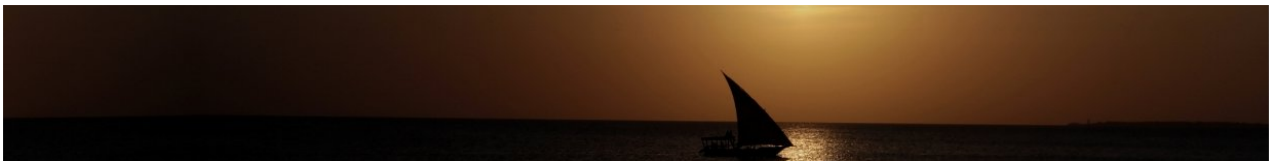

NavIncerta Library

The Lognormal Distribution

Domain:	Probabilistic analysis
Author:	Henk Krijnen
Reviewed by:	Thijs Koeling
Version:	1.0
Date:	July 17, 2017

NavIncerta
Address: Oude Delft 71F
2611BC Delft
The Netherlands
Email: info@navincerta.com
Phone: +31654385214

1 Introduction

In decision analysis and risk and uncertainty evaluations, the lognormal distribution is the most prominent. Our reasons for that are:

- It approximates the distribution of variables that result from multiplicative operations.
- It allows modelling a variable that has upside or downside.
- It has convenient mathematical properties.

Although there are many references on the mathematical properties of lognormal distributions, these often lack handy formulas that allow practical application for decision analysis purposes. Therefore we devote this article to the lognormal distribution and provide useful formulas, including their derivation.

Note: in this document we use the convention of the descending cumulative distribution when quoting percentiles. So P_{90} is the 'low' value and P_{10} is the 'high' value.

2 Summary of key formulas

In this part the mathematical properties of the lognormal distribution are summarized. Various derivations are given in subsequent paragraphs.

2.1 Mathematical basis

In this section we review the basic formulas that are provided in e.g. wikipedia or textbooks on statistics. These are used as a starting point.

Although limited use is made of the density function, it is provided here for completeness:

$$f(X) = \frac{1}{X\sigma\sqrt{2\pi}} e^{-\frac{(lnX-\mu)^2}{2\sigma^2}} \tag{1}$$

In this expression, X is the lognormally distributed variable which should be greater than 0. $x = ln(X)$ has a normal distribution. Furthermore μ is the expectation of x and σ is its standard deviation. In the following we provide the main formulas as you can find them in the textbooks. They are the basis for further derivations.

The mean or expectation, which we denote by $E[X]$ or $\widetilde{X}$:

$$E[X] = e^{\mu+\frac{1}{2}\sigma^2} \tag{2}$$

The variance, which we denote by $VAR[X]$ or Σ^2

$$VAR[X] = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2} \tag{3}$$

The median, which we denote by P_{50}:

$$P_{50} = e^{\mu} \tag{4}$$

The skewness, which we denote by γ:

$$\gamma = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1} \tag{5}$$

The following expressions provide the relationships between the mean and standard deviation of the underlying normal distribution and the lognormal distribution metrics:

$$\mu = ln(E[X]) - \frac{1}{2} ln(1 + \frac{VAR[X]}{(E[X])^2}) \tag{6}$$

$$\sigma^2 = ln(1 + \frac{VAR[X]}{(E[X])^2}) \tag{7}$$

These expressions form the basis for further derivations.

2.2 Practical formulas

The standard lognormal distribution is defined by two parameters. The common way is to specify the expectation and standard deviation. Below you see some examples of two parameter lognormals.

