



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

Print/export

Languages

Čeština
Deutsch
Español
فارسی
Français
Galego
Italiano
עברית
Magyar
Nederlands
日本語
Polski
Português
Русский
Slovenščina
Basa Sunda
Svenska
Türkçe
Українська
中文

Edit links

Article **Talk**

Read **Edit** **View history**

Search

Log-normal distribution

From Wikipedia, the free encyclopedia
(Redirected from **Lognormal distribution**)

In **probability theory**, a **log-normal (or lognormal) distribution** is a continuous **probability distribution** of a **random variable** whose **logarithm** is **normally distributed**. Thus, if the random variable ***X*** is log-normally distributed, then ***Y*** = **log**(***X***) has a normal distribution. Likewise, if ***Y*** has a normal distribution, then ***X*** = **exp**(***Y***) has a log-normal distribution. A random variable which is log-normally distributed takes only positive real values.

The distribution is occasionally referred to as the **Galton distribution** or **Galton's distribution**, after **Francis Galton**.^[1] The log-normal distribution also has been associated with other names, such as McAlister, Gibrat and Cobb–Douglas.^[1]

A variable might be modeled as log-normal if it can be thought of as the multiplicative **product** of many **independent random variables** each of which is positive. (This is justified by considering the **central limit theorem** in the log-domain.) For example, in finance, the variable could represent the compound return from a sequence of many trades (each expressed as its return + 1); or a long-term **discount factor** can be derived from the product of short-term discount factors. In wireless communication, the sas caused by shadowing or slow fading from random objects is often assumed to be log-normally distributed: see **log-distance path loss model**.

The log-normal distribution is the **maximum entropy probability distribution** for a random variate *X* for which the mean and variance of **ln**(***X***) are fixed.^[2]

Contents

- Notation
- Characterization
 - Probability density function
 - Cumulative distribution function
 - Characteristic function and moment generating function
- Properties
 - Location and scale
 - Geometric moments
 - Arithmetic moments
 - Mode and median
 - Coefficient of variation
 - Partial expectation
 - Other
- Occurrence
- Maximum likelihood estimation of parameters
- Multivariate log-normal
- Generating log-normally distributed random variates
- Related distributions
- Similar distributions
- See also
- Notes
- References
- Further reading
- External links

Notation

In a log-normal distribution *X*, the parameters denoted *μ* and *σ* are, respectively, the **mean** and **standard deviation** of the variable's natural **logarithm** (by definition, the variable's logarithm is normally distributed), which means

$$X = e^{\mu + \sigma Z}$$

with *Z* a **standard normal variable**.

This relationship is true regardless of the base of the logarithmic or exponential function. If log_{*a*}(*Y*) is normally distributed, then so is log_{*b*}(*Y*), for any two positive numbers *a*, *b* ≠ 1. Likewise, if *e^X* is log-normally distributed, then so is ***a**^X*, where ***a*** is a positive number ≠ 1.

On a logarithmic scale, *μ* and *σ* can be called the *location parameter* and the *scale parameter*, respectively.

In contrast, the mean, standard deviation, and variance of the non-logarithmized sample values are respectively denoted *m*, *s.d.*, and *v* in this article. The two sets of parameters can be related as (see also **#Arithmetic moments** below):^[3]

$$\mu = \ln \left(\frac{m^2}{\sqrt{v + m^2}} \right), \sigma = \sqrt{\ln \left(1 + \frac{v}{m^2} \right)}$$

Characterization

Probability density function

The **probability density function** of a log-normal distribution is:^[1]

$$f_X(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\,e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}},\quad x>0$$

This follows by applying the **change-of-variables rule** on the density function of a normal distribution.

Cumulative distribution function

The **cumulative distribution function** is

$$F_X(x;\mu,\sigma) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right)\right] = \frac{1}{2}\operatorname{erfc}\left(-\frac{\ln x - \mu}{\sigma\sqrt{2}}\right) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right),$$

where erfc is the **complementary error function**, and Φ is the cumulative distribution function of the **standard normal** distribution.

Characteristic function and moment generating function

All moments of the log-normal distribution exist and it holds that: **E**(***X**ⁿ*) = **e**^{*nμ* + ^{*n*²*σ*²}²}. However, the expected value **E**(***e**^{*tX*}*) is not defined for any positive value of the argument ***t*** as the defining integral diverges. In consequence the **moment generating function** is not defined.^[4] The last is related to the fact that the lognormal distribution is not uniquely determined by its moments.

