Univariate random variables

Nowak, A.S., Collins K.R. Reliability of structures.
McGraw-Hill Higher Education 2000

Any random variable is defined by its cumulative distribution function (CDF), $F_X(x)$.

The probability density function $f_X(x)$ of a continuous random variable is the first derivative of $F_X(x)$.

The most important continuous random variables used in structural reliability analysis are as follows: uniform, normal (Gaussian), lognormal, gamma, extreme type I (Gumbel), extreme type II (Frechet), extreme type III (Weibull).

The binomial and Poisson distributions of discrete random variables are distinguished too.

Uniform distribution

A *uniform random var*iable or *uniform distribution* denotes a constant PDF function in the interval [*a*, *b*]. Thus all numbers in this interval are equally likely to appear.

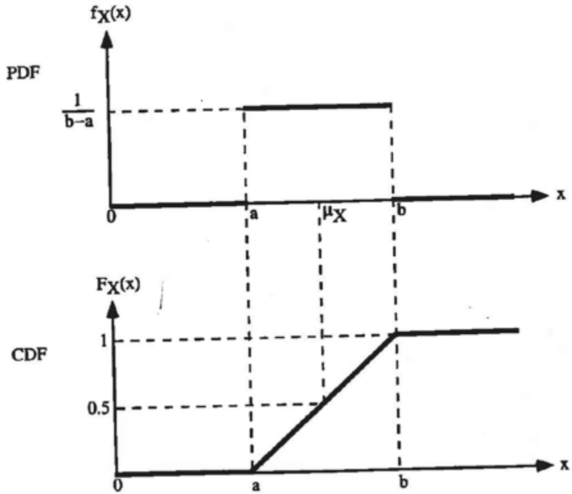
Mathematically the uniform PDF function is defined as follows:

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

a and *b* define the lower and upper bounds of the random variable.

The cumulative distribution function (CDF) for a uniform random variable is

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



PDF and CDF of a uniform random variable

The mean and variance are as follows:

$$\mu_X = \frac{a+b}{2}$$

$$\sigma_X^2 = \frac{(b-a)^2}{12}$$

$$\sigma_X^2 = E\left(X^2\right) - \left[E\left(X\right)\right]^2$$

$$E\left(X\right) = \int_{-\infty}^{\infty} x f\left(x\right) dx = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$$

$$E\left(X^2\right) = \int_{-\infty}^{\infty} x^2 f\left(x\right) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

$$\sigma_X^2 = E\left(X^2\right) - \left(E\left(X\right)\right)^2 = \frac{b^2 + ab + a^2}{3} - \frac{(b+a)^2}{4} = \frac{(b-a)^2}{12}$$

Normal (Gaussian) distribution

The most important distribution is the *normal distribution* also called the *Gaussian distribution.*

It is a two-parameter distribution defined by the density function

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right]$$

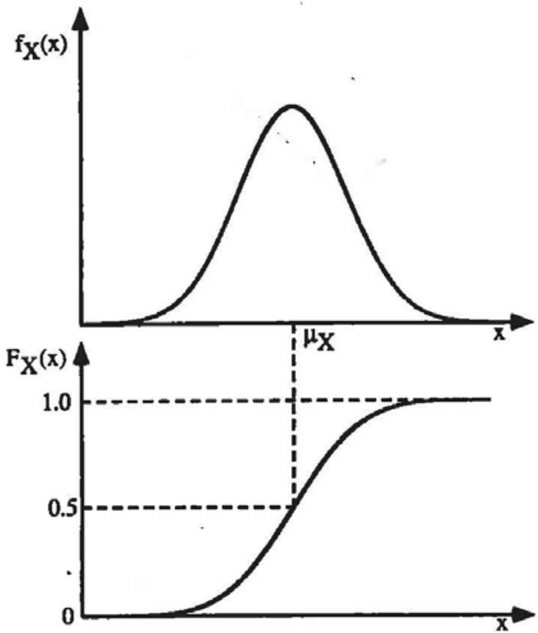
where μ and σ are equal to μ_X *(expected value*, and σ_X (*standard deviation*), respectively.This distribution will be denoted $N(\mu, \sigma)$.

