$\mu = E(Y) = \frac{1}{p}$$

$$\sigma^2 = \mathrm{Var}(Y) = \frac{1-p}{p^2}.$$

As before, the geometric random variable does not quite address the problem at hand either, since it counts the number of trials needed to achieve the *first* success, when what we desire is the number of trials needed to achieve m successes. However, this is our next stop on the road to development of the negative binomial distribution.

Finally, the Negative Binomial Random Variable

At last, we can extend the concept of the geometric random variable to count the number of trials needed to achieve the rth success ($r > 1$).

The negative binomial distribution is defined for an experiment that is very similar to the geometric experiment. Again, it is concerned with identical and independent trials, each of which can result in either success or failure. The probability of success on any given trial is p, and is constant from trial to trial. However, instead of considering the number of the trial on which the *first* success occurs, the negative binomial random variable Y_r is the number of the individual trial on which the rth (second, third, fourth, etc.) success occurs. Thus, the experiment consists of a series of trials, and concludes with the rth success.

The PMF of the negative binomial distribution is given as follows. A random variable Y_r is said to have a *negative binomial probability distribution* if and only if:

$$\mathrm{P}(Y_r = y) = \binom{y-1}{r-1} p^r (1-p)^{y-r},\ y = r, r+1,\ r+2, \ldots,\ 0 \le p \le 1.$$

The mean and variance of the negative binomial distribution are given as follows:

$$\mu = \mathrm{E}(Y_r) = \frac{r}{p}$$

$$\sigma^2 = \mathrm{Var}(Y_r) = \frac{r(1-)p}{p^2}.$$

And, the cumulative distribution function of Y_r is given as follows:

$$P(Y_r \le y) = \sum_{i-r}^{y} \binom{i-1}{i-1} p^r (1-p)^{i-r}.$$

Example 1. In the context we have been discussing, we have a situation in which it takes, on average, 10 attempts to produce one functional unit. Thus, $p = 1/10 = 0.1$ and is assumed to be independent from trial to trial. From the definition above, the expected value of the negative binomial random variable that represents the number of attempts needed to achieve four functional units is

$$\mu = \mathrm{E}(Y_4) = \frac{4}{0.1} = 40.$$

What is the probability that we will achieve four functional units in exactly four attempts?

$$\mathrm{P}(Y_4 = 4) = \binom{3}{3} (0.1)^4 (0.9)^0 = 0.0001.$$

What is the probability that we will achieve four functional units in exactly 10 attempts?

$$\mathrm{P}(Y_4 = 10) = \binom{9}{3} (0.1)^4 (0.9)^6 = 0.0045.$$

What is the probability that we will achieve four functional units in exactly 40 attempts?

$$\mathrm{P}(Y_4 = 40) = \binom{39}{3}(0.1)^4(0.9)^{36} = 0.0206.$$

Each of these probabilities is very small, so it probably doesn't make sense to base a cost estimate on a specific number of attempts. That is, if we were to estimate the cost based on the expected value, 40, we would have a cost estimate of $20M with an associated probability of 0.0206. Why would anyone want to commit $20M on such a small likelihood of success?

Of course, in reality, the manufacturing process would stop once the 4th success was achieved. This could happen at almost any time. It could happen on the 4th attempt, though that is highly unlikely. Or it could happen on the 33rd attempt with substantially higher likelihood. So, a more relevant question is, "What is the probability that we will achieve success in *n or fewer* attempts?"

Example 2. What is the probability that we will achieve four functional units in *no more than* four attempts?

$$\mathrm{P}(Y_4 \le 4) = \sum_{i=4}^{4}\binom{i-1}{3}(0.1)^4(0.9)^{i-4} = 0.0001.$$

Note: you may have noticed that this is a silly question. In this case, we would need to have four successes in a row, so the question is the same as $P(Y_4 = 4)$, and, of course, the answer is the same too!

$$\mathrm{P}(Y_4 = 4) = \binom{3}{3}(0.1)^4(0.9)^{0} = 0.0001.$$

What is the probability that we will achieve four functional units in *no more than* 10 attempts?

