Functions and Transformations

The premise in this document is a set of continuous random variables, y, that has some functional relationship with on another set of continuous random variables, x. It is possible that they are single random variables, y and x, and the situation y and x is also possible. Regardless, two questions may arise:

- If we know the functional relations ship between **x** and **y**, and we know the probability distributions or at least partial descriptors for **x**, what is the distribution of **y**? This is called "analysis of functions" of random variables.
- If we know the probability distribution for **x** and **y**, what is the functional relationship between them? We call this "probability transformations."

Analysis of Functions

First, consider one function, Y, which is a function, $h(\mathbf{X})$, of n random variables, \mathbf{X} . Also assume for now that we have only second-moment information, i.e., mean, variance, and correlation, and that the second-moment information for Y is sought. To solve this problem, the expectation operator is vital. It is defined as

$$\mathbf{E}[Y] = \mathbf{E}[h(\mathbf{X})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, x_2, \cdots, x_n) f(x_1, x_2, \cdots, x_n) dx_1 dx_2 \cdots dx_n$$
(1)

Expectation is a linear operator, having the properties

$$E[a] = a$$

$$E[a \cdot h(\mathbf{X})] = a \cdot E[h(\mathbf{X})]$$

$$E[h_1(\mathbf{X}) + h_2(\mathbf{X})] = E[h_1(\mathbf{X})] + E[h_2(\mathbf{X})]$$

$$\frac{\partial}{\partial \theta} E[h(\mathbf{X}, \theta)] = E[\frac{\partial}{\partial \theta} h(\mathbf{X}, \theta)]$$
(2)

In passing it is noted that the variance, which is the expectation of $(X-\mu_X)^2$ has the properties

$$Var[a + X] = Var[X]$$

$$Var[b \cdot X] = b^{2} \cdot Var[X]$$
(3)

Some of these properties are useful to determine the mean and variance of Y. Consider first a linear function, where $Y=h(\mathbf{X})$ is written, in vector and index notation:

$$Y = a + \mathbf{b}^T \mathbf{X} = a + b_i X_i \tag{4}$$

where a is a constant and **b** is a vector of constants. The expectation of Eq. (4) is conveniently carried out in index notation:

$$\mu_Y = \mathbf{E}[Y] = a + b_i \cdot \mathbf{E}[X_i] = a + \mathbf{b}^T \mathbf{M}_X$$
(5)

Functions and Transformations

Essentially, the mean of the function is obtained by substituting the means for the random variables.

The variance of the linear function in Eq. (4) is also derived in index notation:

$$\sigma_{Y}^{2} = \mathbf{E}\left[\left(Y - \mu_{Y}\right)^{2}\right]$$

$$= \mathbf{E}\left[\left((a + b_{i}X_{i}) - (a + b_{j}\mu_{j})\right)^{2}\right]$$

$$= \mathbf{E}\left[\left(b_{i}X_{i} - b_{j}\mu_{j}\right)^{2}\right]$$

$$= \mathbf{E}\left[\left(b_{i}X_{i} - b_{j}\mu_{j}\right) \cdot \left(b_{k}X_{k} - b_{l}\mu_{l}\right)\right]$$

$$= \mathbf{E}\left[b_{i}X_{i} \cdot b_{k}X_{k} - b_{i}X_{i} \cdot b_{l}\mu_{l} - b_{j}\mu_{j} \cdot b_{k}X_{k} + b_{j}\mu_{j} \cdot b_{l}\mu_{l}\right] \qquad (6)$$

$$= \mathbf{E}\left[b_{i}X_{i} \cdot b_{k}X_{k}\right] - \mathbf{E}\left[b_{i}X_{i} \cdot b_{l}\mu_{l}\right] - \mathbf{E}\left[b_{j}\mu_{j} \cdot b_{k}X_{k}\right] + \mathbf{E}\left[b_{j}\mu_{j} \cdot b_{l}\mu_{l}\right]$$

$$= b_{i}b_{k} \cdot \mathbf{E}\left[X_{i}X_{k}\right] - b_{j}b_{l} \cdot \mu_{j}\mu_{l}$$

