The Log-Rayleigh Distribution for Local Maxima of spectrally Represented Log-normal Processes

Jan Grashorn*1,5, Marius Bittner1,5, Cao Wang4, and Michael Beer1,2,3

1Institute for Risk and Reliability, Leibniz University Hannover, Germany
2Institute for Risk and Uncertainty, University of Liverpool, UK
3Shanghai Institute for Disaster Prevention and Relief, Tongji University, Shanghai, China
4School of Civil, Mining and Environmental Engineering, University of Wollongong, Wollongong, NSW, Australia
5International Research Training Group 2657, Leibniz University Hannover & L.M.T. ENS Paris Saclay, Germany & France

Abstract

We address a novel probability distribution, namely the log-Rayleigh distribution, suited to model the maximum occurring loads in a log-normal process. Log-normal processes in engineering applications can be used to model for example wave or wind loads acting on structures. Usually, to assess probabilistic events, e.g. the probability of failure of structures under log-normal load processes, the generation of time-histories is necessary. With the given probability distribution, the maximum load events can directly be sampled, eliminating this step. Also, since a closed form of the PDF is given, the integrals involved in reliability analysis can directly be evaluated. We show that the proposed log-Rayleigh distribution can accurately model the distribution of local maxima in each log-normal process when compared to samples obtained from a Monte Carlo approach. Furthermore, we conduct a parameter study to evaluate the influence of the parameters in the log-Rayleigh distribution. Details on the generation of a log-normal process and a benchmark of this process are also included. Finally, a mechanical model related to a static structural reliability analysis is evaluated to show suitable utilities of the newly formed log-Rayleigh distribution.

Keywords: Log-Rayleigh distribution, log-normal process, Probability distribution, Local maxima, Spectral representation

1. Introduction

There are many civil engineering problems that involve random quantities. For example, the serviceability of a structure in a future time period cannot be predicted exactly due to the uncertainties arising from both the structural performance and the external actions. This implies the importance of modeling the time-variation of stochastic processes in a probabilistic framework. Under this context, the log-normal processes is among the widely-used stochastic process models. For example, Desmond and Guy [1] used a log-normal process to analyze the annual river flow data. Wu and Ni [2] studied the stochastic fatigue crack growth based on a log-normal process. Li et al. [3] investigated the time-dependent reliability of structures subjected to log-normal load processes. From a multidisciplinary view, the log-normal process can also be used to deal with signal processing as in [4], [5] and to describe the variation of stock prices [6].

The probabilistic behaviour of the local maxima of a log-normal process is an important feature of the process’ statistical characteristics. When employing the log-normal process to model the stochastic load process, the probability model of local maxima can be used for structural reliability assessment [7]. Yet the probabilistic behaviour of the local maxima associated with a log-normal process remains unaddressed in the literature. A relevant result is that, for a Gaussian stochastic process, the local maxima follows a Rayleigh distribution. With this regard, the aim of this paper is to investigate the probability model of local maxima of a log-normal process. In section 2 the Spectral Representation method is introduced to model and approximate stochastic processes. In section 3 the novel representation of log-normal processes with the Log-Rayleigh distribution is derived, a parameter study is carried out to validate the fitting of the distribution and to analyze its parameters. In section 4 a structural reliability analysis is carried out to demonstrate the utility of the newly formed Log-Rayleigh distribution for the approximation of log-normal stochastic processes. In section 5 concluding remarks are given.

2. Generation of log-normal processes

A lognormal process can be generated based on a Gaussian process, where a Gaussian zero-mean process, \( f(t) \), can be represented by a series of cosine functions [8],

\[
    f(t) = \sqrt{2} \sum_{n=0}^{N-1} A_n \cos(\omega_n t + \xi_n) \tag{1}
\]

with

\[
    A_n = (2 S(\omega_n) \Delta \omega)^{\frac{1}{2}} , \quad \omega_n = n \Delta \omega , \quad \Delta \omega = \omega / N
\]

where \( \omega_n \) is the upper cut-off frequency, \( \xi = [\xi_0 \cdots \xi_n]^T \) are independent and identically distributed (i.i.d.) random variables distributed uniformly over \([0, 2\pi]\) and \(S(\omega)\) is the Power Spectral Density (PSD) function. \( \omega_n \) is usually chosen such that it falls into a region where \( S(\omega) \) becomes sufficiently small. The log-normal process \( X(t) \) is then calculated by \( X(t) = \exp \{ f(t) \} \). \( S(\omega) \) was here chosen to be

\[
    S(\omega) = \frac{a}{\omega^b} , \quad a, b > 0 \tag{2}
\]

where \( \omega \) is the frequency and \( a \) and \( b \) are constants which can be chosen according to some predefined values for the process standard deviation \( \sigma_X \) and the standard deviation of the rate of change in the process \( \sigma_X^\prime \)

\[
    a = \frac{6\sqrt{2}}{\pi} \frac{\sigma_X^5}{\sigma_X^3} , \quad b = \frac{2\sigma_X^6}{\sigma_X^3} \tag{3}
\]

