Appendix C: Probability distributions relevant to wind engineering

C1 Introduction

Probability distributions are an essential part of wind engineering as they enable the random variables involved such as wind speeds, wind directions, surface pressures and structural response (e.g deflections and stresses), to be modelled mathematically. Some of these variables are random *processes*, i.e. they have time-varying characteristics, as shown in Figure C1. The probability density describes the distribution of the magnitude or amplitude of the process, without any regard to the time axis.

The appendix will cover firstly some basic statistical definitions. Secondly, a selection of probability distributions for the complete population of a random variable – the normal (Gaussian), lognormal, Weibull, Poisson, will be considered. Thirdly, the three types of Extreme Value distributions and the closely related Generalized Pareto Distributions will be discussed.

C2 Basic definitions

C2.1 Probability density function (p.d.f.)

The probability density function (Figure C2), $f_X(x)$ is the limiting probability that the value of a random variable, X, lies between x and $(x + \delta x)$. Thus the probability that X lies between a and b is:

$$\Pr\{a < x < b\} = \int_a^b f_x(x)dx \qquad \text{(C1)}$$

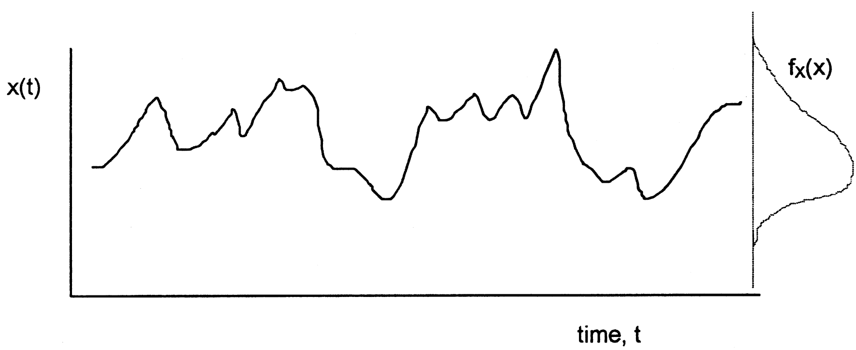

Figure C1 A random process and amplitude probability density.

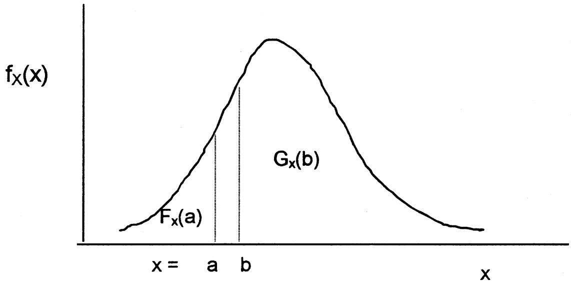

Figure C2 Probability density function and cumulative distribution functions.

Since any value of X must lie between $-\infty$ and $+\infty$:

$$\int_{-\infty}^{\infty} f_x(x)dx = \Pr\{-\infty < X < \infty\} = 1$$

Thus the area under the graph of $f_X(x)$ versus x must equal 1.0.

C2.2 Cumulative distribution function (c.d.f.)

The cumulative distribution function $F_x(x)$ is the integral between $-\infty$ and x of $f_x(x)$.

$$\text{i.e. } F_x(x) = \int_{-\infty}^{x} f_x(x)dx = \Pr\{-\infty < X < x\} = \Pr\{X < x\} \tag{C2}$$

The complementary cumulative distribution function, usually denoted by $G_X(x)$ is:

$$G_X(x) = 1 - F_x(x) = \Pr\{X > x\} \tag{C3}$$

F_x(a) and G_x(b) are equal to the areas indicated on Figure C2.
Note that:

$$f_x(x) = \frac{dF_X(x)}{dx} = -\frac{dG_X(x)}{dx} \tag{C4}$$

The following basic statistical properties of a random variable are defined and their relationship to the underlying probability distribution given.

Mean

$$\bar{X} = (1/N)\Sigma_i x_i = \int_{-\infty}^{\infty} x f_x(x)dx \tag{C5}$$

Thus the mean value is the first moment of the probability density function (i.e. the x coordinate of the centroid of the area under the graph of the p.d.f.), where N is the number of samples.