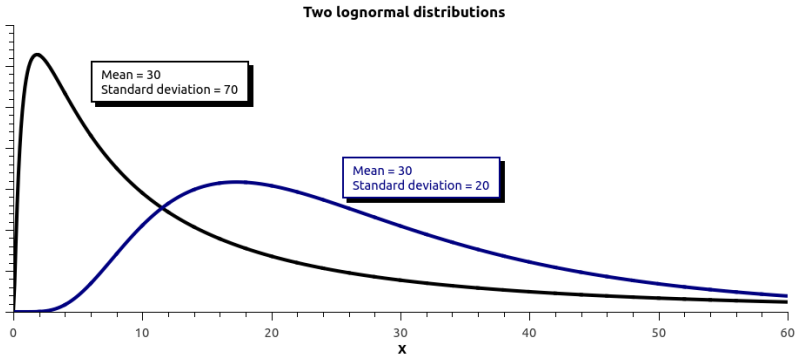

Figure 1: Two lognormals

You note that the two parameter lognormal is right-sided and the curve starts at (0,0).

Often a distribution is characterized by three percentile values: P_{90}, P_{50}, P_{10}. For a two parameter lognormal distribution the following relationship applies:

$$P_{50}^2 = P_{90} \times P_{10} \tag{8}$$

The moments[1] of a lognormal distribution can be calculated using the following formulas:

[1] With the first moment being the mean, the second moment being the square of the standard deviation and the third moment being the skewness times the standard deviation to the third power.

The mean or expectation:

$$E[X] \equiv \widetilde{X} = P_{50} \times exp\left(K_1 \times [ln(P_{10}) - ln(P_{50})]^2\right) \tag{9}$$

K_1 is a constant with $K_1 = 0.30443728...$

The standard deviation:

$$\Sigma = \widetilde{X} \times \sqrt{\frac{\widetilde{X}^2}{P_{50}^2} - 1} \tag{10}$$

The skewness:

$$\gamma = \frac{\Sigma^3}{\widetilde{X}^3} + 3\frac{\Sigma}{\widetilde{X}} \tag{11}$$

Vice versa, the percentile values are calculated as follows:

$$P_{50} = \frac{\widetilde{X}^2}{\sqrt{\widetilde{X}^2 + \Sigma^2}} \tag{12}$$

$$P_{90} \approx P_{50} \times e^{-1.812\sqrt{ln\left(\widetilde{X}\right) - ln(P_{50})}} \tag{13}$$

$$P_{10} \approx P_{50} \times e^{+1.812\sqrt{ln\left(\widetilde{X}\right) - ln(P_{50})}} \tag{14}$$

We can introduce a third parameter: a shift. In that way the lognormal can be positioned elsewhere on the horizontal axis. See figure 2.

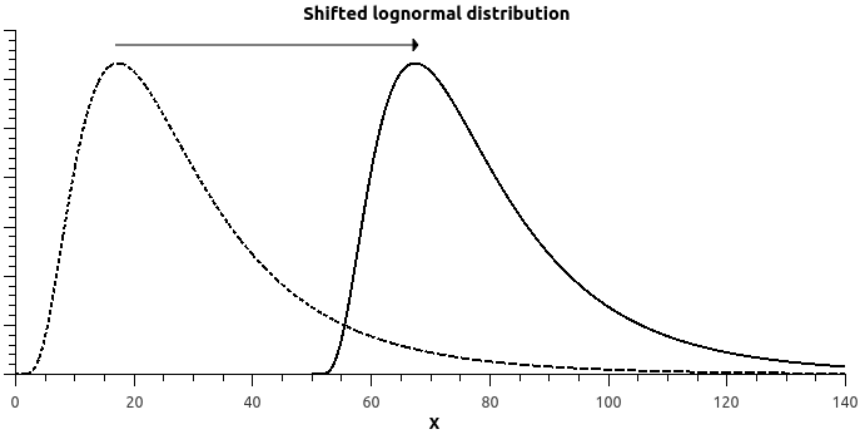

Figure 2: Shifted lognormal

If we have a set of arbitrary percentile numbers P_{90}, P_{50}, P_{10} ($P_{10} > P_{50} > P_{90}$, $P_{10} - P_{50} > P_{50} - P_{90}$) we can always fit a lognormal distribution by making use of a shift.

