Chapter 5
Gaussian Statistics — An Overview

The *Gaussian distribution* (*normal* or *bell-shaped distribution*) is a widely used statistical distribution and it is generally used as the foundation for statistical quality control. Simply measuring the time-zero values of a parameter (resistor values, mechanical tolerances, children heights, class grades on a test, etc.) can result in a distribution of values which can be described by a normal distribution.

5.1 Normal Distribution

The normal distribution *f(x)* shown in Fig. 5.1 is defined by the equation:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\left[\frac{x - x_{50}}{\sigma\sqrt{2}}\right]^2\right\}. \tag{5.1}$$

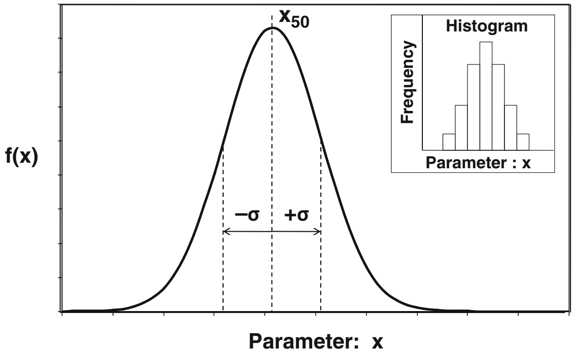

Fig. 5.1 Gaussian (or normal) distribution is illustrated. x_{50} is the mean=mode=median. 68.3% of observations are between (+/−)σ, 95.5% of observations are between (+/−)2σ and 99.7% of observations are between (+/−)3σ.

J.W. McPherson, *Reliability Physics and Engineering*,
DOI 10.1007/978-1-4419-6348-2_5, © Springer Science+Business Media, LLC 2010

For the normal distribution (since it is symmetrical), x_{50} represents the mean = mode = median. In order to be consistent with later chapters in this book, x_{50}[1] will be referred to as the median (50% of the values are below the median value and 50% are above). σ is the standard deviation[2] (represents the spread in the data) and can be approximated by $\sigma = x_{50} - x_{16}$, where x_{16} represents the value and where 16%[3] of the observations are below this value. Once x_{50} and σ are determined from the data, then the full distribution is described by Eq. (5.1).

x_{50} and σ can be determined from a plot of the cumulative fraction F(x):

$$F(x) = \int_0^x f(x)dx. \tag{5.2}$$

The cumulative (cum) fraction F integral in Eq. (5.2) must be numerically evaluated and is given by:

$$\begin{aligned} F(x) &= \frac{1}{2} erfc\left(\frac{x_{50} - x}{\sigma\sqrt{2}}\right) \qquad (for\ x \le x_{50}) \\ and& \\ F(x) &= 1 - \frac{1}{2} erfc\left(\frac{x - x_{50}}{\sigma\sqrt{2}}\right) \qquad (for\ x \ge x_{50}) \end{aligned}, \tag{5.3}$$

where the *erfc* stands for the *error function complement*. Some often used values for the *erfc* are shown in Table 5.1. In the past, such tables were widely used by engineers. Now, however, *ERFC* is a standard *Excel Function* so any arbitrary value is readily available to the engineer.

Shown below (Table 5.2) is an example of a suggested method for data collection that can be used for relatively easy statistical analysis. In this example, 25 measurements were taken on the shear strength (in units of gm-f)[4] of Au ball-bonds to aluminum pads on semiconductor chips. These 25 observed measurements (data points) were then ranked from smallest to largest value. In order to insure that all 25 data points can be used when plotting the data, an unbiased estimate is used for the cumulative fraction failed F.[5] An unbiased estimate used for F in this text is:

---

[1]Mean can be estimated: $x_{50} = \sum_{i=1}^{N} x_i/N$, where N is the sample size.

[2]Standard deviation can be estimated: $\sigma = \left[\sum_{i=1}^{N} (x_i - x_{50})^2/(N-1)\right]^{1/2}$.

[3]A more precise value is 15.87%.

[4]One gm-f equals 9.8×10^{-3} Newton.

[5]A cumulative probability of exactly F=1 cannot be plotted. Therefore, in order to ensure that all 25 data points can be plotted, then an *unbiased estimate* of the cum F is needed. In reliability physics and engineering, Eq. (5.4) is generally used.

