EXTREME Value Distribution

smallest value will have its own probability distribution, which may be an exact distribution or an asymptotic distribution.

Exact Distributions—Consider a random variable X (or population) with known probability distribution $f_X(x)$ or $F_X(x)$. A sample of size n from this population will have a largest and a smallest value; these extreme values will also have their respective distributions which are related to the distribution of the initial variate X.

A sample of size n is a set of observations $(x_1, x_2, \ldots, x_n)$ representing, respectively, the first, second, ..., and nth observed values. Because the observed values are unpredictable, they are a specific realization of an underlying set of random variables $(X_1, X_2, \ldots, X_n)$. Therefore, we are interested in the maximum and minimum of $(X_1, X_2, \ldots, X_n)$; i.e., the random variables

$$Y_n = \max(X_1, X_2, \ldots, X_n)$$

and

$$Y_1 = \min(X_1, X_2, \ldots, X_n)$$

Under certain assumptions, the exact probability distributions of Y_1 and Y_n can be derived as follows.

We may observe first that if Y_n is less than some specified value y, then all the other sample random variables $X_1, X_2, \ldots, X_n$ must individually be less than y. Then with the assumption that $X_1, X_2, \ldots, X_n$ are statistically independent and identically distributed as the initial variate X, i.e.,

$$F_{X_1}(x) = F_{X_2}(x) = \cdots = F_{X_n}(x) = F_X(x)$$

The CDF of Y_n, therefore, is

$$\begin{aligned} F_{Y_n}(y) &= P(X_1 \le y, X_2 \le y, \ldots, X_n \le y) \\ &= [F_X(y)]^n \end{aligned} \tag{4.25}$$

and the corresponding PDF is

$$f_{Y_n}(y) = \frac{dF_{Y_n}(y)}{dy} = n[F_X(y)]^{n-1} f_X(y) \tag{4.26}$$

where $F_X(y)$ and $f_X(y)$ are, respectively, the CDF and PDF of the initial variate X.

Similarly, the exact distribution of Y_1, can be derived as follows. In this case, we observe that if Y_1, the smallest among $X_1, X_2, \ldots, X_n$, is larger than y, then all the sample random variables must be individually larger than y. Hence, the *survival function* which is the complement of the CDF, is

$$\begin{aligned} 1 - F_{Y_1}(y) &= P(X_1 > y, X_2 > y, \ldots, X_n > y) \\ &= [1 - F_X(y)]^n \end{aligned}$$

Thus, the CDF of Y_1 is

$$F_{Y_1}(y) = 1 - [1 - F_X(y)]^n \tag{4.27}$$

and the corresponding PDF is

$$f_{Y_1}(y) = n[1 - F_X(y)]^{n-1} f_X(y) \tag{4.28}$$

Clearly, from Eqs. 4.25 through 4.28, we see that the exact distribution of the largest and smallest values from samples of size n is a function of the distribution of the initial variate.

▶ EXAMPLE 4.17 Consider the initial variate X with the exponential PDF as follows:

$$f_X(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

The corresponding CDF of X is

$$F_X(x) = 1 - e^{-\lambda x}$$

Therefore, the CDF of the largest value from samples of size n, according to Eq. 4.25, is

$$F_{Y_n}(y) = (1 - e^{-\lambda y})^n$$

and the corresponding PDF is

$$f_{Y_n}(y) = \lambda n(1 - e^{-\lambda y})^{n-1} e^{-\lambda y}$$

Graphically, the above PDF and CDF of Y_n are shown (for $\lambda = 1.0$) in Fig. E4.17 for different sample sizes n from 1 to 100.

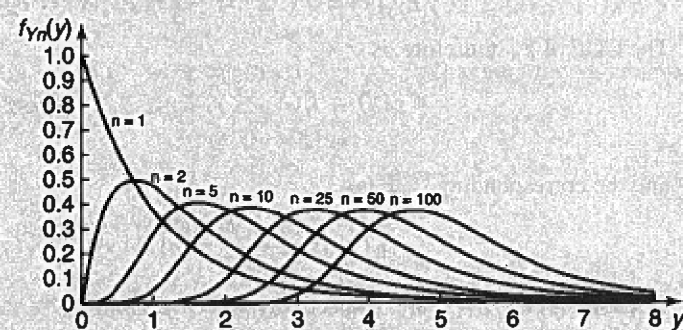
From Fig. E4.17, we can see that the PDF as well as the CDF shift to the right with increasing n. Also, as expected, the mode of the largest value increases with increasing n.

