Important Probability Distributions

Certain discrete distributions describe many natural phenomena and have broad applications in statistical process control. Two of them are the binomial distribution and the Poisson distribution, discussed next. Later, some important continuous probability distributions are introduced.

Binomial Distribution

The **binomial distribution** describes the probability of obtaining exactly x "successes" in a sequence of n identical experiments, called trials. A *success* can be any one of two possible outcomes of each experiment. In some situations, it might represent a defective item, in others, a good item. The probability of success in each trial is a constant value p. The binomial probability function is given by the following formula:

$$
\begin{aligned}
f(x) &= \binom{n}{x} p^x (1-p)^{n-x} \\
&= \frac{n!}{x!(n-x)!}\, p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n
\end{aligned}
$$

where p is the probability of a success, n is the number of items in the sample, and x is the number of items for which the probability is desired (0, 1, 2, . . . , n). The expected value, variance, and standard deviation of the binomial distribution are

$$
\begin{aligned}
E(p) &= \mu = np \\
\sigma^2 &= np(1-p) \\
\sigma &= \sqrt{np(1-p)}
\end{aligned}
$$

Binomial probabilities for selected values of p and n have been tabulated in Appendix D. Naturally, computer programs and Excel functions are also available to make binomial computations easier.

Poisson Distribution

The second discrete distribution often used in quality control is the **Poisson distribution**. The Poisson probability distribution is given by

$$f(x) = \frac{e^{-\mu}\mu^x}{x!}$$

where μ = expected value or average number of occurrences, $x = 0, 1, 2, 3, \ldots$, and $e \approx 2.71828$, a constant.

The Poisson distribution is closely related to the binomial distribution. It is derived by allowing the sample size (n) to become very large (approaching infinity) and the probability of success or failure (p) to become very small (approaching zero) while the expected value (np) remains constant. Thus, when n is large relative to p, the Poisson distribution can be used as an approximation to the binomial. A common rule of thumb is if $p \le 0.05$ and $n \ge 20$, the Poisson will be a good approximation with $\mu = np$. It is also used to calculate the number of occurrences of an event over a specified interval of time or space, such as the number of scratches per square inch on a polished surface.

Note that the previously stated conditions under which the Poisson distribution is a good approximation of the binomial have been met; that is, $n \ge 20$ and $p \le 0.05$. Table PD.1 compares these probability values to the true values using the binomial distribution. The results show that the Poisson distribution does provide a good ap-

Table PD.1 Binomial versus Poisson Probability Values

x	Binomial Probability	Poisson Probability
0	0.21464	0.22313
1	0.33890	0.33467
2	0.25864	0.25102
3	0.12705	0.12551
4	0.04514	0.04707
5	0.01235	0.01412
	0.99672	0.99552

proximation to the binomial probabilities when the specified conditions are met. If the conditions for the Poisson approximation cannot be met, a normal approximation to the binomial, discussed in the next section, may be of use.

Two of the most frequently used continuous probability distributions are the normal distribution and the exponential distribution. They form the basis for many of the statistical analyses performed in quality assurance today.

Normal Distribution

The probability density function of the **normal distribution** is represented graphically by the familiar bell-shaped curve. However, not every symmetric, unimodal curve is a normal distribution, nor can all data from a sample or population be assumed to fit a normal distribution. However, data are often assumed to be normally distributed to simplify certain calculations. In most cases, this assumption makes little difference in the results but is important from a theoretical perspective.

The probability density function for the normal distribution is as follows:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty$$

where

μ = the mean of the random variable x

σ^2 = the variance of x

e = 2.71828 . . .

π = 3.14159 . . .

If a normal random variable has a mean $\mu = 0$ and a standard deviation $\sigma = 1$, it is called a **standard normal distribution**. The letter z is usually used to represent this particular random variable. By using the constants 0 and 1 for the mean and standard deviation, respectively, the probability density function for the normal distribution can be simplified as

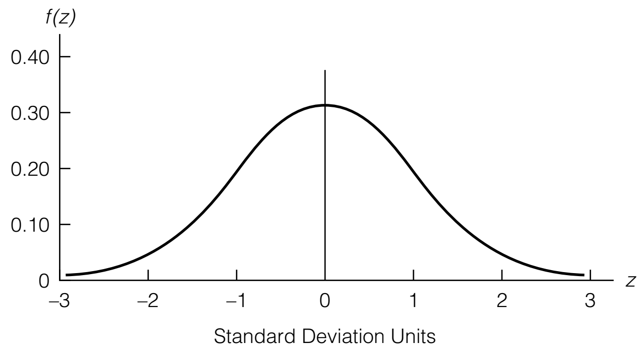
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

This standard normal distribution function is shown in Figure PD.1. Because $\sigma = 1$, the scale on the z axis is given in units of standard deviations. Special tables of areas under the normal curve have been developed as an aid in computing probabilities. Such a table is given in Appendix A.