First we calculate the shift C:

$$C = \frac{P_{50}^2 - P_{90} \times P_{10}}{2P_{50} - P_{90} - P_{10}} \tag{15}$$

We then apply the shift to the percentiles:

$$P'_{90} = P_{90} - C \tag{16}$$
$$P'_{50} = P_{50} - C \tag{17}$$
$$P'_{10} = P_{10} - C \tag{18}$$

We can now proceed with calculating the expectation or mean ($\widetilde{X}'$), standard deviation (Σ') and skew (γ') of the shifted lognormal distribution. However, the shift does not affect the standard deviation and the skew, hence $\Sigma = \Sigma'$ and $\gamma = \gamma'$.

$$\widetilde{X}' = P'_{50} \times exp\left(K_1 \times [ln(P'_{10}) - ln(P'_{50})]^2\right) \tag{19}$$
$$\widetilde{X} = \widetilde{X}' + C \tag{20}$$
$$\Sigma = \Sigma' = \widetilde{X}' \times \sqrt{\frac{\widetilde{X}'^2}{P'^2_{50}} - 1} \tag{21}$$
$$\gamma = \gamma' = \frac{\Sigma^3}{\widetilde{X}'^3} + 3\frac{\Sigma}{\widetilde{X}'} \tag{22}$$

So this means that for any set of three numbers we can calculate an exactly fitting lognormal distribution provided that $P_{10} - P_{50}$ is greater than $P_{50} - P_{90}$.

In the following section you will find the derivations of these formulas.

3 Derivations

3.1 Percentiles

We define x as a stochastic normally distributed variable. Its mean is μ and its standard deviation σ. If X is the corresponding lognormal variable, then we have:

$$X = e^x \tag{23}$$
$$x = ln(X) \tag{24}$$

We define a percentile value for x as $x_{p,\alpha}$ with $P(x > x_{p,\alpha}) = \alpha$, with α the descending cumulative chance.

We can write:

$$x_{p,\alpha} = \mu + q_\alpha \times \sigma \tag{25}$$

q_α is a constant that follows from the standardized normal distribution which depends on the percentile chosen. For example for $\alpha = 10\%$, $q_\alpha \approx 1.282$.[2] The corresponding lognormal percentile is found as follows (we write q instead of q_α for simplicity):

$$X_p = e^{\mu+q\sigma} \tag{26}$$
$$= e^{\mu} e^{q\sigma} \tag{27}$$
$$= P_{50} e^{q\sigma} \tag{28}$$

Going back to the normal distribution of x we know that:

$$x_{p,\alpha} = \mu + q \times \sigma \tag{29}$$
$$x_{p,1-\alpha} = \mu - q \times \sigma \tag{30}$$

$x_{p,\alpha}$ and $x_{p,1-\alpha}$ are symmetric percentiles of the normal distribution for example the P_{90} and P_{10} if we were to take $q = 1.282$. The corresponding percentiles of the lognormal distribution X are:

$$X_{p,\alpha} = P_{50} \times e^{q\times\sigma} \tag{31}$$
$$X_{p,1-\alpha} = P_{50} \times e^{-q\times\sigma} \tag{32}$$

So:

$$X_{p,\alpha} \times X_{p,1-\alpha} = P_{50} \times e^{q\times\sigma} \times P_{50} \times e^{-q\times\sigma} \tag{33}$$
$$= {P_{50}}^2 \tag{34}$$

For example:

$${P_{50}}^2 = P_{10} \times P_{90} \tag{35}$$

[2] More precisely $\alpha = 1.2815515655446$. This value can be calculated in a spreadsheet by using the function NORM.INV(90%,0,1).

