Random Variables

Discrete Random Variables

A discrete random variable has a discrete sample space. For example, the outcomes of a discrete random variable may be damage states 1, 2, 3, and 4. Consider a random variable denoted by uppercase X, with outcomes, i.e., realizations, denoted by lowercase x. The probability of occurrence of each outcome of the discrete random variable is given by the probability mass function (PMF):

$$p_X(x) \equiv P(X = x) \tag{1}$$

The PMF has the following property:

$$\sum_{i=1}^{N} p_X(x_i) = 1$$
 (2)

where N is the number of possible outcomes. An alternative presentation of the probability distribution is the cumulative distribution function (CDF):

$$F_X(x) \equiv P(X \le x) = P(X \le x_i) = \sum_{j=1}^i p_X(x_j)$$
(3)

The CDF has the two properties $F(-\infty)=0$ and $F(\infty)=1$. Yet another representation of the probability distribution is the complementary CDF (CCDF):

$$G_{X}(x) = 1 - F_{X}(x) \tag{4}$$

which has the properties $G(-\infty)=1$ and $F(\infty)=0$.

Partial Descriptors

A random variable is completely defined by its probability distribution. However, "partial descriptors" are useful in lieu of having the complete distribution. The partial descriptors are equal to or related to the parameters of the probability distributions that are listed later in this document. The partial descriptors are also related to the statistical moments of the probability distribution. The first moment of the distribution is the mean of the random variable:

$$\mu_{X} = E[X] = \sum_{i=1}^{N} x_{i} \cdot p_{X}(x_{i})$$
(5)

The second moment is called the mean square of the random variable:

$$E[X^{2}] = \sum_{i=1}^{N} x_{i}^{2} \cdot p_{X}(x_{i})$$
(6)

Conversely, central moments are taken about the mean of the random variable. As a result, the first central moment is zero. The second central moment is the variance of the random variable, which is the square of the standard deviation:

$$\sigma_X^2 = \operatorname{Var}[X] = E[(x - \mu_X)^2] = \sum_{i=1}^N (x_i - \mu_X)^2 \cdot p_X(x_i)$$
(7)

Several concepts for discrete random variables, such as coefficient of variation and coefficient of skewness are the same for continuous random variables. Therefore, further details are provided in the document on continuous random variables.

The Bernoulli Distribution

Consider a discrete random variable, X, with two possible outcomes: failure and success, i.e., 0 and 1, respectively. The probability of success is denoted p. Consequently, the probability of failure is 1-p and the Bernoulli PMF is thus defined:

$$p(x) = \begin{cases} 1-p & \text{for } x=0\\ p & \text{for } x=1 \end{cases}$$
(8)

Using earlier formulas, the mean is p and the variance is p(1-p). To specify that a random variable, X, has the Bernoulli distribution, one writes: X~Bernoulli(p).

The Binomial Distribution

Consider a sequence of mutually independent Bernoulli trials with constant success probability, p. Let the random variable X denote the number of successes in n trials. The PMF for this random variable is the binomial distribution

$$p(x) = \binom{n}{x} \cdot p^{x} \cdot (1-p)^{n-x}$$
(9)

The mean of X is np and its variance is np(1-p). To specify that a random variable, X, has the binomial distribution, one writes: X~Binomial(p,n).

The Geometric Distribution

The number of trials, *S*, until success, and the number of trials between successes, in a Bernoulli sequence is given by the geometric distribution:

$$p(s) = p \cdot (1-p)^{s-1}$$
(10)

The mean recurrence time, sometimes called return period, is 1/p. The variance is $(1-p)/p^2$. To specify that a random variable, *S*, has the geometric distribution, one writes: *S*-Geometric(*p*).

The Negative Binomial Distribution

The number of Bernoulli trials, W, until k occurrences of success is

$$W = S_1 + S_2 + \dots + S_k \tag{11}$$

where S_i is the number of trials between success number *i*-1 and success number *i*. The distribution of S is geometric. Combined with the fact that Eq. (11) is a sum of random variables, the mean and variance of W is

$$\mu_w = k \cdot \mu_s = k \cdot \frac{1}{p} \tag{12}$$

$$\sigma_W^2 = k \cdot \sigma_S^2 = k \cdot \frac{1-p}{p^2} \tag{13}$$

The distribution type for W is the negative binomial distribution:

$$p(w) = \binom{w-1}{k-1} \cdot p^k \cdot (1-p)^{w-k}$$
(14)

To specify that a random variable, W, has the Bernoulli distribution, one writes: W~NegativeBinomial(p,k).

The Poisson Distribution

In situations where the number of Bernoulli trials is infinite, such as when every time instant is considered a trial, the Poisson distribution gives the number of successes, x:

$$p(x) = \frac{(\lambda \cdot T)^x}{x!} e^{-\lambda \cdot T}$$
(15)

where λ is the rate of occurrence of success per unit time and *T* is the time period under consideration. The mean number of occurrences is λT , which is also equal to the variance. To specify that a random variable, *X*, has the Poisson distribution, one writes: *X*~Poisson(λ ,*T*).

Continuous Random Variables

The sample space of a continuous random variable is the whole or part of the real continuous axis. Consider a random variable denoted by uppercase X, with outcomes, i.e., realizations, denoted by lowercase x. A continuous random variable is associated with a probability density function (PDF):

$$f_X(x) \equiv P(x \le X \le x + dx) \tag{16}$$

which has the following property:

$$\int_{-\infty}^{\infty} f_X(x) dx = 1$$
(17)

The CDF for a continuous random variable is

$$F_X(x) \equiv P(X \le x) = \int_{-\infty}^x f_X(x) dx$$
(18)

which has the properties $F(-\infty)=0$ and $F(\infty)=1$. The PDF can be computed from the CDF by differentiation:

$$f(x) = \frac{dF(x)}{dx} \tag{19}$$

Another representation of the probability distribution of X is the complementary CDF (CCDF):

$$G_X(x) = 1 - F_X(x)$$
 (20)

Partial Descriptors

A random variable is completely defined by its probability distribution. However, "partial descriptors" are useful in lieu of having the complete distribution. The partial descriptors are equal to or related to the parameters of the probability distributions that are listed later in this document. The partial descriptors are also related to the statistical moments of the probability distribution. The first moment of the distribution is the mean of the random variable:

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) \, dx \tag{21}$$

In passing, it is noted that the mean can also be calculated as the area underneath the CCDF. Because the PDF is obtained by differentiation of the CCDF, with a negative sign in front, Eq. (21) turns into:

$$\mathbf{E}[X] = -\int_{0}^{\infty} x \cdot \frac{G(x)}{dx} dx$$

Integration by parts yields:

$$\mathbf{E}[X] = -\int_{0}^{\infty} x \cdot \frac{G(x)}{dx} dx = -\left[x \cdot G(x)\right]_{0}^{\infty} + \int_{0}^{\infty} 1 \cdot G(x) dx$$

The boundary term vanishes, hence:

$$\mathrm{E}[X] = \int_{0}^{\infty} G(x) dx$$

The second moment is called the mean square of the random variable:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x) \, dx \tag{22}$$

Conversely, central moments are taken about the mean of the random variable. As a result, the first central moment is zero. The second central moment is the variance of the random variable, which is the square of the standard deviation:

$$\sigma_X^2 = \operatorname{Var}[X] = E[(x - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f_X(x) \, dx$$
(23)

By expanding Eq. (23) one finds that "the variance is equal to the mean square minus the square of the means:"

$$\sigma_X^2 = E[X^2] - \mu_X^2 \tag{24}$$

The coefficient of variation of a random variable is defined as:

$$\delta_{X} = \frac{\sigma_{X}}{|\mu_{X}|} \tag{25}$$

The coefficient of skewness of a random variable is related to the third central moment as follows:

$$\gamma_x = \frac{E[(x - \mu_x)^3]}{\sigma_x^3} \tag{26}$$

The coefficient of Kurtosis provides a measure of the flatness of the distribution and is related to the fourth central moment as follows:



$$\kappa_{\chi} = \frac{E[(x - \mu_{\chi})^4]}{\sigma_{\chi}^4}$$
(27)

Figure 1: Plots of selected continuous PDFs from the GNU Scientific Library Reference.

