An Approximate Method to Represent the Extreme Values of Non-Stationary Gaussian White Noise

H. Morikawa[1)]

Associate Professor, Department of Built Environment, Tokyo Institute of Technology, Japan
morika@enveng.titech.ac.jp

Abstract: The statistics for extreme values of non-stationary processes are critical to designing structures in some engineering fields, such as earthquake engineering, coastal engineering, wind engineering, and so on. However, it is not easy to estimate the extreme values of non-stationary processes whose stochastic properties depend on time, because we have to deal with the i.n.n.i.d. (independent not necessarily identically distributed) random variables to solve problems of this type. Thus, we will discuss a probabilistic distribution for the extreme values of non-stationary Gaussian white noise as the simplest and most primary problem for i.n.n.i.d. random variables. Firstly, the closed form solutions are derived for the extreme values of i.n.n.i.d. Gaussian variables with two different properties and the qualitative properties are determined. Next, we propose an approximate representation of the distribution for the extreme values of i.n.n.i.d. random variables using these obtained properties, and confirm the appropriateness of the result through Monte Carlo simulations.

1. INTRODUCTION

The statistics for extreme values of non-stationary processes are critical to designing structures in some engineering fields, such as earthquake engineering, coastal engineering, and so on. Thus, many researchers have proposed various methods for this purpose. For example, Vanmarcke (1972) has developed a method to estimate the extreme values of a given system's response to random excitation using the spectral moments in a frequency domain. Furthermore, this method was extended by Kiureghian (1980).

While we may deal with this problem in time domain, most research on this type of problem has been limited to stationary processes. Especially, in a case where a time series is stationary Gaussian white noise with zero mean, we can directly apply the asymptotic representation for the extreme values of i.i.d. (independent identically distributed) Gaussian variables. Thus, the closed form solutions are easily obtained. This asymptotic representation of extreme values was introduced to the engineering fields by Gumbel and most classic and basic formulation as known as Gumbel's distribution (Gumbel 1958, Galambos 1978).

However, it is not easy to estimate the extreme values of non-stationary processes whose stochastic properties depend on time, because we have to deal with the i.n.n.i.d. (independent not necessarily identically distributed) random variables in a case of the simplest problem such as white noise. Although general representations for this type of problems can be obtained (Reise 1989, Ahsanullah and Nevzorov 2001), it is difficult to derive the closed form or asymptotic solutions for any specific distributions such as Gaussian distribution, etc. If such solutions

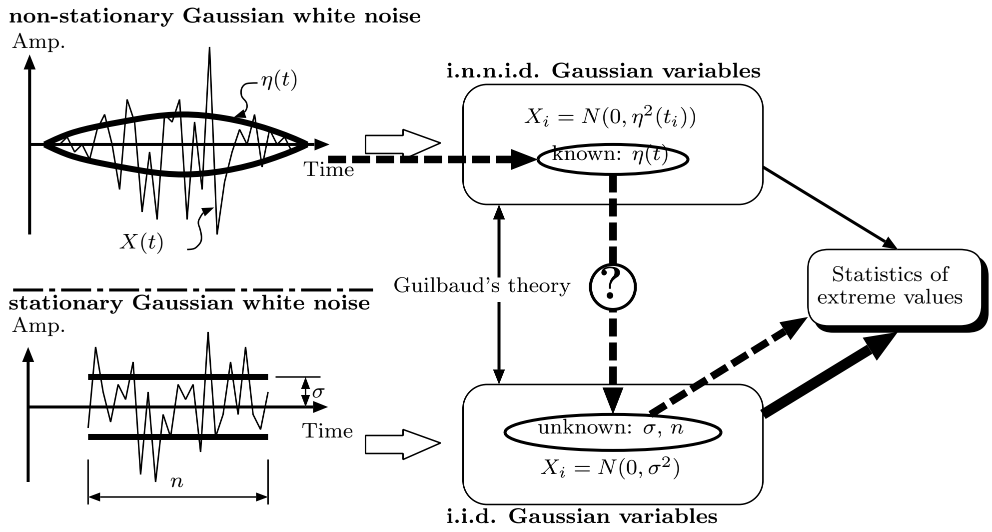

Figure 1 Schematic diagram to show the relationship between i.n.n.i.d. and i.i.d. Gaussian variables and concept to derive the statistics of extreme values of non-stationary Gaussian white noise.

are derived, their representations will be complicated and it is not suitable to apply them to the problems in the engineering fields.