3.2 From percentiles to expectation

So if we have a normally distributed variable x, then for a specific percentile, we have:

$$q_\alpha = \frac{x_{p,\alpha} - \mu}{\sigma} \tag{36}$$

If we now consider that a percentile value $x_p = ln(P_\alpha)$, where P_α is the percentile value of the lognormal distribution that has a chance α of being exceeded. So then:

$$q_\alpha = \frac{ln(P_\alpha) - \mu}{\sigma} \tag{37}$$

Thus:

$$\sigma = \frac{ln(P_\alpha) - \mu}{q_\alpha} \tag{38}$$

We also know that $P_{50} = e^\mu$ hence $\mu = ln(P_{50})$. We then get:

$$\sigma = \frac{ln(P_\alpha) - ln(P_{50})}{q_\alpha} \tag{39}$$

So for example if we take $\alpha = 10\%$:

$$\sigma \approx \frac{ln(P_{10}) - ln(P_{50})}{1.282} \tag{40}$$

So if we now remember that $\widetilde{X} = E[X] = e^{\mu + \frac{1}{2}\sigma^2}$, then (writing $exp(x)$ rather than e^x for clarity):

$$\widetilde{X} = P_{50} \times exp\left(\frac{[ln(P_\alpha) - ln(P_{50})]^2}{2[q_\alpha]^2}\right) \tag{41}$$

$$\widetilde{X} = P_{50} \times exp\left(K_1 \times [ln(P_{10}) - ln(P_{50})]^2\right) \tag{42}$$

K_1 is a constant: $K_1 = \frac{1}{2{q_\alpha}^2} \approx 0.30443728$, for $\alpha = 10\%$

K_1 can be calculated by the spreadsheet function $(0.5/(NORM.INV(90\%, 0, 1)^2)$.

This equation will give the same result if we replace P_{10} by P_{90}. This provides a way to calculate an exact mean of a lognormal distribution from the percentile values.

3.3 The variance and standard deviation

We have:

$$E[X] = \widetilde{X} \tag{43}$$
$$\widetilde{X} = e^{\mu + \frac{\sigma^2}{2}} \tag{44}$$
$$= e^{\mu}\sqrt{e^{\sigma^2}} \tag{45}$$
$$= P_{50}\sqrt{e^{\sigma^2}} \tag{46}$$

Hence:

$$e^{\sigma^2} = \frac{\widetilde{X}^2}{{P_{50}}^2} \tag{47}$$

We have

$$VAR[X] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} \tag{48}$$

$$VAR[X] = (e^{\sigma^2} - 1)(e^{\mu})^2 e^{\sigma^2} \tag{49}$$
$$= \left(\frac{\widetilde{X}^2}{{P_{50}}^2} - 1\right) {P_{50}}^2 \frac{\widetilde{X}^2}{{P_{50}}^2} \tag{50}$$
$$= \widetilde{X}^2 \left(\frac{\widetilde{X}^2}{{P_{50}}^2} - 1\right) \tag{51}$$

Hence:

$$\Sigma = \widetilde{X}\sqrt{\frac{\widetilde{X}^2}{{P_{50}}^2} - 1} \tag{52}$$

3.4 The skewness

$$\gamma = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1} \tag{53}$$
$$= (e^{\sigma^2} + 2)\sqrt{\frac{\Sigma^2}{e^{2\mu + \sigma^2}}} \tag{54}$$
$$= (e^{\sigma^2} + 2)\sqrt{\frac{\Sigma^2}{\widetilde{X}^2}} \tag{55}$$

We remember that:

$$\Sigma^2 = VAR[X] = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2} \tag{56}$$
$$= (e^{\sigma^2} - 1)\widetilde{X}^2 \tag{57}$$

Hence:

$$e^{\sigma^2} = \frac{\Sigma^2 + \widetilde{X}^2}{\widetilde{X}^2} \tag{58}$$

With this we can take the derivation of the skewness further:

$$\gamma = (\frac{\Sigma^2 + \widetilde{X}^2}{\widetilde{X}^2} + 2)\sqrt{\frac{\Sigma^2}{\widetilde{X}^2}} \tag{59}$$
$$= (\frac{\Sigma^2}{\widetilde{X}^2} + 3)\frac{\Sigma}{\widetilde{X}} \tag{60}$$
$$\gamma = \frac{\Sigma^3}{\widetilde{X}^3} + 3\frac{\Sigma}{\widetilde{X}} \tag{61}$$

We have thus expressed the skewness (γ) of the lognormal distribution as a function of the mean ($\widetilde{X}$) and the standard deviation (Σ). This equation is central in the DeltaLogN method.

