



12th International Conference on Vibration Problems, ICOVP 2015

Multivariate Extreme Value Distributions for Vector of Non-stationary Gaussian Processes

Vighnesh Ambetkar^{a,c}, Ramakrishna Kuppa^b, Sayan Gupta^{a,*}

^aDepartment of Applied Mechanics, Indian Institute of Technology Madras, Chennai 600 036, India

^bDepartment of Mechanical Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, 500 047, India

^cDepartment of Ocean Engineering, Indian Institute of Technology Madras, Chennai 600 036, India

Abstract

The focus of this study is on estimating the multivariate extreme value distributions associated with a vector of mutually correlated non-stationary Gaussian processes. This involves computing the joint crossing statistics of the vector processes by assuming the crossings to be Poisson counting processes. A mathematical artifice is adopted to take into account the dependencies that exist between the crossings of the processes. The crux in the formulation lies in the evaluation of a four-dimensional integral, which can be computationally expensive. This difficulty is bypassed by using saddlepoint approximation to reduce the dimension of the integral to be numerically computed to just two. The developments are illustrated through a numerical example and are validated using Monte Carlo simulations.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the organizing committee of ICOVP 2015

Keywords: Multivariate extreme value distribution; Non-stationary Gaussian processes; Poisson counting process; Saddlepoint approximation

1. Introduction

Time variant reliability analysis of structural systems is usually studied in the time invariant format by defining the problem in terms of random variables that represent the extreme values of the response within a specified time interval. The focus therefore is on estimating the extreme value distributions of the response. For structural systems where the response constitutes a vector of correlated processes, estimates of the system reliability can be obtained from the knowledge of the joint multivariate extreme value distributions of the vector processes.

The problem of extreme value distributions for a vector of stationary Gaussian processes was earlier studied in [1]. This involved approximating the multivariate counting process associated with the level crossings as a multivariate Poisson random process. The successful development of the formulation required the evaluation of a six-dimensional integral, which was shown to be reduced to a two dimensional one using simplifying operations. This double integral was numerically evaluated and approximations for the multivariate extreme value distributions were obtained. Efforts to extend this methodology for non-stationary vector Gaussian processes was not possible as the non-stationary nature

* Corresponding author. Tel.: +91-44-22574055.

E-mail address: sayan@iitm.ac.in

of the processes did not afford reduction in the dimension of the integrals [2] and as a result, the analytical formulation was not numerically viable. The present study bypasses these difficulties by taking advantage of the saddlepoint approximation method [3] to bring about a reduction in the dimension of the integrals, making the analytical approach for estimating the joint crossing statistics, computationally efficient even for non-stationary vector Gaussian processes.

2. Problem Statement

Consider $\{X_i(t)\}_{i=1}^k$ to be a vector of correlated non-stationary Gaussian random processes which are expressible in the form,

$$X_i(t) = e_i(t)X_{is}(t), \quad (1)$$

where, $X_{is}(t)$ is a stationary Gaussian random process and $e_i(t)$ is a deterministic envelope function, of the form

$$e_i(t) = a_i[\exp(-b_it) - \exp(-c_it)]. \quad (2)$$

Here, the parameters b_i and c_i determine the shape of $e_i(t)$ and a_i is a normalization factor such that $\max[e_i(t)] = 1.0$. The components $X_{is}(t)$ are assumed to be mutually correlated stationary Gaussian processes, whose spectral properties are defined in terms of the power spectral density (PSD) matrix $\mathbf{S}(\omega)$ or the covariance matrix $\mathbf{R}(\tau)$. For each $X_i(t)$, let us define $N_i(\alpha_i, 0, T)$ to be the number of upcrossings of level α_i in the time interval $[0, T]$ and $X_{mi} = \max_{0 \leq t \leq T} X_i(t)$ is defined to be the maxima of $X_i(t)$ in the time interval $[0, T]$. For a given i , $N_i(\alpha_i, 0, T)$ and X_{mi} are random variables.