Table 5.1 Error Function Complement (*erfc*).

y	erfc(y)	y	erfc(y)
0	1.0000	1	0.1573
0.1	0.8875	1.1	0.1198
0.2	0.7773	1.2	0.0897
0.3	0.6714	1.3	0.0660
0.4	0.5716	1.4	0.0477
0.5	0.4795	1.5	0.0339
0.6	0.3961	1.6	0.0237
0.7	0.3222	1.7	0.0162
0.8	0.2579	1.8	0.0109
0.9	0.2031	1.9	0.0072
1	0.1573	2.0	0.0047

$$F = \frac{Observation\ \# \quad - \quad 0.3}{Sample\ Size \quad + \quad 0.4}, \tag{5.4}$$

where *observation #* is the cumulative number of observations.

The cum fraction F is very useful in that it permits relatively easy plotting of the statistical data and relatively easy parameter (x_{50}, σ) extraction from the data.

In Table 5.2, the Z-value is the number of standard deviations associated with a given cum fraction F and can found from standard lookup tables such as the ones below, or can be easily generated with an EXCEL spreadsheet: to go from Z to F, use the EXCEL function F=NORMSDIST(Z); to go from F to Z, use the EXCEL function: Z=NORMSINV(F).

The plot of the data (from Table 5.2) is shown in Fig. 5.2, as well as the extracted best fitting normal distribution parameters (x_{50}, σ). Using these best fitting normal-distribution parameters (x_{50},σ), shown in Fig. 5.2, the resulting normal distribution is shown Fig. 5.3.

In general, once the normal distribution parameters (x_{50}, σ) are determined, then any other fraction F can be found using the equation:

$$x_F = x_{50} - z_F\sigma. \tag{5.5}$$

The following relations are so frequently used that they are highlighted here:

$$x_{16\%} = x_{50} - 1\sigma; \quad x_{1\%} = x_{50} - 2.33\sigma; \quad x_{0.13\%} = x_{50} - 3\sigma. \tag{5.6}$$

5.2 Probability Density Function

The normal distribution, as defined by Eq. (5.1), is a normalized distribution (which means that the total area under the curve is equal to unity). Thus, f(x) can be thought of as a probability density function such that f(x)dx is the probability of finding a

Table 5.2 Statistical Data for Bond Shear Strengths.

Statistical Data Collection and Analysis Method				
Sample Size	Observation	Ranked Data	Unbiased Estimate of Cum Fraction*	Normal Distribution
25	#	Shear Strength (gm-f)	F	Z-Value
	1	17.07	0.028	−1.918
	2	17.11	0.067	−1.499
	3	18.02	0.106	−1.246
	4	18.20	0.146	−1.055
	5	18.50	0.185	−0.896
	6	18.61	0.224	−0.757
	7	18.70	0.264	−0.632
	8	18.72	0.303	−0.515
	9	18.79	0.343	−0.406
	10	18.96	0.382	−0.301
	11	19.20	0.421	−0.199
	12	19.34	0.461	−0.099
	13	19.42	0.500	0.000
	14	19.44	0.539	0.099
	15	19.46	0.579	0.199
	16	19.55	0.618	0.301
	17	19.61	0.657	0.406
	18	19.75	0.697	0.515
	19	19.81	0.736	0.632
	20	19.88	0.776	0.757
	21	19.96	0.815	0.896
	22	19.98	0.854	1.055
	23	20.03	0.894	1.246
	24	20.25	0.933	1.499
	25	20.26	0.972	1.918

*Unbiased Estimate: F = (Observation # −0.3) / (Sample Size + 0.4)

value between x and x+dx, as illustrated in Fig. 5.4. The probability of finding a value in the range, between x_1 and x_2, is then given by:

$$P(x_1 \ to \ x_2) = \int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1). \tag{5.7}$$

Table 5.3. Conversion Tables for F to Z and Z to F.

