CHAPTER 2
STATISTICAL CONSIDERATIONS

Charles R. Mischke, Ph.D., P.E.
Professor Emeritus of Mechanical Engineering
Iowa State University
Ames, Iowa

NOMENCLATURE

A	Area, constant
a	Constant
B	Constant
b	Constant
C	Coefficient of variation
d	Diameter
F_i	ith failure, cumulative distribution function
$F(x)$	Cumulative distribution function corresponding to x
f_i	Class frequency
$f(x)$	Probability density function corresponding to x
h	Simpson's rule interval
i	failure number, index
LN	Lognormal
N	Normal
n	design factor, sample size, population
$\bar{n}$	mean of design factor distribution
P	Probability, probability of failure
R	Reliability, probability of success or survival

r	Correlation coefficient
S'_{ax}	Axial loading endurance limit
S'_e	Rotary bending endurance limit
S_y	Tensile yield strength
S'_{se}	Torsional endurance limit
S_{ut}	Tensile ultimate strength
x	Variate, coordinate
x_i	ith ordered observation
x_0	Weibull lower bound
y	Companion normal distribution variable
z	z variable of unit normal, $N(0, 1)$
α	Constant
Γ	Gamma function
Δx	Histogram class interval
θ	Weibull characteristic parameter
μ	Population mean
$\hat{\mu}$	Unbiased estimator of population mean
σ	stress
σ	Standard deviation
$\hat{\sigma}$	Unbiased estimator of standard deviation
$\Phi(z)$	Cumulative distribution function of normal distribution, body of Table 2.1
ϕ	Function
$\bar{\phi}$	Fatigue ratio mean
$\boldsymbol{\phi}_{ax}$	Axial fatigue ratio variate
$\boldsymbol{\phi}_b$	Rotary bending fatigue ratio variate
$\boldsymbol{\phi}_t$	Torsional fatigue ratio variate

2.1 INTRODUCTION

In considering machinery, uncertainties abound. There are uncertainties as to the

- Composition of material and the effect of variations on properties
- Variation in properties from place to place within a bar of stock
- Effect of processing locally, or nearby, on properties
- Effect of thermomechanical treatment on properties
- Effect of nearby assemblies on stress conditions
- Geometry and how it varies from part to part
- Intensity and distribution in the loading
- Validity of mathematical models used to represent reality
- Intensity of stress concentrations
- Influence of time on strength and geometry
- Effect of corrosion

- Effect of wear

  ⋮

- Length of any list of uncertainties

The algebra of real numbers produces unique single-valued answers in the evaluation of mathematical functions. It is not, by itself, well suited to the representation of behavior in the presence of variation (uncertainty). Engineering's frustrating experience with "minimum values," "minimum guaranteed values," and "safety as the absence of failure" was, in hindsight, to have been expected. Despite these not-quite-right tools, engineers accomplished credible work because any discrepancies between theory and performance were resolved by "asking nature," and nature was taken as the final arbiter. It is paradoxical that one of the great contributions to physical science, namely the search for consistency and reproducibility in nature, grew out of an idea that was only partially valid. Reproducibility in cause, effect, and extent was only approximate, but it was viewed as ideally true. Consequently, searches for invariants were "fruitful."

What is now clear is that consistencies in nature are a stability, not in magnitude, but in the pattern of variation. Evidence gathered by measurement in pursuit of uniqueness of magnitude was really a mix of systematic and random effects. It is the role of statistics to enable us to separate these and, by sensitive use of data, to illuminate the dark places.

2.2 HISTOGRAPHIC EVIDENCE

Each heat of steel is checked for chemical composition to allow its classification as, say, a 1035 steel. Tensile tests are made to measure various properties. When many heats that are classifiable as 1035 are compared by noting the frequency of observed levels of tensile ultimate strength and tensile yield strength, a histogram is obtained as depicted in Fig. 2.1*a* (Ref. [2.1]). For specimens taken from 1- to 9-in bars from 913 heats, observations of mean ultimate and mean yield strength vary. Simply specifying a 1035 steel is akin to letting someone else select the tensile strength randomly from a hat. When one purchases steel from a given heat, the average tensile properties are available to the buyer. The variability of tensile strength from location to location within any one bar is still present.

The loading on a floorpan of a medium-weight passenger car traveling at 20 mi/h (32 km/h) on a cobblestone road, expressed as vertical acceleration component amplitude in g's, is depicted in Fig. 2.1*b*. This information can be translated into load-induced stresses at critical location(s) in the floorpan. This kind of real-world variation can be expressed quantitatively so that decisions can be made to create durable products. Statistical methods permit quantitative descriptions of phenomena which exhibit consistent patterns of variability. As another example, the variability in tensile strength in bolts is shown in the histogram of the ultimate tensile strength of 539 bolts in Fig. 2.2.

The designer has decisions to make. No decisions, no product. Poor decisions, no marketable product. Historically, the following methods have been used which include varying amounts of statistical insight (Ref. [2.2]):

1. Replicate a previously successful design (Roman method).
2. Use a "minimum" strength. This is really a percentile strength often placed at the 1 percent failure level, sometimes called the ASTM minimum.
3. Use permissible (allowable) stress levels based on code or practice. For example, stresses permitted by AISC code for weld metal in fillet welds in shear are 40 percent of the tensile yield strength of the welding rod. The AISC code for structural

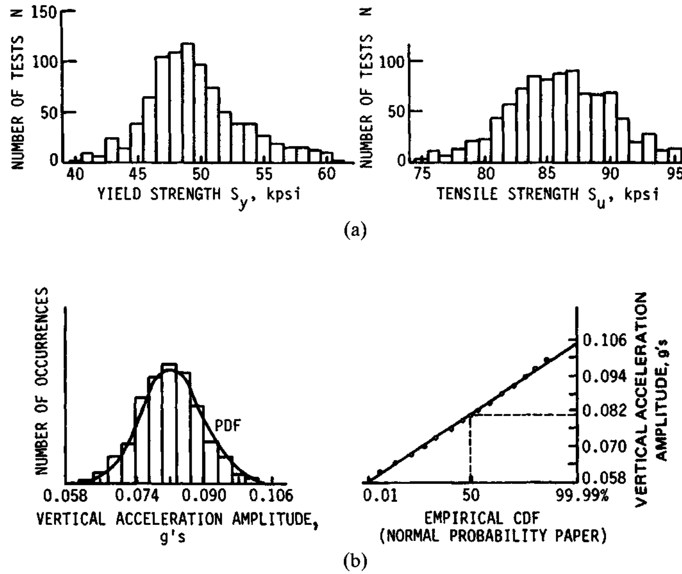


FIGURE 2.1 (*a*) Ultimate tensile strength distribution of hotrolled 1035 steel (1–9 in bars) for 913 heats, 4 mills, 21 classes, $\hat{\mu} = 86.2$ kpsi, $\hat{\sigma} = 3.92$ kpsi, and yield strength distribution for 899 heats, 22 classes, $\hat{\mu} = 49.6$ kpsi, $\hat{\sigma} = 3.81$ kpsi. (*b*) Histogram and empirical cumulative distribution function for loading of floor pan of medium weight passenger car—roadsurface, cobblestones, speed 20 mph (32 km/h).

members has an allowable stress of 90 percent of tensile yield strength in bearing. In bending, a range is offered: $0.45S_y \le \sigma_{all} \le 0.60S_y$.

4. Use an allowable stress based on a design factor founded on experience or the corporate design manual and the situation at hand. For example,

$$\sigma_{all} = S_y/n \tag{2.1}$$

where n is the design factor.

5. Assess the probability of failure by statistical methods and identify the design factor that will realize the reliability goal.

Instructive references discussing methodologies associated with methods 1 through 4 are available. Method 5 will be summarized briefly here.

In Fig. 2.3, histograms of strength and load-induced stress are shown. The stress is characterized by its mean $\bar{\sigma}$ and its upper excursion $\Delta\sigma$. The strength is characterized by its mean $\bar{S}$ and its lower excursion ΔS. The design is safe (no instances of failure will occur) if the stress margin $m = S - \sigma > 0$, or in other words, if $\bar{S} - \Delta S > \bar{\sigma} + \Delta\sigma$, since no instances of strength S are less than any instance of stress σ. Defining the design factor as $n = \bar{S}/\bar{\sigma}$, it follows that

$$n \ge \frac{1 + \Delta\sigma/\bar{\sigma}}{1 - \Delta S/\bar{S}} \tag{2.2}$$

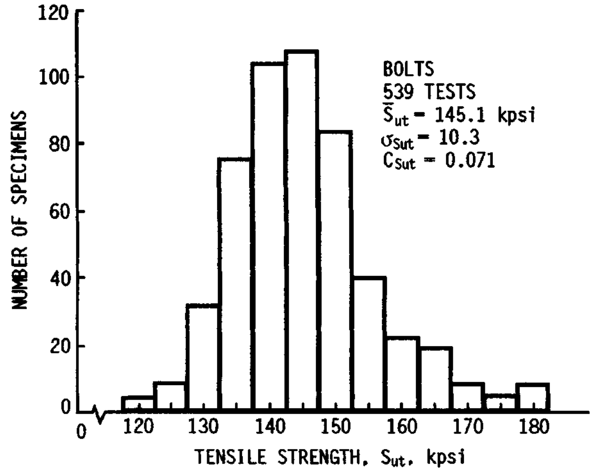


FIGURE 2.2 Histogram of bolt ultimate tensile strength based on 539 tests displaying a mean ultimate tensile strength $\bar{S}_{ut} = 145.1$ kpsi and a standard deviation of $\sigma_{S_{ut}} = 10.3$ kpsi.