The asymptotic representation of extreme values for i.i.d. random variable were derived on the basis of the ingenious ideas. As following this way, it is important to find an approximate representation for the extreme values of i.n.n.i.d. random variables using simple formulations. Unfortunately, nobody can propose appropriate representations for this type of problems, even though the problem is described for i.n.n.i.d. Gaussian variables.

Therefore, we will discuss a probabilistic distribution for the extreme values of non-stationary Gaussian white noise as the simplest and most primary problem: the extreme values of i.n.n.i.d. Gaussian are treated. Furthermore, we will limit the property of non-stationarity to the simple case keeping the application to the earthquake ground motion in mind: specifically, we will deal with the discrete white noise whose mean is zero, and standard deviation depend on time. The standard deviation has one peak and predominates the peak value over the time. Hereafter, we call "white noise" instead of the "discrete white noise" for the simplicity.

2. PROBLEM SETTING

We will deal with the asymptotic distribution, $F_Y(y)$ for maximum value of i.n.n.i.d. Gaussian variables X_i $(i = 1, 2, \ldots)$, as the simplest non-stationary process: that is,

$$X_i \equiv X(t_i) = \eta(t_i) \cdot W(t_i) \tag{1}$$

where, t_i stands for i-th discrete time, $W(t_i)$ for Gaussian white noise with zero mean and unit variance, and $\eta(t_i)$ for standard deviation which depends on time and varies smoothly with one extreme peak and $\eta(t; t \leq 0) = \eta(\infty) = 0$. The variations of $\eta(t_i)$ will be set much smaller than the time increments Δt. It is noted that $\eta(t_i)$ will play the role of a kind of envelop function of $X(t_i)$.

Generally speaking, the order statistics of i.n.n.i.d. random variables can be represented by using that of i.i.d. random variables because of the Guilbaud's theory (Reiss 1989). However, this theory does not give any information how we can find the appropriate i.i.d. random variable corresponding with i.n.n.i.d. random variables with a specific probability distribution of X_i. As shown as question mark, ? in Figure 1, although the asymptotic distribution should be replaced by one of i.i.d. random variables in a case where the function $\eta(t)$, which is the probabilistic characteristics of i.n.n.i.d. random variable X_i, is given, there is no way to find parameters for the corresponding i.i.d. random variables.

To find the appropriate parameters for i.i.d. random variables are easier than to derive directly any asymptotic representation for the extreme values of i.n.n.i.d. random variables, because the asymptotic representation are already obtained for the extreme values for i.i.d. random variables.

From the above discussion, we will consider the approximate representations for the maximum values of non-stationary Gaussian white noise following the thick arrows in Figure 1. In this procedure, most significant problem is to represent the relationships between the parameters of i.i.d. and i.n.n.i.d. random variables. Thus, we will concentrate our concern into this problem, that is, to find the relationships as shown as ? in Figure 1.

Firstly, it will be confirmed that we can replace i.n.n.i.d. variables with i.i.d. variables. Then, we will derive analytically the asymptotic representation for the maximum values of i.n.n.i.d. Gaussian variables with two distributions: we consider a case where $\eta(t_i)$ takes only two values. From this analysis, we discuss the relationships between the parameters for two types of variables and determine the qualitative properties to examine the possibility of the approximate representation for maximum values of i.n.n.i.d. variables. Finally, we will propose an approximate representation for our problem using the properties obtained from the above discussion, and confirm the appropriateness of the result through Monte Carlo simulations.

3. DISTRIBUTION FOR EXTREME VALUES OF GAUSSIAN VARIABLES WITH TWO DIFFERENT PROPERTIES

Let us consider X_i $(i = 1, 2, \ldots, n_1 + n_2)$ which consists of n_j independent Gaussian variables, X_{jk} with zero mean and variance σ_j^2 $(j = 1, 2,\ k = 1, 2, \ldots, n_j)$: that is, X_{jk} is $N(0, \sigma_j^2)$ and X_i should be X_{1k} or X_{2k}. Since X_i $(i = 1, 2, \ldots, n_1 + n_2)$ are independent mutually, we can renumber X_i without loss of generality. Thus, let us set X_{1k} $(k = 1, \ldots, n_1)$ for X_i $(i = 1, \ldots, n_1)$ and X_{2k} $(k = 1, \ldots, n_2)$ for X_i $(i = n_1 + 1, \ldots, n_2)$.