From Cum F to Z-Values		From Z-Values to Cum F	
Cum	Standard Deviations (Z-values)	Standard Deviations	Cum F
F	NORMSINV(F)	Z-Value	NORMSDIST(Z)
0.001	−3.090232306	−3.0	0.0013
0.01	−2.326347874	−2.5	0.0062
0.1	−1.281551566	−2.0	0.0228
0.2	−0.841621234	−1.5	0.0668
0.3	−0.524400513	−1.0	0.1587
0.4	−0.253347103	−0.5	0.3085
0.5	−1.39214E-16	0.0	0.5000
0.6	0.253347103	0.5	0.6915
0.7	0.524400513	1.0	0.8413
0.8	0.841621234	1.5	0.9332
0.9	1.281551566	2.0	0.9772
0.95	1.644853627	2.5	0.9938
0.99	2.326347874	3.0	0.9987
0.999	3.090232306		

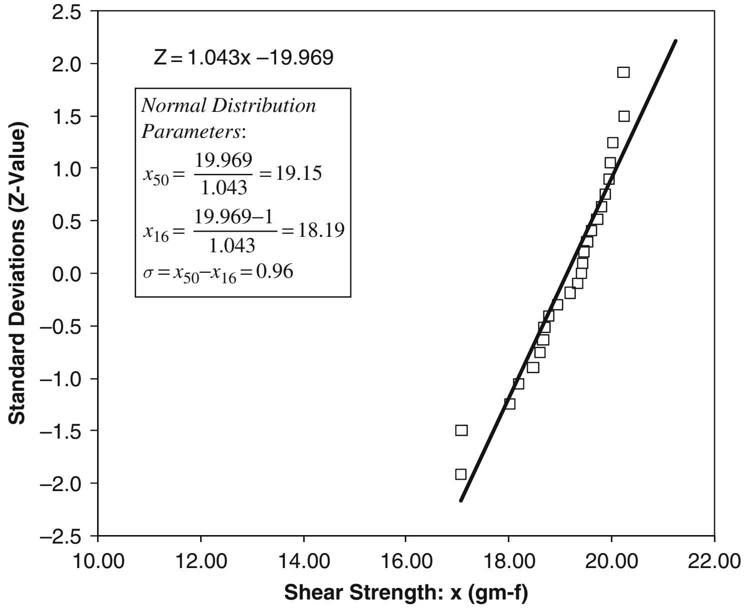

Fig. 5.2 Normal distribution plotting for data found in Table 5.2.

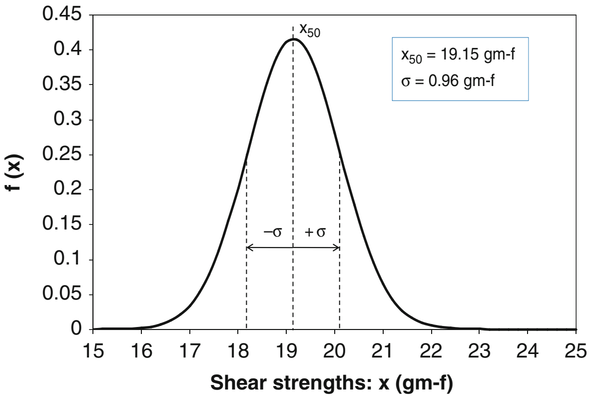

Fig. 5.3 Shear strengths (from Table 5.2) presented as a normal distribution.

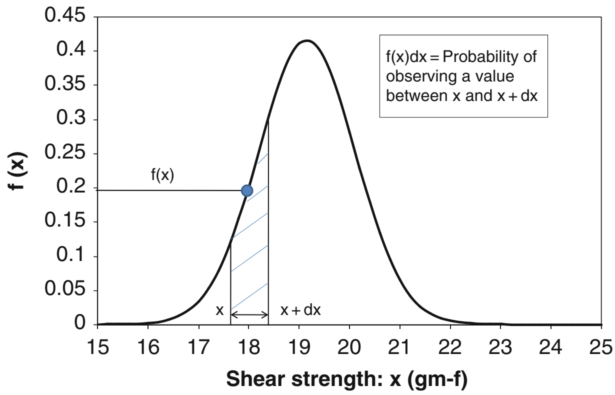

Fig. 5.4 f(x)dx represents the probability of observing a value of shear strength between x and x+dx.

Example Problem 5.1

From Fig. 5.3, normal distribution characteristic parameters that describe the ball-bond shear strengths are:

$$x_{50} = 19.15\ gm - f$$
$$and$$
$$\sigma = 0.96\ gm - f.$$

Find the probability that, if one does a single measurement of the shear strength of the ball bonds, a value between 18.0 and 19.0 gm-f will be obtained.