\*E-Mail: grashorn@irz.uni-hannover.de

793
Let $x^* = \Phi^{-1}(F_X(x))$, and Eq. (6) becomes
\[
F_{X_{\max} | \lambda}(x) = \exp \left\{ - \frac{\sigma_X \Delta}{2\pi} \exp \left\{ - \frac{x^*}{2} \right\} \right\}
\]
(7)

For small $x$ it can be assumed that $e^{-x} = 1 - x$ so that Eq. (7) becomes
\[
F_{X_{\max} | \lambda}(x) = 1 - \frac{\sigma_X \Delta}{2\pi} \exp \left\{ - \frac{[\Phi^{-1}(F_X(x))]^2}{2} \right\}
\]
(8)

A random variable $X$ is deemed to follow a log-Rayleigh distribution with location parameter $\kappa > 0$ and scale parameter $\sigma > 0$ if its CDF takes the form of
\[
F_X(x) = 1 - \kappa \exp \left\{ - \frac{(\ln x)^2}{2\sigma^2_X} \right\}, \quad x \geq 1, \quad \kappa = \frac{\sigma_X \Delta}{2\pi}
\]
(9)

Corresponding, the PDF of $X$ is as follows,
\[
f_X(x) = \exp \left\{ - \frac{(\ln x)^2}{2\sigma^2_X} \right\} \cdot \frac{\kappa \ln x}{x\sigma^2_X}, \quad x \geq 1
\]
(10)

Thus, the expected value of $X$ is determined by
\[
\mathbb{E}(X) = \int_1^\infty x \cdot f_X(x) dx = \int_1^\infty \exp \left\{ - \frac{(\ln x)^2}{2\sigma^2_X} \right\} \frac{\ln x}{x\sigma^2_X} dx
\]
\[
= 1 + \sqrt{2\pi} \sigma_X \exp \left[ \frac{\sigma^2_X}{2} \right] \left[ 1 + \text{erf} \left( \frac{\sigma_X}{\sqrt{2}} \right) \right]
\]
(11)

where $\mathbb{E}(\cdot)$ is the expected value of the variable in the brackets. Furthermore, the expectancy of $X^2$ is
\[
\mathbb{E}(X^2) = \int_1^\infty x^2 \cdot f_X(x) dx
\]
\[
= 1 + \sqrt{2\pi} \sigma_X \exp \left[ 2\sigma^2_X \right] \left[ 1 + \text{erf} \left( \sqrt{2}\sigma_X \right) \right]
\]
(12)

which further yields the variance of $X$ as $\mathbb{E}(X^2) - \mathbb{E}(X)^2$.

More generally, the $n$th raw moment of $X (n = 1, 2, \ldots)$ is
\[
\mathbb{E}(X^n) = \int_1^\infty x^n \cdot f_X(x) dx
\]
\[
= 1 + \sqrt{n} \sigma_X \exp \left[ n\sigma^2_X \right] \left[ 1 + \text{erf} \left( \frac{n\sigma_X}{\sqrt{2}} \right) \right]
\]
(13)

in which $\text{erf}(\cdot)$ is the error function.

For visualization purposes, Fig. 2 shows the dependence of Eq. (8) on $\sigma_X$ and $\sigma_X$. It can be noticed that higher values for both $\sigma_X$ and $\sigma_X$ shift the distribution towards the right, meaning that higher maxima are more likely. Note also that the CDF is not monotonically increasing and can take on values below zero. While this reduces the usefulness of the log-Rayleigh distribution in more general applications we show in the following discussion that it is able to capture the upper end of the distribution (indicated by the grey area in Fig. 2) without compromising the laws of stochastic...
calculus. Being able to calculate this upper tail is very useful in for example reliability estimation because the probability of rare events can be calculated directly without the need for Monte Carlo sampling and therefore without direct model calls.