Similarly, the **characteristic function** **E**[*e^{itX}*] is not defined in the half complex plane and therefore it is not **analytic** in the origin. In consequence, the characteristic function of the log-normal distribution cannot be represented as an infinite convergent series.^[5] In particular, its Taylor **formal series**

∑

n
=
0

∞

i
t

n

n
!

e

n
μ
+

n

2

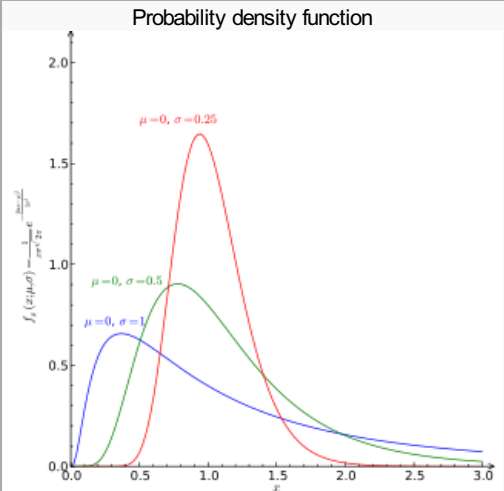
σ

2

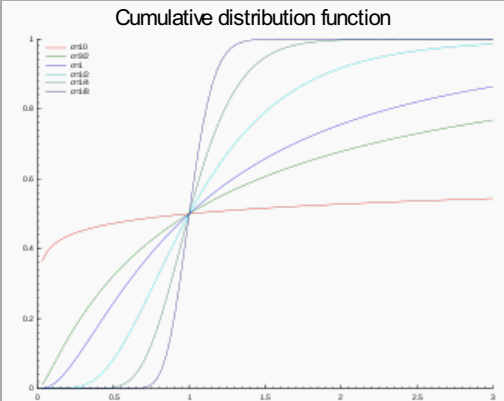
/
2

 diverges. However, a number of alternative

Log-normal



Some log-normal density functions with identical location parameter *μ* but differing scale parameters *σ*



Cumulative distribution function of the log-normal distribution (with *μ* = 0)

Notation	 ln N (μ<!-- μ --> , σ<!-- σ -->)
Parameters	<i>σ</i> > 0 — shape (real), <i>μ</i> ∈ R — log-scale
Support	<i>x</i> ∈ (0, +∞)
pdf	 1 x √<!-- √ --> 2 π<!-- π --> σ<!-- σ --> e −<!-- − --> (ln ⁡<!-- ⁡ --> x −<!-- − --> μ<!-- μ -->) 2 2 σ<!-- σ --> 2
CDF	 1 2 + 1 2 erf ⁡<!-- ⁡ --> [ln ⁡<!-- ⁡ --> x −<!-- − --> μ<!-- μ --> √<!-- √ --> 2 σ<!-- σ -->]
Mean	 e μ<!-- μ --> + σ<!-- σ --> 2 / 2
Median	 e μ<!-- μ -->
Mode	 e μ<!-- μ --> −<!-- − --> σ<!-- σ --> 2
Variance	 (e σ<!-- σ --> 2 −<!-- − --> 1) e 2 μ<!-- μ --> + σ<!-- σ --> 2
Skewness	 (e σ<!-- σ --> 2 + 2) √<!-- √ --> e σ<!-- σ --> 2 −<!-- − --> 1
Ex. kurtosis	 e 4 σ<!-- σ --> 2 + 2 e 3 σ<!-- σ --> 2 + 3 e 2 σ<!-- σ --> 2 −<!-- − --> 6
Entropy	 1 2 + 1 2 ln ⁡<!-- ⁡ --> (2 π<!-- π --> σ<!-- σ --> 2) + μ<!-- μ -->
MGF	(defined only on the negative half-axis, see text)
CF	representation ∑<!-- ∑ --> n = 0 ∞<!-- ∞ --> i t) n n ! e n μ<!-- μ --> + n 2 σ<!-- σ --> 2 / 2 is asymptotically divergent but sufficient for numerical purposes
Fisher information	 (1 σ<!-- σ --> 2 0 0 1 (2 σ<!-- σ --> 4))

divergent series representations have been obtained ^{[5] [6] [7] [8]}

A closed-form formula for the characteristic function

ϕ
(
t
)