Solution

$$P(18.0\ to\ 19.0) = \int_{18.0}^{19.0} f(x)dx = F(19.0) - F(18.0)$$

The cum fail fractions are given by:

$$F(19.0) = \frac{1}{2} erfc\left(\frac{19.15 - 19.0}{0.96\sqrt{2}}\right) = 0.438$$
and
$$F(18.0) = \frac{1}{2} erfc\left(\frac{19.15 - 18.0}{0.96\sqrt{2}}\right) = 0.115$$

This gives:

$$P(18.0\ to\ 19.0) = F(19.0) - F(18.0) = 0.438 - 0.115 = 0.323$$

Therefore, the probability of a single bond-shear measurement producing a value between 18.0 and 19.0 gm-f is 0.323 (or 32.3%).

5.3 Statistical Process Control

Suppose that one knows (maybe from previous experience) that the lower reliable bond strength is 15.5 gm-f (an under-bonding condition). Likewise, when the time-zero bond strength exceeds 24.5 gm-f (an over-bonding condition), the bond is also unreliable. A very natural question to ask is—how does one statistically characterize the bonding process and is this process under control for reliable use? To answer the above questions, *capability* parameters *Cp* and *Cpk* are used.

Cp and *Cpk* are defined quantities:

$$Cp = \frac{(Upper\ Spec\ Limit) - (Lower\ Spec\ Limit)}{6\sigma} \tag{5.8}$$

and

$$Cpk = Cpl = \frac{(x_{50} - Lower\ Spec)}{3\sigma} \tag{5.9a}$$

or

$$Cpk = Cpu = \frac{(Upper\ Spec - x_{50})}{3\sigma}. \tag{5.9b}$$

The value of Cpk is stated based on whether Eq. (5.9a) or Eq. (5.9b) produces a smaller value. For a perfectly *centered* process, note that $Cpk = Cpl = Cpu = Cp$.

Example Problem 5.2

For the ball-bonding process illustrated in Fig. 5.3, with (x_{50}=19.15gm-f, σ=0.96), what is the *capability* (Cp) for this process and how well is it centered (Cpk)? Assume that the lower permitted level is 15.5 gm-f and the upper permitted level is 24.5 gm-f.

Solution

The process *capability* is given by:

$$Cp = \frac{(24.5 - 15.5)gm - f}{6(0.96)gm - f} = 1.56.$$

The centering for the process is given by:

$$Cpk = Cpu = \frac{(24.5 - 19.15)\ gm - f}{3(0.96)gm - f} \quad = 1.86$$
$$or$$
$$Cpk = Cpl = \frac{(19.15 - 15.5)\ gm - f}{3(0.96)gm - f} = 1.27.$$

Therefore, Cpk is 1.27 (note that the smaller of the two Cpk values is used). Cpk is non-symmetrical (since Cpl is not equal to Cpu), and is dominated by the lower-end specification.

Example Problem 5.3

From the previous example problem, it was determined that Cpk=1.27 and was dominated by the lower-end of the distribution relative to the specification (spec). (a) What fraction of the bonds has the potential for reliability problems occurring at the lower-end of the spec? (b) Fraction of bonds above the upper-end spec?

Solution

(a) One will need to find the number of standard deviations (*Z-value*) that corresponds to the lower-end spec. From Fig. 5.2 one obtains:

$$Z = \left(\frac{1.043}{gm - f}\right) x - 19.969.$$

With the lower-end spec at x = 15.5gm-f, this gives: Z = −3.803.
Using the EXCEL NORMSDIST function, one obtains:

$$F = \text{NORMSDIST}(-3.803) = 7.15 \times 10^{-5}.$$

Therefore, the fraction of bonds at reliability risk due to the lower-end spec is 71.5 ppm (parts per million) or 0.00715% of the bonds.

(b) One needs to find the number of standard deviations (Z-value) at the upper-end specification. Again, using:

$$Z = \left(\frac{1.043}{gm - f}\right) x - 19.969,$$

with the upper-end spec of x=24.5gm-f, one obtains: Z=5.585.
Using the EXCEL NORMSDIST function, one obtains:

$$F = \text{NORMSDIST}(5.585) = 0.9999999883.$$

Therefore, the fraction of the bonds at reliability risk due to the upper-end spec is 1−F where:
$1 - F = 1 - 0.9999999883 = 11.7 \times 10^{-9}$ or 11.7 parts per billion (ppb).