As primitive as Eq. (2.2) is, it tells us that we must consider $\bar{S}$, $\bar{\sigma}$, *and* ΔS, $\Delta\sigma$—i.e., not just the means, but the variation as well. As the number of observations increases, Eq. (2.2) does not serve well as it stands, and so designers fit statistical distributions to histograms and estimate the risk of failure from interference of the distributions. Engineers seek to assess the chance of failure in existing designs, or to permit an acceptable risk of failure in contemplated designs.

If the strength is normally distributed, $S \sim N(\mu_S, \sigma_S)$, and the load-induced stress is normally distributed, $\sigma \sim N(\mu_\sigma, \sigma_\sigma)$, as depicted in Fig. 2.4, then the z variable of the standardized normal $N(0, 1)$ can be given by

$$z = -\frac{\mu_S - \mu_\sigma}{(\sigma_S^2 + \sigma_\sigma^2)^{1/2}} \tag{2.3}$$

and the reliability R is given by

$$R = 1 - \Phi(z) \tag{2.4}$$

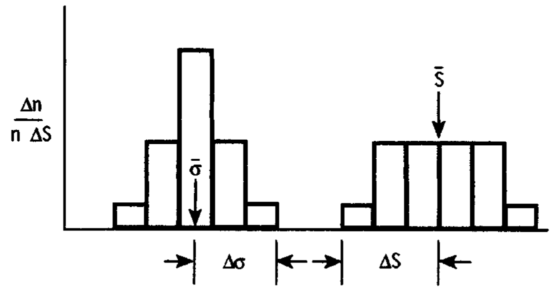


FIGURE 2.3 Histogram of a load-induced stress σ and strength S.

where $\Phi(z)$ is found in Table 2.1. If the strength is lognormally distributed, $\mathbf{S} \sim LN(\mu_S, \sigma_S)$, and the load-induced stress is lognormally distributed, $\boldsymbol{\sigma} \sim LN(\mu_\sigma, \sigma_\sigma)$, then z is given by

$$z = -\frac{\mu_{\ln S} - \mu_{\ln \sigma}}{(\sigma_{\ln S}^2 + \sigma_{\ln \sigma}^2)^{1/2}} = -\frac{\ln\left(\dfrac{\mu_S}{\mu_\sigma}\sqrt{\dfrac{1+C_\sigma^2}{1+C_S^2}}\right)}{\sqrt{\ln\,(1+C_S^2)\,(1+C_\sigma^2)}} \tag{2.5}$$

where $C_S = \sigma_S/\mu_S$ and $C_\sigma = \sigma_\sigma/\mu_\sigma$ are the coefficients of variation of strength and stress. Reliability is given by Eq. (2.4).

Example 1

a. If $\mathbf{S} \sim N(50, 5)$ kpsi and $\boldsymbol{\sigma} \sim N(35, 4)$ kpsi, estimate the reliability R.

b. If $\mathbf{S} \sim LN(50, 5)$ kpsi and $\boldsymbol{\sigma} \sim LN(35, 4)$ kpsi, estimate R.

Solution

a. From Eq. (2.3),

$$z = -\frac{(50-35)}{\sqrt{5^2+4^2}} = -2.34$$

From Eq. (2.4),

$$R = 1 - \Phi(-2.34) = 1 - 0.009\,64 = 0.990$$

b. $C_S = 5/50 = 0.10$, $C_\sigma = 4/35 = 0.114$.

From Eq. (2.5),

$$z = -\frac{\ln\left(\dfrac{50}{35}\sqrt{\dfrac{1+0.114^2}{1+0.100^2}}\right)}{\sqrt{\ln\,(1+0.1^2)\,(1+0.114^2)}} = -2.37$$

and from Eq. (2.4),

$$R = 1 - \Phi(-2.37) = 1 - 0.008\,89 = 0.991$$

It is possible to design to a reliability goal. One can identify a design factor $\bar{n}$ which will correspond to the reliability goal *in the current problem.* A different problem requires a different design factor even for the same reliability goal. If the strength and stress distributions are lognormal, then the design factor $\mathbf{n} = \mathbf{S}/\boldsymbol{\sigma}$ is lognormally distributed, since quotients of lognormal variates are also lognormal. The coefficient of variation of the design factor $\mathbf{n}$ can be approximated for the quotient $\mathbf{S}/\boldsymbol{\sigma}$ as

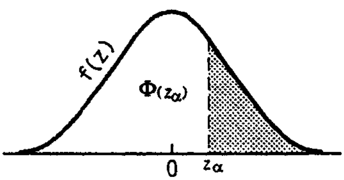
$$C_n = \sqrt{C_S^2 + C_\sigma^2} \tag{2.6}$$

The mean and standard deviation of the companion normal to $\mathbf{n} \sim LN$ are shown in Fig. 2.5 and can be quantitatively expressed as

TABLE 2.1 Cumulative Distribution Function of Normal (Gaussian) Distribution

$$\Phi(z_\alpha) = \int_{-\infty}^{z_\alpha} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

$$= \begin{cases} \alpha & z_\alpha \le 0 \\ 1-\alpha & z_\alpha > 0 \end{cases}$$



z_α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3238	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139

z_α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3	0.00135	0.0^3968	0.0^3687	0.0^3483	0.0^3337	0.0^3233	0.0^3159	0.0^3108	0.0^4723	0.0^4481
4	0.0^4317	0.0^4207	0.0^4133	0.0^5854	0.0^5541	0.0^5340	0.0^5211	0.0^5130	0.0^6793	0.0^6479
5	0.0^6287	0.0^6170	0.0^7996	0.0^7579	0.0^7333	0.0^7190	0.0^7107	0.0^8599	0.0^8332	0.0^8182
6	0.0^9987	0.0^9530	0.0^9282	0.0^9149	$0.0^{10}777$	$0.0^{10}402$	$0.0^{10}206$	$0.0^{10}104$	$0.0^{11}523$	$0.0^{11}260$

z_α	−1.282	−1.645	−1.960	−2.326	−2.576	−3.090	−3.291	−3.891	−4.417
$F(z_\alpha)$	0.10	0.05	0.025	0.010	0.005	0.001	0.0005	0.000 05	0.000 005
$R(z_\alpha)$	0.90	0.95	0.975	0.990	0.995	0.999	0.9995	0.999 95	0.999 995

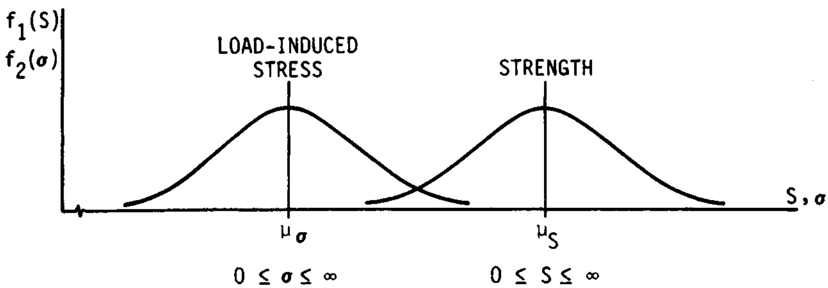


FIGURE 2.4 Probability density functions of load-induced stress and strength.

$$\mu_y = \ln \mu_n - \ln \sqrt{1 + C_n^2}$$
$$\sigma_y = \sqrt{\ln (1 + C_n^2)}$$

The z variable of $z \sim N(0, 1)$ corresponding to the abscissa origin in Fig. 2.5 is

$$z = \frac{y - \mu_y}{\sigma_y} = \frac{0 - \mu_y}{\sigma_y} = \frac{0 - (\ln \mu_n - \ln \sqrt{1 + C_n^2})}{\sqrt{\ln (1 + C_n^2)}}$$

Solving for μ_n, now denoted as $\bar{n}$, gives

$$\mu_n = \bar{n} = \exp\left[-z \sqrt{\ln (1 + C_n^2)} + \ln \sqrt{(1 + C_n^2)}\right] \tag{2.7}$$

Equation (2.7) is useful in that it relates the mean design factor to problem variability through C_n and the reliability goal through z. Note that the design factor $\bar{n}$ is independent of the mean value of **S** or $\boldsymbol{\sigma}$. This makes the geometric decision yet to

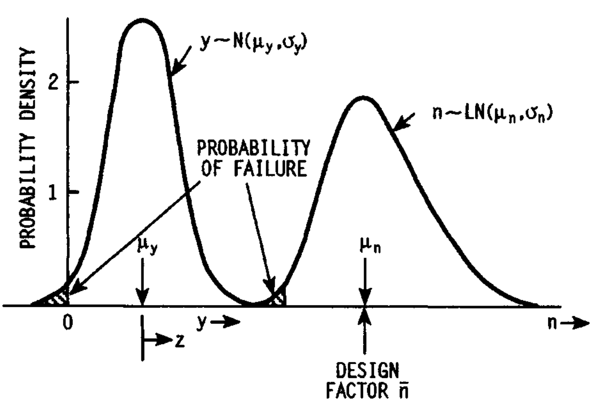


FIGURE 2.5 Lognormally-distributed design factor **n** and its companion normal **y** showing the probability of failure as two equal areas, which are easily quantified from normal probability tables.

be made independent of $\bar{n}$. If the coefficient of variation of the design factor C_n^2 is small compared to unity, then Eq. (2.7) contracts to

$$\bar{n} \doteq \exp\left[C_n(-z + C_n/2)\right] \tag{2.8}$$

Example 2. If $\mathbf{S} \sim LN(50, 5)$ kpsi and $\boldsymbol{\sigma} \sim LN(35, 4)$ kpsi, what design factor $\bar{n}$ corresponds to a reliability goal of 0.990 ($z = -2.33$)?