Variance

$$\sigma_x^2 = (1/N)\Sigma_i[x_i - \bar{X}]^2 \tag{C6}$$

σ_X (the square root of the variance) is called the standard deviation

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \bar{X})^2 f_x(x) dx \tag{C7}$$

Thus the variance is the second moment of the p.d.f. about the mean value. It is analogous to the second moment of area of a cross-section about a centroid.

Skewness

$$s_x = [1/(N\sigma_x^3)]\Sigma_i[x_i - \bar{X}]^3 = (1/\sigma_x^3)\int_{-\infty}^{\infty} (x - \bar{X})^3 f_x(x) dx \tag{C8}$$

The skewness is the normalised third moment of the probability density function. Positive and negative skewness are illustrated in Figure C3. A distribution that is symmetrical about the mean value has a zero skewness.

C3 Parent distributions

C3.1 Normal or Gaussian distribution

For $-\infty < X < \infty$,

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma_x}\exp\left[\frac{-(x - \bar{X})^2}{2\sigma_x^2}\right] \tag{C9}$$

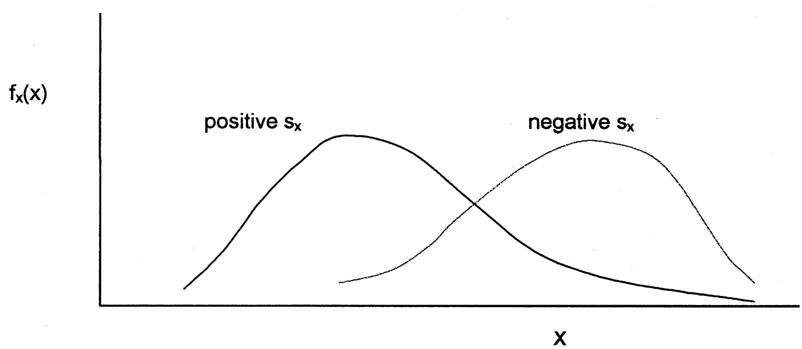

Figure C3 Positive and negative skewness.

where $\bar{X}$, σ_x are the mean and standard deviation.

This is the most commonly used distribution. It is a symmetrical distribution (zero skewness) with the familiar bell-shape (Figure C4).

$$F_x(x) = \Phi\left(\frac{x - \bar{X}}{\sigma_X}\right) \tag{C10}$$

where $\Phi(\ )$ is the cumulative distribution function of a normally distributed variable with a mean of zero and a unit standard deviation,

$$\text{i.e. } \Phi(u) = \left(\frac{1}{\sqrt{2\pi}}\right)\int_{-\infty}^{u} \exp\left(\frac{-z^2}{2}\right)dz \tag{C11}$$

Tables of Φ (u) are readily available in statistics textbooks, etc.

If $Y = X_1 + X_2 + X_3 +$ ....X_N, where X_1, X_2, X_3 ....X_N, are random variables with any distribution, the distribution of Y tends to become normal as N becomes large. If X_1, X_2, ... themselves have normal distributions, then Y has a normal distribution for any value of N.

In wind engineering, the normal distribution is used for turbulent velocity components, and for response variables (e.g. deflection) of a structure undergoing random vibration. It should be used for variables that can take both negative and positive values, so it would not be suitable for scalar wind speeds that can only be positive.

C3.2 Lognormal distribution

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left[\frac{-\left\{\log_e\left(\frac{x}{m}\right)\right\}^2}{2\sigma^2}\right] \tag{C12}$$

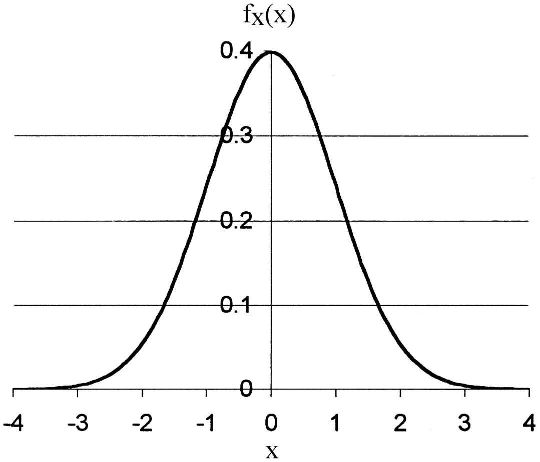

Figure C4 Normal distribution.