Problems

1. O-rings (from a manufacturing line) were randomly selected for diameter measurements. The 25 measurements are shown in the below table (all measurements are in mm). Find the Normal Distribution parameters: median diameter size (x_{50}) and the standard deviation σ.

181.4	173.0	172.2	173.5	180.5
187.8	178.6	170.7	179.5	186.5
171.1	180.0	183.4	177.3	187.0
176.7	186.1	182.5	174.2	188.7
184.0	185.6	190.0	175.4	189.5

Answers: x_{50}= 181.6 mm   σ = 6.8 mm

2. For the O-ring manufacturing process in Problem 1 (x_{50}=181.6 mm, σ=6.8 mm), find the capability parameters: *Cp* and *Cpk*. Assume that the upper spec limit is 215 mm and the lower spec limit is 155 mm.

   Answers: $Cp = 1.47$ $Cpk = 1.30$

3. The breakdown-strength distribution for capacitor dielectrics had a median value of $(E_{bd})_{50}$ =10.50 MV/cm and a σ=1.8 MV/cm.

   a) Find the fraction of caps with a breakdown $\leq$ 8MV/cm.
   b) Find the fraction of caps with a breakdown $\geq$ 12 MV/cm.

   Answers: a) 0.082 b) 0.202

4. The rupture-strength distribution of water pipes had a median value of $(\text{Rupture-Stress})_{50}$ =900 MPa and a σ=120 MPa.

   a) Find the fraction of pipes with a rupture stress of $\leq$ 600 MPa.
   b) Find the fraction of pipes with a rupture stress of $\geq$ 1300 MPa.

   Answers: a) 6.21×10^{-3}= 6210 ppm b) 4.29×10^{-4} = 429 ppm

5. Resistors have a resistance-value distribution with a median value of $(R)_{50}$ =189 ohm and a σ=3.5 ohm.

   a) Find the fraction of resistors with a resistance value of $\leq$160 ohms.
   b) Find the fraction of resistors with a resistance value of $\geq$200 ohms.

   Answers: a) 5.55×10^{-7} = 0.555 ppm b) 8.37×10^{-4} = 837 ppm

6. A group of patients had a heart-rate distribution with a median value $(HR)_{50}$=60 beats/min and a σ=2 beats/min.

   a) Find the fraction of patients with a heart rate of $\leq$ 50 beats/min.
   b) Find the fraction of patients with a heart rate of $\geq$ 70 beats/min.

   Answers: a) 2.87×10^{-7} = 0.287 ppm b) 2.87×10^{-7} = 0.287 ppm

7. Using the breakdown-strength distribution, defined in Problem 3, what are the process capability parameters: *Cp* and *Cpk*? Assume an upper-level limit of 12MV/cm and a lower-level limit of 8 MV/cm.

   Answers: $Cp = 0.37$ $Cpk = 0.28$

8. For the rupture-strength distribution, defined in Problem 4, what are the process capability parameters: *Cp* and *Cpk* ? Assume an upper-level limit of 1300Mpa and a lower-level limit of 800 MPa.

   Answers: $Cp = 0.97$ $Cpk = 0.83$

9. For the resistor distribution, defined in Problem 5, what are the process capability parameters: *Cp* and *Cpk* ? Assume an upper-level limit of 200 ohm and a lower-level limit of 160 ohm.

   Answers: $Cp = 1.90$ $Cpk = 1.05$

10. For the heart-rate distribution, defined in Problem 6, what are the capability parameters: *Cp* and *Cpk* for this group of patients? Assume an upper-level limit of 70 beats/min and a lower-level limit of 50 beats/min.

    Answers: $Cp = 1.67$ $Cpk = 1.67$

References

Ash, C.: **The Probability Tutoring Book**, IEEE Press, (1993).
Bowker, A. and G. Lieberman: **Engineering Statistics**, Prentice-Hall Publishing, (1972).
Dixon, W. and F. Massey: **Introduction to Statistical Analysis**, McGraw-Hill Book Co., (1957).
Fowler, J., L. Cohen and P. Jarvis: **Practical Statistics for Field Biology**, John Wiley & Sons, (1998).
Larsen, R.: **Engineering with EXCEL**, 2nd Ed., Pearson/Prentice Hall Publishing, (2005).
Miller, I. and J. Freund: **Probability and Statistics for Engineers**, Prentice Hall Publishing, (1977).