$$= b_{i}b_{k} \cdot \mathbf{Cov}\left[X_{i}X_{k}\right]$$

$$= \mathbf{b}^{T}\mathbf{\Sigma}_{XX}\mathbf{b}$$

Similarly, one can also find the covariance between two different linear functions, $Y_1=a+\mathbf{b}^T\mathbf{X}$ and $Y_2=c+\mathbf{d}^T\mathbf{X}$:

$$\operatorname{Cov}[Y_1, Y_2] = \operatorname{E}\left[\left(Y_1 - \mu_{Y_1}\right)\left(Y_2 - \mu_{Y_2}\right)\right] = \mathbf{b}^T \boldsymbol{\Sigma}_{XX} \mathbf{d}$$
(7)

One may attempt to derive exact analytical second-moment expressions also for nonlinear functions. However, depending on the complexity of the function, this may not be possible. In that case, one option is to approximate the function(s) by a Taylor expansion about the mean. Keeping the first two terms of the expansion yields the linearized approximation

$$Y = h(\mathbf{X}) \approx h(\mathbf{M}_{X}) + \nabla h(\mathbf{M}_{X})^{T} (\mathbf{X} - \mathbf{M}_{X})$$
(8)

where, from mathematics, we know that $\nabla h(\mathbf{M}_X)$ is the gradient vector of the function, evaluated at the mean. According to the earlier derivations, the linearization in Eq. (8) yields the following second-moment results:

$$\mu_{Y} = h(\mathbf{M}_{X}) \tag{9}$$

$$\sigma_Y^2 = \nabla h(\mathbf{M}_X)^T \mathbf{\Sigma}_{XX} \nabla h(\mathbf{M}_X)$$
(10)

$$\operatorname{Cov}[Y_1, Y_2] = \nabla h_1(\mathbf{M}_X)^T \mathbf{\Sigma}_{XX} \nabla h_2(\mathbf{M}_X)$$
(11)

Obtaining Full Distributions

Reconsider the situation in which a dependent random variable, Y, is related by a known functional relationship to one or more independent random variables, **X**. In this section the entire probability distribution for Y is sought, not only second-moment information. This is referred to as finding the "derived distribution." First consider the case of a

continuous random variable, Y, related to another continuous random variable, X, by the functional relationship:

$$Y = h(X) \tag{12}$$

where h is a monotonically increasing function so that there exists a one-to-one mapping between realizations of x and y. This relationship is shown schematically by the solid line in Figure 1. To obtain the probability distribution for Y when the distribution for X is known, one starting point is the definition for the CDF for Y:

$$F_{Y}(y) = P(Y \le y) \tag{13}$$

Substituting Eq. (12) into Eq. (13) yields

$$F_{y}(y) = P(h(X) \le y) \tag{14}$$

The probability distribution for X is known, while the distribution for Y is sought. The "probability preserving" transformation yields the sought CDF:

$$F_{Y}(y) = F_{X}(x) = F_{X}(h^{-1}(y))$$
(15)

That equation is differentiated to obtain the PDF for *Y*:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dx}{dy} \cdot \frac{dF_X(x)}{dx} = \frac{dx}{dy} \cdot f_X(x) = \frac{dx}{dy} \cdot f_X(h^{-1}(y))$$
(16)

Rearranging yields

$$f_Y(y) \cdot dy = f_X(x) \cdot dx \tag{17}$$

This equality is visualized by shaded areas in Figure 1.



Figure 1: Derived distribution.

In the multi-variate case, depending on the distributions of the independent random variables, \mathbf{X} , and the functional relationship with Y there may not exist an analytical expression for the distribution of Y. However, the following equality holds:

$$f_{\mathbf{Y}}(y_1, y_2, \cdots, y_n) dy_1 dy_2 \cdots dy_n = f_{\mathbf{X}}(x_1, x_2, \cdots, x_n) dx_1 dx_2 \cdots dx_n$$
(18)

which by rearranging is written in terms of the Jacobian determinant:

$$f_{\mathbf{Y}}(y_1, y_2, \cdots, y_n) = f_{\mathbf{X}}(x_1, x_2, \cdots, x_n) \left| \det(\mathbf{J}_{\mathbf{y}, \mathbf{x}}) \right|^{-1}$$
(19)

Note, however, that certain special cases are available: A linear function of normal random variables is always normal. A product function of lognormal random variables is always lognormal.