Expanding the above $F_{Y_n}(y)$ by the binomial series expansion, we obtain the series

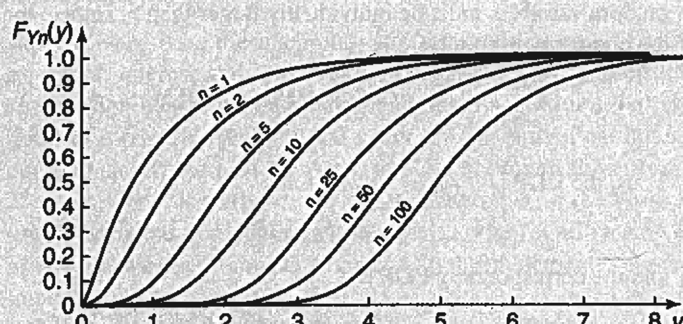
$$(1 - e^{-\lambda y})^n = 1 - ne^{-\lambda y} + \frac{n(n-1)}{2!} e^{-2\lambda y} - \cdots$$

For large n the above series approaches the double exponential $\exp(-ne^{-\lambda y})$. Hence, for large n the CDF of the largest value from an exponential population approaches the double exponential

$$F_{Y_n}(y) = \exp(-ne^{-\lambda y})$$



(a) PDF for Y_n



(b) CDF for Y_n

Figure E4.17 PDFs and CDFs of Y_n for different n ($\lambda = 1.0$). ◀

The Asymptotic Distributions—In Example 4.17, we observed that in the case of the exponential initial distribution, the CDF of the largest value from samples of size n approaches the double exponential distribution as n increases; in this case, this double exponential distribution is the *asymptotic distribution* of the largest value. This characteristic is shown also in Fig. 4.2, illustrating the convergence of the exact distribution of Y_n to the asymptotic double exponential distribution as $n \to \infty$.

The characteristics illustrated in Fig. 4.2 for the initial exponential distribution actually apply also to other initial distributions; i.e., the distribution of an extreme value converges asymptotically in distribution as n increases. According to Gumbel (1954, 1960), there are three types of such asymptotic distributions (although not exhaustive) depending on the tail behavior of the initial PDFs; namely, as follows:

- ***Type I***: The double exponential form.
- ***Type II***: The single exponential form.
- ***Type III***: The exponential form with an upper (or lower) bound.

The extreme value from an initial distribution with an exponentially decaying tail (in the direction of the extreme) will converge asymptotically to the *Type I* limiting form. This was illustrated earlier in Example 4.17 and demonstrated graphically in Fig. 4.2. For an initial variate with a PDF that decays with a polynomial tail, the distribution of its extreme value will converge to the *Type II* limiting form, whereas if the extreme value is limited, the