The distribution function, is given by

$$F_X(x) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma} \right)^2 \right] dt$$

This integral cannot be evaluated on a closed form.

The figure shows general shapes of both the PDF and CDF of a normal random variable.



Standard normal (Gaussian) variable

General remarks: let X be an arbitrary random variable.
The standard form of X, denoted by Z, is defined

$$Z = \frac{X - \mu_X}{\sigma_X}$$

The mathematical expectation (mean value) of an arbitrary function, $g(X)$, of the random variable X is defined

$$\mu_{g(X)} = E\left[g\left(X\right)\right] = \int_{-\infty}^{\infty} g\left(x\right) f_X\left(x\right) dx$$

The formula above, with $Z = \mathrm{g}(X)$ and the variance property prove

$$\mu_Z = E\left[\frac{X - \mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X}\left[E\left(X\right) - E\left(\mu_X\right)\right] = \frac{1}{\sigma_X}\left(\mu_X - \mu_X\right) = 0$$

$$\sigma_Z^2 = E\left(Z^2\right) - \mu_Z^2 = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)^2\right] - 0 = \frac{1}{\sigma_X^2}\left[E\left(X - \mu_X\right)^2\right] = \frac{\sigma_X^2}{\sigma_X^2} = 1$$

Thus mean of any standard random variable is 0, its variance is 1.

Standard random variable – a “zero mean, unit variance” form.
The distribution function of a normal variable

$$F_X(x) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2 \right] dt$$

Substituting $s = \dfrac{t-\mu}{\sigma}, \quad dt = \sigma ds$ the equation (2.46) becomes

$$F_X(x) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} s^2 \right] ds = \Phi_X\left(\frac{x-\mu}{\sigma}\right)$$

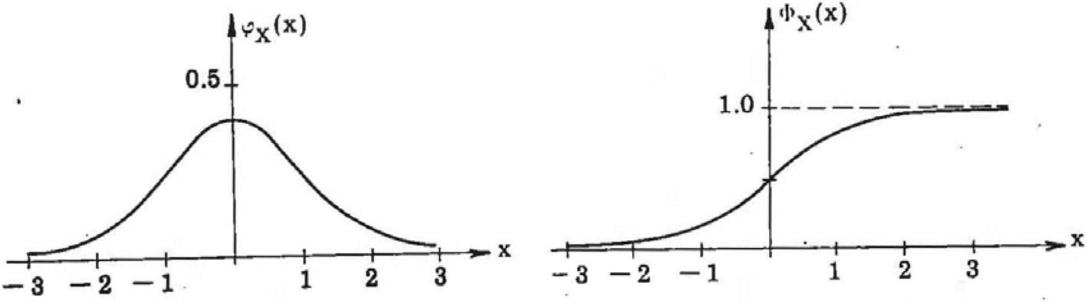
where Φ_X is a *standard normal distribution function* defined by

$$\Phi_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{t^2}{2} \right] dt$$

The corresponding *standard normal density function* is

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

The functions φ_X and Φ_X are shown in figure:



Thus only a standard normal table is required.
Spreadsheet packages include a standard normal CDF function.

Values of $\Phi(z)$ are listed in Table1 for z ranging from 0 to –8.9 (part of the table shown, after Nowak, Collins: Reliability of structures)

Table 1. The CDF of the standard normal variable $\Phi(z)$.