3.5 From moments to percentiles

We have with (51):

$$\Sigma^2 = \widetilde{X}^2 \left(\frac{\widetilde{X}^2}{{P_{50}}^2} - 1 \right) \tag{62}$$

Hence

$$\frac{\Sigma^2}{\widetilde{X}^2} = \frac{\widetilde{X}^2}{{P_{50}}^2} - 1 \tag{63}$$
$$\frac{\widetilde{X}^2}{{P_{50}}^2} = \frac{\Sigma^2}{\widetilde{X}^2} + 1 \tag{64}$$
$$P_{50}^2 = \frac{\widetilde{X}^4}{\Sigma^2 + \widetilde{X}^2} \tag{65}$$

Hence:

$$P_{50} = \frac{\widetilde{X}^2}{\sqrt{\widetilde{X}^2 + \Sigma^2}} \tag{66}$$

If we rewrite (28) as follows:

$$P_\alpha = P_{50} \times exp\left(q_\alpha \sigma\right) \tag{67}$$

Also, for (7) we get:

$$\sigma = \sqrt{ln\left(\frac{\widetilde{X}^2 + \Sigma^2}{\widetilde{X}^2}\right)} \tag{68}$$

and with (66) this becomes:

$$\sigma = \sqrt{ln\left(\frac{\widetilde{X}^2}{P_{50}^2}\right)} \tag{69}$$

$$\sigma = \sqrt{2} \times \sqrt{ln\left(\frac{\widetilde{X}}{P_{50}}\right)} \tag{70}$$

Now we can write for P_{90} and P_{10}, using (67):

$$P_{90} = P_{50} \times exp\left\{q_{90\%} \times \sqrt{2} \times \sqrt{ln\left(\frac{\widetilde{X}}{P_{50}}\right)}\right\} \tag{71}$$

$$P_{90} \approx P_{50} \times e^{-1.812\sqrt{ln\left(\widetilde{X}\right) - ln(P_{50})}} \tag{72}$$

$$P_{10} = P_{50} \times exp\left\{q_{10\%} \times \sqrt{2} \times \sqrt{ln\left(\frac{\widetilde{X}}{P_{50}}\right)}\right\} \tag{73}$$

$$P_{10} \approx P_{50} \times e^{+1.812\sqrt{ln\left(\widetilde{X}\right) - ln(P_{50})}} \tag{74}$$

Remember, q_α is defined as the ordinate value of the standard normal distribution that has a chance α of being exceeded. From the standard normal distribution for $\alpha = 90\%$ then $q = -1.282..$, for $\alpha = 10\%$ then $q = +1.282...$ These numbers need to be multiplied by $\sqrt{2}$ to get -1.812 and $+1.812$ respectively.

3.6 The shifted lognormal distribution

How did we get expression (15) to calculate C?

$$P'^2_{50} = P'_{90} \times P'_{10} \tag{75}$$
$$(P_{50} - C)^2 = (P_{90} - C) \times (P_{10} - C) \tag{76}$$
$$P^2_{50} - 2CP_{50} + C^2 = P_{90}P_{10} - CP_{10} - CP_{90} + C^2 \tag{77}$$
$$2CP_{50} - C(P_{10} + P_{90}) = P^2_{50} - P_{90}P_{10} \tag{78}$$
$$C = \frac{P^2_{50} - P_{90}P_{10}}{2P_{50} - P_{90} - P_{10}} \tag{79}$$

4 Other topics

4.1 Mirror imaged lognormal distributions

Sometimes we would like to model a phenomenon that has more downside than upside. Or we may have a derived variable that has a negative skew. We can use mirror imaged lognormal distributions for this purpose.