For scalar processes where $X_i(t)$ are Gaussian, it has been shown that X_{mi} follows Gumbel distributions. The problem of estimating the joint probability distribution function (PDF) of $\mathbf{X}_m = \{X_{mi}\}_{i=1}^k$, when $\{X_i(t)\}_{i=1}^k$ constitute vector correlated stationary Gaussian processes has been discussed in [1]. For the sake of completion, the salient steps of the formulation are explained in sections 3 and 4.

3. Level Upcrossings

As previously defined, $\{N_i(\alpha_i, 0, T)\}_{i=1}^k$ is assumed to be a vector of multivariate Poisson random variables. For simplicity, let us consider the case of a bivariate process and define three mutually independent Poisson random variables U_1, U_2 and U_3 , and λ_1, λ_2 and λ_3 , respectively represent their respective parameters. Introducing the following transformations,

$$\begin{aligned} N_1(\alpha_1, 0, T) &= U_1 + U_3, \\ N_2(\alpha_2, 0, T) &= U_2 + U_3, \end{aligned} \quad (3)$$

it can be shown that, $N_1(\alpha_1, 0, T)$ and $N_2(\alpha_2, 0, T)$ are Poisson random variables with parameters $(\lambda_1 + \lambda_3)$ and $(\lambda_2 + \lambda_3)$, respectively. It can be further shown that λ_3 is the covariance of $N_1(\alpha_1, 0, T)$ and $N_2(\alpha_2, 0, T)$. Based on this, we can write

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{Bmatrix} = \begin{Bmatrix} \langle N_1(\alpha_1, 0, T) \rangle \\ \langle N_2(\alpha_2, 0, T) \rangle \\ \text{Cov}[N_1(\alpha_1, 0, T), N_2(\alpha_2, 0, T)] \end{Bmatrix}, \quad (4)$$

where, $\langle \cdot \rangle$ denotes the mathematical expectation operator and $\text{Cov}[\cdot]$ denotes the covariance function. The expectation of the counting process, $\langle N_i(\alpha_i, 0, T) \rangle$ can be computed by integrating the mean upcrossing intensity with respect to time. The details of this can be found in section 5.

The covariance of N_1 and N_2 can be expressed as

$$\text{Cov}[N_1, N_2] = \langle N_1(\alpha_1, 0, T)N_2(\alpha_2, 0, T) \rangle - \langle N_1(\alpha_1, 0, T) \rangle \langle N_2(\alpha_2, 0, T) \rangle. \quad (5)$$

The evaluation of joint expectation $\langle N_1(\alpha_1, 0, T)N_2(\alpha_2, 0, T) \rangle$ is central to this study since it involves evaluation of four dimensional integration which can be computationally demanding. A detailed discussion about evaluation of joint expectation is presented sections 6 and 7.

4. Extreme value distribution

The bivariate PDF of level crossings can be derived using the corresponding characteristic function and is given by [4]

$$P[(N_1 = j) \cap (N_2 = l)] = \exp[-(\lambda_1 + \lambda_2 + \lambda_3)] \sum_{i=0}^{\min(j,l)} \frac{\lambda_1^{(j-i)} \lambda_2^{(l-i)} \lambda_3^i}{(j-i)!(l-i)!i!}. \quad (6)$$

For the case of $j = 0$ and $l = 0$, we get

$$P[(N_1 = 0) \cap (N_2 = 0)] = \exp[-(\lambda_1 + \lambda_2 + \lambda_3)]. \quad (7)$$

The above result can be generalized for the case of $k > 2$. The number of mutually independent Poisson random variables required for any k is given by $C_1^k + C_2^k$, where C_r^n is the number of combinations of n random variables taken r at a time. The joint distribution of the extreme values X_{m_i} is related to the probability of the respective number of level crossings of levels α_i being zero and is given by the relation

$$P\left[\bigcap (X_{m_i} \leq \alpha_i)\right] = P\left[\bigcap \{N_i(\alpha_i, 0, T) = 0\}\right]. \quad (8)$$