Then, the probability distributions for maximum value Y_j of X_{jk} $(j = 1, 2)$ can be approximately written for large n_j as follows (Ahsanullah and Nevzorov 2001):

$$F_{Y_j}(y) = P(X_{jk} < y) \approx \exp\left[-\exp[-\alpha_j(y - u_j)]\right], \tag{2}$$

where $P(A)$ denotes the probability of A, and

$$\alpha_j = \sqrt{2\ln n_j}/\sigma_j \tag{3a}$$

$$u_j = \left\{\sqrt{2\ln n_j} - \frac{\ln(\ln n_j) + \ln(4\pi)}{2\sqrt{2\ln n_j}}\right\}\sigma_j. \tag{3b}$$

Thus, the probability distribution for maximum values of X_i is represented as

$$F_Y(y) = P(X_i < y) = \prod_{j=1}^{2} P(X_{jk} < y) = \prod_{j=1}^{2} F_{Y_j}(y) \approx \exp\left[-\sum_{j=1}^{2} \exp[-\alpha_j(y - u_j)]\right]. \tag{4}$$

Eq.(4) gives an approximate representation from the meaning of the asymptotic distribution for a special case of i.n.n.i.d. random variable with large n_j.

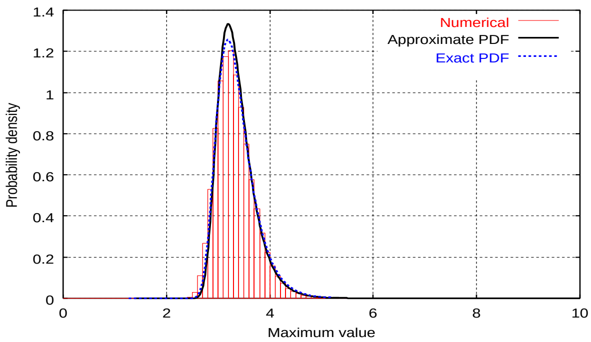

Figure 2 An example of distribution for extreme values of Gaussian white noise ($n_1 = n_2 = 500$, $\sigma_1 = 1.0$, $\sigma_2 = 1.05$).

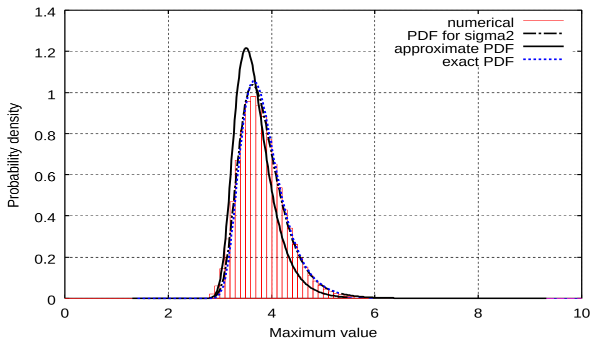

Figure 3 An example of distribution for extreme values of Gaussian white noise ($n_1 = n_2 = 500$, $\sigma_1 = 1.0$, $\sigma_2 = 1.25$).

According to Guilbaud's theory, Eq.(4) can be represented by the asymptotic distribution for the maximum values of i.i.d. random variable: namely, Eq.(4) can be replaced by

$$F_Y(y) \approx \exp[-\exp[-\alpha(y-u)]]. \tag{5}$$

As pointed out above, this theory does not give any information about the relationships between α and u of Eq.(5), and α_j and u_j $(j = 1, 2)$ of Eq.(2). Thus, we will discuss how α and u can be represented by α_j and u_j $(j = 1, 2)$ in this section.

Calculating the Eq.(4) with various values of the parameters n_j and σ_j $(j = 1, 2)$ of Eqs.(3a) and (3b), we obtained the following relationships as $\sigma_1 \approx \sigma_2$ and $n_1 \approx n_2$ by trial and error:

$$\alpha \approx \sqrt{2\ln n}/\sigma \tag{6a}$$

$$u \approx \left\{ \sqrt{2\ln n} - \frac{\ln(\ln n) + \ln(4\pi)}{2\sqrt{2\ln n}} \right\} \sigma, \tag{6b}$$

where

$$n = \sum_{j=1}^{2} n_j \tag{7a}$$

$$\sigma = \frac{\sum_{j=1}^{2} n_j \sigma_j}{n}. \tag{7b}$$

Although this is not the mathematical consequence, these results may be expected instinctively under the above condition of σ_j and n_j. σ of Eq.(7b) is given by the weighted mean of σ_j with respect to n_j. Considering the general characteristics of the standard deviation, σ^2 should be represented by the weighted mean of σ_j^2 $(j = 1, 2)$ with respect to n_j^2, though Eq.(7b) gives good approximation as $\sigma_1 \approx \sigma_2$ as $n_1 \approx n_2$.