Solution. $C_S = 5/50 = 0.100$, $C_\sigma = 4/35 = 0.114$. From Eq. (2.6),

$$C_n = (0.100^2 + 0.114^2)^{1/2} = 0.152$$

From Eq. (2.7),

$$\begin{aligned}\bar{n} &= \exp\left[-(-2.33)\sqrt{\ln(1 + 0.152^2)} + \ln\sqrt{(1 + 0.152^2)}\right] \\ &= 1.438\end{aligned}$$

From Eq. (2.8),

$$\bar{n} \doteq \exp\{0.152\,[-(-2.33) + 0.152/2]\} = 1.442$$

The role of the mean design factor $\bar{n}$ is to separate the mean strength $\bar{S}$ and the mean load-induced stress $\bar{\sigma}$ sufficiently to achieve the reliability goal. If the designer in Example 2 was addressing a shear pin that was to fail with a reliability of 0.99, then $z = +2.34$ and $\bar{n} = 0.711$. The nature of C_S is discussed in Chapters 8, 12, 13, and 37.

For normal strength–normal stress interference, the equation for the design factor $\bar{n}$ corresponding to Eq. (2.7) is

$$\bar{n} = \frac{1 \pm \sqrt{1 - (1 - z^2C_S^2)(1 - z^2C_\sigma^2)}}{1 - z^2C_S^2} \tag{2.9}$$

where the algebraic sign + applies to high reliabilities ($R \geq 0.5$) and the – sign applies to low reliabilities ($R < 0.5$).

2.3 USEFUL DISTRIBUTIONS

The body of knowledge called statistics includes many classical distributions, thoroughly explored. They are useful because they came to the attention of the statistical community as a result of a pressing practical problem. A distribution is a particular pattern of variation, and statistics tells us, in simple and useful terms, the many things known about the distribution. When the variation observed in a physical phenomenon is congruent, or nearly so, to a classical distribution, one can infer all the useful things known about the classical distribution. Table 2.2 identifies seven useful distributions and expressions for the probability density function, the expected value (mean), and the variance (standard deviation squared).

TABLE 2.2 Useful Continuous Distributions

Distribution name	Parameters	Probability density function	Expected value	Variance
Uniform	$b > a$	$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Normal	$-\infty < \mu < \infty$ $\sigma > 0$	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$ $-\infty < x < \infty$	μ	σ
Lognormal	$-\infty < \mu < \infty$ $\sigma > 0$	$f(x) = \dfrac{1}{\sigma_y x\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}\left(\dfrac{\ln x - \mu_y}{\sigma_y}\right)^2\right]$ $\mu_y = \ln \mu_x - \ln \sqrt{1 + C_x^2}$ $\sigma_y = \sqrt{\ln (1 + C_x^2)} \quad x \ge 0$	$\exp\left[\mu_y + \dfrac{\sigma_y{}^2}{2}\right]$	$\exp [2 \mu_y + 2\sigma_y^2] - \exp [2 \mu_y + \sigma_y^2]$ or $\exp [2 \mu_y + \sigma_y^2] (\exp [\sigma_y^2] - 1)$
Gamma	$\lambda > 0$ $\eta > 0$	$f(x) = \begin{cases} \dfrac{\lambda^\eta}{\Gamma(\eta)} x^{\eta-1} \exp(-\lambda x) & x \ge 0 \\ 0 & \text{elsewhere} \end{cases}$	$\dfrac{\eta}{\lambda}$	$\dfrac{\eta}{\lambda^2}$
Exponential	$\theta > 0$ $x_0 > 0$	$f(x) = \begin{cases} \dfrac{1}{\theta - x_0} \exp\left[-\left(\dfrac{x - x_0}{\theta - x_0}\right)\right] & x > x_0 \\ 0 & \text{elsewhere} \end{cases}$	θ	$(\theta - x_0)^2$
Rayleigh	$\sigma > 0$	$f(x) = \begin{cases} \dfrac{x}{\sigma^2} \exp\left(\dfrac{-x^2}{2\sigma^2}\right) & x \ge 0 \\ 0 & \text{elsewhere} \end{cases}$	$\dfrac{(\sigma^2\pi)^{1/2}}{\sqrt{2}}$	$0.429\sigma^2$
Weibull	$x_0 > \theta$ $\theta > 0$ $b > 0$	$f(x) = \begin{cases} \dfrac{b}{\theta - x_0}\left(\dfrac{x - x_0}{\theta - x_0}\right)^{b-1} \exp\left[-\left(\dfrac{x - x_0}{\theta - x_0}\right)^b\right] & x > x_0 \\ 0 & \text{elsewhere} \end{cases}$	$x_0 + (\theta - x_0)\,\Gamma\left(1 + \dfrac{1}{b}\right)$	$(\theta - x_0)^2\left[\Gamma\left(1 + \dfrac{2}{b}\right) - \Gamma^2\left(1 + \dfrac{1}{b}\right)\right]$

A frequency histogram may be plotted with the ordinate $\Delta n/(n\ \Delta x)$, where Δn is the class frequency, n is the population, and Δx is the class width. This ordinate is probability density, an estimate of $f(x)$. If the data reduction gives estimates of the distributional parameters, say mean and standard deviation, then a plot of the density function superposed on the histogram will give an indication of fit. Computational techniques are available to assist in the judgment of good or bad fit. The chi-squared goodness-of-fit test is one based on the probability density function superposed on the histogram (Ref. [2.3]).

One might plot the cumulative distribution function (CDF) vs. the variate. The CDF is just the probability (the chance) of a failure at or below a specified value of the variate x. If one has data in this form, or arranges them so, then the CDF for a candidate distribution may be superposed to see if the fit is good or not. The Kolomogorov-Smirnov goodness-of-fit test is available (Ref. [2.3]). If the CDF is plotted against the variate on a coordinate system which rectifies the CDF–x locus, then the straightness of the data string is an indication of the quality of fit. Computationally, the linear regression correlation coefficient r may be used, and the corresponding r test is available (Ref. [2.3]).

Table 2.3 shows the transformations to be applied to the ordinate (variate) and abscissa (CDF, usually denoted F_i) which will rectify the data string for comparison with a suspected parent distribution.

TABLE 2.3 Transformations which Rectify CDF Data Strings

Distribution	Transformation function to data x	Transformation to cumulative distribution function F
Uniform	x	F
Normal	x	$z(F)$
Lognormal	$\ln(x)$	$z(F)$
Weibull	$\ln(x - x_0)$	$\ln \ln [1/(1-F)]$
Exponential	$x - x_0$	$\ln [1/(1-F)]$

Consider a right cylindrical surface generated with an automatic screw machine turning operation. When the machine is set up to produce a diameter at the low end of the tolerance range, each successive part will be slightly larger than the last as a result of tool wear and the attendant increase in tool force due to dulling wear. If the part sequence number is n and the sequence number is n_f when the high end of the tolerance is reached, a is the initial diameter produced, and b is the final diameter produced, one can expect the following relation:

$$x = a + \frac{(b-a)n}{\nu_f} \tag{2.10}$$

However, suppose one measured the diameter every thousandth part and built a data set, smallest diameter to largest diameter (ordered):

n	n_1	n_2	n_3	...
x	x_I	x_2	x_3	...

If the data are plotted with n as abscissa and x as ordinate, one observes a rather straight data string. Consulting Table 2.2, one notes that the linearity of these untransformed coordinates indicates uniform random distribution. A word of caution: If the parts are removed and packed in roughly the order of manufacture, there is no distribution at all! Only if the parts are thoroughly mixed and we draw randomly does a distribution exist. One notes in Eq. (2.10) that the ratio n/n_f is the fraction of parts having a diameter equal to or less than a specified x, and so this ratio is the cumulative distribution function F. Substituting F in Eq. (2.10) and solving for F yields

$$F(x) = \frac{x - a}{b - a} \qquad a \le x \le b \tag{2.11}$$

From Table 2.2, take the probability density function for uniform random distribution, $f(x) = 1/(b - a)$, and integrate from a to x to obtain Eq. (2.11).