Using Eqs. (7) and (8), the joint extreme value distribution can be given in terms of the parameters λ_i , $i = (1, \dots, k)$ of the independent Poisson random variables. Thus, for a k -dimensional vector of Gaussian random processes $\{X_i(t)\}_{i=1}^k$, the joint distribution is of the form

$$P_{\mathbf{X}_m}(\alpha_1, \dots, \alpha_k) = \exp\left[-\sum_{i=1}^{k(k+1)/2} \lambda_i\right]. \quad (9)$$

The marginal distribution of the extreme values can be derived from the joint distribution and is expressible as

$$P_{X_{m_j}}(\alpha_j) = P_{X_{m_1}, \dots, X_{m_k}}(\infty, \dots, \alpha_j, \dots, \infty). \quad (10)$$

The above formulation is applicable in general for any random processes irrespective of whether they are stationary or Gaussian. The crux however lies in estimating the parameters λ_i , which from the property of Poisson random variables are known to be equal to their mean. In the following section, we discuss how these parameters can be estimated in general, and more specifically, when the processes are non-stationary and Gaussian.

5. Evaluation of mean upcrossings

As discussed in section 3, $\langle N(\alpha, 0, T) \rangle$ represents the expected number of upcrossings for the given random process corresponding to a level α in the interval $[0, T]$. In general, for any non-stationary process, $X(t)$, $\langle N(\alpha, 0, T) \rangle$ is given by,

$$\langle N(\alpha, 0, T) \rangle = \int_0^T v_X^+(\alpha, t) dt, \quad (11)$$

where $v_X^+(\alpha, t)$ is the upcrossing intensity corresponding to level α . $v_X^+(\alpha, t)$ is evaluated by well-known Rice's formula [5, 6] as

$$v_X^+(\alpha, t) = \int_0^\infty \dot{x} p_{X(t)\dot{X}(t)}(\alpha, \dot{x}; t) \cdot d\dot{x} \quad (12)$$

The difficulty in estimating the upcrossing intensity using Eq. (12) is that the knowledge of the joint probability density function (pdf) of the process and its instantaneous time derivative is in general not available. However, when $X(t)$ is Gaussian, $X(t)$ and $\dot{X}(t)$ are independent and the joint pdf in Eq. (12) can be expressed as the product of

their marginal pdf. This enables obtaining the closed form expressions for $\nu_X^+(\alpha, t)$. For a zero mean non-stationary Gaussian random process $X(t)$ with variance $\sigma_X^2(t)$ and its time derivative $\dot{X}(t)$ with variance $\sigma_{\dot{X}}^2(t)$, the upcrossing intensity for $X(t)$ corresponding to level α is given by [7]

$$\nu_X^+(\alpha, t) = \sqrt{\left(\frac{1 - \rho_1^2(t)}{2\pi}\right)\left(\frac{\sigma_X(t)}{\sigma_X(t)}\right)} \exp\left(-\frac{\alpha^2}{2\sigma_X^2(t)}\right) \left[\phi(q(t)) + q(t)\Phi(q(t))\right], \tag{13}$$

where,

$$\rho_1(t)\sigma_X(t)\sigma_{\dot{X}}(t) = \langle X(t)\dot{X}(t) \rangle,$$

$$q(t) = \frac{\rho_1(t)}{\sqrt{1 - \rho_1^2(t)}} \left(\frac{\alpha}{\sigma_X(t)}\right).$$

and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and PDF respectively. $\rho_1(t)$ is the time-varying correlation coefficient between the processes $X(t)$ and $\dot{X}(t)$. Substituting Eq. (13) in Eq. (11), $\langle N(\alpha, 0, T) \rangle$ can be obtained by numerical integration.

