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, $\mathbf{X}$, with known probability distribution, into a vector of random of random variables, $\mathbf{Y}$, also with prescribed probability distribution. Again, the objective is to determine the functional relationship between $\mathbf{X}$ and $\mathbf{Y}$. Another document on analysis of functions addresses the problem of finding the unknown probability distribution of Y or $\mathbf{Y}$ 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:

$$F_X(x) = F_Y(y) \tag{1}$$

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

$$y = F_Y^{-1}\left(F_X(x)\right) \tag{2}$$

and x is written

$$x = F_X^{-1}\left(F_Y(y)\right) \tag{3}$$

where F^{-1} denotes the inverse CDF. As an example application, 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:

$$x = F^{-1}\left(\Phi(y)\right) \tag{4}$$

The transformation that is established in Eq. (1) 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 $\mathbf{\Sigma}_{XX}$. 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:

$$y = \frac{x - \mu}{\sigma} \tag{5}$$

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

$$\mathbf{y} = \mathbf{a} + \mathbf{B}\mathbf{x} \tag{6}$$

were the vector **a** and the square matrix **B** contain unknown constants. Eq. (6) 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**:

$$\mathbf{M}_Y = \mathbf{a} + \mathbf{B}\mathbf{M}_X = \mathbf{0} \tag{7}$$

$$\mathbf{\Sigma}_{YY} = \mathbf{B}\mathbf{\Sigma}_{XX}\mathbf{B}^T = \mathbf{I} \tag{8}$$

B is the only unknown in Eq. (8). 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**:

$$\mathbf{\Sigma}_{XX} = \mathbf{B}^{-1}\mathbf{B}^{-T} \tag{9}$$

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

$$\mathbf{\Sigma}_{XX} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T \tag{10}$$

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. (9) and (10) one finds that

$$\mathbf{B} = \tilde{\mathbf{L}}^{-1} \tag{11}$$

which, substituted into Eq. (7), yields

$$\mathbf{a} = -\tilde{\mathbf{L}}^{-1}\mathbf{M}_X \tag{12}$$

Thus, the sought standardization transformation reads

$$\mathbf{y} = \tilde{\mathbf{L}}^{-1}\left(\mathbf{x} - \mathbf{M}_X\right) \tag{13}$$

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

$$\mathbf{x} = \mathbf{M}_X + \tilde{\mathbf{L}}\mathbf{y} \tag{14}$$

The Cholesky decomposition in Eq. (10) 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

$$\mathbf{\Sigma}_{XX} = \mathbf{D}_X \mathbf{R}_{XX} \mathbf{D}_X = \mathbf{D}_X \mathbf{L}\mathbf{L}^T \mathbf{D}_X \tag{15}$$

where $\mathbf{D}_X$ 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

$$\mathbf{y} = \mathbf{L}^{-1}\mathbf{D}_X^{-1}\left(\mathbf{x} - \mathbf{M}_X\right) \qquad \Leftrightarrow \qquad \mathbf{x} = \mathbf{M}_X + \mathbf{D}_X \mathbf{L}\mathbf{y} \tag{16}$$

When probability transformations like the one in Eq. (6) 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**:

$$\mathbf{J}_{\mathbf{y},\mathbf{x}} \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \tilde{\mathbf{L}}^{-1} = \mathbf{L}^{-1}\mathbf{D}_X^{-1} \tag{17}$$

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. (1) is applied to each random variable at a time. In particular, the transformation into standard normal random variables is sought in reliability analysis:

$$F_i(x_i) = \Phi(y_i) \quad \Leftrightarrow \quad y_i = \Phi^{-1}\left(F_i(x_i)\right) \quad \Leftrightarrow \quad x_i = F_i^{-1}\left(\Phi(y_i)\right) \tag{18}$$

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. (18) with respect to x_i:

$$\frac{\partial}{\partial x_i}\left(F_i(x_i) = \Phi(y_i)\right) \quad \Rightarrow \quad f_i(x_i) = \frac{\partial}{\partial x_i}\Phi(y_i) = \frac{\partial y_i}{\partial x_i}\frac{\partial}{\partial y_i}\Phi(y_i) = \frac{\partial y_i}{\partial x_i}\varphi(y_i) \tag{19}$$

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

$$\frac{\partial y_i}{\partial x_i} = \frac{f_i(x_i)}{\varphi(y_i)} \tag{20}$$

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:

$$z_i = \Phi^{-1}\left(F_i(x_i)\right) \tag{21}$$

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:

$$\rho_{ij} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x_i-\mu_i}{\sigma_i}\right)\left(\frac{x_j-\mu_j}{\sigma_j}\right)\varphi_2(z_i,z_j,\rho_{0,ij})\,dz_i\,dz_j \tag{22}$$

where

$$\begin{aligned} x_i &= F_i^{-1}\left(\Phi(z_i)\right) \\ x_j &= F_j^{-1}\left(\Phi(z_j)\right) \end{aligned} \tag{23}$$

and φ_2 is the bivariate standard normal PDF:

$$\varphi_2(z_i,z_j,\rho_{0,ij}) = \frac{1}{2\pi\sqrt{1-\rho_{0,ij}^2}}\exp\left[-\frac{z_i^2+z_j^2-2\cdot\rho_{0,ij}\cdot z_i\cdot z_j}{2\cdot(1-\rho_{0,ij}^2)}\right] \tag{24}$$