In other cases such as $\sigma_\ell \ll \sigma_j$ $(\ell, j = 1, 2,\ \ell \neq j)$ and $n_1 \approx n_2$, $\alpha \approx \alpha_j$ and $u \approx u_j$ can be applied. This means that the distribution function, $F_Y(y)$ for the maximum values of i.n.n.i.d. Gaussian variables, X_i is approximately rewritten by the distribution for the maximum values of i.i.d. Gaussian variables with larger values of σ_j. Furthermore, in a case where σ_j is more than only 1.1 to 1.2 times of σ_ℓ, the effect from the maximum values of X_i with σ_ℓ is negligible. Thus, it is enough to treat the two cases of $\sigma_1 \approx \sigma_2$ and $\sigma_\ell \ll \sigma_j$ as $n_1 \approx n_2$.

Figures 2 and 3 show the numerical examples of the approximation of probability distribution for maximum value, Y, of X_i as $n_1 = n_2$. In this figures, the histogram of Y, which are obtained from the Monte Carlo simulation of 10000 times, is also shown. Figure 2 gives

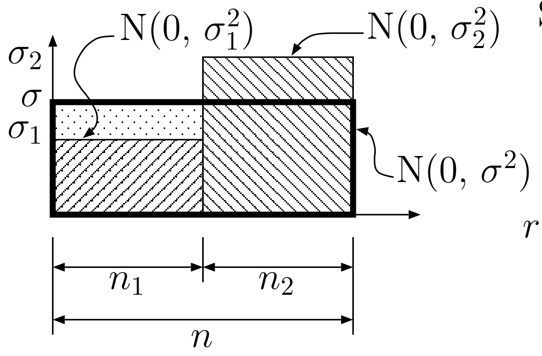

Figure 4 Concept to replace i.n.n.i.d. Gaussian variable with i.i.d. Gaussian variable.

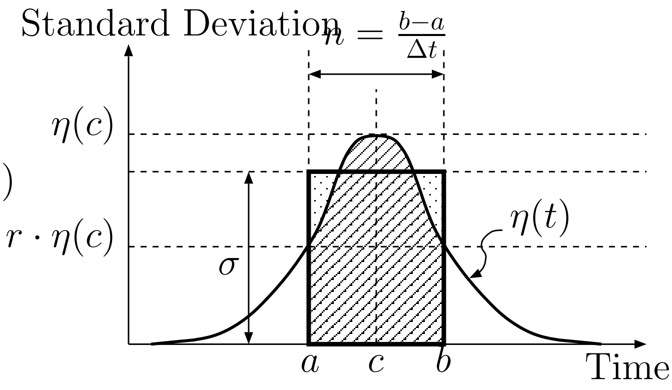

Figure 5 Concept to calculate the distribution of extreme value of non-stationary Gaussian white noise using the i.i.d. Gaussian variable.

the result as $\sigma_1 \approx \sigma_2$ and it is observed that the shapes of Eq.(4) and Eq.(5) obtained by using Eqs.(6a) and (6b) coincide. On the other hand, Figure 3 shows the case for $\sigma_1 \ll \sigma_2$. In this case, Eq.(4) coincides with the distribution for the maximum value of X_i with σ_2: $F_{Y_2}(y)$. However, the distribution estimated by the weighted mean of σ_j $(j = 1, 2)$ fails to represent the histogram.

From the above numerical calculations, we can confirm the Guilbaud's theory: the probability distribution for maximum values of i.n.n.i.d Gaussian variables are replaced by one of i.i.d. Gaussian distribution. Furthermore, the results give an instructions how to determine the values of α and u of Eq.(5).

As a result, we can conclude the method to determine the parameters for substitute i.i.d. distribution as follows. In a case of $\sigma_1 \approx \sigma_2$, we can choose the value of σ to satisfy the equation

$$\sigma \left(\sum_{j=1}^{2} n_j \right) = \sum_{j=1}^{2} (n_j \sigma_j) . \tag{8}$$

This means that the area obtained by σ and $n = \sum_{j=1}^{2} n_j$ should be same as the total area from n_j and σ_j $(j = 1, 2)$ as shown in Figure 4. On the other hand, in a case of $\sigma_\ell \ll \sigma_j$ $(\ell, j = 1, 2;$ $\ell \neq j)$, we can use the probability distribution for the maximum value, Y_j, of i.i.d. Gaussian variable with $N(0, \sigma_j^2)$.