Probability Transformations

Some of this material was first described to me in a course taught by Professor Armen Der Kiureghian at the University of California at Berkeley. In 2005 he made a concise description available in Chapter 14 "First- and second-order reliability methods" of the CRC Engineering Design Reliability Handbook edited by Nikolaidis, Ghiocel and Singhal, published by the CRC Press in Boca Raton, Florida.

This document seeks to determine the functional relationship between two random variables—or two vectors of random variables—given knowledge about both probability distributions. As an illustration, consider a random variable X, which is associated some known marginal probability distribution. The transformation to a random variable Y, which has, say, the standard normal distribution, is sought. More generally, the aim is to transform the vector of random variables, X, with known probability distribution, into a vector of random variables, Y, also with prescribed probability distribution. Again, the objective is to determine the functional relationship between X and Y. Another document on analysis of functions addresses the problem of finding the unknown probability distribution of Y or Y when the functional relationship is known.

Transformation of One Random Variable

It is both pedagogically and practically useful to first consider single-variable transformations. Consider a random variable X with CDF $F_X(x)$. Suppose a transformation to the random variable Y is sought. First, consider the problem where the target distribution for Y is known. In fact, let Y be a random variable with CDF $F_Y(y)$. That is, both F_X and F_Y are known. To establish the transformation, which is referred to as the "probability-preserving transformation," the two CDFs are equated:

$$F_{X}(x) = F_{Y}(y) \tag{20}$$

This states that the probability mass at values below the equivalent thresholds x and y must be equal. As a result, y is written

$$y = F_Y^{-1}(F_X(x))$$
 (21)

and x is written

$$x = F_X^{-1} \left(F_Y(y) \right)$$
 (22)

where F^{-1} denotes the inverse CDF. As an example, consider a value y generated by a random number generator according to the standard normal probability distribution, whose CDF is denoted $\Phi(y)$. To transform that realization into a random variable x with, say, the uniform probability distribution, whose CDF is written F(x), the following calculation is carried out:

$$x = F^{-1}(\Phi(y)) \tag{23}$$

The transformation that is established in Eq. (20) is extensively utilized in reliability analysis to transform a random variable with some distribution into a standard normal variable.

Standardization of Second-Moment Vector

Let **x** denote the realization of a vector of random variables with means \mathbf{M}_X and covariance matrix $\boldsymbol{\Sigma}_{XX}$. The objective in this section is to transform **X** into a vector **Y** of the same number of random variables with zero means and unit covariance matrix. I.e., **Y** is a vector of uncorrelated and "standardized" random variables. Some readers will perhaps recall from elementary statistics courses that for the case of one random variable, the relationship is:

$$y = \frac{x - \mu}{\sigma} \tag{24}$$

where μ is the mean and σ is the standard deviation. In general, a second-moment transformation is written

$$\mathbf{y} = \mathbf{a} + \mathbf{B}\mathbf{x} \tag{25}$$

were the vector **a** and the square matrix **B** contain unknown constants. Eq. (25) represent linear functions of random variables, and we seek **a** and **B**. Thus, according to "analysis of functions," two equations for the unknowns **a** and **B** are established by enforcing zero means and unit covariance matrix for y:

$$\mathbf{M}_{Y} = \mathbf{a} + \mathbf{B}\mathbf{M}_{X} = \mathbf{0} \tag{26}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} = \mathbf{B}\boldsymbol{\Sigma}_{\boldsymbol{\chi}\boldsymbol{\chi}}\mathbf{B}^{T} = \mathbf{I}$$
(27)

