Transactions on Engineering and Computing Sciences - Vol. 12, No. 1
Publication Date: February 25, 2024
DOI:10.14738/tecs.121.15979.
Onyemachi, C. U. (2024). Application With an Event Time Distribution on Spatial Process. Transactions on Engineering and Computing Sciences, 12(1). 206-213.

SERVICES FOR SCIENCE AND EDUCATION

Application With an Event Time Distribution on Spatial Process

Chris Uchechi Onyemachi
Faculty of Physical Sciences, Department of Statistics
Abia State University, Uturu, Nigeria

ABSTRACT

The use of models in sciences to describe phenomena that assume non negative values such as size distributions, flood frequencies, radio wave amplitude fluctuations, air velocity fluctuations and geographical analysis are enhanced with the lognormal distribution, and the likes of Weibull and Exponential distributions as a life time distribution model. In most cases, these are empirically observed and test fits and often other models such as gamma distribution are also fitted. The distribution is specified by saying that the life time of random variable X is lognormal if the logarithm Y = In X of the life time is normally distributed with mean μ and variance σ². The two-parameter lognormal distribution denoted by $\Lambda(\mu,\sigma^2)$; with the corresponding normal distribution $N(\mu,\sigma^2)$ is considered. Lognormal distributions play a central role in human and ecological risk association, its probability plots are invaluable because they provide a powerful way to visualize measured data or simulated results such as in landslides.

Keywords: Non negative value phenomena, lognormal distribution, model, life time, empirically observed.

INTRODUCTION

Spatial statistics analysis has risen to popularity due to demand from a wide spectrum of interests. These include among others: ecology, geography, epidemiology, agronomy, biology and spatial economics. Owing to arduous complexity of atmospheric phenomena and the difficulty of conducting controlled experiments, there is much modeling to derive results numerically or analytically based on a variety of assumptions, which often include lognormal distributions of basic variables according to [4].

A spatial process consists of a set of random variables X_t or $X(t)$ which depend on a parameter t (continuous or discrete and usually indicating time), which is said to define a stochastic process. A set of random variables $\{X(t), t \in T\}$, all defined on the sample space S, is defined as a stochastic process or random process or simply a process, where T is the index set of the process. The index set T may be a countable set, such as {1, 2…} or {…, -2, -1,0,1, 2…} or an uncountable set such as (0, ∞) and (-∞, ∞). The model for such a process is an experiment of chance that continues in time or space [1] posit. The theory of stochastic processes has a number of important applications of which reliability theory is pertinent.

Spatial statistics analysis has focused traditionally on three main components: (1) geostatistics, i.e., a spatial process indexed over a continuous space, (2) lattice data, i.e., spatial data indexed

over a lattice of points, and (3) spatial point patterns, i.e., pertaining to the location of "events" of interest and their spatial patterns. Next to the modeling of stochastic processes of such variables with high volatility in recent time is aided by data mining, a science of extracting useful information from large data sets or databases as [5] observe.

The time to an event is distinguished from other stochastic processes occurring in time, often expressed as differential equations by use of the term "point processes as explained by [2]. Events such as in medical field: the incubation of disease, time to recovery or to death; events in business and economics including: time in and out of job, strikes duration, etc., these processes are time-bound and discrete and follow a Brownian stochastic dynamism. The intention of this paper, however, is not to catalogue all processes under review. Our interest will be limited to using specific hypothetical spatial process to demonstrate the flexibility of this important statistical distribution.

Landslides, also known as landslips, are several forms of mass wasting that may include a wide range of ground movements such as rockfalls, shallow or deep-seated slope failures, mud flows and debris flows. They are typically "time to an event" phenomena. When interest is on the timing of events and the probability of one or more events occurring by a given time, the lognormal is a good fit distribution.

Time is of essence as the metric of processes, according to [7] for the lognormal avails itself to explanation of the underlying process as used with point process data.

THE LOGNORMAL DISTRIBUTION

The lognormal distribution, is flexible to model different processes and have been extensively used to model the non-normal processes. The lognormal distribution is based on a Brownian motion (BM) process according to [6]. A standard BM is a type of Markov stochastic process with a drift rate to zero and variance of one.