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	5.00E-01	4.96E-01	4.92E-01	4.88E-01	4.84E-01	4.80E-01	4.76E-01	4.72E-01	4.68E-01	4.64E-01
−0.1	4.60E-01	4.56E-01	4.52E-01	4.48E-01	4.44E-01	4.40E-01	4.36E-01	4.33E-01	4.29E-01	4.25E-01
−0.2	4.21E-01	4.17E-01	4.13E-01	4.09E-01	4.05E-01	4.01E-01	3.97E-01	3.94E-01	3.90E-01	3.86E-01
−0.3	3.82E-01	3.78E-01	3.74E-01	3.71E-01	3.67E-01	3.63E-01	3.59E-01	3.56E-01	3.52E-01	3.48E-01
−0.4	3.45E-01	3.41E-01	3.37E-01	3.34E-01	3.30E-01	3.26E-01	3.23E-01	3.19E-01	3.16E-01	3.12E-01
−0.5	3.09E-01	3.05E-01	3.02E-01	2.98E-01	2.95E-01	2.91E-01	2.88E-01	2.84E-01	2.81E-01	2.78E-01
−0.6	2.74E-01	2.71E-01	2.68E-01	2.64E-01	2.61E-01	2.58E-01	2.55E-01	2.51E-01	2.48E-01	2.45E-01
−0.7	2.42E-01	2.39E-01	2.36E-01	2.33E-01	2.30E-01	2.27E-01	2.24E-01	2.21E-01	2.18E-01	2.15E-01
−0.8	2.12E-01	2.09E-01	2.06E-01	2.03E-01	2.00E-01	1.98E-01	1.95E-01	1.92E-01	1.89E-01	1.87E-01
−0.9	1.84E-01	1.81E-01	1.79E-01	1.76E-01	1.74E-01	1.71E-01	1.69E-01	1.66E-01	1.64E-01	1.61E-01
−1	1.59E-01	1.56E-01	1.54E-01	1.52E-01	1.49E-01	1.47E-01	1.45E-01	1.42E-01	1.40E-01	1.38E-01
−1.1	1.36E-01	1.33E-01	1.31E-01	1.29E-01	1.27E-01	1.25E-01	1.23E-01	1.21E-01	1.19E-01	1.17E-01
−1.2	1.15E-01	1.13E-01	1.11E-01	1.09E-01	1.07E-01	1.06E-01	1.04E-01	1.02E-01	1.00E-01	9.85E-02
−1.3	9.68E-02	9.51E-02	9.34E-02	9.18E-02	9.01E-02	8.85E-02	8.69E-02	8.53E-02	8.38E-02	8.23E-02
−1.4	8.08E-02	7.93E-02	7.78E-02	7.64E-02	7.49E-02	7.35E-02	7.21E-02	7.08E-02	6.94E-02	6.81E-02
−1.5	6.68E-02	6.55E-02	6.43E-02	6.30E-02	6.18E-02	6.06E-02	5.94E-02	5.82E-02	5.71E-02	5.59E-02
−1.6	5.48E-02	5.37E-02	5.26E-02	5.16E-02	5.05E-02	4.95E-02	4.85E-02	4.75E-02	4.65E-02	4.55E-02
−1.7	4.46E-02	4.36E-02	4.27E-02	4.18E-02	4.09E-02	4.01E-02	3.92E-02	3.84E-02	3.75E-02	3.67E-02
−1.8	3.59E-02	3.51E-02	3.44E-02	3.36E-02	3.29E-02	3.22E-02	3.14E-02	3.07E-02	3.01E-02	2.94E-02
−1.9	2.87E-02	2.81E-02	2.74E-02	2.68E-02	2.62E-02	2.56E-02	2.50E-02	2.44E-02	2.39E-02	2.33E-02
−2	2.28E-02	2.22E-02	2.17E-02	2.12E-02	2.07E-02	2.02E-02	1.97E-02	1.92E-02	1.88E-02	1.83E-02
−2.1	1.79E-02	1.74E-02	1.70E-02	1.66E-02	1.62E-02	1.58E-02	1.54E-02	1.50E-02	1.46E-02	1.43E-02
−2.2	1.39E-02	1.36E-02	1.32E-02	1.29E-02	1.25E-02	1.22E-02	1.19E-02	1.16E-02	1.13E-02	1.10E-02
−2.3	1.07E-02	1.04E-02	1.02E-02	9.90E-03	9.64E-03	9.39E-03	9.14E-03	8.89E-03	8.66E-03	8.42E-03
−2.4	8.20E-03	7.98E-03	7.76E-03	7.55E-03	7.34E-03	7.14E-03	6.95E-03	6.76E-03	6.57E-03	6.39E-03
−2.5	6.21E-03	6.04E-03	5.87E-03	5.70E-03	5.54E-03	5.39E-03	5.23E-03	5.08E-03	4.94E-03	4.80E-03
−2.6	4.66E-03	4.53E-03	4.40E-03	4.27E-03	4.15E-03	4.02E-03	3.91E-03	3.79E-03	3.68E-03	3.57E-03
−2.7	3.47E-03	3.36E-03	3.26E-03	3.17E-03	3.07E-03	2.98E-03	2.89E-03	2.80E-03	2.72E-03	2.64E-03
−2.8	2.56E-03	2.48E-03	2.40E-03	2.33E-03	2.26E-03	2.19E-03	2.12E-03	2.05E-03	1.99E-03	1.93E-03
−2.9	1.87E-03	1.81E-03	1.75E-03	1.69E-03	1.64E-03	1.59E-03	1.54E-03	1.49E-03	1.44E-03	1.39E-03