{\displaystyle \varphi (t)}

 with

t

{\displaystyle t}

 in the domain of convergence is not known. A relatively simple approximating formula is available in closed form and given by

ϕ
(
t
)
≈

exp
⁡
(
−

W

2

(
t

σ

2

e

μ

)
+
2
W
(
t

σ

2

e

μ

)

2

σ

2

)

,

[9]

{\displaystyle \varphi (t)\approx {\frac {\exp \left(-{\frac {W^{2}(t\sigma ^{2}e^{\mu })+2W(t\sigma ^{2}e^{\mu })}{2\sigma ^{2}}}\right)}{\sqrt {1+W(t\sigma ^{2}e^{\mu })}}},\,[9]}

where

W

{\displaystyle W}

 is the **Lambert W function**. This approximation is derived via an asymptotic method but it stays sharp all over the domain of convergence of

ϕ
.

{\displaystyle \varphi .}

Properties ^[edit]

Location and scale ^[edit]

For the log-normal distribution, the location and scale properties of the distribution are more readily treated using the **geometric mean** and **geometric standard deviation** than the **arithmetic mean** and standard deviation.

Geometric moments ^[edit]

The **geometric mean** of the log-normal distribution is

e

μ

{\displaystyle e^{\mu }}

. Because the log of a log-normal variable is symmetric and quantiles are preserved under monotonic transformations, the geometric mean of a log-normal distribution is equal to its median.^[10]

The geometric mean (*m*_g) can alternatively be derived from the arithmetic mean (*m*_a) in a log-normal distribution by:

m

g

=

m

a

e

−

1
2

σ

2

.

{\displaystyle m_{g}=m_{a}e^{-{\frac {1}{2}}\sigma ^{2}}.}

Note that the geometric mean is *less* than the arithmetic mean. This is due to the **AM–GM inequality**, and corresponds to the logarithm being convex down. The correction term

e

−

1
2

σ

2

{\displaystyle e^{-{\frac {1}{2}}\sigma ^{2}}}

 can accordingly be interpreted as a **convexity correction**. From the point of view of **stochastic calculus**, this is the same correction term as in **Itô's lemma for geometric Brownian motion**.

The **geometric standard deviation** is equal to

e

σ

.

{\displaystyle e^{\sigma }.}

^[*citation needed*]

Arithmetic moments ^[edit]

If *X* is a lognormally distributed variable, its **expected value** (E – the **arithmetic mean**), **variance** (Var), and **standard deviation** (s.d.) are

E
[
X
]
=

e

μ
+

1
2

σ

2

,

{\displaystyle \mathrm {E} [X]=e^{\mu +{\frac {1}{2}}\sigma ^{2}},}

Var
[
X
]
=
(

e

σ

2

−
1
)

e

2
μ
+

σ

2

=
(

e

σ

2

−
1
)
(
E
[
X
]

)

2

{\displaystyle \mathrm {Var} [X]=(e^{\sigma ^{2}}-1)e^{2\mu +\sigma ^{2}}=(e^{\sigma ^{2}}-1)(\mathrm {E} [X])^{2}}

s
.
d
.
[
X
]
=

Var
[
X
]

=

e

μ
+

1
2

σ

2

√

e

σ

2

−
1
.

{\displaystyle \mathrm {s.d.} [X]={\sqrt {\mathrm {Var} [X]}}=e^{\mu +{\frac {1}{2}}\sigma ^{2}}{\sqrt {e^{\sigma ^{2}}-1}}.}

Therefore, *s.d.* = *m* × c.v., in terms of the expected value and the **#coefficient of variation**.

Equivalently, parameters *μ* and *σ* can be obtained if the expected value and variance are known; it is simpler if *σ* is computed first:

μ
=
ln
⁡
(
E
[
X
]
)
−

1
2

ln
⁡
(
1
+

Var
[
X
]

(
E
[
X
]

)

2

)
=
ln
⁡
(
E
[
X
]
)
−

1
2

σ

2

,

{\displaystyle \mu =\ln(E[X])-{\frac {1}{2}}\ln\left(1+{\frac {\mathrm {Var} [X]}{(\mathrm {E} [X])^{2}}}\right)=\ln(E[X])-{\frac {1}{2}}\sigma ^{2},}

σ

2

=
ln
⁡
(
1
+

Var
[
X
]

(
E
[
X
]

)

2

)
.