The Normal Distribution

The lines in Figure 1 display several PDFs for a random variable with different probability distributions. The normal distribution is one of them. Among many other applications, the normal distribution emerges in situations where the random variable is a sum of many underlying and independent variables. The central limit theorem states that under such circumstances the probability distribution of the sum approaches the normal distribution, sometimes called the Gaussian distribution. The normal distribution is a two-parameter distribution in which the two parameters directly represent the mean and standard deviation. To specify that a random variable, X, has the normal distribution, one writes: $X \sim N(\mu, \sigma)$. The implementation in **R**t employs the following notation:

PDF	$f(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2\right)$
mean=	μ
stdv=	σ
μ=	mean
σ=	stdv

The Lognormal Distribution

In the same way as the normal distribution arise from summation of random phenomena the lognormal distribution emerge in products of variables. In fact, the product of lognormal random variables is also lognormally distributed. For pedagogical purposes, consider the product of n independent random variables:

$$X = Z_1 \cdot Z_2 \cdots Z_n \tag{28}$$

Taking the natural logarithm on both sides yields

$$\ln(X) = \ln(Z_1) + \ln(Z_2) + \dots + \ln(Z_n)$$
(29)

In accordance with the central limit theorem, ln(X) approaches the normal distribution. By definition, that means that X has the lognormal distribution. In other words, if Y is a normal random variable and

$$Y = \ln(X) \tag{30}$$

then X is said to have the lognormal distribution. To establish the lognormal distribution it is possible to use the normal distribution as a starting point. To that end, consider the probability transformation between Y and X, given the functional relationship between the two variables in Eq. (30):

$$F_{X}(x) = F_{Y}(y) = F_{Y}(\ln(x))$$
 (31)

$$f_X(x) \cdot dx = f_Y(y) \cdot dy \implies f_X(x) = \frac{dy}{dx} \cdot f_Y(y) = \frac{1}{x} \cdot f_Y(\ln(x))$$
 (32)

where f_Y and F_Y are the normal PDF and CDF, respectively. By employing the standard normal distribution, Eqs. (31) and (32) turn into

$$f(x) = \frac{1}{x} \cdot \varphi\left(\frac{\ln(x) - \mu_{Y}}{\sigma_{Y}}\right)$$
(33)

$$F(x) = \Phi\left(\frac{\ln(x) - \mu_{Y}}{\sigma_{Y}}\right)$$
(34)

and solving for x in Eq. (34) yields the inverse lognormal CDF in terms of the inverse normal CDF:

$$x = e^{\Phi^{-1}(p)\sigma_Y + \mu_Y} \tag{35}$$

The remaining question is how to evaluate the parameters μ_Y and σ_Y , which are the mean and standard deviation of the normal random variable $Y=\ln(X)$, not the lognormal random variable X. Several options are possible. One is having μ_X and σ_X . It can then be shown that the sought parameters are

$$\mu_{Y} = \ln(\mu_{X}) - \frac{1}{2} \cdot \ln\left(1 + \left(\frac{\sigma_{X}}{\mu_{X}}\right)^{2}\right)$$
(36)

$$\sigma_{Y} = \sqrt{\ln\left(\left(\frac{\sigma_{X}}{\mu_{X}}\right)^{2} + 1\right)}$$
(37)

Another option is having the median and "dispersion" of X. To understand this, consider first the median of the normal random variable Y, denoted m_Y , which for the normal distribution equals the mean, μ_Y . Because of Eq. (30), it is clear that the sought parameter μ_Y is

$$\mu_Y = \ln(m_X) \tag{38}$$

As an aside note, this implies that the term $\ln(x)-\mu_Y$ in the argument of the distributions above can be written as

$$\ln(x) - \mu_Y = \ln(x) - \ln(m_X) = \ln\left(\frac{x}{m_X}\right)$$
(39)

The so-called dispersion is merely another name for σ_Y . In *Rt* the following symbols are used for μ_Y and σ_Y :

$$\mu_{Y} = \zeta \tag{40}$$

$$\sigma_{Y} = \sigma$$

As a result, to specify that a random variable, X, has the lognormal distribution, one writes: $X \sim LN(\zeta, \sigma)$. The implementation in *Rt* is:

PDF	$f(x,\zeta,\sigma) = \frac{1}{x \cdot \sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{\ln(x) - \zeta}{\sigma}\right)^2\right)$
mean=µ _X =	$\exp\left(\zeta + \frac{\sigma^2}{2}\right)$
stdv= σ_X =	$\sqrt{\exp(\sigma^2)-1} \cdot \exp\left(\zeta + \frac{\sigma^2}{2}\right)$
$\zeta = \mu_{\rm Y} = \ln(m_{\rm X}) =$	$\ln(mean) - \frac{1}{2} \cdot \ln\left(1 + \left(\frac{stdv}{mean}\right)^2\right)$
$\sigma = \sigma_{\rm Y}$ =dispersion=	$\sqrt{\ln\!\left(\left(\frac{stdv}{mean}\right)^2 + 1\right)}$

The Uniform Distribution

The uniform distribution is often thought to represent situations with little knowledge. This lack of information is expressed by distributing the probability density uniformly over a range of outcomes. To specify that a random variable, X, has the uniform distribution, one writes: $X \sim U(a,b)$. In **R**t, the implementation is:

	Rt
PDF	$f(x,a,b) = \frac{1}{b-a}$
mean=	$\frac{a+b}{2}$
stdv=	$\frac{b-a}{\sqrt{12}}$
a=	mean – stdv $\cdot \sqrt{3}$
b=	$mean + stdv \cdot \sqrt{3}$

The Exponential Distribution

Among other applications, the exponential distribution emerges for the time between occurrences, T, which is equal to the time until the first occurrence, in the Poisson process. It is a one-parameter distribution with PDF

$$f(t) = \lambda \cdot \exp(-\lambda \cdot t) \tag{41}$$

with mean and standard deviation $1/\lambda$. To specify that a random variable, X, has the exponential distribution, one writes: $X \sim \text{Exp}(\lambda)$. However, in **R**t the exponential distribution is implemented as a two-parameter distribution that includes a shift

parameter and a different notation, as shown in the table below. Therefore, another way to specify that a random variable, X, has the exponential distribution is $X \sim \text{Exp}(\mu, x_0)$.

	Rt
PDF	$f(x,\mu,x_0) = \frac{1}{\mu} \cdot \exp\left(-\frac{1}{\mu} \cdot \left(x - x_0\right)\right)$
mean=	μ + x_0
stdv=	μ
μ=	stdv
$x_0 =$	mean-µ

The Gamma Distribution

Among other applications, the gamma distribution models the time until occurrence number x in a Poisson process. Thus, it represents the sum of independent exponentially distributed random variables and contains the exponential distribution as a special case. The PDF of this distribution includes the gamma-function, Γ . Like the lognormal distribution, the gamma distribution is limited to positive realizations. To specify that a random variable, X, has this distribution, one writes: X~Gamma(a,b). The implementation in **Rt** is:

	Rt
PDF	$f(x,a,b) = \frac{1}{\Gamma(a) \cdot b^a} \cdot x^{a-1} \cdot \exp\left(-\frac{x}{b}\right)$
mean=	$a \cdot b$
stdv=	$\sqrt{a} \cdot b$
<i>a</i> =	$\left(\frac{mean}{stdv}\right)^2$
<i>b</i> =	$\frac{stdv^2}{mean}$

Viewed as a waiting time, x, in a Poisson process, and in several other situations, the gamma distribution is often written as

$$f(x) = \frac{v(vx)^{k-1}e^{-vx}}{\Gamma(k)}$$
(42)

where a and b are related to the parameters v and k as follows: a=k and b=1/v. Then the mean is k/v and the standard deviation is $\sqrt{k/v}$. Using the distribution with v and k, the postulation that a random variable, X, has the Gamma distribution can also be written

X~Gamma(v,k). Because of the different parameterization options for the Gamma distribution it is necessary to apply great caution in the specification of which form of the distribution is intended.