Engineers often have to identify a distribution from a small amount of data. Data transformations which rectify the data string are useful in recognizing a distribution. First, place the data in a column vector, order smallest to largest. Second, assign corresponding cumulative distribution function values F_i using median rank $(i - 0.3)/(n + 0.4)$ if seeking a median locus, or $i/(n + 1)$ if seeking a mean locus (Ref. [2-4]). Third, apply transformations from Table 2.3 and look for straightness.

Normal distributions are used for many approximations. The most likely parent of a data set is the normal distribution; however, that does not make it common. When a pair of dice is rolled, the most likely sum of the top faces is 7, which occurs in 1/6 of the outcomes, but 5/6 of the outcomes are other than 7.

Properties of materials—ultimate tensile strength, for example—can have only positive values, and so the normal cannot be the true distribution. However, a normal fit may be robust and therefore useful. The lognormal does not admit variate values which are negative, which is more in keeping with reality. Histographic data of the ultimate tensile strength of a 1020 steel with class intervals of 1 kpsi are as follows:

Class frequency f_i	2	18	23	31	83	109	138	151
Class midpoint x_i	56.5	57.5	58.5	59.5	60.5	61.5	62.5	63.5

Class frequency f_i	139	130	82	49	28	11	4	2
Class midpoint x_i	64.5	65.5	66.5	67.5	68.5	69.5	70.5	71.5

Now $\Sigma x_i f_i = 63\,625$ and $\Sigma x_i^2 f_i = 4\,054\,864$, and so $\bar{x}$ and $\hat{\sigma}$ are $\bar{x} = \Sigma x_i f_i/n = 63\,625/1000 = 63.625$ kpsi, and

$$\hat{\sigma} = \sqrt{\frac{\Sigma x_i^2 f_i - (\Sigma x_i f_i)^2/n}{n \quad - \quad 1}}$$

$$\sigma = \sqrt{\frac{4\,054\,864 - (63\,625)^2/1000}{(1000 - 1)}} = 2.5942 \text{ kpsi}$$

From Table 2.2, the mean and standard deviation of the companion normal to a lognormal are (Ref. [2-2])

$$\mu_y = \ln \bar{x} - \ln \sqrt{1 + C_x^2} = \ln 63.625 - \ln \sqrt{1 + 0.040\,773^2}$$
$$= 4.1522$$
$$\sigma_y = \sqrt{\ln (1 + C_x^2)} = \sqrt{\ln (1 + 0.040\,773^2)} = 0.0408$$

The lognormal probability density function of x is

$$g(x) = \frac{1}{x\sigma_y \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu_y}{\sigma_y}\right)^2\right]$$

$$= \frac{1}{0.0408x \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - 4.1522}{0.0408}\right)^2\right]$$

A plot of the histogram and the density is shown in Fig. 2.6. A chi-squared goodness-of-fit test on a modified histogram (compacted somewhat to have 5 or more in each class) cannot reject the null hypothesis of lognormality at the 0.95 confidence level.

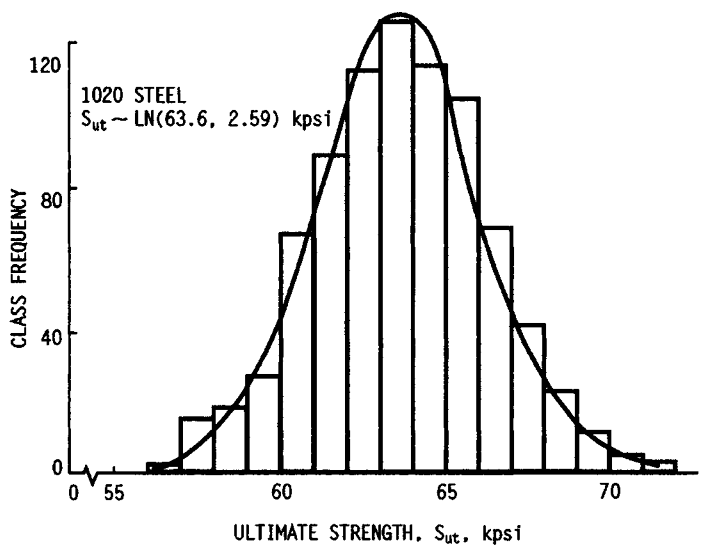


FIGURE 2.6 Histographic report of the results of 1000 ultimate tensile strength tests on a 1020 steel.

2.4 RANDOM-VARIABLE ALGEBRA

Engineering parameters which exhibit variation can be represented by random variables and characterized by distribution parameters and a distribution function. Many distributions have two parameters; the mean and standard deviation (variance) are preferred. It is common to display statistical parameters by roster between curved parentheses as (μ, σ). If the normal distribution is to be indicated, then an N is placed before the parentheses as $N(\mu, \sigma)$; this indicates a normal distribution with a mean of μ and a standard deviation of σ. Similarly, $LN(\mu, \sigma)$ is a lognormal distribution and $U(\mu, \sigma)$ is a uniform distribution. To distinguish a real-number variable w from a random variable **y**, boldface is used. Thus $\mathbf{z} = \mathbf{x} + \mathbf{y}$ displays **z** as the sum of random variables **x** and **y**. With knowledge of **x** and **y**, of interest are the parameters and distribution of **z** (Ref. [2.6]).

For distributional information, various closure theorems and the central limit theorem of statistics are useful. The sums of normal variates are themselves normal. The quotients and products of lognormals are lognormal. Real powers of a lognormal variate are likewise lognormal. Sums of variates from any distribution tend asymptotically to approach normal. Products of variates from any distribution tend asymptotically to lognormal. In some cases a computer simulation is necessary to discover distributions resulting from an algebraic combination of variates. The mean and standard deviation of a function $\phi(x_1, x_2, \ldots, x_n)$ can be estimated by the following rapidly convergent Taylor series of expected values for unskewed (or lightly skewed) distributions (Ref. [2.7, Appendix C]):

$$\mu_\phi = \phi(x_1, x_2, \ldots, x_n)_\mu + \frac{1}{2}\sum_{i=1}^{n} \left.\frac{\partial^2 \phi}{\partial x_i^2}\right|_\mu \sigma_{x_i}^2 + \cdots \tag{2.12}$$

$$\sigma_\phi = \left\{ \sum_{i=1}^{n} \left(\frac{\partial \phi}{\partial x_i}\right)_\mu^2 \sigma_{x_i}^2 + \frac{1}{2}\sum_{i=1}^{n} \left(\frac{\partial^2 \phi}{\partial x_i^2}\right)_\mu^2 \sigma_{x_i}^4 + \cdots \right\}^{1/2} \tag{2.13}$$

Equations (2.12) and (2.13) for simple functions can be used to form Table 2.4 to display the dominant first terms of the series. More expanded information, including correlation, can be found in Refs. [2.3] and [2.7].

Equations (2.12) and (2.13) can be used to propagate the means and standard deviations through functions. The various closure theorems of statistics, or computer simulation, can be used to find robust distributional information.

TABLE 2.4 Means, Standard Deviations, and Coefficients of Variation of Simple Operations with Independent (Uncorrelated) Random Variables*

Function	Mean value μ	Standard deviation σ	Coefficient of variation C
a	a	0	0
$\mathbf{x}$	μ_x	σ_x	σ_x/μ_x
$\mathbf{x}+a$	$\mu_x + a$	σ_x	σ_x/μ_x
$a\mathbf{x}$	$a\mu_x$	$a\sigma_x$	σ_x/μ_x
$\mathbf{x}+\mathbf{y}$	$\mu_x + \mu_y$	$(\sigma_x^2 + \sigma_y^2)^{1/2}$	σ_{x+y}/μ_{x+y}
$\mathbf{x}-\mathbf{y}$	$\mu_x - \mu_y$	$(\sigma_x^2 + \sigma_y^2)^{1/2}$	σ_{x-y}/μ_{x-y}
$\mathbf{xy}$	$\mu_x\mu_y$	$C_{xy}\mu_{xy}$	$(C_x^2 + C_y^2)^{1/2}$
$\mathbf{x}/\mathbf{y}$	μ_x/μ_y	$C_{x/y}\mu_{x/y}$	$(C_x^2 + C_y^2)^{1/2}$
$1/\mathbf{x}$	$1/\mu_x$	C_x/μ_x	C_x
$\mathbf{x}^2$	μ_x^2	$2C_x\mu_x^2$	$2C_x$
$\mathbf{x}^3$	μ_x^3	$3C_x\mu_x^3$	$3C_x$
$\mathbf{x}^4$	μ_x^4	$4C_x\mu_x^4$	$4C_x$

* Tabulated quantities are obtained by the partial derivative propagation method, some results of which are approximate. For a more complete listing including the first two terms of the Taylor series, see Charles R. Mischke, *Mathematical Model Building,* 2d rev. ed., Iowa State University Press, Ames, 1980, appendix C.

