CEE 260/MIE 273: Probability and Statistics in Civil Engineering

Lecture M3c: Lognormal and Exponential Distributions

Jimi Oke

UMass Amherst

College of Engineering

September 28, 2025

Outline

Recap of normal distribution

- The PDF of the normal distribution (parameters: mean μ and variance σ^2) is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{1}$$

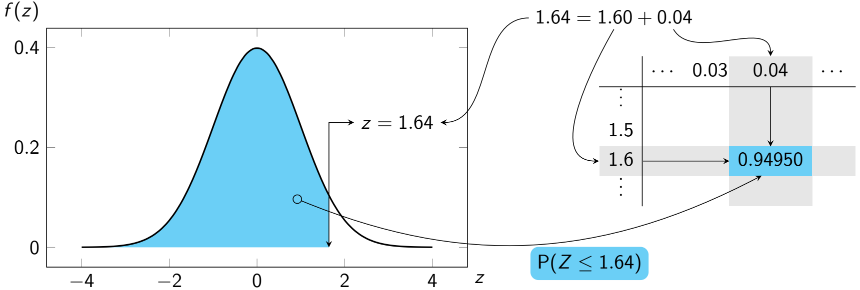
- There is no closed-form solution to the integral of the normal CDF
- Instead, it is customary to standardize a normal variable to its "Z-score":

$$Z = \frac{X-\mu}{\sigma} \tag{2}$$

- The mean and variance of the standard normal distribution are 0 and 1, respectively.
- The symbol Φ ("phi") is used to represent the CDF of the *standard* normal distribution, whose values can be looked up in a table.
- In Python, the `norm.cdf(x, mu, sigma)` and `norm.ppf(p, mu, sigma)` can be used to compute probabilities and inverse CDFs of the normal distribution, respectively.

Using the standard normal CDF probability table

- First convert the random variable to its Z-score
- Find the corresponding value in the table



Objectives of today's lecture

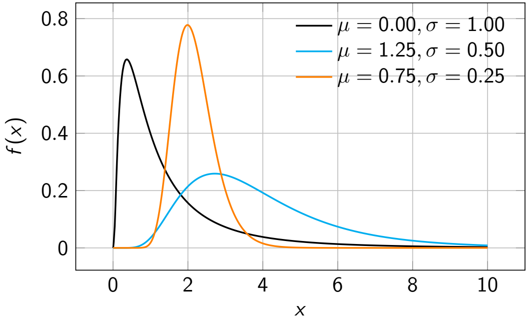
- Lognormal distribution
- Relationship between lognormal and normal distributions
- Exponential distribution
- Memoryless property of exponential distribution

Lognormal distribution

PDF

A random variable X that is lognormally distributed with the parameters μ and σ (denoted $X \sim \mathcal{LN}(\mu, \sigma)$) has the PDF:

$$f_X(x) = \frac{1}{(\sigma x)\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \mu}{\sigma}\right)^2\right] \quad x \geq 0 \tag{3}$$



Mean, median and variance of a lognormal distribution

Let X be a lognormally distributed random variable with parameters μ and σ: $X \sim \mathcal{LN}(\mu, \sigma^2)$

- The **mean** of X is given by

$$\mathbb{E}(X) = \exp\left[\mu + \frac{1}{2}\sigma^2\right] \tag{4}$$

- The **median** of X is:

$$\text{Median}(X) = \exp(\mu) \tag{5}$$

- The **variance** of X is given by:

$$\mathbb{V}(X) = \left(\exp\left[\sigma^2\right] - 1\right)\exp\left[2\mu + \sigma^2\right] \tag{6}$$

Notes

- μ and σ are the mean and standard deviation of the associated normal distribution of $\ln(X)$.
- Thus, if $X \sim \mathcal{LN}(\mu, \sigma)$, then $\ln(X) \sim \mathcal{N}(\mu, \sigma)$

Example 1: Mean and variance of lognormal distribution (1)

The incubation period of the COVID-19 infection is assumed to be lognormally distributed with a median of about 5 days and $\sigma^2 = 0.42$. What are the mean and variance of its distribution?

Solution

First, we find the parameter μ:

$$
\begin{aligned}
\text{Median}(X) &= e^{\mu} \\
5 &= e^{\mu} \\
\Longrightarrow \ln(5) &= \mu \\
\therefore \text{The mean is given by } \mathbb{E}(X) &= e^{\mu + \frac{1}{2}\sigma^2} = e^{\ln(5) + 0.21} = e^{\ln(5)} \times e^{0.21} \\
&= 5(e^{0.21}) = \boxed{6.17 \text{ days}}
\end{aligned}
$$

Example 1: Mean and variance of lognormal distribution (2)

Solution (cont.)