The Beta Distribution

The original beta distribution is defined in the interval 0 to 1, i.e., it is a bounded distribution. To specify that a random variable, X, has this basic beta distribution, one writes: X~Beta(a,b). The following table provides the information for this basic zero-to-one distribution.

PDF	$f(x,a,b) = \frac{\Gamma(a+b)}{\Gamma(a) \cdot \Gamma(b)} \cdot x^{a-1} \cdot (1-x)^{b-1}$
mean=	$\frac{a}{a+b}$
stdv=	$\frac{1}{a+b} \cdot \sqrt{\frac{a \cdot b}{a+b+1}}$

A more versatile beta distribution is obtained by letting the user specify the interval. Instead of being defined in the interval 0 to 1, the beta distribution that is available in **Rt** is defined in the interval *min* to *max*. This yields a particularly handy distribution, but the versatility comes at the cost of having to specify four distribution parameters. To specify that a random variable, X, has this full beta distribution, one writes: X~Beta(a,b,min,max). Because there are four distribution parameters, the distribution parameters cannot be determined uniquely from the mean and standard deviation. The implementation in **Rt** is:

	Rt
PDF	$f(x,a,b,\min,\max) = \frac{1}{\max - \min} \cdot \frac{\Gamma(a+b)}{\Gamma(a) \cdot \Gamma(b)} \cdot \left(\frac{x-\min}{\max - \min}\right)^{a-1} \cdot \left(1 - \left(\frac{x-\min}{\max - \min}\right)\right)^{b-1}$
mean=	$min + \frac{a}{a+b} \cdot (max - min)$
stdv=	$\frac{1}{a+b} \cdot \sqrt{\frac{a \cdot b}{a+b+1}} \cdot (max - min)$

The Laplace Distribution

To specify that a random variable, X, has the Laplace distribution, one writes: $X \sim \text{Laplace}(a)$ or $X \sim \text{Laplace}(a, x_0)$. Implementation in **R**t:

	Rt
PDF	$f(x,a,x_0) = \frac{1}{2a} \cdot \exp\left(-\left \frac{x-x_0}{a}\right \right)$
mean=	<i>x</i> ₀
stdv=	$\sqrt{2} \cdot a$
a=	$\frac{stdv}{\sqrt{2}}$
$x_0 =$	mean

The Chi-square Distribution

Let Y be a collection of ν normally distributed random variables with unit variances. The sum of the squares of these random variables, i.e.,

$$X = \sum_{i=1}^{\nu} Y_i^2$$
 (43)

is a random variable that has the chi-squared distribution with v degrees of freedom. The implementation in Rt is:

	Rt
PDF	$f(x, v, x_0) = \frac{1}{2 \cdot \Gamma(v/2)} \cdot \left(\frac{x - x_0}{2}\right)^{\frac{v}{2} - 1} \cdot \exp\left(\frac{x - x_0}{2}\right)$
mean=	$v + x_0$
stdv=	$\sqrt{2v}$
V=	$\frac{stdv^2}{2}$
$x_0 =$	mean – v

The t-distribution Distribution

Let Y_1 be a normally distributed random variable and let Y_2 be a chi-squared random variable with $x_0=0$ and ν degrees of freedom. Then the ratio

$$X = \frac{Y_1}{\sqrt{\frac{Y_2}{v}}} \tag{44}$$

	Rt
PDF	$f(x, v, x_0) = \frac{1}{2 \cdot \Gamma(v/2)} \cdot \left(\frac{x - x_0}{2}\right)^{\frac{v}{2} - 1} \cdot \exp\left(\frac{x - x_0}{2}\right)$
mean=	<i>x</i> ₀
stdv=	$\sqrt{\frac{v}{v-2}}$ for $v \ge 2$
V=	$\frac{2 \cdot stdv^2}{stdv^2 - 1}$
$x_0 =$	mean

is a random variable that has the t-distribution. The implementation in *Rt* is:

The Logistic Distribution

Implementation in *Rt*:

	Rt
PDF	$f(x,a,x_0) = \frac{\exp\left(-\frac{x-x_0}{a}\right)}{a \cdot \left(1 + \exp\left(-\frac{x-x_0}{a}\right)\right)^2}$
mean=	$a + x_0$
stdv=	$a \cdot \pi/\sqrt{3}$
a=	$\frac{stdv\cdot\sqrt{3}}{\pi}$
$x_0 =$	mean – a

The Rayleigh Distribution

The original Rayleigh distribution is a one-parameter distribution that is defined only for positive values of the random variable, *X*, with PDF:

$$f(x) = 2 \cdot \lambda \cdot x \cdot e^{-\lambda x^2} \tag{45}$$

Sometimes a parameter conversion from λ to σ is employed, where $\lambda = 1/2\sigma^2$ so that the PDF is written

$$f(x) = \frac{x}{\sigma^2} \cdot e^{-\frac{x^2}{2\sigma^2}}$$
(46)

in which case a random variable, X, with the Rayleigh distribution is written X-Rayleigh(σ). Because of the different parameterization options for the Gamma distribution it is necessary to apply great caution in the specification of which form of the distribution is intended. Furthermore, the implementation in **R**t involves a shift parameter, which means that a random variable, X, with the Rayleigh distribution is written X-Rayleigh(σ , x_0):

	Rt
PDF	$f(x,\sigma,x_0) = \frac{x}{\sigma^2} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{x-x_0}{\sigma}\right)^2\right)$
mean=	$x_0 + \sigma \cdot \sqrt{\frac{\pi}{2}}$
stdv=	$\sigma \cdot \sqrt{\frac{4-\pi}{2}}$
σ=	$\frac{stdv}{\sqrt{\frac{4-\pi}{2}}}$
x ₀ =	$mean - \sigma \cdot \sqrt{\frac{\pi}{2}}$

The Gumbel Distribution

Although another document addresses extreme value distributions and power law models the implementation of the Gumbel and Weibull distributions in Rt are documented here, starting with the Gumbel distribution:

	Rt
PDF	$f(x,a,b) = a \cdot b \cdot \exp(-(b \cdot \exp(-a \cdot x) + a \cdot x))$
mean=	$\frac{\ln(b) + \gamma}{a}$
stdv=	$\frac{\pi}{a \cdot \sqrt{6}}$
a=	$\frac{\pi}{stdv\cdot\sqrt{6}}$
<i>b</i> =	$\exp(a \cdot mean - \gamma)$

where γ is the Euler constant, which is approximately equal to 0.577215665.

The Weibull Distribution

Implementation in *Rt*:



Extreme Value Models

This section addresses continuous random variables that represent the maximum of several outcomes of another continuous random variable. For the sake of the following derivations, consider a continuous random variable Z. In any one realization the probability that the outcome is less than z is the CDF, $F_Z(z)$. Now consider a situation where there are n independent realizations of the random variable. In this situation it is often the maximum realization that is of interest, perhaps because it represents a loading value that may exceed some capacity. To address the probability that the maximum realization will not exceed a threshold, x, first define X as a new random variable, where

$$X = \max(Z_1, Z_2, \cdots, Z_n) \tag{47}$$

Under the condition that the random variables Z_i are independent and identically distributed the probability that the maximum X does not exceed the threshold x is

$$F_{X}(x) = P(X \le x)$$

$$= P(Z_{1} \le x \cap Z_{2} \le x \cap \dots \cap Z_{n} \le x)$$

$$= P(Z_{1} \le x) \cdot P(Z_{2} \le x) \cdots P(Z_{n} \le x)$$

$$= F_{Z}(x)^{n}$$
(48)

where it is emphasized that F_Z is the CDF for the original random variable Z. This is the CDF for the extreme value of Z in n "experiments." The corresponding PDF is obtained by differentiation:

$$f_X(x) = \frac{d}{dx} F_X(x)$$

= $\frac{d}{dx} F_Z(x)^n$ (49)
= $n \cdot F_Z(x)^{n-1} \cdot f_Z(x)$

where it is reiterated that it is the CDF and PDF of the original random variable Z that enter the expression.