The first terms of Eqs. (2.12) and (2.13) are often sufficient as a first-order estimate; thus

$$\mu_\phi = \phi(\mu_{x_1}, \mu_{x_2}, \ldots, \mu_{x_n}) \tag{2.14}$$

$$\sigma_\phi = \left\{ \sum_{i=1}^{n} \left(\frac{\partial \phi}{\partial x_i} \right)_\mu^2 \sigma_{x_i}^2 \right\}^{1/2} \tag{2.15}$$

and if ϕ is of the form $\alpha\, x_1{}^a\, x_2{}^b\, x_3{}^c \ldots$, then C_ϕ is given by

$$C_\phi = \left(a^2 C_{x_1}^2 + b^2 C_{x_2}^2 + \cdots \right)^{1/2} \tag{2.16}$$

Equations (2.14), (2.15), and (2.16) are associated with the partial derivative estimation method. These equations are very important in what they suggest in general about engineering computations in stochastic situations. The estimate of the mean in a functional relationship comes from substituting mean values of the variates. This suggests that deterministic and familiar engineering computations are still useful in stochastic problems if mean values are used. Calculations such as the quotient of *minimum* strength divided by *maximum* load-induced stress are not appropriate when chance of failure is being considered.

Equation (2.15) says that the variance of ϕ is simply the sum of the weighted variances of the parameters, with the weighting factors depending on the functional relationship involved. In terms of the standard deviation, it is a weighted Pythagorean combination.

The good news is that engineering's previous deterministic experience is useful in stochastic problems provided one uses mean values. The bad news is that there is additional effort associated with propagating the variation through the same relationships and identifying the resulting distributions. The other element of bad news is that Eqs. (2.14) and (2.15) are approximations, but the corresponding good news is that they are robust approximations. In summary,

1. A random variable or function of random variables can be characterized by statistical parameters, often the mean and variance, and a distribution function, whether assumed or goodness-of-fit tested.
2. Ordinary deterministic algebra using means of variates is useful in estimating means and standard deviations of functions of variates.
3. The distribution of a function of random variables can often be determined from closure theorems.
4. Computer simulation techniques can address cases not covered (see Chap. 5).

Example 3. If 12 random selections are made from the uniform random distribution $U[0, 1]$ and the real number 6 is subtracted from the sum of the 12, what are the mean, the standard deviation, and the distribution of the result?

Solution. Note the square brackets in $U[0, 1]$. These denote parameters other than the mean and standard deviation, in this case range numbers a and b—i.e., there are no observations less than a nor more than b. The sum ϕ is defined by

$$\phi = x_1 + x_2 + \cdots + x_{12} - 6$$

From Table 2.2,

$$\mu_x = (a + b)/2 = (0 + 1)/2 = 1/2$$

$$\sigma_x^2 = (b - a)^2/12 = (1 - 0)^2/12 = 1/12$$

From Table 2.4, the mean is the sum of the means:

$$\bar{\phi} = \bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_{12} - 6 = 1/2 + 1/2 + \cdots + 1/2 - 6 = 0$$

From Table 2.4, the standard deviation of the sum of independent random variables is the square root of the sum of the variances:

$$\sigma_\phi = (1/12 + 1/12 + \cdots + 1/12 + 0)^{1/2} = 1$$

From the central limit theorem, the sum of random variables asymptotically approaches normality. The sum of 12 variates cannot be rejected using a null hypothesis of normality. Thus, $\phi \sim N(\mu_\phi, \sigma_\phi) = N(0, 1)$. Computing machinery manufacturers supply a machine-specific pseudo-random number generator $U[0, 1]$. The reason the program is supplied is the machine specificity involved. Such a program is the building block from which other random numbers can be generated with software.

Example 3 is the basis for a Fortran subroutine to generate pseudo-random numbers from a normal distribution N(xbar, sigmax). If RANDU is the subprogram name of the uniform random number generator in the interval [0, 1], and IX and IY are seed integers, then

```
      SUBROUTINE GAUSS(IX,IY,XBAR,SIGMAX,X)
      SUM=0.
      DO 100 I=1,12
      CALL RANDU(IX,IY,U)
      SUM=SUM+U
  100 CONTINUE
      X=XBAR+(SUM-6.)*SIGMAX
      RETURN
      END
```

2.5 STOCHASTIC ENDURANCE LIMIT BY CORRELATION AND BY TEST

Designers need rational approaches to meet a variety of situations. A product can be produced in such large quantities (or be so dangerous) that elaborate testing of materials, components, and prototypes is justified. Smaller quantities can be produced and the product can be of modest value, so that less comprehensive testing of materials—perhaps only ultimate tensile strength testing—is economically justified. Or so few items can be produced that no testing of materials is done at all.

For an R. R. Moore rotating beam bending endurance test, approximately 60 specimens in a staircase test matrix method of testing are employed to find the endurance limit of a steel. Considerable time and expense is involved, using a standard specimen and a procedure that will remove the effects of surface finish, size, loading, temperature, stress concentration, et al. Since such testing is not always possible, engineers with an interest in bending, axial (push-pull), and torsional fatigue use correlations of endurance limit to mean tensile strength as a first-order estimate as follows:

$$\mathbf{S}'_e = \phi_b \bar{S}_{ut} \tag{2.17}$$

$$\mathbf{S}'_{ax} = \phi_{ax} \bar{S}_{ut} \tag{2.18}$$

$$\mathbf{S}'_{se} = \phi_t \bar{S}_{ut} \tag{2.19}$$

where ϕ_b, ϕ_{ax}, and ϕ_t are called *fatigue ratios.* Data reported by Gough are shown in Fig. 2.7. It is clear that the bending fatigue ratio ϕ_b is not constant in a class of materials and varies widely; that is to say, it is a random variable. The mean of ϕ is called the fatigue ratio, and in bending in steel it is about 0.5, which is conservative about half the time. Table 2.5 shows the mean and standard deviation of ϕ_b for classes of materials. From 133 full-scale R. R. Moore tests on steels, ϕ_b is found to be lognormally distributed.

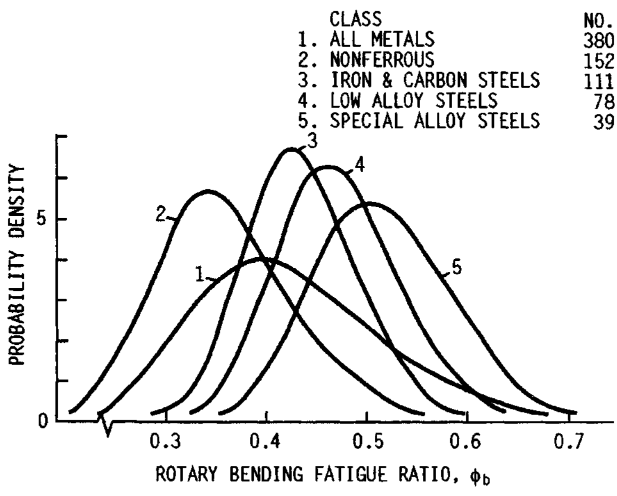


FIGURE 2.7 Probability density functions of fatigue ratio ϕ_b reported by Gough for five classifications of metals.

TABLE 2.5 Stochastic Parameters of Fatigue Ratio ϕ*

Class of metals	Number of tests	$\hat{\mu}_\phi$	$\hat{\sigma}_\phi$
All metals	380	0.44	0.10
Nonferrous	152	0.37	0.075
Irons and carbon steels	111	0.44	0.060
Low-alloy steels	78	0.475	0.063
Special-alloy steels	39	0.52	0.070

* Data from Gough reported in J. A. Pope, *Metal Fatigue,* Chapman and Hall, London, 1959 and tabulated in C. R. Mischke, "Prediction of Stochastic Endurance Strength," *Transactions of the American Society of Mechanical Engineers, Journal of Vibration, Acoustics, Stress and Reliability in Design,* vol. 109, no. 1, Jan. 1987, pp. 113–122.

$$\boldsymbol{\phi}_b = 0.445d^{-0.107}(1, 0.138) \tag{2.20}$$

When the standard specimen diameter of 0.30 in is substituted in Eq. (2.20), one obtains $\boldsymbol{\phi}_{0.30} = 0.506(1, 0.138)$, which is still lognormally distributed. Multiplying the 0.506 by the mean and standard deviation, one can write $\boldsymbol{\phi}_{0.30} \sim LN(0.506, 0.070)$. The coefficient of variation is 0.138. Table 2.6 shows approximate mean values of $\boldsymbol{\phi}_b$ for several material classes.