A positive random variable X is said to be lognormally distributed with two parameters μ and σ^2 if $Y = InX$ is normally distributed with mean μ and variance σ^2, its probability density function (pdf) is given by

$$f(x) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{In(x)-\mu}{\sigma}\right)^2} & , \quad 0 < x < \infty \\ 0 & \quad Otherwise \end{cases} \tag{1}$$

If X has a lognormal distribution with parameters μ and σ^2, then the denotation $X \sim \Lambda(\mu, \sigma^2)$ is used.

METHODOLOGY

In this section, some statistical and mathematical properties of the lognormal distribution including the moments, moment generating function, mode, and median are presented.

1. The products and quotients of lognormal variates is lognormal. This result demonstrates both a fundamental property of independent lognormal distributions and the felicity of parameterizing the distributions in terms of the mean and standard deviations of the underlying normal distribution.
2. The product of a large number of independently and identically distributed variables is asymptotically lognormal i.e., if $X_1, \quad X_2, \quad . \quad . \quad ., \quad X_n$ is a set of *i.i.d* random variables, such that the mean of each $\log X_i = \mu$ and its variance $= \sigma^2$, then the product $X_1 X_2 ... X_n$ is asymptotically distributed in accordance with lognormal distribution with mean μ and variance $n\sigma^2$.
3. Moments:

The r^{th} moment about origin is given by

$$\mu_r^I = E(X^r) = \int_0^\infty x^r f(x) dx$$

$$= \int_0^\infty x^r \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{In(x)-\mu}{\sigma}\right)^2} dx \qquad (2)$$

Let $z = In(x)$, substituting in (3)

$$E(X^r) = \int_0^\infty e^{rz} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz$$

$$= M_z(r)$$

Where $M_z(r)$ denotes the moment generating function of the random variable $Z \sim N\left(\mu, \sigma^2\right)$. Thus

$$M_z(r) = e^{\mu r + \frac{1}{2}\sigma^2 r^2}$$

If $r = 1$, then $E(X) = e^{\mu + \frac{1}{2}\sigma^2}$

Similarly, if $r = 2$, then $E(X^2) = e^{2\mu + 2\sigma^2}$

and the variance is $E(X^2) - (E(X))^2$

$$V(X) = e^{\left(2\mu+\sigma^2\right)}\{e^{\sigma^2} - 1\}$$

The median is

$$Med(X) = e^{\mu}$$

while the mode is

$$Mode(x) = e^{(\mu-\sigma^2)}$$

The cumulative distribution function for X can be obtained given the transformation $W \sim N(\mu, \sigma^2)$ as

$$F(x) = P[X < x] = P[\exp(W) \leq x]$$

$$= P[W \leq In(x)]$$

$$= P\left[Z \leq \frac{In(x) - \mu}{\sigma}\right]$$

$$= \Phi\left[\frac{In(x) - \mu}{\sigma}\right] \quad (3)$$

APPLICATIONS

Two examples of spatial processes are used to demonstrate the applicability of lognormal distributions to modeling event-time phenomena.

1. In a study undertaken by the author, information concerned with predicting survival dynamics in patients undergoing a certain type of abdominal operation in a hospital surgical unit was obtained from the records of a random sample of 50 patients in pre-operation evaluation. In this example, response variables are associated with the "survival time". The model presents several restrictions on the distribution of the variables Y_i. such that a restriction to normal distribution is obvious, but with apparent specification that each random variable has the same expected value (μ) and same variance (σ^2). Such restriction provides a fundamental purpose which is conditioned on multiple variables from the same statistical population. The abridged data is annexed, and the MINITAB software was used in the analysis. The variables of interest include: X = Age, X = Gender: 0 = Male, 1 = Female, Y = Survival Time (minutes)

RESULTS

Descriptive Statistics: In (survival time)

Variable	N N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
In (surv time)	50 0	6.4418	0.0717	0.5071	5.1980	6.1753	6.4345	6.7040	7.7590