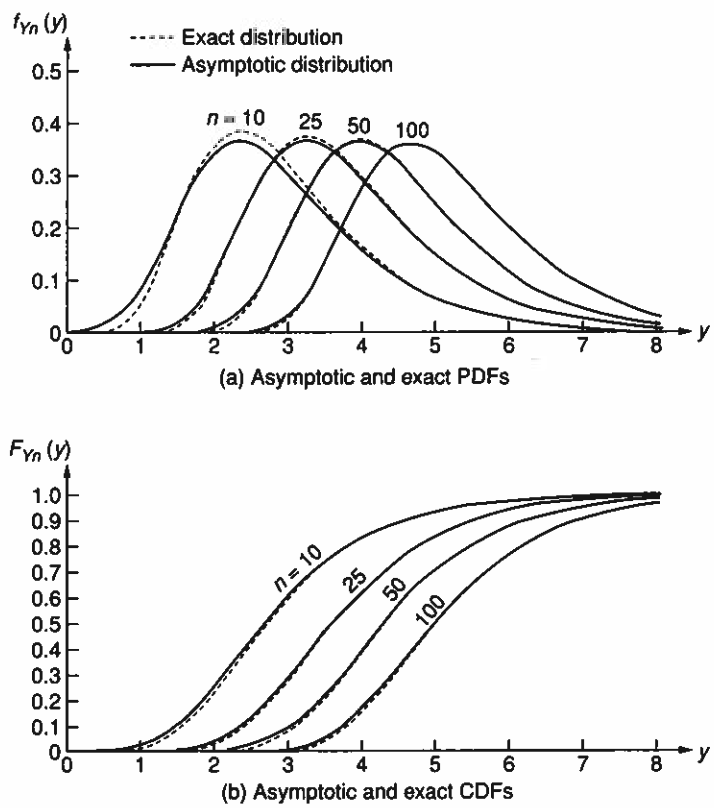


Figure 4.2 Exact and asymptotic distributions of Y_n from an exponential initial distribution.

corresponding extremal distribution will converge asymptotically to the *Type III* asymptotic form.

Parameters of the Asymptotic Distributions—Even though the forms of the asymptotic distributions do not depend on the distributions of the initial variates, the parameters of the asymptotic distributions, such as u_n and α_n of the *Type I* asymptote, will depend on the distribution of the initial variate. We shall limit our illustrations below to the parameters of the *Type I* asymptotic form (a more complete discussion of the other asymptotic forms can be found in Ang and Tang, Vol. 2, 1984).

The Gumbel Distribution—The CDF of the *Type I* asymptotic form for the largest value, well known as the Gumbel distribution (Gumbel, 1958), is

$$F_{Y_n}(y) = \exp\left[-e^{-\alpha_n(y-u_n)}\right] \tag{4.29a}$$

and its PDF is

$$f_{Y_n}(y) = \alpha_n e^{-\alpha_n(y-u_n)}\exp\left[-e^{-\alpha_n(y-u_n)}\right] \tag{4.29b}$$

in which

$u_n =$ the most probable value of Y_n

$\alpha_n =$ an inverse measure of the dispersion of values of Y_n

Moreover, the mean and variance of the largest value, Y_n, and smallest value, Y_1, are related to the respective parameters as follows (see Ang and Tang, Vol. 2, 1984):

$$\mu_{Y_n} = u_n + \frac{\gamma}{\alpha_n} \tag{4.30a}$$

in which $\gamma = 0.577216$ (the Euler number); and

$$\sigma^2_{Y_n} = \frac{\pi^2}{6\alpha_n^2} \tag{4.30b}$$

whereas for the smallest value, the corresponding mean and variance are

$$\mu_{Y_1} = u_1 - \frac{\gamma}{\alpha_1} \qquad \text{and} \qquad \sigma^2_{Y_1} = \frac{\pi^2}{6\alpha_1^2}$$

▶ EXAMPLE 4.18

Consider an initial variate X with the standard normal distribution, i.e., $N(0,1)$, with the PDF

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

In this case, the tail of the PDF is clearly exponential; hence, the asymptotic distribution of the largest value is of the double exponential form (*Type I*); specifically, the CDF of Y_n is

$$F_{Y_n}(y) = \exp[-e^{-\alpha_n(y-u_n)}]$$

and its PDF is

$$f_{Y_n}(y) = \alpha_n e^{-\alpha_n(y-u_n)}\exp[-e^{-\alpha_n(y-u_n)}]$$

with parameters

$$u_n = \sqrt{2\ln n} - \frac{\ln \ln n + \ln 4\pi}{2\sqrt{2\ln n}}$$

and

$$\alpha_n = \sqrt{2\ln n}$$

Explanation of the derivation of the above parameters may be found in Ang and Tang, Vol. 2 (1984). The mean and standard deviation of the largest value can then be obtained, respectively, from Eqs. 4.30a and 4.30b.