When the number of experiments is large then the distribution for the extreme value, X, is mostly dependent on the tail behaviour of the underlying probability distribution for Z. It is rather insensitive to the overall behaviour of the actual underlying probability distribution. For these situations several asymptotic extreme-value distributions are developed. They cannot be synthesized into one distribution because the result is different for minimum and maximum values. Furthermore, the result is different for different types of tail-behaviour in the underlying probability distribution.

Type I Distributions (Gumbel)

This distribution addresses the maximum value of many experiments. The "Type I" assumption is that the tail of the underlying distribution varies exponentially:

$$F_{z}(z) = 1 - \exp(-h(z)) \tag{50}$$

The tails are unbounded. This type of tail is found in the normal, exponential, and gamma distributions. Application of this underlying tail distribution in extreme value theory yields the Type I Largest and Type I Smallest distributions, for the maximum and minimum of many realizations, respectively. The resulting distributions are named after Gumbel:

$$F(x) = \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right)$$
(51)

Type II Distribution (Frechet)

Here the left tail of the underlying distribution is bounded at zero, while the upper tail varies according to

$$F_{Z}(z) = 1 - c_{1} \cdot \left(\frac{1}{z}\right)^{c_{2}}$$
(52)

where c_1 and c_2 are constants. The derived Type II extreme value distribution is named after Frechet:

$$F(x) = \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}\right) \quad x \ge \mu$$
(53)

Type III Distribution (Weibull)

The fundamental assumption for Type III distributions is that the tail of the underlying random variable is bounded by a value z_0 :

$$F_{Z}(z) = 1 - c_{1} \cdot (z_{0} - z)^{c_{2}}$$
(54)

This results in the extreme value distributions named after Weibull. A simple sign-change in Frechet's distribution yields the reversed Type III Weibull distribution:

$$F(x) = \exp\left(-\left(-\frac{\mu - x}{\sigma}\right)^{\alpha}\right) \qquad x < \mu \tag{55}$$

The Weibull distribution is written:

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^k\right) \quad \text{for } x \ge 0 \tag{56}$$

$$f(x) = \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda}\right)^{k-1} \cdot \exp\left(-\left(\frac{x}{\lambda}\right)^k\right) \quad \text{for } x \ge 0$$
(57)

where k > 0 is the shape parameter and $\lambda > 0$ is the scale parameter.

Generalized Extreme Value Distribution

$$F(x) = \exp\left(-\left(1 + \xi \cdot \left(\frac{x - \mu}{\sigma}\right)\right)^{-1/\xi}\right)$$
(58)

where μ is the location parameter, σ is the scale parameter, and ξ is the shape parameter.

Theorem of Total Probability

Recall that the rule of total probability is applicable only when the conditioning is on mutually exclusive and collectively exhaustive events. The rule of total probability to obtain the probability of an event, having probability values conditioned upon the outcomes of a continuous random variable:

$$P(A) = \int_{-\infty}^{\infty} P(A \mid x) \cdot f(x) dx$$
(59)

The rule of total probability to obtain a continuous probability distribution, having the distribution conditioned on some events is:

$$f(x) = \sum_{i=1}^{N} f(x \mid E_i) \cdot P(E_i)$$
(60)

where N is the number of mutually exclusive and collectively exhaustive events, which could be the outcomes of a discrete random variables. The rule of total probability to obtain a probability distribution, having the distribution conditioned on the outcomes of another continuous random variable is:

$$f(x) = \int_{-\infty}^{\infty} f(x \mid y) \cdot f(y) dy$$
(61)

For discrete random variables, the rule of total probability to obtain the probability of an event, having probability values conditioned upon the outcomes of the random variable:

$$\mathbf{P}(A) = \sum_{i=1}^{N} \mathbf{P}(A \mid x_i) \cdot p(x_i)$$
(62)

The rule of total probability to obtain a probability distribution, having the distribution conditioned on the outcomes of another random variable or some other discrete events:

Random Variables

terje.civil.ubc.ca

$$p_X(x) = \sum_{i=1}^{N} p_X(x \mid y_i) \cdot p_Y(y_i)$$
(63)

Joint and Conditional Distributions

While the probability distribution for an individual random variable is called "marginal," the probability distribution for multiple random variables is called a "multivariate" or "joint" distribution. The joint PDF of two continuous random variables *X* and *Y* is defined as

$$f(x,y) \cdot dx \cdot dy = \mathbf{P}\left(x < X \le x + dx \cap y < Y \le y + dy\right)$$
(64)

The joint PDF has the following properties:

$$f(x,y) \ge 0$$

$$f(y) = \int_{-\infty}^{\infty} f(x,y) dx$$

$$f(x) = \int_{-\infty}^{\infty} f(x,y) dy$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = 1$$

(65)

The relationship between the joint PDF and the joint CDF is

$$F(x,y) = P(X \le x \cap Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y) dx dy$$
(66)

which implies that

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$
(67)

Having the joint distribution, conditional distributions are defined in accordance with the conditional probability rule for events:

$$f(x \mid y) \cdot dx = \frac{f(x, y) \cdot dx \cdot dy}{f(y) \cdot dy} \Longrightarrow f(x \mid y) = \frac{f(x, y)}{f(y)}$$
(68)

As a result, the joint distribution can be expressed in terms of conditional distributions as

$$f(x,y) = f(x \mid y) \cdot f(y) = f(y \mid x) \cdot f(x)$$
(69)

Two random variables, X and Y, are said to be statistically independent if

$$f(x | y) = f(x)$$
 or $f(y | x) = f(y)$ (70)

Statistical independence implies that the joint distribution for two statistically independent random variables, *X* and *Y*, is the product of the marginals:

$$f(x,y) = f(y) \cdot f(x) \tag{71}$$

Partial Descriptors

In the context of joint distributions, partial descriptors include the mean product:

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot f(x,y) dx dy$$
(72)

and the covariance:

$$Cov[X,Y] = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X) \cdot (y - \mu_Y) \cdot f(x,y) dx dy$$
(73)

Expansion of the integrand in Eq. (73) reveals that the covariance is equal to the mean product minus the product of the means:

$$\operatorname{Cov}[X,Y] = \operatorname{E}[XY] - \mu_X \mu_Y \tag{74}$$

This echoes the fact that the variance of a marginal distribution equals the mean square minus the square of the means. As discussed later in this document, the covariance between two random variables is a measure of linear dependence between them. A normalized, i.e., dimensionless measure of this linear dependence is the correlation coefficient, which is defined as:

$$\rho_{XY} = \frac{\operatorname{Cov}[X,Y]}{\sigma_X \sigma_Y} = \frac{\operatorname{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = \operatorname{E}\left[\frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_Y}\right]$$
(75)

In the same way as relative frequency diagrams are instructive visualizations of the realizations of a single random variable, a scatter diagram is valuable when two random variables are observed simultaneously. The scatter diagram visualizes the outcomes of one variable along one axis versus the outcomes of the other variable along the other axis. The plot gives a sense of the dependence between the two variables. Statistical dependence between random variables may take different forms. For example, one form of dependence is that one variable varies exponentially with the other. Yet another example is linear dependence, in which the realizations of one random variable tend to be proportional to the outcomes of another random variable. Correlation, defined in Eq. (75), measures linear dependence. In other words, two random variables can be uncorrelated but statistically dependent. It is also emphasized that when statistical dependence is specified by means of correlation then the possibility of a non-positive definite correlation matrix is present. In reliability analysis, this prevents the transformation into standard normal random variables. As a result, some correlation structures are impractical and/or unphysical. Importantly, the range of possible correlation depends upon the marginal probability distributions of the random variables. Hence, in reliability analysis applications, the specification of correlation must be made with care and with knowledge about the marginal probability distributions.