{\displaystyle \sigma ^{2}=\ln\left(1+{\frac {\mathrm {Var} [X]}{(\mathrm {E} [X])^{2}}}\right).}

For any real or complex number *s*, the *s*th **moment** of log-normal *X* is given by^[1]

E
[

X

s

]
=

e

s
μ
+

1
2

s

2

σ

2

.

{\displaystyle \mathrm {E} [X^{s}]=e^{s\mu +{\frac {1}{2}}s^{2}\sigma ^{2}}.}

A log-normal distribution is not uniquely determined by its moments

E

[

X

k

]

{\displaystyle \mathrm {E} [X^{k}]}

 for *k* ≥ 1, that is, there exists some other distribution with the same moments for all *k*.^[1] In fact, there is a whole family of distributions with the same moments as the log-normal distribution.^[*citation needed*]

Mode and median ^[edit]

The **mode** is the point of global maximum of the probability density function. In particular, it solves the equation (ln *f*)' = 0:

Mode
[
X
]
=

e

μ
−

σ

2

.

{\displaystyle \mathrm {Mode} [X]=e^{\mu -\sigma ^{2}}.}

The **median** is such a point where *F*_{*X*} = 1/2:

Med
[
X
]
=

e

μ

.

{\displaystyle \mathrm {Med} [X]=e^{\mu }.}

Coefficient of variation ^[edit]

The **coefficient of variation** c.v. is the ratio *s.d.* over *m* (on the natural scale) and is equal to:

c
.
v
.
=

√

e

σ

2

−
1

{\displaystyle \mathrm {c.v.} ={\sqrt {e^{\sigma ^{2}}-1}}

Contrary to the *s.d.*, the *c.v.* is constant for log-normally distributed data, independent of the data expected value.

Partial expectation ^[edit]

The partial expectation of a random variable *X* with respect to a threshold *k* is defined as

g
(
k
)
=

∫

k

∞

x
f
(
x
)
d
x

{\displaystyle g(k)=\int _{k}^{\infty }xf(x)\,dx}

 where *f*(*x*) is the probability density function of *X*. Alternatively, and using the definition of **conditional expectation**, it can be written as *g*(*k*)=

E
[
X
|

X
>
k
]
P
(
X
>
k
)

{\displaystyle \mathrm {E} [X|X>k]P(X>k)}

. For a log-normal random variable the partial expectation is given by:

g
(
k
)
=

∫

k

∞

x
f
(
x
)
d
x
=

e

μ
+

1
2

σ

2

Φ
(

μ
+

σ

2

−
ln
⁡
k

σ

)
.

{\displaystyle g(k)=\int _{k}^{\infty }xf(x)\,dx=e^{\mu +{\frac {1}{2}}\sigma ^{2}}\Phi \left({\frac {\mu +\sigma ^{2}-\ln k}{\sigma }}\right).}

Where Phi is the **normal cumulative distribution function**. The derivation of the formula is provided in the discussion of this Wikipedia entry. The partial expectation formula has applications in insurance and economics, it is used in solving the partial differential equation leading to the **Black–Scholes formula**.

Other ^[edit]

A set of data that arises from the log-normal distribution has a symmetric **Lorenz curve** (see also **Lorenz asymmetry coefficient**).^[11]

The harmonic (*H*), geometric (*G*) and arithmetic (*A*) means of this distribution are related;^[12] such relation is given by

H
=

G

2

A

.

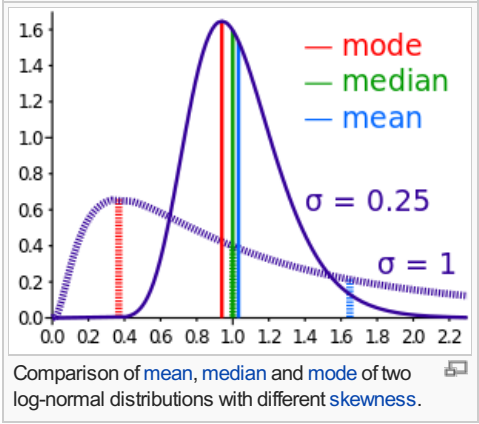
{\displaystyle H={\frac {G^{2}}{A}}.}

Log-normal distributions are **infinitely divisible**.^[1]

Occurrence ^[edit]