Values $\Phi(z)$ for $z > 0$ can also be obtained from Table 1 by applying the symmetry property of the normal distribution:

$$\Phi(z) = 1 - \Phi(-z)$$

The probability information for the standard normal random variable allows for the CDF and PDF values for any normal random variable by a simple coordinate transformation.
Let X be anynormal random variable and Z be a standard form of X. We can show that

$$X = \mu_X + Z\sigma_X$$

The definition of CDF implies

$$F_X(x) = P(X \le x) = P(\mu_X + Z\sigma_X \le x) = P\left(Z \le \frac{x - \mu_X}{\sigma_X}\right)$$

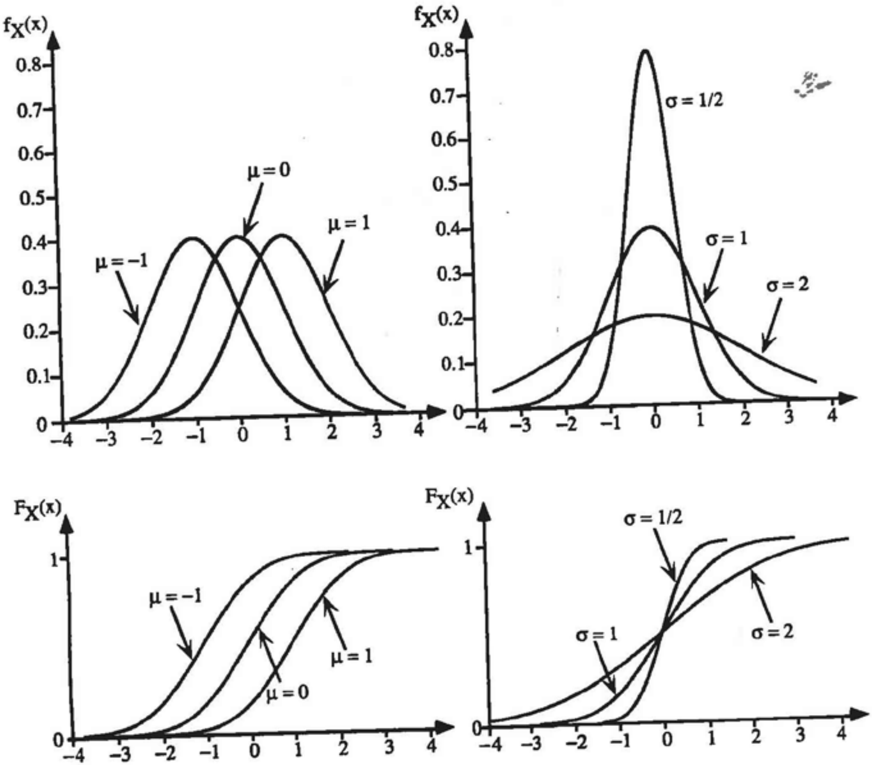
$$F_X(x) = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right) = F_Z(z)$$

The relationship between the PDF of any normal random variable, $f_X(x)$, with the PDF of the standard normal variable, $\phi(x)$:

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} \Phi\left(\frac{x-\mu_X}{\sigma_X}\right) = \frac{1}{\sigma_X} \phi\left(\frac{x-\mu_X}{\sigma_X}\right)$$

Using above eqs. the distribution functions for an arbitrary normal random variable (given μ_X and σ_X) may be derived, using information in Table 1.