6. Estimation of joint crossings

As discussed in section 3, the evaluation of the covariance term, and thus expectation of joint crossings is central to the study of multivariate extreme value distributions. For the bivariate case, the joint expectation $\langle N_1 N_2 \rangle$, corresponding to upcrossings of levels α_i , is expressed as

$$\langle N_1 N_2 \rangle = \int_0^T \int_0^T \int_0^\infty \int_0^\infty \left\{ \dot{x}_1 \dot{x}_2 p_{X_1 X_2 \dot{X}_1 \dot{X}_2}(\alpha_1, \alpha_2, \dot{x}_1, \dot{x}_2; t_1, t_2) \right\} d\dot{x}_1 d\dot{x}_2 dt_1 dt_2. \tag{14}$$

Let the inner double integral in Eq. (14) be $I(t_1, t_2)$. Assuming that the processes and their time derivatives are jointly Gaussian and non-stationary, the joint pdf can be expressed as

$$p_{X_1 X_2 \dot{X}_1 \dot{X}_2}(\alpha_1, \alpha_2, \dot{x}_1, \dot{x}_2; t_1, t_2) = \frac{1}{4\pi^2 \sqrt{\det(\Delta)}} \exp\left\{-0.5\{Y^T \Delta^{-1} Y\}\right\}, \tag{15}$$

where, $Y^T = [\alpha_1, \alpha_2, \dot{x}_1, \dot{x}_2]^T$ and $\Delta(t_1, t_2)$ is the 4×4 covariance matrix. Here, $\det(\cdot)$ denotes the determinant of the square matrix Δ representing the covariance matrix and is a function of t_1 and t_2 . Upon simplification, Eq. (15) reduces to

$$p_{X_1 X_2 \dot{X}_1 \dot{X}_2}(\alpha_1, \alpha_2, \dot{x}_1, \dot{x}_2; t_1, t_2) = \frac{1}{4\pi^2 \sqrt{\det(\Delta)}} \exp\left\{-0.5\{q_1 + 2q_2\dot{x}_1 + 2q_3\dot{x}_2 + q_4\dot{x}_1^2 + q_5\dot{x}_2^2 + 2q_6\dot{x}_1\dot{x}_2\}\right\}, \tag{16}$$

where,

$$\begin{aligned} q_1(t_1, t_2) &= (J_{11}\alpha_1^2 + 2J_{12}\alpha_1\alpha_2 + J_{22}\alpha_2^2), \\ q_2(t_1, t_2) &= (J_{13}\alpha_1 + J_{23}\alpha_2), \\ q_3(t_1, t_2) &= (J_{14}\alpha_1 + J_{24}\alpha_2), \\ q_4(t_1, t_2) &= J_{33}, \\ q_5(t_1, t_2) &= J_{44}, \\ q_6(t_1, t_2) &= J_{34}. \end{aligned}$$

Here, $J_{ij}(t_1, t_2)$ are the elements of $J(t_1, t_2) = \Delta^{-1}(t_1, t_2)$.

7. Saddlepoint approximation

Analytical evaluation of the four dimensional integral in Eq. (14) is not possible and one needs to adopt approximate numerical methods to evaluate $\langle N_1 N_2 \rangle$. One can use quadrature schemes and other related methods- however, the computational costs associated can be quite high. The four dimensional integration required for computing $\langle N_1 N_2 \rangle$ can be partially simplified by using the symbolic manipulation discussed in [2]. However, this leads to $I(t_1, t_2)$ being expressed as a single integral, which implies that the four dimensional integral in Eq. (14) can now be written as a triple integral. It turns out that the computational costs for this method are of the same order or more as that of Monte Carlo simulations [2]. Instead, in this study, we adopt a multidimensional saddlepoint approximation to approximate the inner double integral $I(t_1, t_2)$ as follows [8, 9]:

$$\begin{aligned} I(t_1, t_2) &= \int_0^\infty \int_0^\infty \left\{ \dot{x}_1 \dot{x}_2 p_{X_1 X_2}(\alpha_1, \alpha_2, \dot{x}_1, \dot{x}_2; t_1, t_2) \right\} d\dot{x}_1 d\dot{x}_2 \\ &= \frac{1}{4\pi^2 \sqrt{\det(\Delta)}} \int_0^\infty \int_0^\infty f(\dot{x}_1, \dot{x}_2) \exp[-\lambda S(\dot{x}_1, \dot{x}_2)] d\dot{x}_1 d\dot{x}_2 \\ &\approx \frac{1}{4\pi^2 \sqrt{\det(\Delta)}} \left(\frac{2\pi}{\lambda} \right) \frac{f(\dot{x}_1^*, \dot{x}_2^*)}{\sqrt{\det(H(S)(\dot{x}_1^*, \dot{x}_2^*))}} \exp[-\lambda S(\dot{x}_1^*, \dot{x}_2^*)], \end{aligned} \quad (17)$$