$$\mathrm{P}(Y_4 \le 10) = \sum_{i=4}^{10}\binom{i-1}{3}(0.1)^4(0.9)^{i-4} = 0.0128.$$

What is the probability that we will achieve four functional units in *no more than* 40 attempts?

$$\mathrm{P}(Y_4 \le 40) = \sum_{i=4}^{40}\binom{i-1}{3}(0.1)^4(0.9)^{i-4} = 0.5769.$$

Now we see that if we plan for 40 attempts, then there is nearly a 58% chance that we will have our four successes sometime within those 40 trials. Unfortunately, we just don't know on which trial we will be able to stop. At the same time, it is quite possible that 40 trials are not enough. We still have more than a 42% chance that it will take *more than* 40 trials in order to achieve four functional units! What is a cost analyst to do?

Determining the Appropriate Number of Units to Build

What the cost analyst would really like to know is, "On which trial will we achieve the *r*th success?" But, as we have just shown via the negative binomial distribution, the number of the trial on which the *r*th success will occur is a random variable, and cannot be known with certainty. However, there are a few good guesses that we can make. One way to do this would be to select the *most likely* trial number. Another would be to select the *mean* trial number, and a third would be to select the trial number that corresponds to a *given level of confidence*.

The Most Likely Trial Number

Consider the PMF of the Negative Binomial distribution for our previous example shown in Figure 1. This is the PMF for values of y from 4 to 100.

Notice that the PMF is maximized at a value near 30, which is less than the expected value of 40. In fact, it can be shown that the negative binomial distribution's PMF is maximized at:

$$y_{\text{mode}} = \left\lfloor 1 + \frac{r-1}{p} \right\rfloor.$$

In this example, the PMF is maximized when $y = 31$:

$$y_{\text{mode}} = \left\lfloor 1 + \frac{4-1}{0.1} \right\rfloor = 31.$$

This is shown in Table 1, if you look specifically at the number located in the column headed "P(Y = y)" corresponding to value 31 of the "Y" column. Therefore, using the *most likely* trial number approach, we can say that the *most likely* trial number on which the 4th success will occur is trial number 31 and, thus, estimate the cost of building 31 units. The drawback to this method, however, is that there is still a very low probability that it will take exactly 31 trials to achieve four successful units, and the probability that we will achieve four successes at some point up to and including the 31st trial is provided in Table 1 and is also significantly small.

$$\mathrm{P}(Y_4 <= 31) = \binom{30}{3}(0.1)^4(0.9)^{27} = 0.0236,$$

$$\mathrm{P}(Y_4 = 31) = \sum_{i=4}^{31}\binom{i-1}{3}(0.1)^4(0.9)^{i-4} = 0.3762.$$

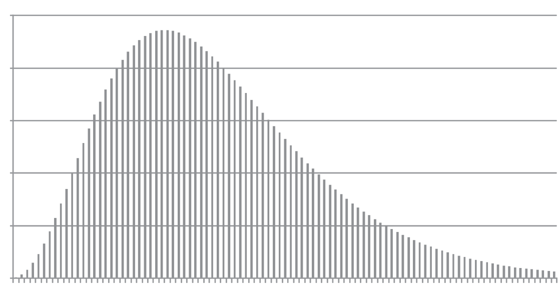

FIGURE 1 Negative binomial PMF.

TABLE 1 Negative binomial table of probabilities

Avg. Number of attempts needed, *n*: 10 to achieve one success
Number of Widgets needed, *m*: 4
Expected number of trials needed: 40
Std Dev: 19.0