CDFs and PDFs for normal random variables are shown in Fig

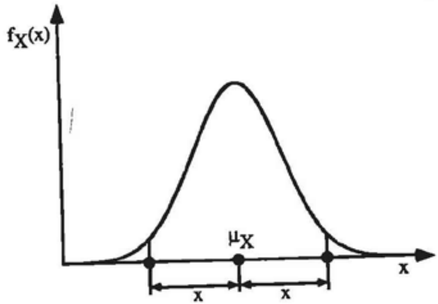


Important properties of the PDF and CDF functions for normal random variables are summarized as follows:

1. The PDF $f_X(x)$ is symmetric about the mean μ_X

$$f_X(\mu_X + x) = f_X(\mu_X - x)$$

The illustration is shown in Figure



2. The symmetry property of 1. yields $F_X(\mu_X + x) + F_X(\mu_X - x) = 1$

Logarithmic normal distribution

Let the random variable $Y = \ln X$ be normally distributed $N(\mu_Y, \sigma_Y)$.

The random variable X follows the *logarithmic normal distribution*, with parameters $\mu_Y \in R$, $\sigma_Y > 0$

The log-normal density function is stated, for $x > 0$

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

Let X be log-normally distributed with the parameters μ_Y and σ_Y.

Note that μ_Y and σ_Y are not equal to μ_X and σ_X.

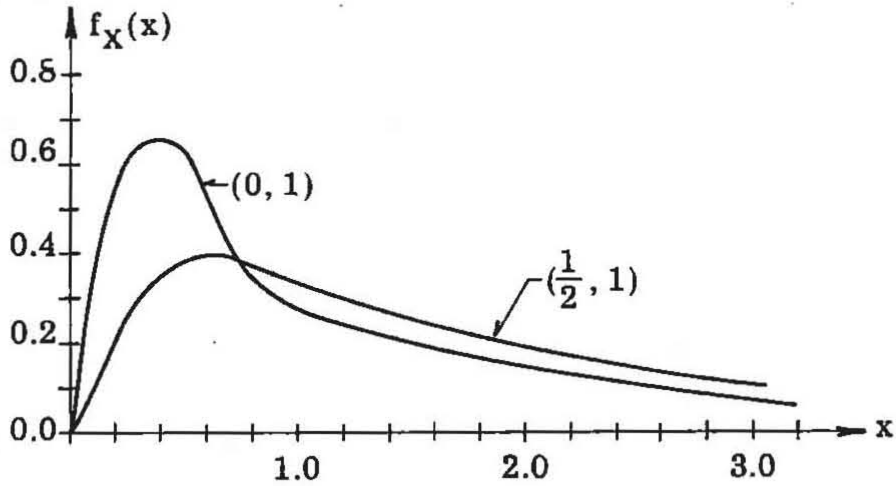
It can be shown that

$$\mu_X(x) = \exp\left(\mu_Y + \frac{1}{2}\sigma_Y^2\right)$$

$$\sigma_X = \sqrt{\mu_X^2\left(e^{\sigma_Y^2} - 1\right)}$$

$$F_X(x) = P(X \le x) = \Phi\left(\frac{\ln x - \mu_Y}{\sigma_Y}\right)$$

The log-normal density functions with the parameters $(\mu_Y, \sigma_Y) = (0,1)$ and (1/2, 1) are presented in Fig..



Example

Let the compressive strength X for concrete be log-normally distributed with the parameters $(\mu_Y, \sigma_Y) = (3\text{ MPa}, 0.2\text{ MPa})$.