Matrix Notation

When dealing with second-moment information, i.e., mean, variance, and correlation of multiple random variables it is convenient to use matrix notation. As an illustration, let **X**

be a vector of random variables, while \mathbf{x} is the vector of realizations of \mathbf{X} . Then, the vector of means is

$$\mathbf{M}_{\mathbf{X}} = \left\{ \begin{array}{c} \boldsymbol{\mu}_{1} \\ \boldsymbol{\mu}_{2} \\ \vdots \\ \boldsymbol{\mu}_{n} \end{array} \right\}$$
(76)

where m_i is the mean of random variable number *i*. The covariance matrix is a symmetric matrix that contains the variances of the random variables, and the covariance between them:

$$\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} = \begin{bmatrix} \boldsymbol{\sigma}_{1}^{2} & \boldsymbol{\rho}_{12} \cdot \boldsymbol{\sigma}_{1} \cdot \boldsymbol{\sigma}_{2} & \boldsymbol{\rho}_{13} \cdot \boldsymbol{\sigma}_{1} \cdot \boldsymbol{\sigma}_{3} \\ \boldsymbol{\rho}_{12} \cdot \boldsymbol{\sigma}_{1} \cdot \boldsymbol{\sigma}_{2} & \boldsymbol{\sigma}_{2}^{2} & \boldsymbol{\rho}_{23} \cdot \boldsymbol{\sigma}_{2} \cdot \boldsymbol{\sigma}_{3} \\ \boldsymbol{\rho}_{13} \cdot \boldsymbol{\sigma}_{1} \cdot \boldsymbol{\sigma}_{3} & \boldsymbol{\rho}_{23} \cdot \boldsymbol{\sigma}_{2} \cdot \boldsymbol{\sigma}_{3} & \boldsymbol{\sigma}_{3}^{2} \end{bmatrix}$$
(77)

By defining the matrix D_{XX} to be a square matrix with the standard deviations on the diagonal the covariance matrix is written as the decomposition

$$\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} = \mathbf{D}_{\mathbf{X}\mathbf{X}}\mathbf{R}_{\mathbf{X}\mathbf{X}}\mathbf{D}_{\mathbf{X}\mathbf{X}}$$
(78)

where $\mathbf{R}_{\mathbf{X}\mathbf{X}}$ is the correlation matrix, which is also symmetric:

$$\mathbf{R}_{\mathbf{X}\mathbf{X}} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$
(79)

The Joint Normal Distribution

Unlike the situation for univariate distributions, only a few standard multivariate distribution types are encountered. By far the most common is the joint Normal distribution. The joint normal PDF is

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n} \cdot \sqrt{\det(\mathbf{\Sigma}_{XX})}} \cdot \exp\left(\frac{1}{2} (\mathbf{x} - \mathbf{M}_X)^T \mathbf{\Sigma}_{XX}^{-1} (\mathbf{x} - \mathbf{M}_X)\right)$$
(80)

where n is the number of random variables. In the bi-variate case it reads

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{z}{2(1-\rho)^2}\right)$$
(81)

where

$$z = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2}$$
(82)

A special case is the standard normal distribution, which is characterized by zero means, unit variances, and zero covariances. This PDF is denoted by the symbol φ and takes the form

$$\varphi(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^n}} \cdot \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{y}\right)$$
(83)

This multivariate distribution has several properties that are important in reliability analysis:

- 1. The multivariate standard normal PDF is rotationally symmetric and it decays exponentially in the radial and tangential directions
- 2. The probability content outside a hyper-plane distanced β from the point y=0 is:

$$p = \Phi(-\beta) \tag{84}$$

which is employed in the document on FORM.

3. The probability content outside a hyper- paraboloid with apex distanced β from the point y=0 is also available, as described in the document on SORM.

Copulas

Copulas represent an alternative technique for specifying statistical dependence between random variables. Currently, its use is more widespread in economics than engineering, but that may change. Copulas extend the options for prescribing statistical dependence beyond the use of the correlation coefficient, which only provides linear statistical dependence. The correlation coefficient is convenient and popular for a few reasons. First, it appears prominently in second-moment theory, together with means and standard deviations. Second, the correlation coefficient appears as a parameter in the powerful joint normal probability distribution, as described earlier in this document. However, the convenience of the correlation coefficient diminishes in the general of circumstances. Consider the example when the joint distribution is sought for a set of random variables with mixed marginal distributions and perhaps nonlinear dependence tendencies. This problem is important in reliability analysis where the Nataf or Rosenblatt transformations are usually applied. Under such circumstances the copulas represent an alternative, although it has yet to become popular in reliability analysis. The key feature of the copula technique is that a variety of dependence structures are possible. One example is stronger dependence in the distribution tails. An interesting class of copulas is the generalized elliptical distributions that are generalizations of the joint normal distribution. The joint normal distribution is also elliptical, but it is a special case of the "infinite" possibilities provided by copulas.

From a philosophical viewpoint, the need to specify statistical dependence between random variables is, in some sense, a symptom of imperfect models. The source of correlation is due to hidden phenomena behind the random variables. If the underlying phenomena were modelled then the need to prescribe statistical dependence might vanish. Consider the example of prescribing correlation between the earthquake intensity at two nearby sites. The need to estimate this correlation disappears if the modelling is expanded to include the hypocentre location, the earthquake magnitude, and the attenuation of the intensity to each site. It is those underlying phenomena that cause correlation in intensity between sites. This philosophical discussion is somewhat akin to the discussion on whether aleatory uncertainty exists. It does, unless all models are perfect, which they are not. However, this paragraph is intended to foster a strong focus on modelling and careful examination of the need to prescribe statistical dependence.

Sklar's Theorem

Sklar's theorem is the foundation for the use of copulas. It states that the joint CDF of some random variables, \mathbf{X} , can be written in terms of a copula, C, which is a function of the marginal CDFs of the random variables:

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$$
(85)

That is, the joint distribution is composed of the marginal distributions and the copula function. In other words, the copula is a function that couples the marginal distribution functions. This is the means by which dependence is introduced. It is also observed that copulas express dependence on a "quantile scale," namely along the random variable axes. In this manner, the dependence at 10% probability of exceedance can be different from the dependence at 90% probability of exceedance. Several other interpretations of Eq. (85) are possible. First, it is observed that a copula is what remains of a joint cumulative distribution once the action of the marginal cumulative distribution functions has been removed. In other words, the marginals provide the probability distributions, while the sole purpose of the copula is to provide statistical dependence. Furthermore, Sklar's theorem can be written

$$C(p_1, p_2, \dots, p_n) = F(F_1^{-1}(p_1), F_2^{-1}(p_2), \dots, F_n^{-1}(p_n))$$
(86)

where p_i are probabilities. This form of Eq. (85) is used to "extract" copulas from existing joint distributions, as described shortly. It is noted that a copula is invariant with respect to strictly increasing transformations of the random variables, such as that of transforming random variables from normal to standard normal.

Explicit and Implicit Copulas

The simplest example of a copula is the one that yields no dependence at all. That is, the copula for independent random variables is:

$$F(x_1, x_2, \dots, x_n) = F_1(x_1) \cdot F_2(x_2) \cdots F_n(x_n)$$
(87)

This expression, which corresponds to the definition of statistical independence between random variables, is an example of an explicit copula. Copulas are either implicit or explicit. Implicit copulas are extracted from known joint distributions. For example, the Gauss copula is extracted from the joint normal probability distribution. Specifically, from Sklar's theorem in Eq. (85) it is understood that when the random variables have the joint CDF F then the copula C is the CDF of the marginal distributions. This is what is emphasized in Eq. (86). Consider two correlated normal random variables, here standard normal for simplicity:

$$F(x_1, x_2) = \Phi(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{s_1^2 + s_2^2 - 2\rho s_1 s_2}{2(1-\rho^2)}\right) ds_1 ds_2$$
(88)

The Gauss copula is extracted from this joint CDF by substituting the random variables in the original distribution with the marginal CDFs:

$$C(p_1, p_2) = \Phi(x_1, x_2) = \int_{-\infty}^{\Phi^{-1}(p_2)} \int_{-\infty}^{\Phi^{-1}(p_1)} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{s_1^2 + s_2^2 - 2\rho s_1 s_2}{2(1-\rho^2)}\right) ds_1 ds_2 \quad (89)$$

Thus, the normal copula is extracted.