Trial no., Y	P(Y = y)	P(Y <= y)	Trial no., Y	P(Y = y)	P(Y <= y)	Trial no., Y	P(Y = y)	P(Y <= y)	Trial no., Y	P(Y = y)	P(Y <= y)
4	0.0001	0.0001	21	0.0190	0.1520	38	0.0216	0.5352	55	0.0115	0.8130
5	0.0004	0.0005	22	0.0200	0.1719	39	0.0211	0.5563	56	0.0110	0.8240
6	0.0008	0.0013	23	0.0208	0.1927	40	0.0206	0.5769	57	0.0104	0.8344
7	0.0015	0.0027	24	0.0215	0.2143	41	0.0200	0.5969	58	0.0099	0.8443
8	0.0023	0.0050	25	0.0221	0.2364	42	0.0195	0.6164	59	0.0094	0.8537
9	0.0033	0.0083	26	0.0226	0.2591	43	0.0189	0.6352	60	0.0089	0.8626
10	0.0045	0.0128	27	0.0230	0.2821	44	0.0182	0.6534	61	0.0084	0.8710
11	0.0057	0.0185	28	0.0233	0.3054	45	0.0176	0.6711	62	0.0080	0.8790
12	0.0071	0.0256	29	0.0235	0.3290	46	0.0170	0.6881	63	0.0076	0.8866
13	0.0085	0.0342	30	0.0236	0.3526	47	0.0164	0.7044	64	0.0071	0.8937
14	0.0100	0.0441	31	0.0236	0.3762	48	0.0157	0.7201	65	0.0067	0.9004
15	0.0114	0.0556	32	0.0235	0.3997	49	0.0151	0.7352	66	0.0064	0.9068
16	0.0129	0.0684	33	0.0234	0.4231	50	0.0145	0.7497	67	0.0060	0.9128
17	0.0142	0.0826	34	0.0231	0.4462	51	0.0139	0.7636	68	0.0056	0.9184
18	0.0156	0.0982	35	0.0228	0.4690	52	0.0133	0.7768	69	0.0053	0.9238
19	0.0168	0.1150	36	0.0225	0.4915	53	0.0127	0.7895	70	0.0050	0.9288
20	0.0180	0.1330	37	0.0221	0.5136	54	0.0121	0.8015	71	0.0047	0.9335

However, there is nearly a 38% chance that we will be successful within 31 trials, so it may be worth the risk. In this case, our cost estimate would be \$500K × 31 = \$15.5M.

The Mean Trial Number

One might choose to use this approach in order to be more confident that a sufficient number of trials will be performed in order to achieve four successful units. We have been here before. The *mean* trial number is simply the expected value of the negative binomial random variable:

$$\mu = \mathrm{E}(Y_r) = \frac{r}{p}.$$

For our example, this equates to 40 trials, as shown previously. Using this approach, we can say that the mean number of trials in which we can achieve four successful units is 40 and, thus, estimate the cost of building 40 units. Again, however, there is still a very low probability that it will take exactly 40 trials to achieve four successful units, and the probability that we will achieve four successes at some point up to and including the 40th trial is also fairly small:

$$\mathrm{P}(Y_4 = 40) = \binom{39}{3}(0.1)^4(0.9)^{36} = 0.0206,$$

$$\mathrm{P}(Y_4 <= 40) = \sum_{i=4}^{40}\binom{i-1}{3}(0.1)^4(0.9)^{i-4} = 0.5769,$$

as we have shown earlier. Therefore, there is nearly a 58% chance that we will be successful within 40 trials, and our cost estimate in this case would be \$500K × 40 = \$20M.

The Trial Number That Corresponds to a Given Level of Confidence

A third approach may be to specify a level of confidence in advance, and then determine the number of trials needed to achieve the *r*th success based on that confidence level. Examples of informative confidence levels might be 30%, 50%, 80%, and so on.

Using our example, suppose management is comfortable with an 80% confidence level that the number of units built will deliver four successes. One can refer to the cumulative distribution function (CDF) in Figure 2 to determine the number of trials necessary.

Here we can see that the number of trials that corresponds to the 80th percentile is 54 trials. Another way to view this is to develop a table of probabilities, such as that shown in Table 1, to which we have referred twice before.

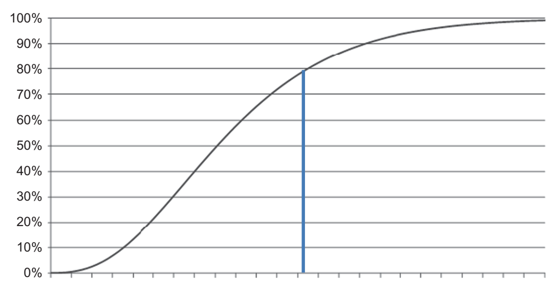

FIGURE 2 Negative binomial CDF (color figure available online).