where the mean value $\bar{X}$ is equal to $m \exp(\sigma^2/2)$ and the variance σ_X^2 is equal to $m^2 \exp(\sigma^2)\,[\exp(\sigma^2) - 1]$. $\log_e x$ in fact has a normal distribution with a mean value of $\log_e m$ and a variance of σ^2.

If a random variable $Y = X_1 . X_2 . X_3 \ldots\ldots\ldots X_N$, where $X_1, X_2, X_3 \ldots\ldots X_N$, are random variables with any distribution, the distribution of Y tends to become lognormal as N becomes large. Thus the lognormal distribution is often used for the distribution of a variable that is itself the product of a number of uncertain variables – for example, wind speed factored by multipliers for terrain, height, shielding, topography, etc.

The lognormal distribution has a positive skewness equal to $[\exp(\sigma^2) + 2][\exp(\sigma^2) - 1]^{1/2}$.

C3.3 'Square-root-normal' distribution

Now consider the distribution of $z = x^2$, where x has the normal distribution.

$$f_Z(z) = \frac{1}{2\left(\frac{\sigma_X}{\bar{X}}\right)\sqrt{2\pi z}}\left\{\exp\left[-\left(\frac{1}{2}\right)\left(\frac{\sqrt{z}-1}{\left(\frac{\sigma_X}{\bar{X}}\right)}\right)^2\right] + \exp\left[-\left(\frac{1}{2}\right)\left(\frac{\sqrt{z}+1}{\left(\frac{\sigma_X}{\bar{X}}\right)}\right)^2\right]\right\} \tag{C13}$$

and the c.d.f. is:

$$F_Z(z) = \Phi\left(\frac{\sqrt{z}-1}{\left(\frac{\sigma_X}{\bar{X}}\right)}\right) + \Phi\left(\frac{\sqrt{z}+1}{\left(\frac{\sigma_X}{\bar{X}}\right)}\right) - 1 \tag{C14}$$

This distribution is useful for modelling the pressure fluctuations on a building which are closely related to the square of the upwind velocity fluctuations, which can be assumed to have a normal distribution (e.g. Holmes, 1981).

C3.4 Weibull distribution

$$f_X(x) = \left(\frac{kx^{k-1}}{c^k}\right)\exp\left[-\left(\frac{x}{c}\right)^k\right] \tag{C15}$$

$$F_X(x) = 1 - \exp\left[-\left(\frac{x}{c}\right)^k\right] \tag{C16}$$

where c (>0) is known as the scale parameter, with the same units as x, and k (>0) is the shape parameter (dimensionless).

The shape of the p.d.f. for the Weibull distribution is quite sensitive to the value of the shape factor, k, as shown in Figure C5. The Weibull distribution can only be used for random variables that are always positive. It is often used as the parent distribution for wind speeds, with k in the range of about 1.5 to 2.5. The Weibull distribution with $k = 2$ is a special case known as the Rayleigh distribution. When $k = 1$, it is known as the Exponential distribution.

C3.5 Poisson distribution

The previous distributions are applicable to *continuous* random variables, i.e. x can take any value over the defined range. The Poisson distribution is applicable only to positive *integer* variables, e.g. number of cars arriving at an intersection in a given time, number of exceedences of a defined pressure level at a point on a building during a windstorm.

In this case, there is no probability density function but instead a probability function:

$$p_X(x) = \lambda^x \frac{\exp(-\lambda)}{x!} \tag{C17}$$

where λ is the mean value of X. The standard deviation is $\lambda^{1/2}$.