If the initial variate X has a general Gaussian distribution, $N(\mu, \sigma)$, we observe from Example 4.2 that $(X-\mu)/\sigma$ will be $N(0,1)$. Then, the asymptotic distribution of the largest value from $(X-\mu)/\sigma$ will be of the double exponential form with the parameters u_n and α_n obtained above. Therefore, if Y_n' is the largest value from the initial Gaussian variate X, then it follows that

$$Y_n = \frac{Y_n' - \mu}{\sigma}$$

is the largest from $\dfrac{X-\mu}{\sigma}$, and the CDF of Y_n' is, therefore,

$$F_{Y_n'}(y') = F_{Y_n}\left(\frac{y'-\mu}{\sigma}\right) = \exp\left[-e^{-\alpha_n\left(\frac{y'-\mu-\sigma u_n}{\sigma}\right)}\right] = \exp\left[-e^{-\frac{\alpha_n}{\sigma}(y'-\mu-\sigma u_n)}\right]$$

Hence, the CDF of Y_n' is of the same double exponential form as Y_n, with the parameters

$$u_n' = \sigma u_n + \mu$$
$$\alpha_n' = \alpha_n/\sigma$$ ◄

In the case of the smallest value from samples of size n, the corresponding distributions, PDF and CDF, would shift to the left as n increases. Similar to the largest value, the distribution of the smallest value will also converge (in distribution) to one of the three types of asymptotic distributions depending on the tail behavior (in the direction of the smallest value) of the PDF of the initial variate.

► **EXAMPLE 4.19**

Consider an initial variate with the standard Gaussian distribution $N(0, 1)$. The tail behavior in the lower end of this PDF is obviously also exponential; thus, the CDF of the smallest value from this initial variate will also converge to the *Type I* asymptotic form as follows:

$$F_{Y_1}(y) = 1 - \exp[-e^{\alpha_1(y-u_1)}]$$

and the corresponding PDF is

$$f_{Y_1}(y) = \alpha_1 e^{\alpha_1(y-u_1)} \exp[-e^{\alpha_1(y-u_1)}]$$

in which the parameters u_1 and α_1 are

$$u_1 = -\sqrt{2\ln n} + \frac{\ln \ln n + \ln 4\pi}{2\sqrt{2\ln n}}; \quad \text{and}$$
$$\alpha_1 = \sqrt{2\ln n}$$ ◄

► **EXAMPLE 4.20**

If the initial variate X is described by the Rayleigh distribution with PDF,

$$f_X(x) = \frac{x}{\alpha^2} e^{-(1/2)(x/\alpha)^2} \qquad x \geq 0$$

in which $\sigma =$ the modal value. Although the tail behavior of this PDF may not be obvious, it is exponential (see Ang and Tang, Vol. 2, 1984), and thus the distribution of the largest value will converge asymptotically to the *Type I* form with the following parameters:

$$u_n = \alpha\sqrt{2\ln n}$$

and

$$\alpha_n = \frac{\sqrt{2\ln n}}{\alpha}$$

Then, according to Eqs. 4.30a and 4.30b, the mean and standard deviation of Y_n are, respectively,

$$\mu_{Y_n} = \alpha\sqrt{2\ln n} + 0.5772\left(\frac{\alpha}{\sqrt{2\ln n}}\right)$$

and

$$\sigma_{Y_n} = \frac{\pi}{\sqrt{6}}\frac{\alpha}{\sqrt{2\ln n}} = \frac{\pi\alpha}{2\sqrt{3\ln n}}$$ ◀

The Fisher–Tippett Distribution—The *Type II* asymptotic distribution of the largest value is often referred to as the Fisher–Tippett distribution (Fisher and Tippett, 1928). Its CDF is

$$F_{Y_n}(y) = \exp\left[-\left(\frac{v_n}{y}\right)^k\right] \tag{4.31}$$

The corresponding PDF is

$$f_{Y_n}(y) = \frac{k}{v_n}\left(\frac{v_n}{y}\right)^{k+1}\exp\left[-\left(\frac{v_n}{y}\right)^k\right] \tag{4.32}$$

where the parameters

v_n = the most probable value of Y_n

k = the shape parameter, which is an inverse measure of the dispersion of values of Y_n