where, $(\dot{x}_1^*, \dot{x}_2^*)$ is the saddle point obtained as a solution of the equation, $\nabla S(\dot{x}_1, \dot{x}_2) = 0$, where ∇ is the gradient operator $\nabla = \left[\frac{\partial}{\partial \dot{x}_1} \quad \frac{\partial}{\partial \dot{x}_2} \right]^T$; $f(\dot{x}_1, \dot{x}_2) = \dot{x}_1 \dot{x}_2$; $\lambda = 0.5$; $S(\dot{x}_1, \dot{x}_2) = q_1 + 2q_2 \dot{x}_1 + 2q_3 \dot{x}_2 + q_4 \dot{x}_1^2 + q_5 \dot{x}_2^2 + 2q_6 \dot{x}_1 \dot{x}_2$. $H(S)(\dot{x}_1, \dot{x}_2)$ is the Hessian Matrix of $S(\dot{x}_1, \dot{x}_2)$. This approximation is valid only when

$$\det(H(S)(\dot{x}_1^*, \dot{x}_2^*)) > 0 \quad (18)$$

where, $\det(H(S)(\dot{x}_1, \dot{x}_2)) = 4(q_4 q_5 - q_6^2)$. The outer double integral in Eq. (14) has to be carried out numerically. Thus, using saddlepoint approximation, the four-dimensional integral in Eq. (14) can be simplified to a double integral which can be numerically evaluated.

8. Numerical example

A numerical example is considered to illustrate the proposed analytical model. The effect of correlation of the parent processes on the joint probability distribution function for the extremes is studied. The analytical results are verified by comparing with those obtained from Monte Carlo simulations.

8.1. Problem description

Consider two zero mean, non-stationary Gaussian processes $X_1(t)$ and $X_2(t)$ defined by,

$$\begin{aligned} X_1(t) &= e_1(t) X_{1s}(t), \\ X_2(t) &= e_2(t) X_{2s}(t), \end{aligned} \quad (19)$$

where, $e_i(t)$ are the envelope functions, $X_{1s}(t)$ and $X_{2s}(t)$ are two zero mean, stationary Gaussian random processes. The autocorrelation functions for the processes $X_{1s}(t)$ and $X_{2s}(t)$ are, respectively,

$$R_{11}(\tau) = S_1^2 \exp[-\alpha \tau^2], \quad (20)$$

$$R_{22}(\tau) = S_2^2 \exp[-\beta \tau^2], \quad (21)$$

with $\alpha, \beta > 0$. The corresponding auto-PSD functions are given by,

$$S_{11}(\omega) = \frac{S_1^2}{2\sqrt{\pi\alpha}} \exp\left(-\frac{\omega^2}{4\alpha}\right), \quad (22)$$

$$S_{22}(\omega) = \frac{S_2^2}{2\sqrt{\pi\beta}} \exp\left(-\frac{\omega^2}{4\beta}\right). \quad (23)$$

The cross-PSD $S_{12}(\omega)$ is given by,

$$S_{12}(\omega) = c(\omega)|S_{11}(\omega)S_{22}(\omega)|^{0.5} \exp[i\gamma(\omega)], \tag{24}$$

where, $\gamma(\omega)$ is the phase spectrum and $c(\omega)$ is the coherence spectrum, which takes values in $[0, 1]$. In this example, it is assumed that $c(\omega)$ has a constant value c for all ω and $\gamma(\omega) = \omega/\psi$, with $\psi = 8$. Clearly, if $c = 0$, $X_1(t)$ and $X_2(t)$ are independent and if $c = 1.0$, they are fully coherent. Both S_1 and S_2 are assumed to be equal to $\sqrt{2}$. α and β are taken to be 100 and 150 respectively. The envelope functions $e_i(t)$ is taken to be of the form given in Eq. (2). For simplification, it is assumed that $e_1(t) = e_2(t) = e(t)$. The parameters a , b and c , respectively, taken to be 2.0000, 0.0436 and 0.1918. The values for the parameters are selected such that the envelope function attains a maximum value within the total duration. The total duration for both the processes $X_1(t)$ and $X_2(t)$ is assumed to be 40 seconds.