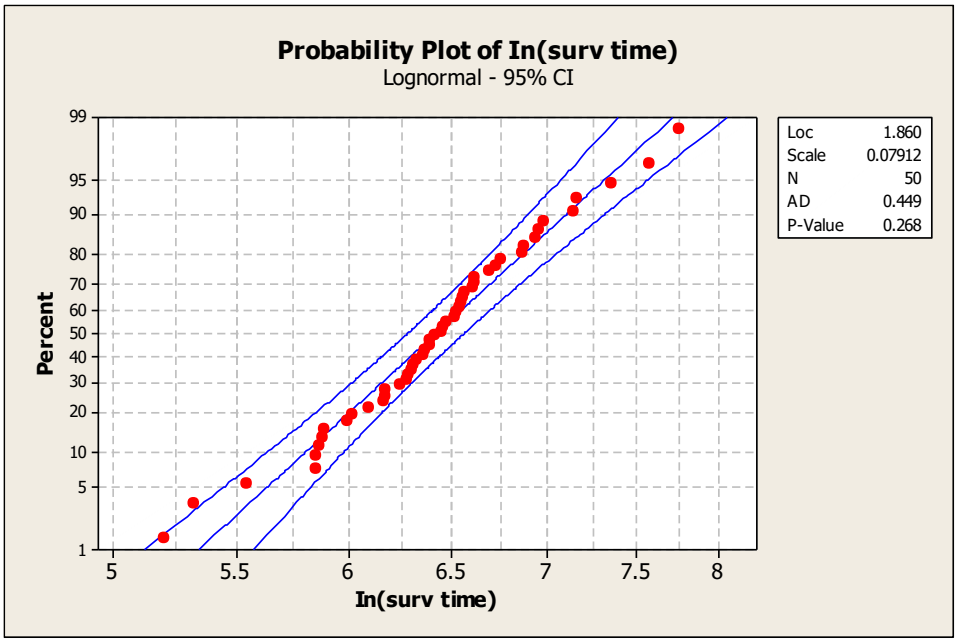


Fig 1: A Time Plot of the Survival time

The time plot visualizes the set of data in comparison to a full lognormal distribution as it combines exploratory data analysis accompanied with parameter estimation. With the data points falling on a straight line which implies that a lognormal distribution will fit the data with high fidelity.

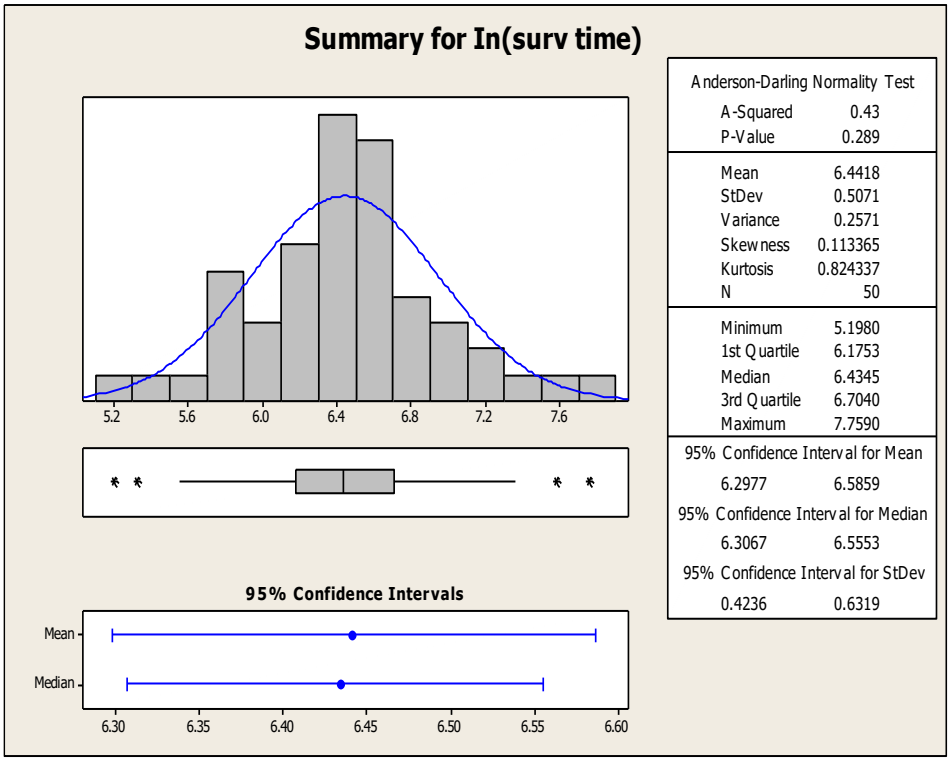


Fig 2: Goodness of fit test

Fig 2 shows the lognormal distribution is a continuous distribution of a random variable whose logarithm is normally distributed. By design, each quantile on a lognormal probability plot depends on the lower one, so that the plotting positions of the independent variable in the linear regression are not independent.