Again we see that 54 trials will give us a level of confidence of 80%. However, this will require a lot more money in the budget, as $500K × 54 = $27M. Perhaps a more reasonable level of confidence would be 50%, in which case we would need to plan for 37 trials, or $18.5M.

So far, we have shown that it is impossible to determine with precision the exact number of trials needed to produce m successful units, and that the best we can do is to make an educated guess when estimating the number of trials necessary. Perhaps a more interesting proposition would be to use the uncertainty discussed above in a cost-risk analysis. Recall that when performing a cost-risk analysis, we are ultimately attempting to capture the estimating uncertainty in a cost probability distribution. As part of this process, we model the uncertainty of both the CERs and the CER input variables. Since quantity is an input variable at some point in the estimating process, then it stands to reason that quantity can be modeled as a random variable. Therefore, where the quantities of units that are yield-constrained are concerned, the quantity can be modeled as a negative binomial random variable.

Number of Units as a Random Variable in a Cost-Risk Analysis

As should be apparent by this point, it is difficult to arrive at a "best answer" when determining how many units will be required. We have considered the "most likely" value, the "mean" value, and the "number that corresponds to a given level of confidence." These are useful methods if one is trying to plan a budget or develop a point cost estimate. But, when developing an independent cost estimate, it is far more interesting to understand the range of possible outcomes for use in a cost-risk/uncertainty analysis.

For someone who performs cost-risk analyses, it is preferable to model the "number to build" as a random variable rather than to use a deterministic value derived through any of the means described previously.

Example 3. Suppose a certain Work Breakdown Structure (WBS) element in a cost estimate is "1.2.4 Widgets." Suppose also that the (fictional) cost estimating relationship (CER) for widgets (whether they work or not!) is as shown in Figure 3:

$$\text{Unit Cost (FY11\$K)} = 57.17 + 0.3(\text{Area(sq.in)})^{1.82},$$
$$SPE = 22\%,\ \text{Pearson's}\, R^2 = 84\%,\ \text{Bias} = 0\%.$$

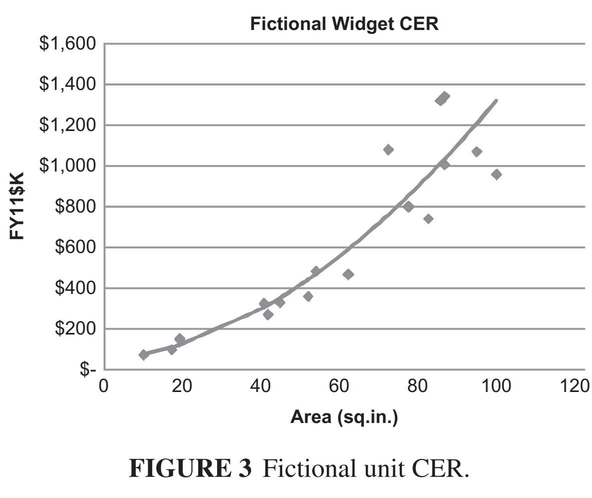

FIGURE 3 Fictional unit CER.

In other words, we have a CER that estimates the cost of an individual widget as a function of its area in square inches. As is typical of CERs, this one also has uncertainty measured by the standard percent error (SPE). In keeping with our original example, we assume that there is no significant learning curve, so the cost of the *n*th unit is the same as the cost of the first unit. This is not necessarily a realistic assumption in all cases, though.