4. APPROXIMATE DISTRIBUTION FOR EXTREME VALUES OF NON-STATIONARY GAUSSIAN WHITE NOISE

We discussed the relationships between the parameters of i.n.n.i.d. and i.i.d. Gaussian variable for a special case in the previous section. Then, we will apply the obtained properties to approximate the probabilistic characteristics for maximum values of Eq.(1). Since we assume the standard deviation, $\eta(t)$ varies smoothly over the time as shown in Figure 5, we can approximate $\eta(t_i) \approx \eta(t_{i+1})$ at t_i と $t_{i+1} = t_i + \Delta t$, respectively, where Δt stands for the small increment of the discrete time. This suggests the possibility that the probability distribution for the maximum value, $F_Y(y)$, of X_i of Eq.(1) will be replaced with the asymptotic distribution for an i.i.d. Gaussian variable of Eq.(5) with parameters given by Eqs.(6a) and (6b).

On the basis of this daring (and mathematically baseless) assumption, we will determine the parameters for an i.i.d. Gaussian variable substituting the i.n.n.i.d. Gaussian variable. The parameters to determine are σ and n of Eqs.(6a) and (6b). Since we considered the area formed by the number of variables and standard deviation to determine σ as shown in Figure 4, the same concept will be introduced as shown in Figure 5. The remaining part of this section is devoted to explain the procedure to obtain the probability distribution approximately using the Figure 5.

Let us consider $\eta(t)$ takes maximum value $\eta(c)$ at $t = c$. Then, introducing a real number r $(0 < r < 1)$, we will determine parameters a and b which satisfy $r \cdot \eta(c) = \eta(a) = \eta(b)$, where $a < c < b$. The area of $\eta(t)$, S_r, are obtained as a function of r at $[a, b]$. To replace the probability distribution for the maximum value of i.n.n.i.d. Gaussian variable with one of i.i.d. Gaussian variable, we will consider a Gaussian variable with constant standard deviation at $[a, b]$. For the standard deviation of this i.i.d. Gaussian variable, we adopt the height σ of the rectangle whose area and length of the base are S_r and $b - a$, respectively. Applying the obtained σ and $n = (b - a)/\Delta t$ to Eqs.(6a) and (6b) and using Eq.(5), we can obtain the approximate probability distribution for the maximum values of i.n.n.i.d. Gaussian variable X_i.

The above procedure is rewritten mathematically as follows: the parameters a and b are determined by

$$a = \sup_{t<c}\{t; \eta(t) = r\eta(c)\} \tag{9a}$$

$$b = \inf_{t>c}\{t; \eta(t) = r\eta(c)\}, \tag{9b}$$

where $\eta(a) = \eta(b)$. Then, the area surrounded by $\eta(t)$ at $[a, b]$ is

$$S_r = \int_a^b \eta(t)dt. \tag{10}$$

n is obtained from $n = (b - a)/\Delta t$, and σ is determined by

$$\sigma = \frac{S_r}{n}. \tag{11}$$

Substituting n and σ for Eq.(5) derives the approximate probability distribution for the maximum value of X_i.

In this procedure, we did not mention the value of r. Generally speaking, since any function can be used for $\eta(t)$, we cannot examine the sensitivity of Eq.(5) with respect to r, exhaustively. We can say, in our calculations, Eq.(5) is not sensitive toward r. However, in a case where r is too small, the approximation of $\eta(t_i) \approx \eta(t_{i+1})$ is not assured. Thus, we can use the value of 0.8 to 0.9 as r. Since this suggests that the optimal value of r may depend on the time increment, Δt, Δt should be small with respect to the variance of $\eta(t)$ to satisfy the approximation $\eta(t_i) \approx \eta(t_{i+1})$.

5. A NUMERICAL EXAMPLE

To confirm the availability of the proposed method, we carried out a simple numerical calculation using the Monte Carlo simulation. The pseudo-random numbers are generated by the Mersenne Twister (Matsumoto and Nishimura 1998) and the Fortran code based on this method (Matsumoto 2002) are used. Since the generated random values follow the uniform distribution, they are transformed to Gaussian distribution (Evans et al. 1993). In this calculation, $\Delta t = 0.01$ are used as small value regarding to the variance of $\eta(t)$.