8.2. Results and Discussions

The spectral bandwidth parameters of the processes $X_1(t)$ and $X_2(t)$ are 0.6030 and 0.6032 respectively. The threshold levels for the processes, α_1 and α_2 are normalized with respect to their standard deviations and are denoted by x_1 and x_2 respectively.

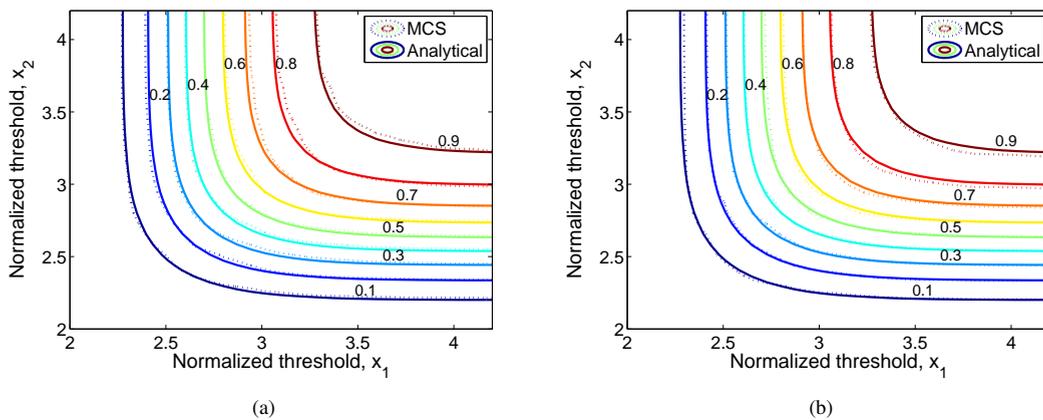


Fig. 1: Contours of $P_{X_{1m}, X_{2m}}(x_1, x_2)$ for 1000 samples in MCS (a) $c = 0.05$; (b) $c = 0.95$ (the numbers on the figure indicate probability levels).

The contour plots for bivariate extreme value distribution for $c = 0.05$ and $c = 0.95$ are shown in Fig. 1. The analytical results are observed to show good agreement with those obtained from Monte Carlo Simulations, carried out on 1000 samples, especially when the threshold levels are high. This can be explained by the fact that the assumption of Poisson upcrossings is more accurate for high threshold levels.

A comparison of conditional distribution $P_{X_{1m}|X_{2m}}(x_1|X_{2m} = 3)$ with univariate probability distribution function $P_{X_{1m}}(x_1)$, is shown in Fig. 2, three values of c . It can be concluded from this plot that the coherence between the parent random processes is an important parameter in the conditional distribution function.

Using the knowledge of extreme value distribution, probability of failure can be evaluated as

$$P_f(x_1, x_2) = 1 - P_{X_{1m}, X_{2m}}(x_1, x_2). \tag{25}$$

The contour plots of probability of failure for $c = 0.50$ obtained using MCS (dotted lines) and analytical method (solid lines) are shown in Fig. 3. From these plots, it can be seen that lower the number of samples, less accurate are the results for lower probability levels. For 1×10^3 samples, results for probability levels less than 10^{-3} do not show good agreement. Similarly for 1×10^4 samples, results show good agreement for the probability levels higher than 10^{-4} .

Importantly, the limitation of this method pertains to the spectral bandwidth of the processes. As discussed in [2], this formulation leads to better results when the bandwidth parameters of the processes lie in the interval $[0.3, 0.7]$.