Then

$$\mu_X = \exp\left(3 + \frac{1}{2} \cdot 0.04\right) = 20.49\text{ MPa}$$

$$\sigma_X^2 = 20.49^2(1.0408 - 1) = 17.14\ (\text{MPa})^2$$

$$\sigma_X = 4.14\text{ MPa}$$

and

$$P(X \leq 10\text{ MPa}) = \Phi\left((\ln 10 - 3)/2\right) = \Phi(-3.487) = 2.4 \cdot 10^{-4}$$

The PDF and CDF may be calculated using functions $\phi(z)$ and $\Phi(z)$ for a standard normal random variable Z as follows:

$$F_X(x) = P(X \leq x) = P(\ln X \leq \ln x) = P(Y \leq y) = F_Y(y)$$

Since Y is normally distributed standard normal functions apply.

Specifically

$$F_X(x) = F_Y(y) = \Phi\left(\frac{y-\mu_Y}{\sigma_Y}\right)$$

where $y = \ln(x)$, $\mu_Y = \mu_{\ln(X)}$ = mean value of $\ln(X)$,

and $\sigma_Y = \sigma_{\ln(X)}$ = standard deviation of $\ln(X)$.

These parameters are functions of μ_X, σ_X and V_X by the following formulas:

$$\sigma^2_{\ln(X)} = \ln\left(V_X^2 + 1\right), \qquad \mu_{\ln(X)} = \ln(\mu_X) - \frac{1}{2}\sigma^2_{\ln(X)}$$

If V_X is less than 0.2, the following approximations are valid:

$$\sigma^2_{\ln(X)} \approx V_X^2, \qquad \mu_{\ln(X)} \approx \ln(\mu_X)$$

For the PDF function Eq. 2.12 gives

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{d}{dx}\Phi\left(\frac{\ln(x)-\mu_{\ln(X)}}{\sigma_{\ln(X)}}\right) = \frac{1}{x\sigma_X}\phi\left(\frac{\ln(x)-\mu_{\ln(X)}}{\sigma_{\ln(X)}}\right)$$

Example

Let X be a lognormal random variable whose mean value is 250, standard deviation is 30. Find $F_X(200)$ and $f_X(200)$.

$$V_X = \frac{\sigma_X}{\mu_X} = \frac{30}{250} = 0.12$$

$$\sigma^2_{\ln(X)} = \ln\left(V_X^2 + 1\right) = 0.0143, \qquad \sigma_{\ln(X)} = 0.01196$$

$$\mu_{\ln(X)} = \ln(\mu_X) - \frac{1}{2}\sigma^2_{\ln(X)} = \ln(250) - 0.5(0.0143) = 5.51$$

$$F_X(200) = \Phi\left(\frac{\ln(x) - \mu_{\ln(X)}}{\sigma_{\ln(X)}}\right) = \Phi\left(\frac{\ln(200) - 5.51}{0.1196}\right) = 0.0384$$

$$f_X(200) = \frac{1}{x\sigma_{\ln(X)}}\phi\left(\frac{\ln(x) - \mu_{\ln(X)}}{\sigma_{\ln(X)}}\right) = \frac{\Phi(-1.77)}{200(0.1196)} = \frac{0.0833}{23.92} = 0.00384$$

Distributions – Extreme

Nowak, A.S., Collins K.R. Reliability of structures.
McGraw-Hill Higher Education 2000

Extreme type I (Gumbel distribution, Fisher-Tippett type I)

Extreme value distributions depict well probabilistic nature of the extreme values (largest or smallest) of some phenomenon over time.

Consider n time intervals, e.g. years.
There is a maximum value of some phenomenon (e.g. wind speed) during each interval (year).

Determine the random model for those largest annual wind speeds.
Let $W_1, \ldots , W_n$ be the largest wind speeds in n years.
Then $X = \max\left(W_1, W_2, \ldots , W_n\right)$ is an extreme Type I random variable.

The CDF and PDF for this random variable are

$$F_X(x) = e^{-e^{-\alpha(x-u)}} \quad \text{for} \quad -\infty \leq x \leq \infty$$

$$f_X(x) = \alpha e^{-e^{-\alpha(x-u)}} e^{-\alpha(x-u)}$$

where u and a are distribution parameters.