Fortunately, *any* normal distribution involving a random variable x with a known (or estimated) mean and standard deviation is easily transformed into a standard normal distribution using the following formula:

$$z = \frac{x - \mu}{\sigma}$$

Figure PD.1 Standard Normal Distribution



This formula takes the value of the variable of interest (x), subtracts the mean value (μ), and divides by the standard deviation (σ). This calculation yields a random variable z, which has a standard normal distribution. Probabilities for this variable can then be found in the table in Appendix A.

The area under the curve that corresponds to one standard deviation from the mean is 0.3413; therefore, the probability that the value of a normal random variable falls within one standard deviation ($\pm 1\sigma$) from the mean is 0.6826. The corresponding x-values are often called *1-sigma limits* in statistical quality control terminology. Two standard deviations on one side of the mean correspond to 0.4772 area under the curve, so the probability that a normal random variable falls within a *2-sigma limit* is twice that figure, or 0.9544. Three standard deviations encompass 0.4986 area under the curve on either side of the mean, or a total area of 0.9972. Hence, the *3-sigma limit* encompasses nearly all of the normal distribution. These concepts form the basis for control charts discussed in Chapter 14.

Normal Approximation to the Binomial Although the binomial distribution is extremely useful, it has a serious limitation when dealing with either small probabilities or large sample sizes—it is tedious to calculate. The discussion of the Poisson approximation to the binomial showed that when the probability of success or failure becomes small, the Poisson distribution permits calculation of similar probability values more easily than the binomial. Also, as the sample size gets large (approaches infinity), the binomial distribution approaches the normal distribution as a limit. Hence, for large sample sizes, good approximations of probabilities that would have been calculated using the binomial distribution can be obtained by using the normal distribution. The normal approximation holds well when $np \geq 5$ and $n(1 - p) \geq 5$.

Exponential Distribution

Another continuous distribution commonly used in quality assurance is the **exponential distribution**. The exponential distribution is used extensively in reliability estimation, discussed in Chapter 12. The probability density function for the exponential distribution is much simpler than the one for the normal distribution. Therefore, direct evaluation is easier, although tabulated values for the exponential distribution are also readily available (see Appendix F). The formula for the exponential probability density function is

$$f(x) = \frac{1}{\mu} e^{-x/\mu}, \quad x \geq 0$$

where

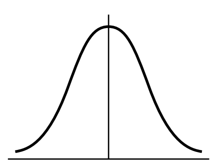
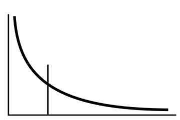
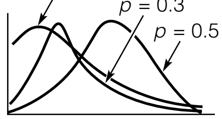
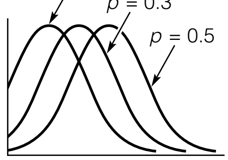
μ = mean value for the distribution
x = time or distance over which the variable extends
e = 2.71828 . . .

The cumulative distribution function is

$$F(x) = 1 - e^{-x/\mu}$$

Figure PD.2 summarizes the four important distributions reviewed in this appendix.

Figure PD.2 Summary of Common Probability Distributions Used in Quality Assurance

Distribution	Form	Probability Function	Comments on Application
Normal		$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ μ = Mean σ = Standard deviation	Applicable when a concentration of observations falls about the average and when observations are equally likely to occur above and below the average. Variation in observations is usually the result of many small causes.
Exponential		$y = \frac{1}{\mu} e^{-\frac{x}{\mu}}$	Applicable when more observations are likely to occur below the average than above.
Poisson		$y = \frac{e^{-\mu}\mu^x}{x!}$ n = Number of trials p = Probability of occurrence x = Number of occurrences $\mu = np$	Same as binomial but particularly applicable when many opportunities for occurrence of an event are possible but have a low probability (less than 0.10) on each trial.
Binomial		$y = \frac{n!}{x!(n-x)!} p^x q^{n-x}$ n = Number of trials x = Probability of occurrence p = Number of occurrences $q = 1 - p$	Applicable in defining the probability of x occurrences in n trials of an event that has a constant probability of occurrence on each independent trial.

Source: Adapted from J. M. Juran and F. M. Gryna, Jr., Instructor's Manual to accompany *Quality Planning and Analysis* (New York: McGraw-Hill, 1980), 125. © 1980 by McGraw-Hill, Inc. Used with permission.