B is the only unknown in Eq. (27). Multiplying through by \mathbf{B}^{-1} from the left and \mathbf{B}^{-T} from the right yields the following expression for the covariance matrix of **X**:

$$\boldsymbol{\Sigma}_{XX} = \mathbf{B}^{-1} \mathbf{B}^{-T} \tag{28}$$

Hence, the unknown matrix \mathbf{B}^{-1} is the one that decomposes Σ_{XX} into a matrix multiplied by its transpose. This is known as the Cholesky decomposition:

$$\boldsymbol{\Sigma}_{XX} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T \tag{29}$$

where a tilde identifies the lower-triangular Cholesky decomposition of the covariance matrix. The tilde will later be removed to identify the Cholesky decomposition of the correlation matrix. Comparing Eqs. (28) and (29) one finds that

$$\mathbf{B} = \tilde{\mathbf{L}}^{-1} \tag{30}$$

which, substituted into Eq. (26), yields

$$\mathbf{a} = -\tilde{\mathbf{L}}^{-1}\mathbf{M}_{\chi} \tag{31}$$

Thus, the sought standardization transformation reads

$$\mathbf{y} = \tilde{\mathbf{L}}^{-1} \left(\mathbf{x} - \mathbf{M}_X \right) \tag{32}$$

Solving for x yields the transformation back to the original vector:

$$\mathbf{x} = \mathbf{M}_{X} + \tilde{\mathbf{L}}\mathbf{y} \tag{33}$$

The Cholesky decomposition in Eq. (29) may be difficult because the covariance matrix contains components with dimensions associated with the dimensions of the random variables. In other words, it may contain numbers with different orders of magnitude. Therefore it is often more accurate to decompose the dimensionless correlation matrix. For this purpose, the covariance matrix is written

$$\boldsymbol{\Sigma}_{XX} = \boldsymbol{D}_{X}\boldsymbol{R}_{XX}\boldsymbol{D}_{X} = \boldsymbol{D}_{X}\boldsymbol{L}\boldsymbol{L}^{T}\boldsymbol{D}_{X}$$
(34)

where D_X is a diagonal matrix with standard deviations on the diagonal and L is the Cholesky decomposition of the correlation matrix. According to the derivations above, the standardization transformation now reads

$$\mathbf{y} = \mathbf{L}^{-1} \mathbf{D}_X^{-1} \left(\mathbf{x} - \mathbf{M}_X \right) \qquad \Longleftrightarrow \qquad \mathbf{x} = \mathbf{M}_X + \mathbf{D}_X \mathbf{L} \mathbf{y}$$
(35)

When probability transformations like the one in Eq. (25) is applied in reliability analysis it is often necessary to also compute the Jacobian matrix associated with the transformation. In other words, the derivative of **y** with respect to **x**, or its inverse, is sought. For the second-moment transformation outlined in this section it is obtained by differentiating the expression for **y**:

$$\mathbf{J}_{\mathbf{y},\mathbf{x}} \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \tilde{\mathbf{L}}^{-1} = \mathbf{L}^{-1} \mathbf{D}_{X}^{-1}$$
(36)

Transformation of Independent Random Variables

The previous derivations are now extended to cases where the entire probability distribution of the random variables, X, is known. For now, suppose they are uncorrelated. As a result, the joint PDF is the product of the marginal PDFs. In this case, the probability preserving transformation in Eq. (20) is applied to each random variable at a time. In particular, the transformation into standard normal random variables is sought in reliability analysis:

$$F_i(x_i) = \Phi(y_i) \quad \Leftrightarrow \quad y_i = \Phi^{-1} \left(F_i(x_i) \right) \quad \Leftrightarrow \quad x_i = F_i^{-1} \left(\Phi(y_i) \right)$$
(37)

where F_i is the CDF of random variable number *i*. The Jacobian matrix for this transformation is obtained by differentiating the left-most equation in Eq. (37) with respect to x_i :