The variance is given by:

$$
\begin{aligned}
\mathbb{V}(X) &= \left(\exp(\sigma^2) - 1\right)\left(\exp[2\mu + \sigma^2]\right) \\
&= (\exp(0.42) - 1)(\exp(2\ln(5) + 0.42)) \\
&= \boxed{19.86 \text{ days}^2}
\end{aligned}
$$

Relationship between normal and lognormal distributions

- A random variable X is lognormally distributed with the **parameters** μ and σ^2 if $\ln(X)$ is normally distributed with the same parameters.

$$X \sim \mathcal{LN}(\mu, \sigma^2) \implies \ln(X) \sim \mathcal{N}(\mu, \sigma^2) \tag{7}$$

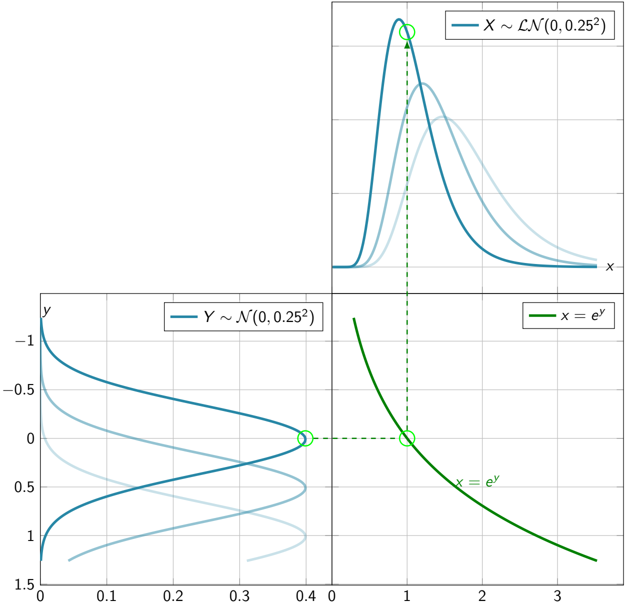
- Conversely, a random variable X is normally distributed with the parameters μ and σ^2 then e^X is lognormally distributed with the same parameters.

$$X \sim \mathcal{N}(\mu, \sigma^2) \implies e^X \sim \mathcal{LN}(\mu, \sigma^2) \tag{8}$$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathbb{V}(X)$

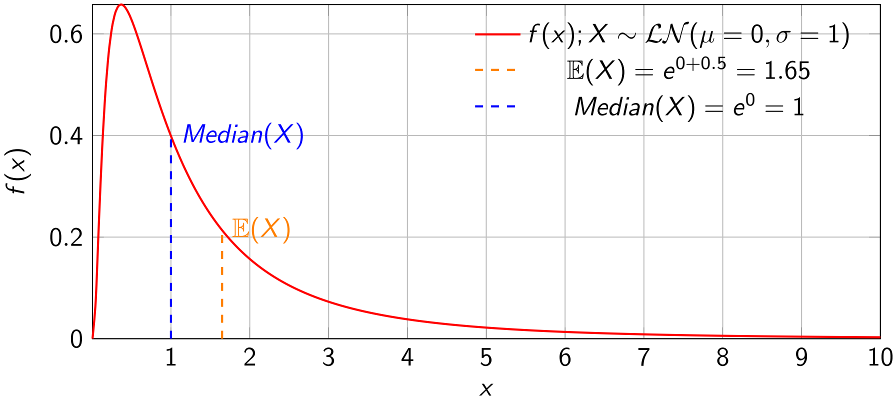
- However, $X \sim \mathcal{LN}(\mu, \sigma^2)$, then $\mu = \mathbb{E}(\ln(X))$ and $\sigma^2 = \mathbb{V}(\ln(X))$

Relationship between normal and lognormal (cont.)



Positive skewness of lognormal distribution

- The lognormal distribution is positively skewed
- Its mean is always greater than its median



Probability of a lognormal random variate

Given a r.v. X that is lognormally distributed with parameters μ and σ^2:

$$P(a < X \leq b) = \frac{1}{\sigma x\sqrt{2\pi}} \int_a^b \exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \mu}{\sigma}\right)^2\right] dx \tag{9}$$

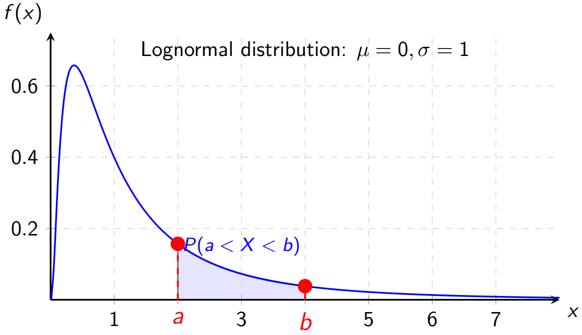


Figure: Lognormal distribution with $\mu = 0$, $\sigma = 1$, showing $P(a < X < b)$ where $a = 2$ and $b = 4$

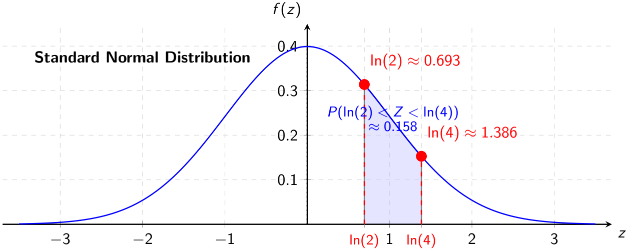
Probability of a lognormal random variate (cont.)