If X is a negatively skewed stochastic variable defined by P_{90}, P_{50}, P_{10} with $(P_{50} - P_{90}) > (P_{10} - P_{50})$ then we can define X' such that it conforms to a two-parameter lognormal distribution by applying this transformation:

$$X' = -X + C \tag{80}$$

One can easily verify that expression (15) applies in this case as well. Once we have calculated C, we can calculate the percentile values of distribution of X' as follows:

$$P'_{90} = -P_{10} + C \tag{81}$$
$$P'_{50} = -P_{50} + C \tag{82}$$
$$P'_{10} = -P_{90} + C \tag{83}$$

From these percentiles (primed) we can calculate the moments of the transformed distribution using the set of equations (19) to (22). For the moments of the original left skewed distribution of X we simply have:

$$\widetilde{X} = -\widetilde{X}' + C \tag{84}$$
$$\Sigma^2 = \Sigma'^2 \tag{85}$$
$$\gamma = -\gamma' \tag{86}$$

This mirror imaging together with the shift option as per equation 15 means that **it is possible to fit a lognormal distribution to any three numbers**.

4.2 Asymmetry Ratio

One of the characteristics of the lognormal distribution is also that it can accommodate high skews. In this respect it is more flexible than for example the triangular or pert distribution. At the same time, this is a disadvantage. One may be tempted to model some uncertain variable with a highly asymmetric distribution to show that there is a significant probability of a substantial outlier. This may lead to modelling artefacts.

We introduce the *asymmetry ratio*. This is not an established indicator but we define and use it in the NavIncerta courses.

The asymmetry ratio is defined as:

$$AR = \frac{P_{10} - P_{50}}{P_{50} - P_{90}} \tag{87}$$

Hence, AR is always greater than 1.

Another way to interpret the asymmetry ratio is upside divided by downside. The higher the AR, the more skewed the distribution.

In the graph below we plot the normalised standard deviation[3] as a function of AR. The curve starts at the left for a distribution with $P_{90} = 1$, P_{50} = 2 and $P_{10} = 4$. We keep the P_{90} and P_{50} fixed and increase the P_{10} from left to right (and thus the AR). What we see is that as the AR increases, the standard deviation tends to increase rapidly. This is caused by the infinite tail of the lognormal distribution, which gets a higher weight at higher asymmetry ratios. This effect is even greater for the skewness.

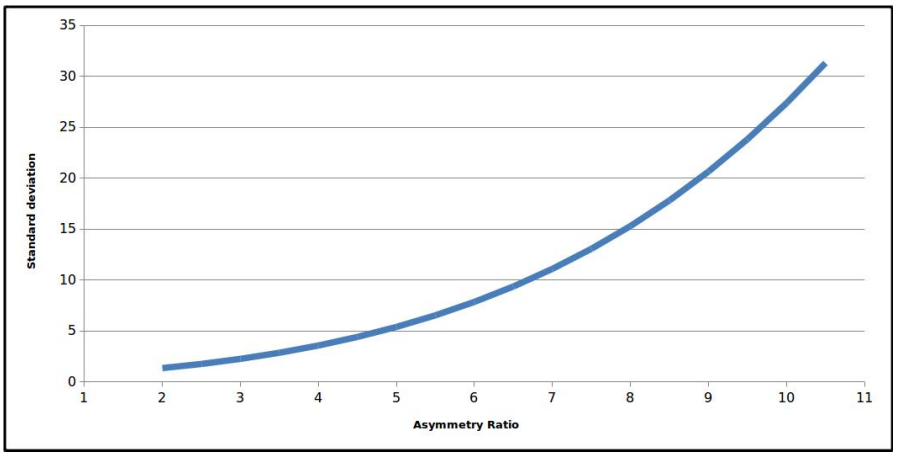

Figure 3: Normalised standard deviation as a function of AR

The implication is that if we use highly skewed (asymmetric) lognormal distributions, the standard deviation and skewness become very (unrealistically) large. A Monte Carlo simulation would become unstable. High skews are therefore to be avoided.

[3] We mean the standard deviation divided by $[P_{50} - P_{90}]$.

It is nearly always an unrealistic proposition to model an uncertain variable with a high AR. If this is proposed it would be appropriate to investigate the reason and consider a different modelling approach. To avoid such issues, it is also possible to use a truncated distribution, with for example the last 1% of the distribution ignored. This approach will be discussed in another article.