In addition, suppose we know that it takes, on average, 10 attempts to build one functional unit, and again, we need four functional units. The typical cost-risk approach is to model (a) the uncertainty of the independent variable (sq. in.), (b) the uncertainty of the CER, and (c) the uncertainty of the number of units that need to be built in order to result in four functional units. Let us suppose that the area is fixed, so there is no need to model that uncertainty. Then, using, say, the commercially available software product @RISK, one would model CER uncertainty as (usually) a lognormal distribution, and the number required, M, as a negative binomial distribution.

$$\text{Widget Cost} \sim \text{Lognormal}\left(\mu = 57.17 + 0.3(\text{Area (sq.in)})^{1.82}, \sigma = 22\%\right)$$
$$M \sim \text{Neg.Bin.}\,(r = 4,\ p = 0.1).$$

If the area is 55.12 square inches, then the spreadsheet implementation would look as shown in Figure 4.

The resulting point estimate is \$20.0M (FY11) for 40 attempts at building the units. But, now take a look at the probability distribution shown in Figure 5 that results from running a Monte Carlo simulation.

With this result we see the combined probability distribution in which both the numbers of attempts necessary, and the cost of each attempt, are modeled as random variables. This distribution represents a convolution of the negative binomial quantity and the lognormal CER, and has the typical long right tail that we would expect.

By examining this distribution, we see that there is a 58.4% chance that the cost to build four functioning units will not exceed \$20.0M. But, we also see that there is substantial likelihood of costing more than that, with even \$50M not being out of the realm of possibility. However, using this approach we can also see the range of possible costs. More interesting yet, we can see in Figure 6 that there is an 80% chance that the cost to build four functioning units will not exceed \$27.6M.

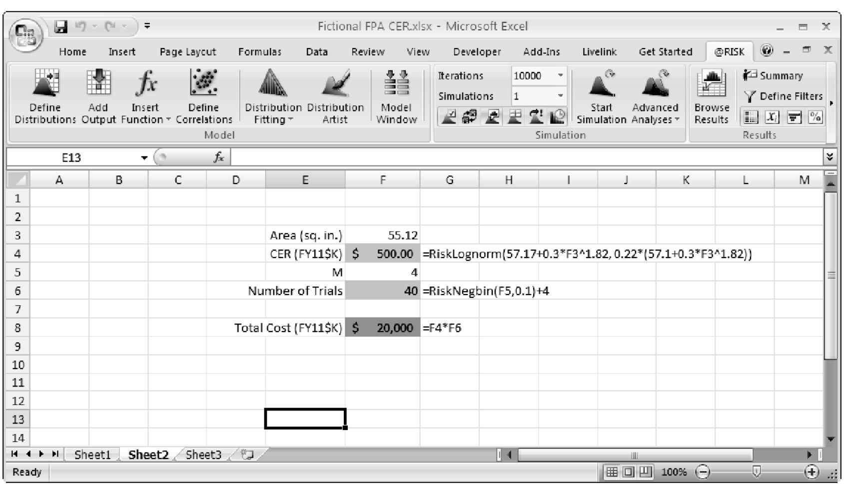

FIGURE 4 Monte Carlo implementation using @RISK.

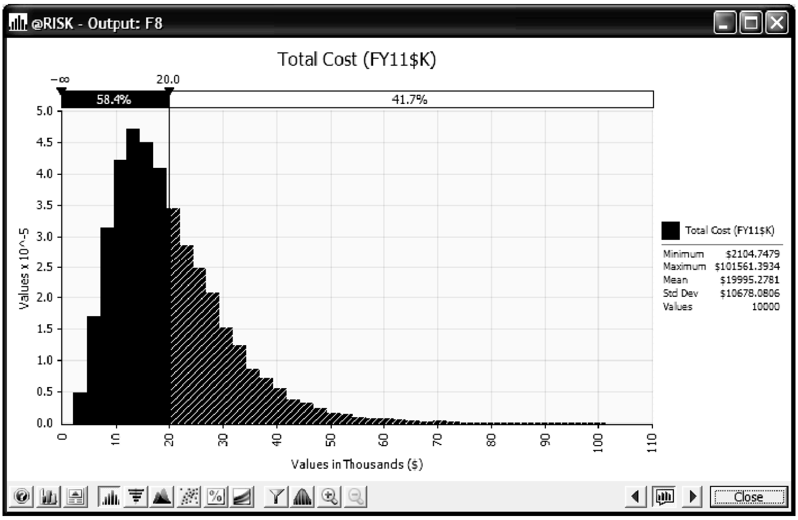

FIGURE 5 Resulting cost histogram using @RISK showing location of point estimate.