Functions and Transformations

$$\frac{\partial}{\partial x_i} \left(F_i(x_i) = \Phi(y_i) \right) \quad \Rightarrow \quad f_i(x_i) = \frac{\partial}{\partial x_i} \Phi(y_i) = \frac{\partial y_i}{\partial x_i} \frac{\partial}{\partial y_i} \Phi(y_i) = \frac{\partial y_i}{\partial x_i} \varphi(y_i) \quad (38)$$

where f and φ are the PDFs corresponding to F and Φ , respectively. As a result, the Jacobian matrix is a diagonal matrix with components

$$\frac{\partial y_i}{\partial x_i} = \frac{f_i(x_i)}{\varphi(y_i)} \tag{39}$$

Transformation of Dependent Random Variables: Nataf

The previous section is now extended to include correlation between the random variables. As a first step, consider the transformation of each random variable x_i according to the transformation in the previous section, i.e., disregarding correlation:

$$\mathbf{z}_i = \Phi^{-1} \left(F_i(x_i) \right) \tag{40}$$

where the variables z are normally distributed with zero means and unit variances. However, they are correlated. To facilitate the sought transformation it is assumed that the random variables z_i are jointly normal. This is called the Nataf assumption. Under this assumption it can be shown (Liu and Der Kiureghian 1986) that the correlation coefficient $\rho_{0,ij}$ between z_i and z_j is related to the correlation coefficient ρ_{ij} between x_i and x_j by the equation:

$$\rho_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{x_i - \mu_i}{\sigma_i} \right) \left(\frac{x_j - \mu_j}{\sigma_j} \right) \varphi_2(z_i, z_j, \rho_{0, ij}) dz_i dz_j$$
(41)

where

$$x_{i} = F_{i}^{-1} \left(\Phi(z_{i}) \right)$$

$$x_{j} = F_{j}^{-1} \left(\Phi(z_{j}) \right)$$
(42)

and φ_2 is the bivariate standard normal PDF:

$$\varphi_{2}(z_{i}, z_{j}, \rho_{0, ij}) = \frac{1}{2\pi\sqrt{1 - \rho_{0, ij}^{2}}} \exp\left[-\frac{z_{i}^{2} + z_{j}^{2} - 2 \cdot \rho_{0, ij} \cdot z_{i} \cdot z_{j}}{2 \cdot (1 - \rho_{0, ij}^{2})}\right]$$
(43)

The Nataf joint distribution model is valid under the lax conditions that the CDFs of x_i be strictly increasing and the correlation matrix of **x** and **z** be positive definite. It is an appealing transformation because it is invariant to the ordering of the random variables and a wide range of correlation values is acceptable. The downside is that Eq. (41) must be solved for each correlated pair of random variables. Once this is done the transformation from **z** to **y** must be addressed. Both are associated with zero means and a unit covariance matrix. In accordance with Eq. (25), but now with zero mean, the transformation reads

$$\mathbf{y} = \mathbf{B}\mathbf{z} \tag{44}$$

where **B** is sought. Similar to Eq. (27) the covariance matrix for **y** is written

$$\boldsymbol{\Sigma}_{YY} = \mathbf{B}\boldsymbol{\Sigma}_{ZZ}\mathbf{B}^T = \mathbf{I}$$
(45)

which yields

$$\boldsymbol{\Sigma}_{\boldsymbol{Z}\boldsymbol{Z}} = \mathbf{B}^{-1}\mathbf{B}^{-T} \tag{46}$$

It is observed that **B** is the inverse of the Cholesky decomposition of the covariance matrix of z. That covariance matrix is equal to the correlation matrix because the standard deviations are all zero. Hence, the Nataf transformation is

$$\mathbf{y} = \mathbf{L}^{-1}\mathbf{z} \quad \Leftrightarrow \quad \mathbf{z} = \mathbf{L}\mathbf{y} \tag{47}$$

where **L** is the Cholesky decomposition of the correlation matrix of **z**, i.e., it contains the correlation coefficients $\rho_{0,ij}$. The Jacobian matrix for the Nataf transformation combines Eqs. (36) and (39):

$$\mathbf{J}_{\mathbf{y},\mathbf{x}} \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{L}^{-1} \left[\frac{f_i(x_i)}{\varphi(y_i)} \right]$$
(48)

where the brackets imply a diagonal matrix. Conversely:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \mathbf{L} \left[\frac{\varphi(y_i)}{f_i(x_i)} \right]$$
(49)