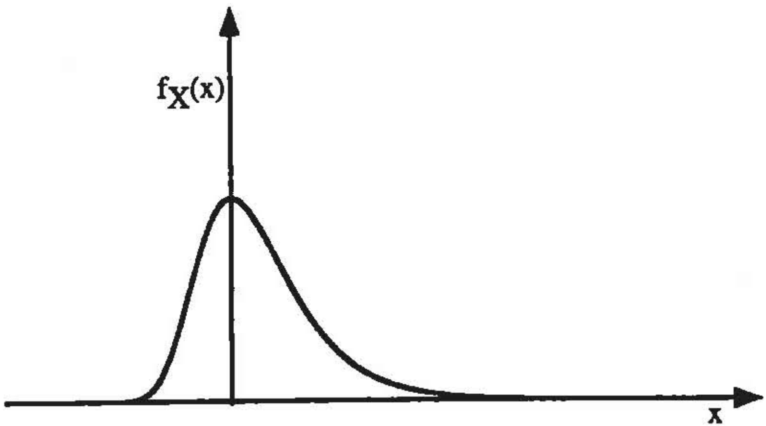


FIGURE. PDF of an extreme Type I random variable.

The mean and standard deviation for this variable may be obtained by the following approximations (Benjamin and Cornell, 1970):

$$\mu_X \approx u + \frac{0.577}{\alpha} \qquad \sigma_X \approx \frac{1.282}{\alpha}$$

Thus if the mean and standard deviation are known:

$$\alpha \approx \frac{1.282}{\sigma_X}, \qquad u \approx \mu_X - 0.45\sigma_X$$

Extreme type II (Frechet distribution, Fisher-Tippett type II)

Extreme Type II variable may model the maximum seismic load applied to a structure. The CDF and PDF are

$$F_X(x) = e^{-e^{-(u/x)^k}} \quad \text{for} \quad 0 \le x \le \infty, \qquad f_X(x) = \frac{k}{u}\left(\frac{u}{x}\right)^{k+1} e^{-(u/x)^k}$$

where u and k are distribution parameters.

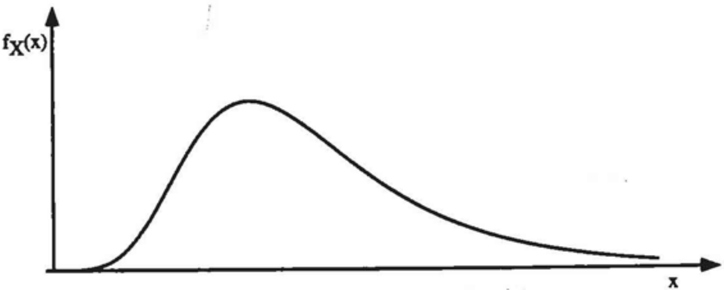


FIGURE. PDF for an extreme Type II random variable

The mean and standard deviation is given:

$$\mu_X = u\Gamma\left(1-\frac{1}{k}\right) \quad \text{for} \quad k>1$$

$$\sigma_X^2 = u^2\left[\Gamma\left(1-\frac{2}{k}\right)-\Gamma^2\left(1-\frac{1}{k}\right)\right] \quad \text{for} \quad k>2$$

The coefficient of variation, V_X is a function of *k* only.

Graphs exist to calculate V_X for any *k* (see, for example, Ang and Tang, 1984)

Extreme Type III (Weibull Distribution)

The extreme Type III distribution is defined by three parameters. Two variants, for the largest and the smallest values exist. The CDF of the largest values is defined by

$$F_X(x) = e^{-\left(\frac{w-x}{w-u}\right)^k} \quad \text{for} \quad x \le w$$

where w, u, and k are parameters.
The mean and variance are

$$\mu_X = w - (w-u)\Gamma\left(1+\frac{1}{k}\right)$$

$$\sigma_X^2 = (w-u)^2\left[\Gamma\left(1-\frac{2}{k}\right) - \Gamma^2\left(1-\frac{1}{k}\right)\right]$$