An important parameter in the log-Rayleigh distribution is \( \Delta \) which controls the width of the observed time interval. Higher values for \( \Delta \) increase the probability of observing larger maxima. Fig. 3 shows the influence of \( \Delta \) on the detection of local maxima (in this case \( \Delta = 5 \) s). The generated log-normal process is divided into intervals of length \( \Delta \) and in each interval the maximum is marked with a cross. If \( \Delta \) is chosen too small (i.e. \( \Delta = \Delta_t \) in a time discrete process with time step length \( \Delta_t \)) this would capture every value in \( X(t) \) and thus be a log-normal distribution. On the other hand, if a process is observed in a large interval \([0, T]\) and \( \Delta \) is chosen such that \( \Delta = T \) only the “global” maximum over the whole duration \( T \) is considered. \( F_{X_{\text{max}}|\Delta}(x) \) then describes the probability of observing \( x < X_{\text{max}|\Delta} \) in the interval \([0, T]\) which is equivalent to the probability of failure in first passage problems if \( X_{\text{max}|\Delta} \) is the system’s resistance. Fig. 3 further shows the idea of only using the upper end of the log-Rayleigh distribution. All occurring maxima below a certain threshold \( x \leq F_{X_{\text{max}}|\Delta}^{-1}(\gamma) \), where \( \gamma \) is a chosen value above which the distribution is valid, are ignored. In addition, \( \gamma \) needs to be chosen in a way that the assumption of small \( x \) holds.

3.2. Validation

We show that the proposed distribution is able to capture local maxima of a log-normal process by comparing the CDF to samples drawn by Monte Carlo simulation (MCS). We also show that sampling from the log Rayleigh distribution is much more efficient than to sample by means of MCS. By use of (Eq. (1)) multiple realizations from a log-normal process with \( \sigma_X = 1 \) and \( t \in [0, 100] \) are simulated and the local maxima are calculated according to Fig. 3. From these values the empirical CDF is constructed which can be compared to a zero-mean log-normal CDF with standard deviation \( \sigma_X \) and the log-Rayleigh CDF (Eq. (9)) with parameters \( \sigma_X \) and \( \sigma_Y \). Fig. 4 shows the result for different values of \( \Delta \). For \( \Delta = 1 \) there is a large discrepancy between samples and the log-Rayleigh distribution, however this vanishes if \( \Delta \) is chosen to be larger. It can also be seen that the log-Rayleigh distribution captures the true distribution at the upper tail of \( F_{X_{\text{max}}|\Delta}(x) \).

Sampling from the log-Rayleigh distribution can be achieved with the inverse CDF method [12]. Drawing a new sample \( X \) is done by letting \( X = F_{X_{\text{max}}|\Delta}^{-1}(u) \) where \( u \) is a uniformly distributed random variable in the interval \([0, 1]\). Samples drawn this way can again be compared to samples obtained from a log-normal process. \( \gamma \) needs to be taken into account here since we only want to sample from the valid region. A comparison of the resulting normalized histograms is shown in Fig. 5. Also pictured is the PDF (Eq. (10)) to show that the samples follow the proposed distribution. Table 1 shows the first three statistical moments of both samples as well as the calculation time.

4. Example

In this section, we will use an illustrative example to show the accuracy of using the log-Rayleigh distribution to model the local maxima of a stationary log-normal process. A simplified offshore structure subjected to environmental wind loads is regarded. It is assumed that the dynamic properties of the wind do not contain a high energy level in the frequency domains near the structure’s eigenfrequencies. Thus it is possible to perform a fully static analysis of the force on the structure induced by wind. Following calculations are based on the DNV standards [13] for the

![Figure 2](image2.png)  
**Figure 2.** CDF of log-Rayleigh distribution for different sets of \( \sigma_X \) and \( \sigma_Y \). The grey box indicates the area where \( F_{X_{\text{max}}|\Delta} \) becomes relevant

![Figure 3](image3.png)  
**Figure 3.** Example of log-normal process, local maxima of each interval are indicated with crosses. The dashed line represents the cutoff threshold \( Pr\{X_{\text{max}}|\Delta < \gamma\} = 0.9 \)

<table>
<thead>
<tr>
<th>( \mu_{\text{emp}} )</th>
<th>( \sigma_{\text{emp}} )</th>
<th>skewness</th>
<th>time [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MCS</strong></td>
<td>27.401</td>
<td>10.207</td>
<td>3.994</td>
</tr>
<tr>
<td><strong>CDF</strong></td>
<td>27.493</td>
<td>10.416</td>
<td>3.827</td>
</tr>
</tbody>
</table>