In methods like SORM the second-order derivative of the transformation is also needed. The double derivative of \mathbf{y} with respect to \mathbf{z} , and vice versa, is zero. The double derivative of \mathbf{x} with respect to \mathbf{z} yields a slightly more complex diagonal matrix:

$$\frac{\partial}{\partial z_{j}} \left(\frac{\partial x_{i}}{\partial z_{i}} = \frac{\varphi(z_{i})}{f(x_{i})} \right)
\Rightarrow \frac{\partial^{2} x_{i}}{\partial z_{i} \partial z_{j}} = \frac{\partial \varphi(z_{i})}{\partial z_{j}} \cdot \frac{1}{f(x_{i})} + \frac{\partial}{\partial z_{j}} \left(\frac{1}{f(x_{i})} \right) \cdot \varphi(z_{i})
\Rightarrow \frac{\partial^{2} x_{i}}{\partial z_{i} \partial z_{j}} = \frac{\partial \varphi(z_{i})}{\partial z_{j}} \cdot \frac{1}{f(x_{i})} + \frac{\partial}{\partial x_{j}} \left(\frac{1}{f(x_{i})} \right) \cdot \frac{\partial x_{j}}{\partial z_{j}} \cdot \varphi(z_{i})
\Rightarrow \frac{\partial^{2} \mathbf{x}}{\partial \mathbf{z}^{2}} = \left[\frac{\partial \varphi(z_{i})}{\partial z_{i}} \cdot \frac{1}{f(x_{i})} - \frac{1}{f(x_{i})^{2}} \cdot \frac{\partial f(x_{i})}{\partial x_{i}} \cdot \frac{\partial x_{i}}{\partial z_{i}} \cdot \varphi(z_{i}) \right]$$
(50)

Transformation of Dependent Random Variables: Rosenblatt

An alternative to the Nataf approach is to consider the joint PDF of \mathbf{x} as a product of conditional PDFs:

$$f(\mathbf{x}) = f_1(x_1) \cdot f_2(x_2 | x_1) \cdots f_n(x_n | x_1, x_2, \dots, x_{n-1})$$
(51)

As a result of this sequential conditioning in the PDF the conditional CDFs are written

Having these CDFs facilitates the triangular transformation that is referred to as Rosenblatt transformation:

$$y_{1} = \Phi^{-1}(F_{1}(x_{1}))$$

$$y_{2} = \Phi^{-1}(F_{2}(x_{2}|x_{1}))$$

$$y_{3} = \Phi^{-1}(F_{3}(x_{3}|x_{1},x_{2}))$$

$$\vdots$$
(53)

To obtain the inverse transformation it is necessary to solve nonlinear equations for x_i , starting at the top of Eq. (53). The Jacobian matrix for this transformation is

$$\mathbf{J}_{\mathbf{y},\mathbf{x}} \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{f_{1}(x_{1})}{\varphi(y_{1})} & 0 & 0 & 0\\ \frac{1}{\varphi(y_{2})} \frac{\partial F_{2}(x_{2}|x_{1})}{\partial x_{1}} & \frac{f_{2}(x_{2}|x_{1})}{\varphi(y_{2})} & 0 & 0\\ \frac{1}{\varphi(y_{3})} \frac{\partial F_{3}(x_{3}|x_{1},x_{2})}{\partial x_{1}} & \frac{1}{\varphi(y_{3})} \frac{\partial F_{3}(x_{3}|x_{1},x_{2})}{\partial x_{2}} & \frac{f_{3}(x_{3}|x_{1},x_{2})}{\varphi(y_{3})} & 0\\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$
(54)

where the term in brackets is a diagonal matrix. The result of the transformation depends somewhat on the ordering of the random variables.

References

Liu, P.-L., and Der Kiureghian, A. (1986). "Multivariate distribution models with prescribed marginals and covariances." *Probabilistic Engineering Mechanics*, 1(2), 105–112.