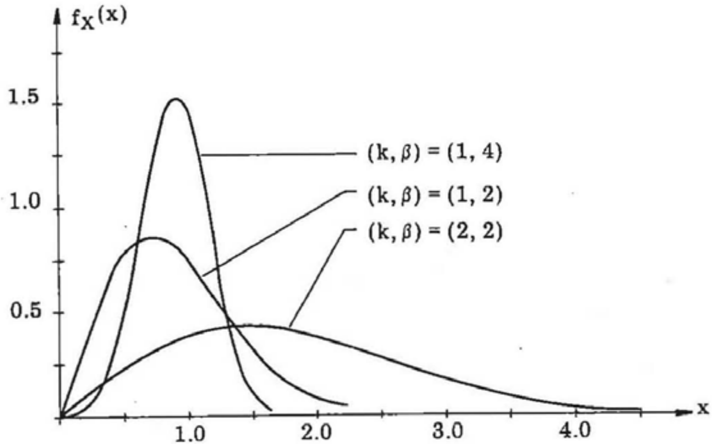
The CDF of the smallest values is defined by

$$F_X(x) = 1 - e^{-\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^k} \quad \text{for} \quad x \ge \varepsilon$$

where u, ε, and k are the parameters.
The mean and variance of the smallest values may be calculated by the following formulas:

$$\mu_X = \varepsilon - (u-\varepsilon)\Gamma\left(1+\frac{1}{k}\right)$$

$$\sigma_X^2 = (u-\varepsilon)^2\left[\Gamma\left(1-\frac{2}{k}\right) - \Gamma^2\left(1-\frac{1}{k}\right)\right]$$



Poisson distribution

The Poisson variable of a discrete probability distribution may be used to determine the PMF *(probability mass function)*for the number of occurrences of an event in a time or space interval (0, t).

Examples: the number of earthquakes within a certain time interval or the number of defects in a certain length of rod.

The following assumptions behind the Poisson distribution must be checked prior to its use:

• event occurrences are independent, i.e. occurrence or nonoccurrence of an event in a prior time interval has no effect on the occurrence of this event in the time interval considered,

• Two or more events cannot occur simultaneously.

Let a discrete random variable *N* represents the number of event occurrences within a prescribed time (or space) interval (0, *t*). Let *v* represent the mean occurrence rate of the event. This parameter is usually obtained from statistical data.

The Poisson PMF function is defined

$$P\left(N = n \text{ in time } t\right) = \frac{\left(vt\right)^n}{n!} e^{-vt} \qquad n = 0, 1, 2, \ldots, \infty$$

The mean and standard deviation of the random variable *N* are

$$\mu_N = vt \qquad\qquad \sigma_N = \sqrt{vt}$$

Return period (or interval) τ also depicts a Poisson variable.

It is simply the reciprocal of the mean occurrence rate *v:* $\tau = \frac{1}{v}$

Return period is a deterministic average time interval between occurrences of events. The actual time interval is random.

EXAMPLE

The average occurrence rate of earthquakes (5 to 8 magnitudes) in a given region is 2.14 earthquakes/year. Determine
(a) The return period for earthquakes in this magnitude range.
(b) The probability of exactly three earthquakes (magnitude between 5 and 8) in the next year.
(c) The annual probability of an earthquake of 5 – 8 magnitude.

Solution

(a) The return period

$$\tau = \frac{1}{v} = \frac{1}{2.14} = 0.47 \text{ year}$$

One earthquake of a given range occurs approx. every six months.

(b) The probability of three earthquakes precisely in the year after a given one is determined using with t = 1 and n = 3:

$$P\left(N = 3 \text{ in } 1 \text{ year}\right) = \frac{\left[(2.14)(1)^3\right]}{3!} e^{-(2.14)(1)} = 0.192$$

(c) We derive the annual probability of *at least one* earthquake. Therefore P (at least one earthquake) = 1 –P(no earthquakes), so

$$P\left(N \geq 1\right) = 1 - \frac{\left[(2.14)(1)\right]^0}{0!} e^{-(2.14)(1)} = 1 - e^{-(2.14)(1)} = 0.88$$