The mean and standard deviation of the variate Y_n are related to the above parameters as follows:

$$\begin{aligned} \mu_{Y_n} &= v_n\Gamma\left(1-\frac{1}{k}\right) \\ \sigma_{Y_n} &= v_n\sqrt{\Gamma\left(1-\frac{2}{k}\right)-\Gamma^2\left(1-\frac{1}{k}\right)} \end{aligned} \tag{4.33}$$

The *Type I* and *Type II* asymptotic forms are related through a logarithmic transformation as follows: If Y_n has the *Type II* asymptotic distribution of Eq. 4.31, with parameters v_n and k, the distribution of $\ln Y_n$ will have the *Type I* asymptotic form with parameters

$$u_n = \ln v_n \quad \text{and} \quad \alpha_n = k$$

The above logarithmic transformation applies also to the smallest value of Y_1 and $\ln Y_1$.

▶ EXAMPLE 4.21

If the initial variate X is lognormal, with parameters λ_X and ζ_X, $\ln X$ is normal with parameters $\mu = \lambda_X$ and $\sigma = \zeta_X$. Then, according to the results of Example 4.17, the largest value of $\ln X$ will converge asymptotically to the *Type I* distribution with parameters

$$u_n = \zeta_X\left(\sqrt{2\ln n} - \frac{\ln\ln n + \ln 4\pi}{2\sqrt{2\ln n}}\right) + \lambda_X \quad \text{and} \quad \alpha_n = \frac{\sqrt{2\ln n}}{\zeta_X}$$

Therefore, according to the above logarithmic transformation, the largest value of the initial lognormal variate X will converge to the *Type II* asymptotic form with parameters

$$v_n = e^{u_n} \quad \text{and} \quad k = \alpha_n$$

◀

The Weibull Distribution—For the *Type III* asymptotic form, the asymptotic distribution of the smallest value is of greater interest. In engineering, it is well known as the Weibull distribution (Weibull, 1951), which was discovered by Weibull for modeling the fracture strength of materials. Its CDF is given by

$$F_{Y_1}(y) = 1 - \exp\left[-\left(\frac{y-\varepsilon}{w_1-\varepsilon}\right)^k\right]; \qquad y \geq \varepsilon \tag{4.34}$$

where

$w_1 =$ the most probable smallest value
$k =$ the shape parameter
$\varepsilon =$ the lower bound value of y

The mean and standard deviation of Y_1 is related to the above parameters as follows:

$$\mu_{Y_1} = \varepsilon + (w_1 - \varepsilon)\Gamma\left(1 + \frac{1}{k}\right);$$

and (4.35)

$$\sigma_{Y_1} = (w_1 - \varepsilon)\sqrt{\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)}$$

▶ EXAMPLE 4.22

Suppose the lower-bound fracture strength of a welded joint is 4.0 ksi. If the actual strength of the joint Y_1 is modeled with a *Type III* smallest asymptotic distribution, Eq. 4.34, with parameters $w_1 = 15.0$ ksi and $k = 1.75$, the probability that the strength of the joint will be at least 16.5 ksi is

$$P(Y_1 \geq 16.5) = \exp\left[-\left(\frac{16.5-4}{15-4}\right)^{1.75}\right] = 0.286$$

The mean and standard deviation of the joint strength are, according to Eq. 4.35, respectively,

$$\begin{aligned}\mu_{Y_1} &= 4 + (15-4)\Gamma\left(1 + \frac{1}{1.75}\right) = 4 + 11 \times \Gamma(1.5714) = 4 + 11 \times 0.8906 \\ &= 13.80\ \text{ksi};\end{aligned}$$

and

$$\begin{aligned}\sigma_{Y_1} &= (15-4)[\Gamma(1+2/1.75) - \Gamma^2(1+1/1.75)]^{1/2} = 11 \times [\Gamma(1.1429) - \Gamma^2(1.5714)]^{1/2} \\ &= 11[0.9354 - (0.8906)^2]^{1/2} = 4.15\ \text{ksi}\end{aligned}$$

The values of the respective gamma functions indicated above were evaluated through *MATLAB.* ◀