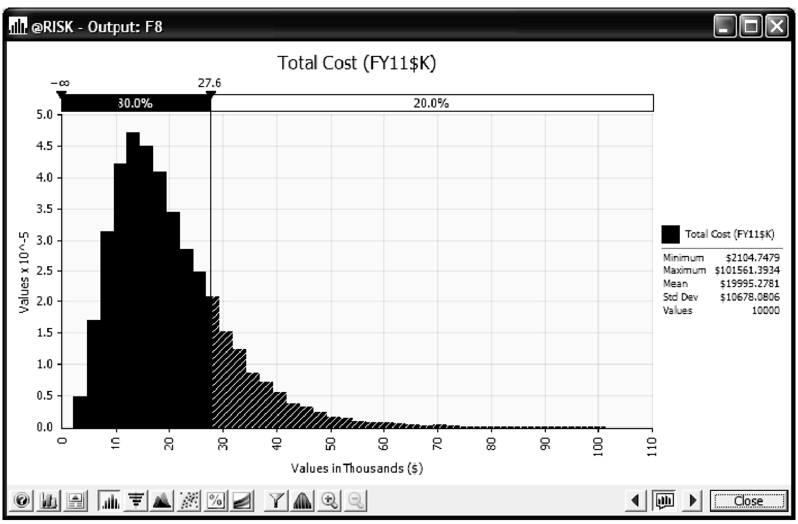

FIGURE 6 Resulting cost histogram using @RISK showing location of 80th percentile.

Recall that when we assumed a fixed cost of $500K per unit, the 80% confidence level result was to build 54 units at a cost of $27.0M. The cost distribution gives a reasonably consistent answer, but does not require us to worry about "how many to build." This distribution includes the combined effects of uncertainty in the cost of each unit and the uncertainty in the number to build. If we simply budget $27.6M, we have an 80% chance that we will meet our goal.

Modeling Uncertainty in a Source Selection Evaluation

Finally, we consider the use of the negative binomial distribution as one means of determining the most probable cost (MPC) in a Government source selection. Here, one must be careful. In an industry cost proposal, especially when two or more competitors exist, an offeror's proposed cost is likely to be lower than it should be. The reason for this is that, in a competition, offerors are motivated to bid the lowest price possible while still being plausibly reasonable and realistic. The usual result is that the ultimate winner of the competition has bid a price that is too low, and must rely on inevitable subsequent engineering changes in order to recoup the "real" cost of the contract. Therefore, in a Government source selection, the Government ideally computes an MPC that theoretically represents the "real" cost of the contract, so that it will be prepared for future cost increases.

Continuing on with the earlier example, suppose an offeror in a competitive source selection bids a price in which it is assumed that four functional units can be produced in a minimum amount of trials. It is unlikely that the offeror would be bold enough to assume it will take only four trials, with success on each trial. But it would not be out of the realm of possibility for the offeror to bid, say, the 20th percentile solution. Furthermore, suppose the offeror is bidding an optimistic price for the units, for example, \$400K. This would correspond to 24 trials, yielding a cost estimate of $24 \times \$400\text{K} = \$9.6\text{M}$ for the four units. The offeror would, of course, come up with some plausible-sounding justification for being able to produce the four units in only 24 trials, and for the discounted price of a build. But, using the assumptions in our example, there is only about a one in five chance that the offeror would be successful given 24 attempts.

Now look at this from the Government cost evaluator's perspective. For the MPC, the cost evaluator should refer to the techniques described previously in determining the appropriate number of units that will need to be built. In addition, the cost evaluator would tend to use the more realistic price of \$500K per build. Moreover, to be consistent, suppose the MPC is indeed intended to represent the "most probable cost." In that event, the cost evaluator should use the "Most Likely Trial Number" method.