**Table 1.** Empirical statistical moments for \( 10^5 \) samples of log-Rayleigh CDF and MCS
design of offshore structures. A simple mono pile above water level is regarded (see Fig. 6a). The average wind speed value at the structure’s height $z$ from sea surface level for a specific time $t$ is given to be

$$v(z, t, \theta) = aV(t, \theta) \frac{z}{10}^{\beta},$$  \hspace{1cm} (14)$$

where $V(t, \theta) = H_{exp}(t, \theta) + V_m$. Here $H_{exp}(t, \theta)$ is a log-normal process with respect to the time domain $t$, approximated by Eq. (1) given the PSD function in Eq. (2) with parameters $a = 10, b = 60$. Then $H_{exp} = \exp\{f(t)\}$. $\alpha = 1$ and $\beta = 0.15$ are wind speed averaging time factors, $V_m = 9$ m/s is the assumed dominant average wind speed. $\theta$ are realizations of the random variables in $\zeta$ and indicate the random variables inherent in the function $V(t, \theta)$, which can be traced back to the method introduced in Eq. (1). Given these considerations it is possible to describe the air pressure as

$$q(z, t, \theta) = \frac{1}{2} C \rho v(z, t, \theta)^2,$$  \hspace{1cm} (15)$$

with the geometric shape coefficient $C = KC_{\infty}, K = 0.884, C_{\infty} = 0.61$ and wind density $\rho = 1.225$ kg/m$^3$. The Force dependant on air pressure, diameter $D$ and assumed angle of incoming wind towards the pile. $\zeta = 90^\circ$ is then stated as

$$F(z, t, \theta) = q(z, t, \theta) D \sin(\zeta).$$  \hspace{1cm} (16)$$

The offshore structure is cylindrical in geometry and 90 m high. In Fig. 6c the largest load is always acting on the
highest point of the structure, therefore \( z = H_z = 90 \text{ m} \) is assumed for the static bending analysis. The offshore structure is fixed at \( z = 0 \text{ m} \), i.e. \( \mathcal{F}(0,t) = 0 \). Only bending and wind load in one dimension is regarded. Therefore the top bending can be calculated via beam theory and is given as

\[
x(t,\theta) = \frac{\mathcal{F}(90,t,\theta)90^3}{3EI},
\]

with Young’s modulus \( E = 2 \cdot 10^{11} \text{ Nm}^{-2} \) and the geometric moment of inertia \( I = \frac{\pi D^4}{32} \). The simulation time is \( T = 1000 \text{ h} \) with a discretization of 12 min per time step.

### 4.1. Reliability

A standard first-passage probability criterion is constructed to analyze the offshore structure’s performance. Target quantity of interest for the structure’s performance is the top displacement described by Eq. (17). The probability that the target quantity \( x(t,\theta) \) exceeds a pre-defined resistance value \( R \) once can be described by

\[
p_f = \Pr\{x(t,\theta) \geq R \mid t \in [0,T]\}.
\]  

To analyze the exceedance, the following performance function is introduced,

\[
g(\theta,t) = R - x(t,\theta).
\]

The failure space is associated to the performance function value, once it is below zero a failure is detected. An indicator function \( I[g(\theta,t)] \) is introduced such that

\[
\begin{align*}
g(\theta,t) \leq 0 & \rightarrow \theta \in \Omega_f \rightarrow I[g(\theta,t)] = 1, \\
g(\theta,t) > 0 & \rightarrow \theta \notin \Omega_f \rightarrow I[g(\theta,t)] = 0.
\end{align*}
\]

Failure can only occur once in time, only the first exceedance is regarded. The following generalized integral must be solved to analytically estimate the structure’s probability of failure.

\[
p_f = \int_{\Omega_f} g(\theta,t) d\theta.
\]  

However, since the failure space is not directly describable in the sample space the above integral is expanded with the indicator function and the performance function.

\[
p_f = \int_{-\infty}^{\infty} I[g(\theta,t)]f(\theta)d\theta = \mathbb{E}[I(g(\theta,t))].
\]

For the offshore structure the demand value, i.e. the maximum allowable displacement of the top is \( R = 2 \cdot 10^{-4} \text{ m} \).