The Poisson distribution is used quite widely in wind engineering to model exceedences or upcrossings of a random process such as wind speed, pressure or structural response, or events such as number of storms occurring at a given location. It can also be written in the form:

$$p_X(x) = (\nu T)^x \frac{\exp(-\nu T)}{x!} \tag{C18}$$

where ν is now the mean rate of occurrence per unit time, and T is the time period of interest.

C4 Extreme value distributions

In wind engineering, as in other branches of engineering, we are often concerned with the largest values of a random variable (e.g. wind speed) rather than the bulk of the population.

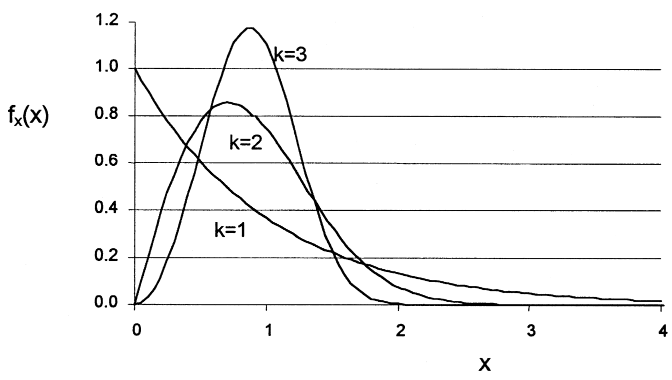

Figure C5 Probability density functions for Weibull distributions ($c = 1$).

If a variable Y is the maximum of n random variables, X_1, X_2, ...X_n and the X_i are all independent,

$$F_y(y) = F_{x1}(y).\ F_{x2}(y).........\ F_{xn}(y),$$

since $P[Y < y] = P[\text{all } n \text{ of the } X_i < y] = P[X_1 < y].\ P[X_2 < y].....\ P[X_n < y]$.

In the special case that all the X_i are identically distributed with c.d.f. $F_X(x)$,

$$F_y(y) = [F_X(x)]^n \tag{C19}$$

If the assumptions of common distribution and independence of the X_i hold, the shape of the distribution of Y is insensitive to the exact shape of the distribution of the X_i. In this case, three limiting forms of the distributions of the largest value Y, as n becomes large may be identified (Fisher and Tippett, 1928; Gumbel, 1958). However, they are all special cases of the Generalized Extreme Value Distribution.

C4.1 Generalized extreme value distribution

The c.d.f. may be written,

$$F_y(y) = \exp\left\{-\left[1 - \frac{k(y-u)}{a}\right]^{1/k}\right\} \tag{C20}$$

In this distribution, k is a shape factor, a is a scale factor, and u is a location parameter. There are thus three parameters in this generalised form.

The three special cases are:

- Type I ($k = 0$). This is also known as the Gumbel distribution.
- Type II ($k < 0$). This is also known as the Frechet distribution.
- Type III ($k > 0$). This is a form of the Weibull distribution.

The Type I can also be written in the form:

$$F_Y(y) = \exp\{-\exp[-(y-u)/a]\} \tag{C21}$$

The G.E.V. is plotted in Figure 2.1 in Chapter 2, with k equal to –0.2, 0 and 0.2 such that the Type I appears as a straight line, with a reduced variate, z, given by:

$$z = -\log_e\{-\log_e[F_Y(y)]\}$$

As can be seen the Type III (k = +0.2) curves in a way to approach a limiting value at high values of the reduced variate (low probabilities of exceedence). Thus the Type III Distribution is appropriate for phenomena that are limited in magnitude for geophysical reasons, including many applications wind engineering. The Type I can be assumed to be a conservative limiting case of the Type III, and it has only two parameters (a and u), since k is predetermined to be 0. For that reason the Type I (Gumbel distribution) is easy to fit to actual data, and is very commonly used as a model of extremes for wind speeds, wind pressures and structural response.