2. Using a reliability study, the modeling of the lifetime of a product can be lognormal process (random variable), the case distribution of semiconductor laser is considered: Given that the lifetime of a semiconductor laser has a lognormal distribution as $X \sim \Lambda(10,1.5)$, it is desired to find:
   i. the probability that the lifetime exceeds 10,000 hours.
   ii. What lifetime is exceeded by 99% of lasers?
   iii. The mean and standard deviation of lifetime

   i. From (2), $P[X > 10{,}000] = 1 - P[\exp(W) \le 10{,}000]$

$$= 1 - P[W \le In(10{,}000)]$$

$$= 1 - \Phi\left[\frac{In(10{,}000) - 10}{1.5}\right]$$

$$= 1 - \Phi(-0.52)$$
$$= 1 - 0.30$$
$$= 0.70$$

   ii. The question is to determine x such that

$$P(X > x) = 0.99$$

   Therefore,

$$P(X > x) = P[\exp(W) > x]$$

$$= P[W > In(x)]$$

$$= 1 - \Phi\left(\frac{In(x) - 10}{1.5}\right) = 0.99$$

   From table $= 1 - \Phi(Z) = 0.99$ when $Z = -2.33$

   Therefore $x = \exp(6.505) = 668.48 hrs$

   iii. To determine the mean and standard deviation

$$E(X) = e^{\mu + \frac{\sigma^2}{2}} = e^{[10+1.125]} = 67{,}846.3$$

$$V(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) = e^{[20+2.25]}\left(e^{2.25-1}\right) = 39{,}070{,}059{,}886.6$$

Showing that the standard deviation $\sqrt{39{,}070{,}059{,}886.6} = 197{,}661.5 hrs$ an indication that standard deviation of lognormal random variable can be large relative to the mean as a distinct property.

CONCLUSION

In this study, the lognormal distribution (with two parameters) is considered, where a random variable x is lognormally distributed with another variable y, the log of x being normally distributed. With ease of transformation x and y are readily related. The lognormal distribution is an ideal in modeling event-to-time phenomena and specifically plays a central role in human and ecological risk, for reasons that many physical, chemical, biological and statistical processes tend to generate random variables that follow lognormal distributions. Such processes defined by the type of geologic materials which are common landslide types are otherwise called "event-to-time" phenomena. These can be handled using the Lognormal distribution which has intrinsic properties.

ACKNOWLEDGEMENT

The author gladly acknowledges the assistance of the Management and Staff of Kelex Hospital and Sicotex Electricals both in Umuahia Abia State Nigeria.

References

[1] Baisnab, A. P and Jas, M. (1993). Elements of Probability and Statistics. McGraw-Hill Publishing Company Limited. New delhi

[2] Cox, D. R and Miller, H. D (1965). Theory of Stochastic Processes. Wiley. NY

[3] David N. B (1991). Statistical Tests of the Lognormal Distribution as a Basis for Interest Rate Charges. Transactions of Society of Actuaries Vol.43, 7 -35

[4] Giorgi, F. (1986). Development of an Atmospheric Aerosol Model for Studies of Global budgets and Effects of Airborne Particulate Materials. A Technical Report. United States: N.p.,

[5] Hand, M M., Simms, D A., Fingersh, L J., Jager, D W., Cotrell, J R., Schreck, S., and Larwood, S M. (2001). Unsteady Aerodynamics Experiment Phase VI: Wind Tunnel Test Configurations and Available Data Campaigns. United States: N. p., https://doi:10.2172/15000240

[6] Pin-Monareez, M. R., Ortiz-Yanez, J. F. and Rodriguez-Borbon, M. I (2015). Non-normal Capability Indices for the Weibull and Lognormal Distributions. Wileyonlinelibrary.com DOI:10.1002/qre.1832

[7] Thomopoulos, N. t (2017). Statistical Distributions. Springer International Publishing. DOI 10.1007/978-319-65112-5_9

Appendices

Distribution of Patients Survival Time Following a Surgical Operation

S/N	Age X_1	Gender X_2	Survival Time Y_i	InY_i
1	69	0	749	6.619
2	48	0	349	5.854
3	65	0	348	5.852
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
48	49	1	518	6.250
49	45	1	2343	7.759
50	58	0	1056	6.962