#### 4.1.1. Monte Carlo estimator

Because usually it is not feasible to estimate the probability of failure stated in Eq. (23) the Monte Carlo estimator is applied to approximate the probability of failure.

\[
\hat{p}_f = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} I[\theta_i,\nu_i].
\]

A number of samples \( N_{MC} \) are generated from our input probability density functions, utilizing again the indicator function simply count the samples leading to a failure defined by the performance function. The second order moment and the coefficient of variation (COV) can be estimated by

\[
\text{Var}[\hat{p}_f] = \frac{\hat{p}_f^2 - \hat{p}_f^2}{N_{MC}}, \quad v_{pf} = \frac{\text{Var}[\hat{p}_f]}{\hat{p}_f}.
\]

Especially for small failure probabilities a large number of samples are needed to estimate the probability of failure accurately, following rule of thumb can be used to estimate the necessary number of samples.

\[
N_{MC} = \frac{1 - p_f}{p_f v_{pf}^2}.
\]

### 4.2. Log-Rayleigh estimator

For this example, where we are regarding a linear relationship between the random input in Eq. (14) and the final calculation of the relevant quantity, i.e. the displacement \( \mathcal{F}(0,t) \) for the performance function Eq. (19), it is possible to directly calculate the probability of failure from the Log-Rayleigh distribution without the need of generating samples. From the given example equivalently to the resistance value \( R \) the stochastic process associated resistance value \( R_\text{exp} \) can be calculated. \( R_\text{exp} \) falls directly in the sampled space of the generated log-normal process. Assume that the probability of failure can also be expressed as

\[
\hat{p}_f = Pr\{H_\text{exp}(t,\theta) \geq R_\text{exp} \mid t \in [0,T]\}.
\]

By utilizing the description of the complementary event it is possible to directly estimate the probability of failure by the CDF stated in Eq. (8):

\[
\hat{p}_f = 1 - F_{\text{exmp}}(R_\text{exp}).
\]

Here the underlying generated stochastic process must suffice the relation in Eq. (3). This ensures the definition of the moments for Eq. (8).

#### 4.3. Results

Results of the reliability analysis for the estimation of the probability of failure are shown in Table 2 and Fig. 7, where Table 2 shows two different MC-simulations \( \hat{p}_{f,1} \) and \( \hat{p}_{f,2} \). Two different system capacity values \( R_\text{exp} \) are assumed. For the lower resistance value \( \hat{p}_f \) and both results for \( \hat{p}_{f} \) show some discrepancy, however if a larger \( R_\text{exp} \) is considered both results are very close, indicating that the log-Rayleigh distribution is converging to the real distribution of maxima in its upper tail. Fig. 7 shows the dependency of \( \hat{p}_f \) on the resistance value \( R_\text{exp} \) and the difference between the Monte Carlo and the log-Rayleigh estimators. For the MC simulation 20 estimations of \( \hat{p}_f \) for the corresponding \( R_\text{exp} \) have been calculated, each estimation contains 15000 samples, the resulting statistics of \( \hat{p}_f \) are shown as boxplots. It can be seen that \( \hat{p}_f \) matches
Figure 6. Example setup and intermediate results for the mechanical analysis of a simplified monopile under wind load in an offshore environment.

Figure 7. Estimations for $\hat{p}_f$ and $\tilde{p}_f$ with respect to $R_{\text{Hexp}}$, for each $R_{\text{Hexp}}$ a total number of $3 \cdot 10^5$ samples are generated these statistics with a slight underestimation for larger failure probabilities. This error is decreasing with a larger resistance value $R_{\text{Hexp}}$ and the following smaller probability of failure.

Generally it has been shown that the estimation of the probability of failure using the log-Rayleigh estimator is in agreement with the value estimated by MC.

5. Concluding remarks

In this paper, a new distribution type, namely the log-Rayleigh distribution on the space $[1, \infty)$, is presented to describe the local maxima of a stationary log-normal stochastic process. The statistical properties of the log-Rayleigh distribution is discussed. On an illustrative example the applicability of the newly formed log-Rayleigh distribution to estimate small failure probabilities associated to log-normal stochastic processes is shown.

The log-Rayleigh estimator for failure probabilities derived from the log-Rayleigh distribution is of superior performance than the MC estimator, as shown in the example. For the MC estimator which is of the type of a stochastic simulation method, multiple system evaluations are necessary to estimate the probability of failure, which leads to a larger variation depending on the number of samples used.
and a higher computation time needed. For the log-Rayleigh estimator it is only necessary to calculate the demand value once, then a direct estimation of the failure probability for the given system is possible utilizing the parametric distribution function. The log-Rayleigh estimator is especially suitable for linear or linearized systems and is applicable for systems excited by a log-normal stochastic process in the input space/physical space. It is also generally suitable to calculate the probability of rare events in a log-normal process.

6. Acknowledgements
The first two authors wish to thank the German research foundation (DFG) for funding the GRK 2657 - International Research Training Group 2657 - Computational Mechanics Techniques in High Dimensions, this support is highly appreciated. Additionally, we would like to thank the Humboldt foundation which enabled the cooperation with Dr. Cao Wang who received the prestigious Humboldt research fellowship.

References