The log-normal distribution is important in the description of natural phenomena. The reason is that for many natural processes of growth, **growth rate** is independent of size. This is also known as **Gibrat's law**, after Robert Gibrat (1904–1980) who formulated it for companies. It can be shown that a growth process following Gibrat's law will result in entity sizes with a log-normal distribution.^[13] Examples include:

- In **biology** and **medicine**,
 - Measures of size of living tissue (length, skin area, weight);^[14]
 - For highly communicable epidemics, such as SARS in 2003, if publication intervention is involved, the number of hospitalized cases is shown to satisfy the lognormal distribution with no free parameters if an entropy is assumed and the standard deviation is determined by the principle of maximum rate of entropy production.^[15]
 - The length of inert appendages (hair, claws, nails, teeth) of biological specimens, in the direction of growth;^[*citation needed*]



- Certain physiological measurements, such as blood pressure of adult humans (after separation on male/female subpopulations)^[16]

Consequently, **reference ranges** for measurements in healthy individuals are more accurately estimated by assuming a log-normal distribution than by assuming a symmetric distribution about the mean.

- In **colloidal chemistry**,
 - Particle size distributions
- In **hydrology**, the log-normal distribution is used to analyze extreme values of such variables as monthly and annual maximum values of daily rainfall and river discharge volumes.^[17]
 - The image on the right illustrates an example of fitting the log-normal distribution to ranked annually maximum one-day rainfalls showing also the 90% **confidence belt** based on the **binomial distribution**. The rainfall data are represented by **plotting positions** as part of a **cumulative frequency analysis**.
- in social sciences and demographics
 - In **economics**, there is evidence that the **income** of 97%–99% of the population is distributed log-normally.^[18]
 - In **finance**, in particular the **Black–Scholes model**, changes in the *logarithm* of exchange rates, price indices, and stock market indices are assumed normal^[19] (these variables behave like compound interest, not like simple interest, and so are multiplicative). However, some mathematicians such as **Benoît Mandelbrot** have argued ^[20] that **log-Lévy distributions** which possesses **heavy tails** would be a more appropriate model, in particular for the analysis for **stock market crashes**. Indeed stock price distributions typically exhibit a **fat tail**.^[21]
 - city sizes**
- technology
 - In **reliability** analysis, the lognormal distribution is often used to model times to repair a maintainable system.^[22]
 - In **wireless communication**, "the local-mean power expressed in logarithmic values, such as dB or neper, has a normal (i.e., Gaussian) distribution." ^[23]
 - It has been proposed that coefficients of friction and wear may be treated as having a lognormal distribution ^[24]
 - In spray process, such as droplet impact, the size of secondary produced droplet has a lognormal distribution, with the standard deviation : $\sigma = \frac{\sqrt{6}}{6}$ determined by the principle

of maximum rate of entropy production^[25] If the lognormal distribution is inserted into the Shannon entropy expression and if the rate of entropy production is maximized (principle of maximum rate of entropy production), then σ is given by : $\sigma = \frac{1}{\sqrt{6}}$ ^[25] and with this parameter the droplet size distribution for spray process is well predicted. It is an open

question whether this value of σ has some generality for other cases, though for spreading of communicable epidemics, σ is shown also to take this value.^[15]

Maximum likelihood estimation of parameters [edit]

For determining the **maximum likelihood** estimators of the log-normal distribution parameters μ and σ, we can use the **same procedure** as for the **normal distribution**. To avoid repetition, we observe that

$$f_L(x;\mu,\sigma)=\prod_{i=1}^n\left(\frac{1}{x_i}\right)f_N(\ln x;\mu,\sigma)$$

where by *f_L* we denote the probability density function of the log-normal distribution and by *f_N* that of the normal distribution. Therefore, using the same indices to denote distributions, we can write the log-likelihood function thus:

$$\begin{aligned}\ell_L(\mu,\sigma|x_1,x_2,\ldots,x_n)&=-\sum_k\ln x_k+\ell_N(\mu,\sigma|\ln x_1,\ln x_2,\ldots,\ln x_n)\\&=\text{constant}+\ell_N(\mu,\sigma|\ln x_1,\ln x_2,\ldots,\ln x_n).\end{aligned}$$