Common Copulas

The following is a set of common copulas, where the notation, p=F(x) is employed for convenience:

Independent:

$$C(p_1, p_2) = p_1 \cdot p_2$$
(90)

Normal:
$$C(p_1, p_2) = \int_{-\infty}^{\Phi^{-1}(p_2)} \int_{-\infty}^{\Phi^{-1}(p_1)} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{s_1^2 + s_2^2 - 2\rho s_1 s_2}{2(1-\rho^2)}\right) ds_1 ds_2$$
 (91)

Student:
$$C(p_1, p_2) = \int_{-\infty}^{T_v^{-1}(p_2)} \int_{-\infty}^{T_v^{-1}(p_1)} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{s_1^2 + s_2^2 - 2\rho s_1 s_2}{2(1-\rho^2)}\right)^{-\left(\frac{\nu+2}{2}\right)} ds_1 ds_2$$
 (92)

$$C(p_1, p_2) = -\frac{1}{\theta} \cdot \ln\left(1 + \frac{\left(e^{-\theta \cdot p_1} - 1\right) \cdot \left(e^{-\theta \cdot p_2} - 1\right)}{e^{-\theta} - 1}\right)$$
(93)

Clayton:

Frank:

$$C(p_1, p_2) = \left(p_1^{-\theta} + p_2^{-\theta} - 1\right)^{-\frac{1}{\theta}}$$
(94)

Gumbel:

$$C(p_{1}, p_{2}) = \exp\left(-\left(\left(-\ln(u_{1})\right)^{\theta} + \left(-\ln(u_{2})\right)^{\theta}\right)^{\frac{1}{\theta}}\right)$$
(95)

Meta Distributions

A potentially interesting aspect of the use of copulas is the possibility of creating entirely new "meta" distributions. This is achieved by first employing Eq. (86) to extract an implicit copula, followed by utilization of Eq. (85) to substitute "arbitrary" CDFs into the copula. Clearly, a great number of possible joint probability distributions—perhaps more or less useful—then become available. To generate realization of random variables from meta distributions the following sampling procedure may be helpful:

- 1. Generate outcomes of some random variables \mathbf{x} from the fundamental distribution, say the normal
- 2. Obtain the marginal CDF value for each random variable, i.e., p=F(x)
- 3. Transform according to some marginal distribution: $x=F^{-1}(p)$

Copula Densities

$$f(x_1, x_2, \dots, x_n) = c(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \cdot \prod_{i=1}^n f_i(x_i)$$
(96)

where c is the derivative of C:

$$c(p_1, p_2, \dots, p_n) = \frac{\partial^n C(p_1, p_2, \dots, p_n)}{\partial p_1 \partial p_2 \cdots \partial p_n}$$
(97)

Measures of Tail Dependence

Because copulas facilitate the specification of different statistical dependence at different quantiles, it is of interest to introduce some generic measure of dependence, particularly in the tail of the distributions. This is particularly useful because, in certain applications, it is the extreme outcomes of dependent random variables that are of interest. The coefficient of upper tail dependence is defined as the probability that the random variable X_i exceeds the value associated with its inverse CDF of q, i.e., "the quantile of order q", given that the other random variable X_j exceeds the value associated with its inverse CDF of q, i.e., "the coefficient is:

$$\lambda_{upper} = \lim_{q \to 1} P\left(X_i > F_i^{-1}(q) \mid X_j > F_j^{-1}(q)\right) = \lim_{q \to 1} \frac{1 - 2q + C(q,q)}{1 - q}$$
(98)

Similarly,

$$\lambda_{lower} = \lim_{q \to 1} P\left(X_i < F_i^{-1}(q) \mid X_j < F_j^{-1}(q)\right) = \lim_{q \to 1} \frac{C(q,q)}{q}$$
(99)

It is noted that for the normal copula, $\lambda_{upper} = \lambda_{lower} = 0$. Hence, it is not possible to take into account tail dependence with this copula, contrary to, say, the Student copula. When using copulas, there are also other measures of dependence other than the measures of tail dependence in Eqs. (98) and (99). These include the "rank-dependent correlation coefficient," such as Kendall's tau and Spearman's rho. For Archimedean copulas, there is a strong connection between Kendall's tau and the parameter of the copula function.

Classical Inference

Classical statistical inference for random variables attempts to determine point estimates for the distribution parameters. In other words, values are sought for the mean and standard deviation of the random variable, and perhaps other distribution parameters. In the classical approach, such point estimates are sometimes complemented by confidence intervals to gage the uncertainty in the point estimates. This document provides the most basic formulas, but starts with an exposure of diagrams that should always be plotted before computations are made.

Diagrams

Certain diagrams are helpful to visualize the characteristics of a probability distribution. Three plots are particularly popular:

- Histogram: In these plots the abscissa axis shows the outcome space for the random variable. To generate a histogram, this axis is divided into "bins" and the number of observed realizations within each bin is plotted on the ordinate axis.
- Frequency diagram: This diagram is a normalized version of the histogram. In particular, the area underneath the frequency diagram is unity, which means that it

can be visually compared with standard PDFs. The frequency diagram is normalized by dividing the ordinate values of the histogram by the total area of the histogram. The total area of the histogram equals the total number of observations multiplied by the bin size.

• Cumulative frequency diagram: While the frequency diagram is comparable to a PDF, the cumulative frequency diagram is comparable with the CDF. This plots also has the random variable along the abscissa axis. Ordinate values are computed at every observed realization of the random variable. Each ordinate value equals the number of realizations at and below that abscissa value, divided by the total number of observations.

Second-moment Statistics

Given *n* observations x_i of the random variable X the sample mean is:

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \tag{100}$$

The sample variance, i.e., the sample standard deviation squared, is:

$$s^{2} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_{i} - \overline{x})^{2}$$
(101)

In situations with many observations Eq. (101) is somewhat cumbersome because the mean of the random variable must be pre-computed before looping through the data again to compute s_2 . This is remedied by the following manipulations:

$$s^{2} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_{i} - \overline{x})^{2}$$

= $\frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} x_{i}^{2} + \sum_{i=1}^{n} \overline{x}^{2} - \sum_{i=1}^{n} 2x_{i}\overline{x} \right)$
= $\frac{1}{n-1} \cdot \left(\left(\sum_{i=1}^{n} x_{i}^{2} \right) + n\overline{x}^{2} - 2n\overline{x}^{2} \right)$
= $\frac{1}{n-1} \cdot \left(\left(\sum_{i=1}^{n} x_{i}^{2} \right) - n \cdot \overline{x}^{2} \right)$ (102)

This expression is more computationally convenient because the data can be looped over only once, to compute the sum of x_i and the sum of x_i^2 . In passing, it is noted that the reason for the denominator (*n*-1) instead of simply *n* is as follows: Consider the sample and the sample variance to be random variables in their own right. Then, the expectation of the sample mean is

$$E[\overline{x}] = \frac{1}{n} \cdot \sum_{i=1}^{n} E[x_i] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mu_x = \mu_x$$
(103)

where the second-last term recognizes that $E[x_i]$ is the mean of the random variable. This provides comfort that the expectation of the sample mean equals the mean of the underlying random variable. Next, consider the expectation of the sample variance:

$$E[s^{2}] = \frac{1}{n-1} \cdot E\left[\left(\sum_{i=1}^{n} x_{i}^{2}\right) - n \cdot \overline{x}^{2}\right]$$

$$= \frac{1}{n-1} \cdot \left(\left(\sum_{i=1}^{n} E[x_{i}^{2}]\right) - n \cdot E[\overline{x}^{2}]\right)$$
(104)