TABLE 2.6 Typical Mean Fatigue Ratios for Several Material Classes

Material class	$\bar{\phi}_{0.30}$
Wrought steel	0.50
Cast steel	0.40
Gray cast iron	0.35
Nodular cast iron	0.40
Normalized nodular cast iron	0.33

Example 4. The results of an ultimate tensile test on a heat-treated 4340 steel (382 Brinell) consisting of 10 specimens gave an estimate of the ultimate tensile strength of $\mathbf{S}_{ut} \sim LN(190, 6.0)$ kpsi. Estimate the mean, standard deviation, and 99th-percentile bending endurance limit for (*a*) the case of no further testing and (*b*) an additional R. R. Moore test resulting in $\mathbf{S}'_e \sim LN(90, 5.3)$ kpsi.

Solution. *a.* The expected fatigue strength is, from Eqs. (2.17) and (2.20),

$$\begin{aligned}\mathbf{S}'_e &= \boldsymbol{\phi}_b \bar{S}_{ut} = 0.445(0.30)^{-0.107}(1, 0.138)190 \\ &= 0.506(1, 0.138)190 \text{ kpsi}\end{aligned}$$

The estimated mean of the endurance limit $\bar{S}'_e$ is given by

$$\bar{S}'_e = 0.506(1)(190) = 96.1 \text{ kpsi}$$

The standard deviation $\sigma_{S'_e}$ is

$$\sigma_{S'_e} = 0.506(0.138)(190) = 13.3 \text{ kpsi}$$

The coefficient of variation is $C_{S'_e} = 13.3/96.1 = 0.138$, as expected. The distribution of $\mathbf{S}'_e$ is lognormal because $\boldsymbol{\phi}_b$ is lognormal. The 99th-percentile endurance limit is found from the companion normal to the endurance limit distribution as follows:

$$\begin{aligned}\mu_y &= \ln \bar{S}'_e - \ln \sqrt{1 + C^2_{S'_e}} = \ln 96.1 - \ln \sqrt{1 + 0.138^2} \\ &= 4.556\end{aligned}$$

$$\sigma_y = \sqrt{\ln (1 + C^2_{S'_e})} = \sqrt{\ln (1 + 0.138^2)} = 0.137$$

Now

$$_{0.99}y = \mu_y - {}_{0.99}z\sigma_y = 4.556 - 2.33(0.137) = 4.237$$

and ${}_{0.99}S'_e$ is given by

$${}_{0.99}S'_e = \exp\,({}_{0.99}y) = \exp\,(4.237) = 69.2 \text{ kpsi}$$

without fatigue testing from the history of the 133 steel materials ensemble embodied in ϕ_b. One can expect 99 percent of the instances of endurance limit to exceed 69.2 kpsi given that the mean tensile strength is 190 kpsi.

b. The results of R. R. Moore testing of the 4340 gave $\mu_{S'_e} = 90$ kpsi and $\sigma_{S'_e} = 5.3$ kpsi. The coefficient of variation is 5.3/90, or 0.059. The 99th-percentile endurance limit is found from the companion normal as follows:

$$\mu_y = \ln 90 - \ln \sqrt{1 + 0.059^2} = 4.498$$

$$\sigma_y = \sqrt{\ln\,(1 + 0.059^2)} = 0.059$$

$${}_{0.99}y = 4.498 - 2.33(0.059) = 4.361$$

$${}_{0.99}S'_e = \exp\,(4.361) = 78.3 \text{ kpsi}$$

It is instructive to plot the density functions. The lognormal density function for part *a* is

$$g_1(S) = \frac{1}{0.137S\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln S - 4.556}{0.137}\right)^2\right]$$

and that for part *b* is

$$g_2(S) = \frac{1}{0.059S\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln S - 4.498}{0.059}\right)^2\right]$$

Figure 2.8 graphically depicts the two and one-half times dispersion resulting from use of the correlation rather than R. R. Moore testing. Testing is costly in money and time. It costs money to reduce dispersion, and one is never without dispersion. However, in designing to a reliability goal, dispersion in strength, loading, and geometry increases the size of parts. Using part *a* strength information results in a larger part than using part *b* information.

2.6 INTERFERENCE

In Eqs. (2.5) and (2.9), one has a way of relating geometric decisions to a reliability goal. The fundamental tactic is to separate the mean strength from the mean stress sufficiently to achieve the reliability goal through geometric decisions. The equation $\mathbf{n} = \mathbf{S}/\boldsymbol{\sigma}$ can be generalized. The denominator is some threatening stimulus which is resisted by some response which has a limited potential (the numerator). Defining the design factor as the quotient of the response potential divided by the stimulus is more general and useful. The stimulus might be a distortion and the response potential the deflection which compromises function. The tools discussed so far have broader application.

Interference of normal-normal and lognormal-lognormal distributions has been presented. There is need for a general method for interference of other distribution

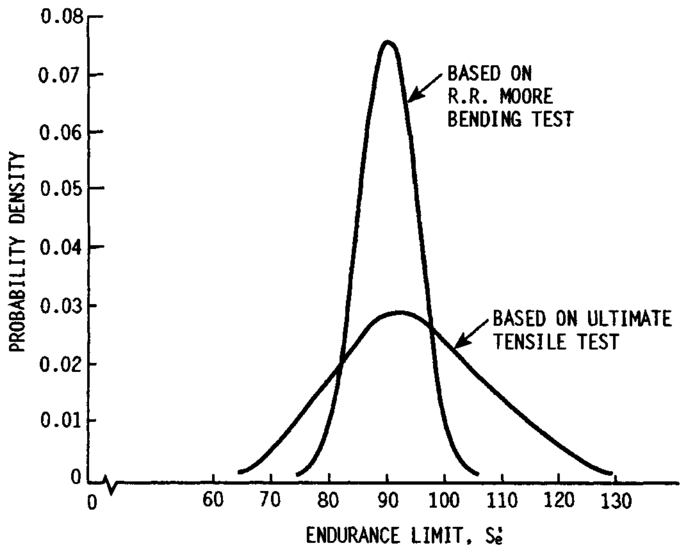


FIGURE 2.8 Probability density functions of rotary bending endurance limit based on historical knowledge of an ensemble of 133 steels, plus tensile testing on a 4340 steel, and based on R. R. Moore endurance limit testing on 4340.

combinations. In Fig. 2.9*a* the probability density of the response potential **S** is $f_1(S)$, and in Fig. 2.9*b* the density function of the stimulus $\boldsymbol{\sigma}$ is $f_2(\sigma)$. The probability that the strength exceeds a stress level x is $dP(S > x)$, which is the differential reliability dR, or

$$dR = R_1(x)\, dF_2(x) = -R_1(x)\, dR_2(x)$$

which integrates to

$$R = -\int_{x=-\infty}^{x=\infty} R_1(x)\, dR_2(x) = -\int_{R_2=1}^{R_2=0} R_1(x)\, dR_2 = \int_0^1 R_1\, dR_2 \tag{2.21}$$

where

$$R_1(x) = \int_x^{\infty} f_1(S)\, dS \qquad \text{and} \qquad R_2(x) = \int_x^{\infty} f_2(\sigma)\, d\sigma$$

which is given geometric interpretation in Fig. 2.9*c*.

An alternative view is that the probability that the stress is less than the strength is expressible as $dP(\sigma < x)$, which is the differential reliability dR, or, from Fig. 2.9*d* and *e*,

$$dR = F_2(x)\, dF_1(x) = -[1 - R_2(x)]\, dR_1(x)$$

which integrates to

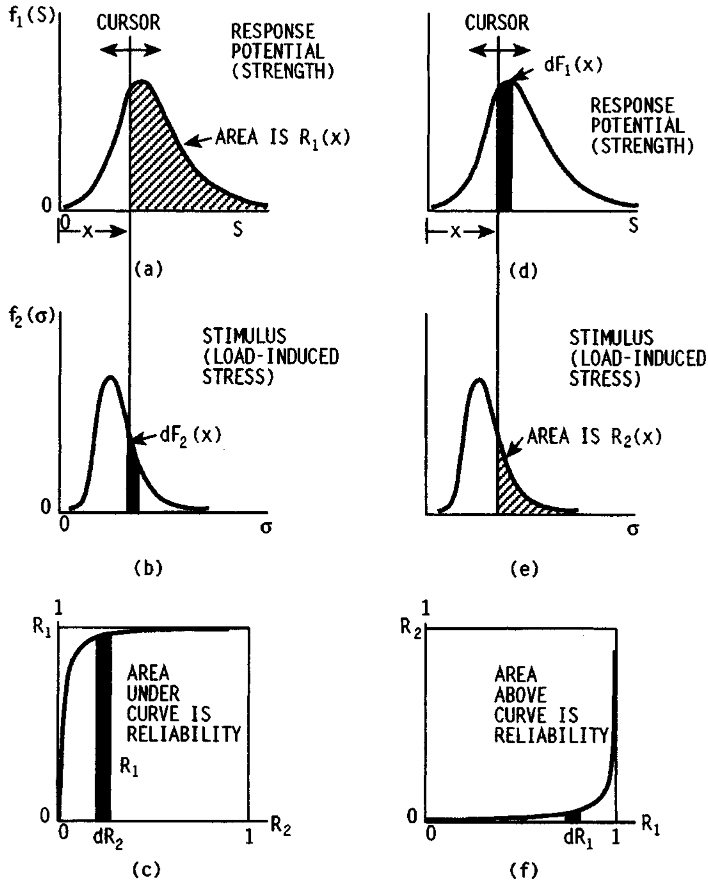


FIGURE 2.9 (*a*), (*b*), and (*c*) Development of the general reliability equation $\int_0^1 R_1 dR_2$ by interference; (*d*), (*e*), and (*f*) development of general reliability equation $1 - \int_0^1 R_2 dR_1$ by interference.

$$R = -\int_{x=-\infty}^{x=\infty} [1 - R_2(x)]\, dR_1(x) = -\int_{R_1=1}^{R_1=0} (1 - R_2)\, dR_1$$

$$= -\int_1^0 dR_1 + \int_1^0 R_2\, dR_1 = 1 - \int_0^1 R_2\, dR_1 \qquad (2.22)$$

where $R_1(x)$ and $R_2(x)$ have the definitions above. Equation (2.22) is given geometric interpretation in Fig. 2.9*f*. When dealing with distributions with lower bounds, such as Weibull, Eq. (2.22) is easier to integrate than Eq. (2.21).