Example 4. The offeror has bid a price that corresponds to 24 builds at a price of \$400K per build, for a total price of \$9.6M. Now, for the purposes of the MPC, the Government cost evaluator assumes the more realistic price of \$500K per build, and assumes that the "most likely trial number" is

$$y_{\text{mode}} = \left\lfloor 1 + \frac{r-1}{p} \right\rfloor = \left\lfloor 1 + \frac{4-1}{0.1} \right\rfloor = 31.$$

The resulting "most probable cost" for the units is $31 \times \$500\text{K} = \$15.5\text{M}$. As shown previously, however, the resulting probability that 31 trials will be enough is just a little over 37%. It is also a fact that the MPC, despite its acronym, is not actually constrained to be the "most probable" cost, but can actually be defined *ad hoc* as desired. Therefore, for a given source selection activity, the source selection team can choose to use one of the other "more probable" methods, such as the "mean trial number" method or the "trial number that corresponds to a given level of confidence" method.

Summary and Conclusions

This article describes the nature of the problem of determining the number of units to build in a yield-constrained manufacturing process. The number of units to build is modeled as a negative binomial random variable. The example of a unit manufacturing process is used to illustrate how to determine the appropriate number of units to plan for under various assumptions. We conclude that there is no "best answer" to the question, but we were able to develop the methodology needed to arrive at alternative answers under the assumptions of the "most likely" number, the "mean" number, and the number that corresponds to a "given level of confidence."

Then, we discuss how a cost analyst can use the methods described to model the number of units as a random variable in a cost-risk analysis, and we show how to apply the methods in a Government source selection evaluation—with the goal of determining the Government's MPC position.

This article will be of use to cost analysts who are in the position of estimating costs when certain hardware elements are yield-constrained, as well as to those who use cost uncertainty analysis techniques. In addition, the article will be useful to source selection cost evaluators in support of MPC development.

Potential Future Research

We have finessed the question of what to do if the manufacturing process "improves" over time, meaning that the probability of success increases with increased numbers of trials. Future research on this topic should include a study of how to determine the number of units to build if the success probability is not constant.

References

Devore, J. L. (1995). *Probability and statistics for engineering and the sciences* (4th ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Mendenhall, W., Wackerly, D. D., & Scheaffer, R. L. (1990). *Mathematical statistics with applications* (4th ed.). Boston, Massachusetts: PWS-Kent.

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Belmont, CA: Wadsworth Publishing Company.

About the Author

Timothy P. Anderson is Systems Director for NASA Program Assessments with The Aerospace Corporation, Arlington VA, and a professional cost analyst and operations research analyst with over 17 years experience, primarily in the context of Department of Defense (DoD) weapon systems and national security space acquisition. After retiring from the U.S. Navy in 2001 after 20 years service, he joined The Aerospace Corporation, serving from 2001 to 2008 and again beginning in 2010, with a stint as a Technical Manager at MCR, LLC in the intervening period. Among his areas of expertise are cost analysis, cost uncertainty analysis, operations research, and decision analysis. His introduction to the estimating field occurred in 1994 when the Navy assigned him to the Naval Center for Cost Analysis. His next, and last, Navy assignment was teaching cost estimation, operations research, and other technical courses as a military professor at the Naval Postgraduate School, Monterey, CA. He is a SCEA Certified Cost Estimator/Analyst (CCEA®), a board member of the Washington DC Area chapter of SCEA, a former adjunct professor in the Systems Engineering/Operations Research Department of George Mason University, and an adjunct professor in the Systems Engineering and Operations Research departments of the Naval Postgraduate School. Mr. Anderson was recognized as the 2010 SCEA National Estimator/Analyst of the Year for Technical Achievement and was awarded the 2010 Wayne E. Meyer Award for Teaching Excellence by the Naval Postgraduate School. He is a frequent presenter of topics related to cost estimating and cost uncertainty analysis at forums, including SCEA, the Military Operations Research Society (MORS), the DoD Cost Analysis Symposium (DoDCAS), the Integrated Program Management (IPM) Conference, and the Space Systems Cost Analysis Group (SSCAG). Mr. Anderson has a B.S. in Industrial and Operations Engineering from the University of Michigan and an M.S. in Operations Research from the Naval Postgraduate School.