To proceed, it is made use of the fact that "the variance is equal to the mean square minus the square of the means," so that:

$$\sum_{i=1}^{n} E[x_i^2] = \sum_{i=1}^{n} \left(E[x_i]^2 + Var[x_i] \right) = \sum_{i=1}^{n} \left(\mu_x^2 + \sigma_x^2 \right)$$
(105)

and

$$E[\overline{x}^2] = E[\overline{x}]^2 + Var[\overline{x}]$$
(106)

where the mean of the sample mean is provided in Eq. (103) and the variance of the sample mean is:

$$Var[\bar{x}] = \frac{1}{n^2} \cdot \sum_{i=1}^n Var[x_i] = \frac{1}{n^2} \cdot \sum_{i=1}^n \sigma_x^2 = \frac{\sigma^2}{n}$$
(107)

Substitution of Eqs. (103) and (107) into Eq. (106) and substitution of Eqs. (105) and (106) into Eq. (104) yields

$$E[s^{2}] = \frac{1}{n-1} \cdot \left(\left(\sum_{i=1}^{n} \left(\mu_{x}^{2} + \sigma_{x}^{2} \right) \right) - n \cdot \left(\mu_{x}^{2} + \frac{\sigma_{x}^{2}}{n} \right) \right)$$
$$= \frac{1}{n-1} \cdot \left(n \cdot \left(\mu_{x}^{2} + \sigma_{x}^{2} \right) - n \cdot \left(\mu_{x}^{2} + \frac{\sigma_{x}^{2}}{n} \right) \right)$$
$$= \frac{1}{n-1} \cdot \left(n \cdot \sigma_{x}^{2} - \sigma_{x}^{2} \right)$$
$$= \frac{\sigma_{x}^{2}}{n-1} \cdot (n-1)$$
$$= \sigma_{x}^{2}$$
(108)

which shows that the denominator (n-1) is necessary to have the expectation of the sample variance match to the underlying random variable.

Correlation

The formulas for sample mean and sample variance of individual random variables are valid for the inference on joint random variables. In addition, the sample correlation coefficient is:

$$\rho = \frac{1}{n-1} \cdot \left(\frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{s_x \cdot s_y} \right)$$
(109)

Inference by Geometrical Considerations

Usually, observed realizations of a random variable form the basis for making inferences about its underlying distribution. However, in some situations it is possible to determine the exact probability distribution of a random variable by geometrical considerations. In particular, this is sometimes possible when the random variable represents a distance. For this purpose, consider an uncertain distance X, which is a random variable with CDF F(x). By definition, the CDF represents the probability that the distance is less than x. This can sometimes be computed as

$$F(x) = P(X \le x) = \begin{cases} \left(\frac{L_{X \le x}}{L_{total}}\right) \\ \left(\frac{A_{X \le x}}{A_{total}}\right) \\ \left(\frac{V_{X \le x}}{V_{total}}\right) \end{cases}$$
(110)

where L, A, and V represent length, area, and volume; L_{total} is the total length of the outcome space and $L_{X \le x}$ is the length of the outcome space where X is less than x, and so forth. For problems that are amenable to this approach, some deal with lengths, others with area, and yet others with volume

Bayesian Inference

Contrary to classical statistics, where point estimates are provided for distribution parameters, the Bayesian approach provides probability distributions. For example, for a Normal random variable, the Bayesian analysis provides the probability distribution for the mean and standard deviation. All other inference statements are made from these distributions. The availability of these distributions is also advantageous because they can be included in subsequent reliability analysis. In the following, let X denote a random variable and let θ denotes a generic parameter in the probability distribution for X. The key objective is to determine the probability distribution of θ given observations of X, collected in the vector **x**. The following formula synthesizes the essence of the Bayesian approach (Box and Tiao 1992; Carlin and Louis 2009):

$$f''(\theta) = \frac{L(\theta)}{c} \cdot f'(\theta) \tag{111}$$

where $f''(\theta)$ is the posterior PDF, *c* is a constant explained shortly, $L(\theta)$ is the likelihood function, and $f'(\theta)$ is the prior PDF. The constant, *c*, serves the purpose of normalizing the posterior, which implies the following definition:

$$c = \int_{-\infty}^{\infty} L(\theta) \cdot f'(\theta) d\theta$$
 (112)

To understand the workings of Eq. (111) it is helpful to first relate it to Bayes' Rule for events, which reads:

$$P(E_1 | E_2) = \frac{P(E_2 | E_1)}{P(E_2)} \cdot P(E_1)$$
(113)

where E_1 is the event for which the probability is sought and E_2 is the event that has occurred. It is noted that the probability of the occurred event, conditioned upon E_1 , serves the role as likelihood, and that the unconditional probability of the observed event serves as normalizing factor in the denominator. This pattern also emerges when Eq. (111) is written in the following more complete form:

$$f''(\boldsymbol{\theta}|\mathbf{x}) = \frac{L(\mathbf{x}|\boldsymbol{\theta})}{c(\mathbf{x})} \cdot f'(\boldsymbol{\theta})$$
(114)

where

$$c(\mathbf{x}) = \int_{-\infty}^{\infty} L(\mathbf{x}|\boldsymbol{\theta}) \cdot f'(\boldsymbol{\theta}) d\boldsymbol{\theta}$$
(115)

which is sometimes called the marginal density of the data. This formulation clarifies that the posterior distribution for θ is directly linked to the observed data via the likelihood function. The formulation of the likelihood function, as well as the prior distribution, is a central topic in this document.

Discrete Problems

This document primarily addresses the case where θ and X are continuous random variables. However, in passing, a few other cases are noted. First, if one or both of the variables are discrete, then the PDFs are simply replaced by PMFs. In fact, if the observed random variable is discrete then the normalization constant takes on a direct meaning, as shown here for the case where both variables are discrete:

$$p(x \mid y) = \frac{p(y \mid x)}{p(y)} \cdot p(x)$$
(116)

Bayes' theorem for updating the probability distribution of a *continuous* random variable, θ , given the occurrence of an event, *E*, is:

$$f(\theta|E) = \frac{P(E|\theta)}{P(E)} \cdot f(\theta)$$
(117)

Similarly, Bayes' theorem for updating the probability distribution of a *discrete* random variable, θ , given the occurrence of an event, *E*, is:

$$p(\theta|E) = \frac{P(E|\theta)}{P(E)} \cdot p(\theta)$$
(118)

Bayes' rule to update the probability of an event, E, given the outcome of a random variable, θ , is:

$$P(A \mid \theta) = \begin{cases} \frac{f(\theta \mid A)}{f(\theta)} \cdot P(A) & \text{(if } \theta \text{ is continuous)} \\ \frac{p(\theta \mid A)}{p(\theta)} \cdot P(A) & \text{(if } \theta \text{ is discrete)} \end{cases}$$
(119)

Prior Distribution

Eq. (111) shows that the prior distribution is one of two key ingredients in the Bayesian approach, with the likelihood function being the other one. The prior distribution is sometimes a point of contention, because it allows subjective information to enter the calculations. On one hand, this is an advantage because it gives more flexibility to the analyst; on the other hand it may seem problematic because the prior assumptions may seem arbitrary. To expose the matter, the following subsections list the options that are available as prior distributions.

Previous Posterior

When a probability distribution for θ is already available, for example from earlier applications of Eq. (111), then it is natural to employ it as prior. In circumstances where this choice leads to an unusually complicated expression for the posterior, then the use of a conjugate prior may be explored, as explained shortly.

Uniform and Non-informative Priors

When little or now prior information is available about θ it is desirable to select a prior that is uniform over the "range of interest." This is either the exact uniform distribution, or a distribution that is approximately uniform over the important range of θ -values. This is intended to express complete a priori uncertainty about θ .

Conjugate Prior

A prior is called conjugate if the distribution type of the posterior is the same as that of the prior. The selection of a conjugate prior is convenient, because it often leads to simple updating rules for the parameters of the distribution for θ .