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. (22) 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. (6), but now with zero mean, the transformation reads

$$\mathbf{y} = \mathbf{B}\mathbf{z} \tag{25}$$

where **B** is sought. Similar to Eq. (8) the covariance matrix for **y** is written

$$\mathbf{\Sigma}_{YY} = \mathbf{B}\mathbf{\Sigma}_{ZZ}\mathbf{B}^T = \mathbf{I} \tag{26}$$

which yields

$$\mathbf{\Sigma}_{ZZ} = \mathbf{B}^{-1}\mathbf{B}^{-T} \tag{27}$$

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

$$\mathbf{y} = \mathbf{L}^{-1}\mathbf{z} \qquad \Leftrightarrow \qquad \mathbf{z} = \mathbf{L}\mathbf{y} \tag{28}$$

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. (17) and (20):

$$\mathbf{J}_{\mathbf{y},\mathbf{x}} \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{L}^{-1}\left[\frac{f_i(x_i)}{\varphi(y_i)}\right] \tag{29}$$

where the brackets imply a diagonal matrix. Conversely:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \mathbf{L}\left[\frac{\varphi(y_i)}{f_i(x_i)}\right] \tag{30}$$

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

$$\begin{aligned}
&\frac{\partial}{\partial z_j}\left(\frac{\partial x_i}{\partial z_i} = \frac{\varphi(z_i)}{f(x_i)}\right) \\
&\Rightarrow \frac{\partial^2 x_i}{\partial z_i \partial z_j} = \frac{\partial \varphi(z_i)}{\partial z_j}\cdot\frac{1}{f(x_i)} + \frac{\partial}{\partial z_j}\left(\frac{1}{f(x_i)}\right)\cdot\varphi(z_i) \\
&\Rightarrow \frac{\partial^2 x_i}{\partial z_i \partial z_j} = \frac{\partial \varphi(z_i)}{\partial z_j}\cdot\frac{1}{f(x_i)} + \frac{\partial}{\partial x_j}\left(\frac{1}{f(x_i)}\right)\cdot\frac{\partial x_j}{\partial z_j}\cdot\varphi(z_i) \\
&\Rightarrow \frac{\partial^2 \mathbf{x}}{\partial \mathbf{z}^2} = \left[\frac{\partial \varphi(z_i)}{\partial z_i}\cdot\frac{1}{f(x_i)} - \frac{1}{f(x_i)^2}\cdot\frac{\partial f(x_i)}{\partial x_i}\cdot\frac{\partial x_i}{\partial z_i}\cdot\varphi(z_i)\right]
\end{aligned} \tag{31}$$

Transformation of Dependent Random Variables: Rosenblatt

An alternative to the Nataf approach is to consider the joint PDF of **x** as a product of conditional PDFs:

$$f(\mathbf{x}) = f_1(x_1)\cdot f_2(x_2 | x_1)\cdots f_n(x_n | x_1, x_2, \ldots, x_{n-1}) \tag{32}$$

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

$$
\begin{aligned}
F(x_1) &= \int_{-\infty}^{x_1} f_1(x_1)\,dx_1 \\
F(x_2|x_1) &= \int_{-\infty}^{x_2} f_2(x_2|x_1)\,dx_2 \\
F(x_3|x_1,x_2) &= \int_{-\infty}^{x_3} f_3(x_3|x_1,x_2)\,dx_3 \\
&\vdots
\end{aligned}
\tag{33}
$$

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

$$
\begin{aligned}
y_1 &= \Phi^{-1}\left(F_1(x_1)\right) \\
y_2 &= \Phi^{-1}\left(F_2(x_2|x_1)\right) \\
y_3 &= \Phi^{-1}\left(F_3(x_3|x_1,x_2)\right) \\
&\vdots
\end{aligned}
\tag{34}
$$

To obtain the inverse transformation it is necessary to solve nonlinear equations for x_i, starting at the top of Eq. (34). The Jacobian matrix for this transformation is

$$
\mathbf{J}_{\mathbf{y},\mathbf{x}} \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix}
\dfrac{f_1(x_1)}{\varphi(y_1)} & 0 & 0 & 0 \\
\dfrac{1}{\varphi(y_2)}\dfrac{\partial F_2(x_2|x_1)}{\partial x_1} & \dfrac{f_2(x_2|x_1)}{\varphi(y_2)} & 0 & 0 \\
\dfrac{1}{\varphi(y_3)}\dfrac{\partial F_3(x_3|x_1,x_2)}{\partial x_1} & \dfrac{1}{\varphi(y_3)}\dfrac{\partial F_3(x_3|x_1,x_2)}{\partial x_2} & \dfrac{f_3(x_3|x_1,x_2)}{\varphi(y_3)} & 0 \\
\vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\tag{35}
$$

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.