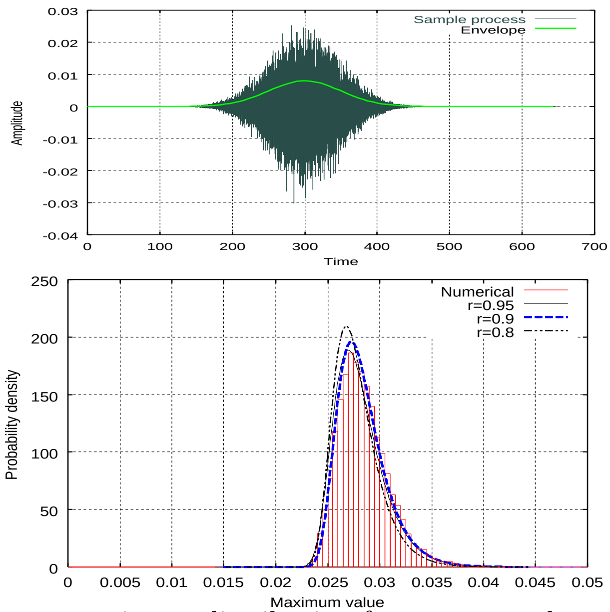

Figure 6 An example of the approximate distribution for extreme values of non-stationary Gaussian white noise. The upper panel shows a sample process and $\eta(t) = \frac{1}{\sqrt{2\pi}50} \exp\left[-\frac{1}{2}\left(\frac{t-300}{50}\right)^2\right]$. The approximate distributions for $r = 0.95$, 0.9, and 0.8 are compared with the result of numerical simulation in the lower panel.

Table 1 Comparison of the parameters for Gumbel's distribution for Eq.(5)

	Numerical	$r = 0.95$	$r = 0.9$	$r = 0.8$
u	0.02733	0.02699	0.027270	0.02678
α	530.74	512.65	532.90	569.92

Figure 6 shows an example of the probability distribution for maximum value of X_i with $\eta(t)$ which is Gaussian type function. The upper panel of this figure compares a realization with the shape of $\eta(t)$. In the lower panel, the estimated distribution are shown with the histogram obtained from Monte Carlo simulation of 10000 times. The lines show the results from the different value of r of Eqs.(9a) and (9b): 0.8, 0.9, 0.95. It is noted that these lines approximate the histogram well. This means that the proposed method gives good approximation to represent the probability distribution for maximum value of i.n.n.i.d. Gaussian value.

Furthermore, to examine the accuracy of the approximation, the values of parameters α and u of Eq.(5) are listed in Table 1. From this, it is observed that the statistical parameters are consistent values with the results from Monte Carlo simulation. This means that the we can set the value of r roughly, because the approximation is not so sensitive to r.

6. CONCLUSIONS

We have analytically derived the asymptotic representation for maximum values of Gaussian variable with two different properties and found qualitative properties. Using these properties, an approximate representation were proposed for maximum values of non-stationary Gaussian white noise and the appropriateness was also confirmed through the numerical simulation.

Acknowledgments:

The author wish to express our special thanks to Professor A. Zerva (Drexel University, USA) for reading the draft and making a number of valuable advice and helpful suggestions. This research was partially supported by the Assistance to International Exchange of Scientist Fund of the Kajima Foundation, and the Ministry of Education, Science, Sports and Culture, Grants-in-Aid for the Encouragement of Young Scientists (A), No. 15681010, 2003–2005.

References:

Ahsanullah, M. and Nevzorov, V.B. (2001), "*Ordered Random Variables*," Nova Science Publishers.

Evans, M., Hastings, N., and Peacock, B. (1993), "*Statistical Distributions*," 2nd edition, John Wiley & Sons.

Galambos, J. (1978), "*The Asymptotic Theory of Extreme Order Statistics*," John Wiley & Sons.

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

Kiureghian, A.D. (1980), "Structural response to stationary excitation," *Journal of Engineering Mechanics Division*, Proc. of ASCE, **106**(EM6), 1195–1213.

Matsumoto, M. and Nishimura, T. (1998), "Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator," *ACM Trans. on Modeling and Computer Simulation*, **8**(1), 3–30.

Matsumoto, M. (2002), `http://www.math.keio.ac.jp/~matumoto/mt.html`

Reiss, R.-D. (1989), "*Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*," Springer Verlag, p.36.

Vanmarcke, E.H. (1972), "Properties of spectral moments with applications to random vibration," *Journal of Engineering Mechanics Division*, Proc. of ASCE, **98**(EM2), 425–446.