C4.2 Generalized Pareto distribution

The complementary cumulative distribution function is:

$$G_X(x) = \left[1 - \left(\frac{kx}{\sigma}\right)\right]^{\frac{1}{k}} \tag{C22}$$

The p.d.f. is:

$$\mathrm{f}_X(x) = \left(\frac{1}{\sigma}\right)\left[1 - \left(\frac{kx}{\sigma}\right)\right]^{\left(\frac{1}{k}\right) - 1} \tag{C23}$$

k is the shape parameter and σ is a scale parameter. The range of X is $0 < X < \infty$ when $k < 0$ or $k = 0$. When $k > 0$, $0 < X < (\sigma/k)$. Thus positive values of k only apply when there is a physical upper limit to the variate, X. The mean value of X is as follows:

$$\bar{X} = \frac{\sigma}{k + 1} \tag{C24}$$

The special case of the shape factor, k, equal to zero, results in the exponential distribution:

$$G_X(x) = \exp(-x/\sigma) \tag{C25}$$

$$F_X(x) = 1 - \exp(-x/\sigma) \tag{C26}$$

$$f_X(x) = (1/\sigma)\exp(-x/\sigma) \tag{C27}$$

The probability density functions for various values of k are shown in Figure C6.

The Generalized Pareto has a close relationship with the Generalized Extreme Value Distribution (Hosking and Wallis, 1987), so that the three types of the G.E.V. are the distributions for the largest of a group of N variables, that have a Generalized Pareto parent distribution with the same shape factor, k. It also transpires that the Generalized Pareto distribution is the appropriate one for the excesses of independent observations above a defined threshold (Davison and Smith, 1990). This distribution is used for the

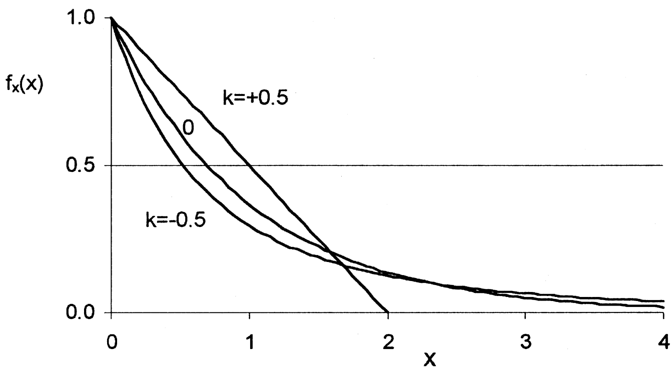

Figure C6 Probability density function for Generalized Pareto distributions ($\sigma = 1$).

excesses of maximum windspeeds in individual storms over defined thresholds (Holmes and Moriarty, 1999, Section 2.4). From the mean rate of occurrence of these storms, which are assumed to occur with a Poisson distribution, predictions can be made of wind speeds with various annual exceedence probabilities.

C5 Other probability distributions

There are many other probability distributions. The properties of the most common ones are listed by Hastings and Peacock (1974).

The general application of probability and statistics in civil and structural engineering is discussed in specialised texts by Benjamin and Cornell (1970) and Ang and Tang (1975).

References

Ang, A. H.-S. and Tang, W. H. (1975) *Probability Concepts in Engineering Planning and Design – Volume I – Basic Principles*, New York: John Wiley.

Benjamin, J. R. and Cornell, C. A. (1970) *Probability, Statistics and Decision for Civil Engineers*, New York: McGraw-Hill.

Davison, A. C. and Smith, R. I. (1990) 'Models for exceedances over high thresholds', *Journal of the Royal Statistical Society, Series B* 52: 393–442.

Fisher, R. A. and Tippett, L. H. C. (1928) 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Proc Cambridge Phil Soc* 24(2): 180–90.

Gumbel, E. J. (1958) *Statistics of Extremes*, Columbia University Press.

Hastings, N. A. J. and Peacock, J. B. (1974) *Statistical Distributions*, New York: J. Wiley.

Holmes, J. D. (1981) 'Non-gaussian characteristics of wind pressure fluctuations', *Journal of Wind Engineering and Industrial Aerodynamic*s 7: 103–8.

Holmes, J. D. and Moriarty, W. W. (1999) 'Application of the generalized Pareto distribution to extreme value analysis in wind engineering', *Journal of Wind Engineering and Industrial Aerodynamic*s 83: 1–10.

Hosking, J. R. M. and Wallis, J. R. (1987) 'Parameter and quantile estimation for the generalized Pareto distribution', *Technometrics* 29: 339–49.