Likelihood Function

The likelihood function is the crucial means by which the observed data affects the posterior. Eq. (111) illustrates that the likelihood function is a function of θ , and Eq. (114) clarifies that it takes as input the vector of observed realizations, **x**. To further understand the meaning of the likelihood function, a strict comparison between Eq. (111) and Bayes' Rule in Eq. (113) could lead to the impression that $L(\theta)$ is the "probability of **x** given θ ." However, this is misleading, both because the probability of any realization **x** is zero, and because $L(\theta)$ does not have to be interpreted as a probability. Instead, because of the normalizing constant, c, in Eq. (111), it is only necessary that the likelihood function is *proportional* to the probability of observing **x**. In short, the

likelihood function requires a probabilistic distribution for X, which is a function of θ , but it is unnecessary to enter a proper probability into Eq. (111). As a generic example, suppose X has the PDF f(x), in which θ enters, and one realization, x, has been observed. Then the likelihood function, $L(\theta)$, equals the value of f(x), evaluated at the observed realization, x, with θ as a free variable. Importantly, f(x) can be scaled by whichever expression that is constant with respect to θ . Similarly, if it has been observed that X < x, then the likelihood function is proportional the CDF, F(x), with θ as a free variable.

Posterior Statistics

Once the posterior distribution is determined in accordance with Eq. (111) it is often of interest to determine, at least, the second-moment statistics of the model parameter(s). From the definition of expectation, the mean is:

$$\mu_{\theta} = \int_{-\infty}^{\infty} \theta \cdot f''(\theta) d\theta \tag{120}$$

and the variance is

$$\sigma_{\theta}^{2} = \int_{-\infty}^{\infty} (\theta - \mu_{\theta})^{2} \cdot f''(\theta) d\theta$$
 (121)

For problems with more than one model parameter, the mean vector is:

$$\mathbf{M}_{\boldsymbol{\theta}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \boldsymbol{\theta} \cdot f''(\boldsymbol{\theta}) d\boldsymbol{\theta}$$
(122)

and the covariance matrix is:

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\boldsymbol{\theta} - \mathbf{M}_{\boldsymbol{\theta}}\right) \cdot \left(\boldsymbol{\theta} - \mathbf{M}_{\boldsymbol{\theta}}\right)^{T} \cdot f''(\boldsymbol{\theta}) d\boldsymbol{\theta}$$
(123)

Predictive Distribution

While the posterior in Eq. (111) provides the probability distribution for the model parameter(s), the so-called predictive distribution addresses the original random variable, X. Specifically, the predictive distribution is a distribution for X that incorporates the uncertainty in the model parameter(s), θ , by using the expectation integral in the following way:

$$f(x) = \int_{-\infty}^{\infty} f(x|\theta) \cdot f''(\theta) d\theta$$
(124)

Computational Methods

The Bayesian approach is philosophically appealing, but it has been hindered by computational challenges for all but simple problems, e.g., problems with conjugate priors. This has changed with the advent of high computer power and new sampling algorithms. It is the evaluation of two integrals that are of particular importance in Bayesian analysis:

1. The integral to obtain the normalizing constant, c, in Eq. (112).

2. The integrals to obtain second-moment statistics in Eqs. (120) through (123), or higher moments and quantiles.

Recently, numerical integration algorithms, implemented on powerful computers, have emerged to solve these problems (Carlin and Louis 2009). While Gaussian quadrature used to be popular, this approach suffers from the curse of dimensionality when the number of model parameters is large. Instead, various Monte Carlo methods have become prevalent. In particular, use of Markov Chain Monte Carlo (MCMC) methods, such as the Metropolis-Hastings algorithm and the Gibbs sampler, have become widespread. Such algorithms do not require a known analytical expression for the sampling distribution, which is advantageous in light of the fact that few posterior distributions come in closed form. Also, while classical Monte Carlo sampling creates independent realizations, the realizations from MCMC methods are dependent. Nominally, this means that more samples are required to reach the same precision as with Monte Carlo sampling, but the "effectiveness" of each sample is greater with MCMC. Several computer programs are now available to carry out advanced Bayesian inference:

- R (www.r-project.org) is the free version of the commercial package S-plus
- WinBUGS (www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml) has syntax similar to R
- RBugs allows WinBUGS to be called from within R
- OpenBUGS (http://mathstat.helsinki.fi/openbugs) is the successor to WinBUGS
- BayesX (www.stat.uni-muenchen.de/~bayesx) has similarities with WinBUGS, with a more limited modelling range

List of Conjugate Priors

The table below gives an overview of conjugate priors, and the convenient updating rules that follow. The notation matches the one used in the documents that give the detailed expressions for the distributions. For problems where notation may conflict, an underline is employed to distinguish the parameter of the distribution of X and the parameter of the prior distribution.

Random variable	Observation	Prior	Rule
$X \sim \text{Binomial}(p,n)$	<i>x</i> occurrences in <i>n</i> trials	$p \sim \text{Beta}(a,b)$	a''=a'+x b''=b'+n-x
<i>X</i> ~Geometric(<i>p</i>)	<i>x</i> trials until first occurrence	$p \sim \text{Beta}(a,b)$	a''=a'+1 b''=b'+x-1
X~NegativeBinomial(p,k)	x trials to k^{th} occurrence	$p \sim \text{Beta}(a,b)$	$a^{"}=a^{+}k$ $b^{"}=b^{+}x-k$
$X \sim \operatorname{Poisson}(\lambda, T)$	x occurrences in T	λ -Gamma(ν , k)	$k''=k'+x$ $\nu''=\nu'+T$
$X \sim \operatorname{Exp}(\lambda)$	<i>n</i> observations of <i>x</i>	λ -Gamma(ν , k)	$k^{\prime\prime} = k^{\prime} + n$ $\nu^{\prime\prime} = \nu^{\prime} + \Sigma x_i$

rable 1. Optiating rules for conjugate priors	Table	1:	Updating	rules for	· conjugate	priors.
---	-------	----	----------	-----------	-------------	---------

$X \sim \operatorname{Exp}(\lambda)$	n obs. of x and m obs. of $X > y$	λ -Gamma(ν , k)	$k^{\prime\prime} = k^{\prime} + n (?)$ $v^{\prime} = v^{\prime} + \sum x_i + \sum y_i$
<i>X</i> ~Gamma(<u><i>v</i>,<i>k</i></u>)	<i>n</i> observations of <i>x</i>	<u>v</u> ~Gamma(<i>v</i> , <i>k</i>)	$k''=k'+n\underline{k}$ $v''=v'+\Sigma x_i$
X ~Rayleigh(λ)	<i>n</i> observations of <i>x</i>	λ ~Gamma(<i>v</i> , <i>k</i>)	$k''=k'+n$ $\nu''=\nu'+\Sigma x_i^2$

<i>X</i> ~Normal(<u><i>µ</i>,<i>о</i></u>)	<i>n</i> observations of <i>x</i>	<u>μ</u> ~Normal(μ,σ)	$\mu'' = \frac{\overline{x} \cdot \sigma'^2 + \mu'\left(\frac{\overline{\sigma}^2}{n}\right)}{\sigma'^2 + \frac{\overline{\sigma}^2}{n}}$ $\sigma'' = \sqrt{\frac{\sigma'^2 \cdot \left(\frac{\overline{\sigma}^2}{n}\right)}{\sigma'^2 + \left(\frac{\overline{\sigma}^2}{n}\right)}}$
X~Lognormal(<u>ζ,σ</u>)	nobservations of x with average \overline{x}	ζ~Normal(μ,σ)	$\mu'' = \frac{\overline{\ln(x)} \cdot \sigma'^2 + \mu'\left(\frac{\underline{\sigma}^2}{n}\right)}{\sigma'^2 + \frac{\underline{\sigma}^2}{n}}$ $\sigma'' = \sqrt{\frac{\sigma'^2 \cdot \left(\frac{\underline{\sigma}^2}{n}\right)}{\sigma'^2 + \left(\frac{\underline{\sigma}^2}{n}\right)}}$

References

Ang, A. H.-S., and Tang, W. H. (2007). Probability concepts in engineering: emphasis on applications in civil & environmental engineering. Wiley.

Box, G. E. P., and Tiao, G. C. (1992). Bayesian inference in statistical analysis. Wiley. Carlin, B. P., and Louis, T. A. (2009). Bayesian methods for data analysis. CRC Press.