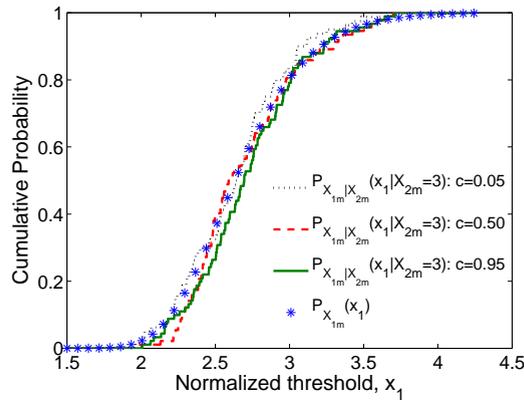


Fig. 2: Comparison of conditional PDF $P_{X_{1m}|X_{2m}}(x_1|X_2 = 3)$ with univariate PDF $P_{X_{1m}}(x_1)$.

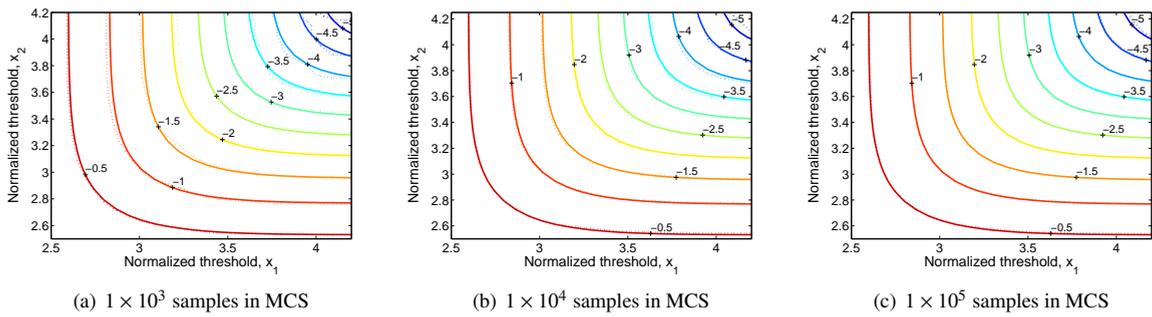


Fig. 3: Contours of $P_f(x_1, x_2)$ (the numbers on the figure indicate $\log_{10}(P_f)$).

9. Conclusions

An analytical method based on saddlepoint approximation has been developed for estimating the joint extreme value distribution of a vector of correlated non-stationary Gaussian random processes. For lower threshold levels, probability of failure estimated by analytical method is close to that using MCS. However for higher threshold levels, more number of samples are required in MCS to accurately estimate the failure probability. Hence it can be concluded that the proposed analytical method is computationally efficient and can be used to estimate the failure probability to the desired order of accuracy by choosing appropriate number of levels.

Acknowledgements

The work presented in this paper has been funded by the Hydrodynamics panel, Naval Research Board, DRDO, Government of India.

References

[1] S. Gupta, C.S. Manohar, Multivariate extreme value distributions for random vibration applications, Journal of Engineering Mechanics. 131(7) (2005), 712–720.
 [2] K.S. Rao, S. Gupta, C.S. Manohar, Multivariate extreme value distributions for structure response to non-stationary random excitations, unpublished report (2005).
 [3] H.E. Daniels, Saddlepoint approximations in statistics, The Annals of Mathematical Statistics. 25(4) (1954), 631–650.

- [4] N.L. Johnson, S. Kotz, Discrete Distributions, Wiley, New York, 1969.
- [5] S.O. Rice, Mathematical analysis of random noise, Bell System Technical Journal. 23(3) (1944), 282–332.
- [6] S.O. Rice, Mathematical analysis of random noise, Bell System Technical Journal. 24(1) (1945), 46–156.
- [7] J.N. Yang, Non-stationary envelope process and first excursion probability, Journal of Structural Mechanics. 1(2) (1972), 231–248.
- [8] N. Bleistein, R.A. Handelsman, Asymptotic Expansions of Integrals, Courier Corporation, 1975.
- [9] R. Wong, Asymptotic Approximation of Integrals, Vol. 34, SIAM, 2001.