Since the first term is constant with regard to μ and σ, both logarithmic likelihood functions, *ℓ_L* and *ℓ_N*, reach their maximum with the same μ and σ. Hence, using the formulas for the normal distribution maximum likelihood parameter estimators and the equality above, we deduce that for the log-normal distribution it holds that

$$\hat{\mu}=\frac{\sum_k\ln x_k}{n},\hat{\sigma}^2=\frac{\sum_k(\ln x_k-\hat{\mu})^2}{n}.$$

Multivariate log-normal [edit]

If ***X* ~ *N*(**μ**, **Σ**)** is a **multivariate normal distribution** then ***Y* = exp(*X*)** has a multivariate log-normal distribution^[26] with mean

$$\mathrm{E}[\mathbf{Y}]_i=e^{\mu_i+\frac{1}{2}\Sigma_{ii}},$$

and **covariance matrix**

$$\mathrm{Var}[\mathbf{Y}]_{ij}=e^{\mu_i+\mu_j+\frac{1}{2}(\Sigma_{ii}+\Sigma_{jj})}(e^{\Sigma_{ij}}-1).$$

Generating log-normally distributed random variates [edit]

Given a random variate Z drawn from the **normal distribution** with 0 mean and 1 standard deviation, then the variate

$$X=e^{\mu+\sigma Z}$$

has a log-normal distribution with parameters **μ** and **σ**.

Related distributions [edit]

- If *X* ~ *N*(μ, σ²) is a **normal distribution**, then exp(*X*) ~ Log-*N*(μ, σ²).
- If *X* ~ Log-*N*(μ, σ²) is distributed log-normally, then ln(*X*) ~ *N*(μ, σ²) is a normal random variable.
- If *X_j* ~ Log-*N*(μ_{*j*}, σ²_{*j*}) are *n* **independent** log-normally distributed variables, and *Y* = ∏_{*j*=1}^{*n*} *X_j*, then Y is also distributed log-normally:

$$Y\sim\mathrm{Log}\text{-}\mathcal{N}\left(\sum_{j=1}^n\mu_j,\;\sum_{j=1}^n\sigma_j^2\right).$$

- Let *X_j* ~ Log-*N*(μ_{*j*}, σ²_{*j*}) be independent log-normally distributed variables with possibly varying σ and μ parameters, and *Y* = ∑_{*j*=1}^{*n*} *X_j*. The distribution of Y has no closed-form expression, but can be reasonably approximated by another log-normal distribution Z at the right tail.^[27] Its probability density function at the neighborhood of 0 has been characterized^[28] and it does not resemble any log-normal distribution. A commonly used approximation due to L.F. Fenton (but previously stated by R.I. Wilkinson and mathematical justified by Marlow^[29]) is obtained by matching the mean and variance of another lognormal distribution:

$$\sigma_Z^2=\log\left[\frac{\sum e^{2\mu_j+\sigma_j^2}(e^{\sigma_j^2}-1)}{(\sum e^{\mu_j+\sigma_j^2/2})^2}+1\right],$$

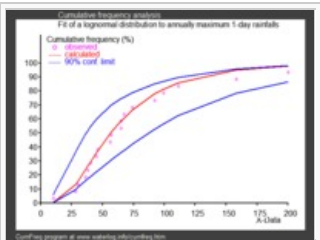
$$\mu_Z=\log\left[\sum e^{\mu_j+\sigma_j^2/2}\right]-\frac{\sigma_Z^2}{2}.$$

In the case that all *X_j* have the same variance parameter σ_{*j*} = σ, these formulas simplify to

$$\sigma_Z^2=\log\left[(e^{\sigma^2}-1)\frac{\sum e^{2\mu_j}}{(\sum e^{\mu_j})^2}+1\right],$$

$$\mu_Z=\log\left[\sum e^{\mu_j}\right]+\frac{\sigma^2}{2}-\frac{\sigma_Z^2}{2}.$$

- If *X* ~ Log-*N*(μ, σ²), then *X* + *c* is said to have a *shifted log-normal* distribution with support x ∈ (c, +∞). E[*X* + *c*] = E[*X*] + *c*, Var[*X* + *c*] = Var[*X*].
- If *X* ~ Log-*N*(μ, σ²), then *aX* ~ Log-*N*(μ + ln *a*, σ²).
- If *X* ~ Log-*N*(μ, σ²), then 1⁄*X* ~ Log-*N*(−μ, σ²).



Fitted cumulative log-normal distribution to annually maximum 1-day rainfalls, see **distribution fitting**