The following example is couched in terms of geometrically simple distributions to avoid obscuring the ideas.

Example 5. If strength is distributed uniformly, $\mathbf{S} \sim U[60, 70]$ kpsi, and stress is distributed uniformly, $\boldsymbol{\sigma} \sim U[58, 63]$ kpsi, find the reliability (*a*) using Eq. (2.22), (*b*) using the geometry of Fig. 2.9*f*, (*c*) using numerical integration based on Fig. 2.9*f*, and (*d*) generalizing part *a* for $\mathbf{S} \sim U[A, B]$ and $\boldsymbol{\sigma} \sim U[a, b]$ for one-tailed overlap.

Solution. *a.* Define R_1 as a function of the cursor position x:

$$R_1 = \begin{cases} 1 & x < 60 \\ (70 - x)/10 & 60 \le x \le 70 \\ 0 & x > 70 \end{cases}$$

Define R_2 as a function of the cursor position x:

$$R_2 = \begin{cases} 1 & x < 58 \\ (63 - x)/5 & 58 \le x \le 63 \\ 0 & x > 63 \end{cases}$$

From Eq. (2.22),

$$R = 1 - \int_0^1 R_2\, dR_1 = 1 - \int_{R_1 = 0}^{R_1 = 1} R_2(x)\, dR_1(x) = 1 - \int_{x = 70}^{x = 60} R_2(x)\, dR_1(x)$$

$$= 1 - \int_{x = 63}^{x = 60} R_2(x)\, dR_1(x) = 1 - \int_{63}^{60} \frac{63 - x}{5} \frac{dx}{10} = 0.91$$

b. Geometrically, the area of the triangle in Fig. 2.10 is 0.6(1 − 0.7)/2, which equals 0.09, and the ones complement is the reliability $R = 1 - 0.09 = 0.91$.

c. Examination of Fig. 2.9*f* shows that the largest contribution to the area under the curve is near $R_1 = 1$; consequently, the tabular method will begin with $R_1 = 1$ at the top of the table. Table 2.7 lists values of R_1 beginning with unity and decreasing in steps of 0.05 ($h = 0.05$ in Simpson's method). Column 2 contains the values of the cursor location x corresponding to R_1. This is obtained by solving the expression R_1 for x, namely $x = 70 - 10R_1$. Column 3 consists of the values of R_2 corresponding to the cursor location x, namely $R_2 = (63 - x)/5$. The ordinates to the curve are in the R_2 column, and values other than zero contribute to the area. At $R_1 = 0.70$, the area contributions cease. The Simpson's rule multipliers m are in column 4. The sum, ΣmR_2, is 5.4. The area under the curve is

$$A = (h/3)\, \Sigma\, mR_2 = (0.05/3)(5.4) = 0.09$$

and the reliability is

$$R = 1 - A = 1 - 0.09 = 0.91$$

d. The survival function R_1 is given by

$$R_1 = \begin{cases} 1 & x < A \\ (B - x)/(B - A) & A \le x \le B \\ 0 & x > B \end{cases}$$

and the survival function R_2 is given by

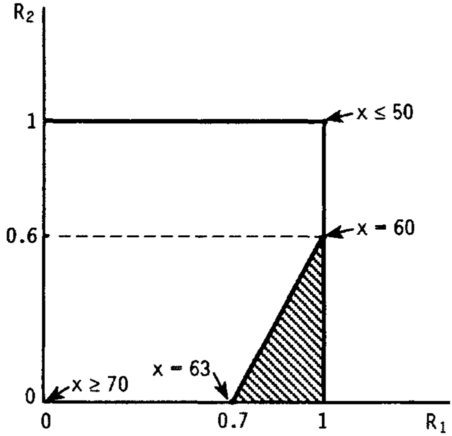


FIGURE 2.10 Assessment of reliability in Example 2.6(b) by geometric interpretation of area.

$$R_2 = \begin{cases} 1 & x < a \\ (b-x)/(b-a) & a \le x \le b \\ 0 & x > b \end{cases}$$

For one-tailed overlap, from Eq. (2.22),

$$R = 1 - \int_0^1 R_2 \, dR_1 = 1 - \int_{R_1=0}^{R_1=1} R_2(x)\, dR_1(x) = 1 - \int_{x=B}^{x=A} R_2(x)\, dR_1(x)$$

Noting that $R_2(x)$ is zero when $x < b$ allows the lower limit to be changed to b.

$$R = 1 - \int_{x=b}^{x=A} R_2(x)\, dR_1(x) = 1 - \int_b^A \frac{b-x}{b-a}\frac{dx}{B-A}$$

$$= 1 - \frac{1}{(b-a)(B-A)} \int_A^b (b-x)\, dx = 1 - \frac{(b-A)^2}{2(b-a)(B-A)}$$

TABLE 2.7 Reliability by Simpson's Rule Interference

R_1	x	R_2	Multiplier m	mR_2
1.00	60.0	0.6	1	0.6
0.95	60.5	0.5	4	2.0
0.90	61.0	0.4	2	0.8
0.85	61.5	0.3	4	1.2
0.80	62.0	0.2	2	0.4
0.75	62.5	0.1	4	0.4
0.70	63.0	0.0	1	0.0
				$\Sigma mR_2 = 5.4$

Note that the reliability declines from unity as the square of the overlap $(b - A)$. For $a = 58$ kpsi, $b = 63$ kpsi, $A = 60$ kpsi, and $B = 70$ kpsi,

$$R = 1 - \frac{(63 - 60)^2}{2(63 - 58)(70 - 60)} = 0.91$$

and when distributions touch, $b = A$ and $R = 1$.

More complicated functions yield to tabular procedures along the lines of Example 5c. Computer programs can be written to carry out tedious work.

A very useful three-parameter distribution is the Weibull, which is expressed in terms of the parameters, the lower bound x_0, the characteristic parameter θ, and the shape parameter b, displayed as $\mathbf{x} \sim W[x_0, \theta, b]$. The mean and standard deviation are found from the parameters as

$$\mu_x = x_0 + (\theta - x_0)\,\Gamma\,(1 + 1/b)$$

$$\sigma_x = (\theta - x_0)[\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)]^{1/2}$$

The Weibull has the advantage of being a closed-form survival function.

$$R = \exp\,\{-[(x - x_0)/(\theta - x_0)]^b\}$$

For interference of a Weibull strength $\mathbf{S} \sim W[x_{01}, \theta_1, b_1]$ with a Weibull stress $\boldsymbol{\sigma} \sim W[x_{02}, \theta_2, b_2]$, use a numerical evaluation of the integral in Eq. (2.22). Write the strength distribution survival equation in terms of the cursor location x as

$$R_1 = \exp\,\{-\,[(x - x_{01})/(\theta_1 - x_{01})]^{b_1}\}$$

and solve for x, which results in

$$x = x_{01} + (\theta_1 - x_{01})[\ln\,(1/R_1)]^{1/b_1}$$

Noting that the survival equation for the stress distribution in terms of the cursor location x is

$$R_2 = \exp\,\{-[(x - x_{02})/(\theta_2 - x_{02})]^{b_2}\}$$

one forms a table such as Table 2.8 to integrate the integral portion of Eq. (2.22). If $\mathbf{S} \sim W[40, 50, 3.3]$ kpsi and $\boldsymbol{\sigma} \sim W[30, 40, 2]$ kpsi, then Table 2.8 follows. The sum ΣmR_2 is 1.443 413, making the area under the R_1R_2 curve by Simpson's rule

$$A = (h/3)\Sigma mR_2 = (0.1/3)(1.433\,413) = 0.048\,114$$

and

$$R = 1 - A = 1 - 0.048\,114 = 0.952$$

TABLE 2.8 Weibull-Weibull Interference by Simpson's Rule, $\mathbf{S} \sim W[40, 50, 3.3]$ kpsi, $\boldsymbol{\sigma} \sim W[30, 40, 2]$ kpsi

R_1	x	R_2	Multiplier m	mR_2
1.0	40.000 000	0.367 879	1	0.367 879
0.9	45.056 404	0.103 627	4	0.414 508
0.8	46.347 480	0.069 086	2	0.138 172
0.7	47.316 865	0.049 850	4	0.199 400
0.6	48.158 264	0.036 986	2	0.073 972
0.5	48.948 810	0.027 582	4	0.110 328
0.4	49.738 564	0.020 321	2	0.040 642
0.3	50.578 627	0.014 483	4	0.057 932
0.2	51.551 239	0.009 614	2	0.019 228
0.1	52.875 447	0.005 338	4	0.021 352
0.0		0	1	0
				$\Sigma mR_2 = 1.443\,413$

The means of the strength **S** and the stress $\boldsymbol{\sigma}$ are

$$\bar{S} = 40 + (50 - 40)\,\Gamma(1 + 1/3.3) = 40 + (50 - 40)(0.8970) = 48.97 \text{ kpsi}$$

$$\bar{\sigma} = 30 + (40 - 30)\,\Gamma(1 + 1/2) = 30 + (40 - 30)(0.8862) = 38.86 \text{ kpsi}$$

The design factor associated with a reliability of 0.952 is $\bar{n} = 48.57/38.86 = 1.25$. Since the distribution of the design factor as a quotient of two Weibull variates is not known, discovering the design factor corresponding to a reliability goal of (say) 0.999 becomes an iterative process, with the previous tabular integration becoming part of a root-finding process, quite tractable using a computer.

The strength distribution reflects the result of data reduction and distributional description found to be robust. Strength distributions from historical ensembles, particularly in fatigue, tend to be lognormal. Stress distributions reflect loading and geometry. Machine parts often exhibit geometries with coefficients of variation that are very small compared with that of the load. Additional useful information is to be found in the technical content of more specialized chapters and in the literature.

2.7 NUMBERS

Engineering calculations are a blend of

- Mathematical constants, such as π or e
- Toleranced dimensions
- Measurement numbers
- Mathematical functions (themselves approximate)
- Unit conversion constants
- Mechanically generated digits from calculators and computers
- Rule-of-thumb numbers

A mixture of all types of numbers can be present. It is prudent to treat all numbers as incomplete numbers, avoid serious loss of precision by using sufficient computational digits, round for brevity, and make no significant number inferences. It is well to review the kinds of numbers.

The set of all *integers* is the set of the counting numbers 1, 2, 3, . . . , augmented by negative integers and zero. The set of all *rational numbers m/n* is constructed from the integers (dividing by zero excepted). The set of all *real numbers* is constructed by adding limits of all bounded monotone increasing sequences (*irrational numbers*) to the set of rational numbers. Each set of numbers contains the previous set. Each point on a *number line* corresponds to a real number. The display of a real number often has the problem of economy of notation. If the true number to be expressed is $\sqrt{2} = 1.414\ 213\ 562\ldots$, an *approximate number*, say 1.414, has a value which approximates the true number $\sqrt{2}$. It is given without qualification and is useful to someone for some purpose. A *significant number* is a number that does not differ from the true number by more than one-half in the last recorded digit. The number 1.41 is a significant number corresponding to $\sqrt{2}$ and is bounded by *range numbers* computable from the $\pm$ 0.005 implied, that is, $\sqrt{2}$ is contained in the interval $1.41 - 0.005 \leq \sqrt{2} \leq 1.41 + 0.005$ *for certain.* If the first digit to be dropped is 0 through 4, the number is simply truncated. If the first digit to be dropped is 5 through 9, the preceding digit is increased by 1. Thus 1.414 as a significant number representing $\sqrt{2}$ states that the true value of $\sqrt{2}$ lies in the interval 1.414 ± 0.0005. The numbers generated by carrying out the addition and subtraction are range numbers between which the true value of $\sqrt{2}$ is certain to lie.

A significant number is a special form of an *approximation-error number.* An approximation to the true value of $\sqrt{2}$ can be written as 1.414 ± 0.0003 with the error explicitly declared. Again, the range numbers generated by carrying out the explicitly displayed addition and subtraction bound the true value.

Another form of number is the *incomplete number.* It can be formed from the true value of the number simply by truncating after a prescribed number of digits. Nearly all who take their arithmetic seriously use incomplete numbers. Every incomplete number encountered in a calculation is treated as exact. Computational precision is assured by selecting the number of computational digits appropriately larger (double-precision rather than single-precision, for example) than the results require.

When one encounters random variables, certain properties of the distribution will be useful, such as the mean, standard deviation, etc. Range numbers as such do not exist here. The use of significant or approximation-error numbers is inappropriate and misleading. Again, incomplete numbers are used for the distribution parameters and in describing their *confidence bounds.*

Many estimation procedures used by engineers are replete with numbers. Here are some used to represent π:

3.14	(an approximate number)
3.142	(a significant number)
$3.142 \pm 0.000\ 41$	(an approximation-error number)
3.141	(an incomplete number)

The incomplete number, the significant number, and the approximate number have no tag or flag that identifies them as to type, and they are indistinguishable without qualification. In the area of machine computation and in the area of numbers resulting from measurement, significant, approximate, and approximation-error numbers are not useful. It is instructive to note that mathematical tables such as trigonometric tables are arrays of significant numbers; however, significant numbers had no role in their development, as incomplete numbers were used. Numbers that arise from measurement, such as the ultimate tensile strength of a 1020 cold-rolled steel, are properly expressed as $N(\mu, \sigma)$ or $LN(\mu, \sigma)$. Now μ is the incomplete number representing an unbiased estimate of the population mean, and σ is the incomplete number representing an unbiased estimate of the population standard deviation. The distribution is denoted by the symbol preceding the parentheses. There are no range numbers, for they contradict the entire notion of a distributed variable. Any qualification on the central tendency (the mean μ) is addressed by the standard deviation and distributional information, which give rise to confidence limits at some stated probability. These confidence limits perform the role of range numbers, in a way. All are incomplete numbers.

Since incomplete numbers are best for serious arithmetic and are used as random-variable descriptors, thoughtful engineers use them almost exclusively. They presume that any unqualified number encountered is an incomplete number. The hand-held calculator uses incomplete numbers for its arithmetic. The display option allows the user to show a *rounded* number representing the incomplete number in the register. No significance is implied, and none should be inferred. Calculations are often displayed in steps, and intermediate rounded results are recorded on paper. These results should not be reentered for the next calculation step; instead, they should be retained in calculator memory so that the full incomplete numbers are used in every calculation, and the number of correct digits in the resulting incomplete number is maximized.

Equation (2.20) particularized for the standard R. R. Moore specimen diameter gives $\phi_{0.30} \sim LN(0.506, 0.070)$. The companion normal has a mean of $\bar{y}$ and a standard deviation s_y of

$$\bar{y} = \ln 0.506 - \ln \sqrt{1 + (0.070/0.506)^2} = -0.690\,697$$

$$s_y = \sqrt{\ln\,[1 + (0.070/0.506)^2]} = 0.137\,685$$

The two-tailed confidence interval on $\bar{y}$ at the 0.99 confidence level is $\bar{y} \pm 2.576(0.137\,685)/\sqrt{133}$ or $-0.721\,451 \le \bar{y} \le -0.659\,943$ and $\exp(-0.721\,451) \le \bar{\phi}_{0.030} \le \exp(-0.659\,943)$ or $0.4860 \le \bar{\phi}_{0.030} \le 0.5169$. It would be misleading to consider the mean of the fatigue ratio as 0.506 to three digits or even to call it 0.5 because it is not known to three significant digits. Indeed, the digit 5 may not be correct. Dispersion in the mean is addressed by the standard deviation of 0.070. The unbiased estimator of the mean is 0.506, a rounded incomplete number, and the meaningful qualification is $0.4860 \le \bar{\phi}_{0.030} \le 0.5169$ at the 0.99 confidence level.

REFERENCES

2.1 *Metals Handbook,* vol. 1, 8th ed., American Society for Testing and Materials, Philadelphia, 1961.

2.2 J. E. Shigley and C. R. Mischke, *Mechanical Engineering Design,* 5th ed., McGraw-Hill, New York, 1989.

2.3 J. B. Kennedy and A. M. Neville, *Basic Statistical Methods for Engineers and Scientists,* 3d ed., Harper and Row, Publishers, New York, 1986.

2.4 C. R. Mischke, "Fitting Weibull Strength Data and Applying It to Stochastic Mechanical Design," *Transactions of the A.S.M.E., Journal of Mechanical Design,* vol. 114, March 1992, pp. 35–41.

2.5 C. R. Mischke, "A Distribution-Independent Plotting Rule for Ordered Failures," *Transactions of the A.S.M.E., Journal of Mechanical Design,* vol. 104, no. 3, July 1982, pp. 593–597.

2.6 E. B. Haugen, *Probabilistic Mechanical Design,* Wiley-Interscience, New York, 1980.

2.7 C. R. Mischke, *Mathematical Model Building,* 2d rev. ed., Iowa State University Press, Ames, 1980.