Substituting $z = \frac{\ln(x)-\mu}{\sigma} \implies dx = \sigma x dz$, we obtain:

$$P(a < X \le b) = \frac{1}{\sqrt{2\pi}} \int_{(\ln(a)-\mu)/\sigma}^{(\ln(b)-\mu)/\sigma} \exp\left[-\frac{1}{2}z^2\right] dz \tag{10}$$

Recognizing that the integrand is the PDF of the **standard normal distribution**, we have:

$$P(a < X \le b) = \Phi\left(\frac{\ln b - \mu}{\sigma}\right) - \Phi\left(\frac{\ln a - \mu}{\sigma}\right) \tag{11}$$



Example 2: Equipment breakdown

The lifetime X of a major oil platform equipment is lognormally distributed with a Median$(X) = 6$ months and $\sigma = 0.30$. To ensure 95% reliability, determine the desired interval x_0 for maintenance.

Given: $\mu = \ln 6 = 1.792$ and $\sigma = 0.30$, we want to find x_0 such that:

$$P(X > x_0) = 1 - P(X \leq x_0) = 0.95$$

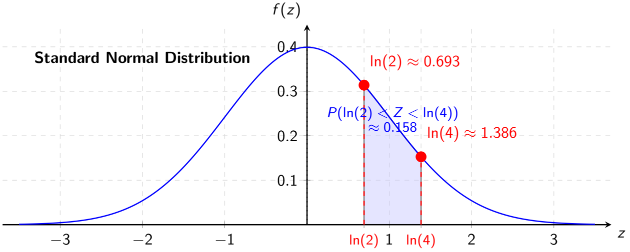


Figure: Standard normal distribution showing $P(\ln(2) < Z < \ln(4))$

Example 2: Probability of incubation period

The incubation period of the COVID-19 infection is assumed to be lognormally distributed with a median of about 5 days and $\sigma^2 = 0.42$. What is the probability that a randomly selected person will show symptoms within 7 days of exposure (i.e., $P(X \leq 7)$)?

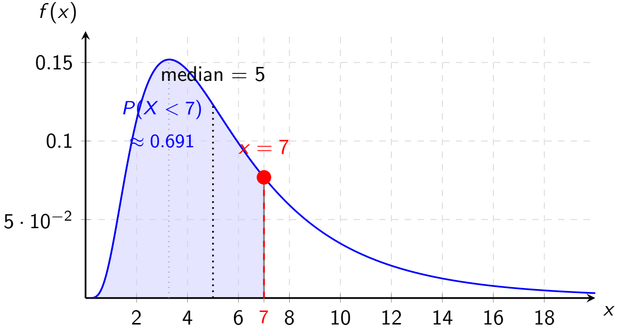


Figure: Lognormal distribution with $\mu = \ln(5)$, $\sigma^2 = 0.42$, showing $P(X \leq 7)$

Example 2: (cont.)

Solution (using Python)

We want to find: $P(X \leq 7) = F(7)$; thus we use the `scipy.stats.lognorm.cdf` function:

```
from scipy import stats
import numpy as np
p = stats.lognorm.cdf(7, s=0.648, scale=5)
```

where:

- `s` $= \sigma = 0.648 = \sqrt{0.42}$ (shape parameter/standard deviation of underlying normal)
- `scale` $= \exp(\mu) = \exp(\ln(5)) = 5$ (scale parameter/median)

The result is: `p = 0.691`, i.e., about 69.1% of the people will show symptoms within 7 days of exposure.

Example 2: (cont.)

Solution (using tables)

We want to find: $P(X \leq 7) = F(7)$; thus we use the relationship:

$$P(X \leq x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

where $\mu = \ln(5) = 1.609$ and $\sigma = \sqrt{0.42} = 0.648$. Thus:

$$\begin{aligned} P(X \leq 7) &= \Phi\left(\frac{\ln 7 - 1.609}{0.648}\right) \\ &= \Phi(0.539) \\ &\approx 0.7054 \text{ (from standard normal table)} \end{aligned}$$

The result is: $P(X \leq 7) \approx 0.7054$, i.e., about 70.54% of the people will show symptoms within 7 days of exposure.

Example 3: Equipment breakdown

The lifetime X of a major oil platform equipment is lognormally distributed with a Median$(X) = 6$ months and $\sigma = 0.30$. To ensure 95% reliability, determine the desired interval x_0 for maintenance.

Given: $\mu = \ln 6 = 1.792$ and $\sigma = 0.30$, we want to find x_0 such that:

$$P(X > x_0) = 1 - P(X \leq x_0) = 0.95 \implies P(X \leq x_0) = 0.05$$

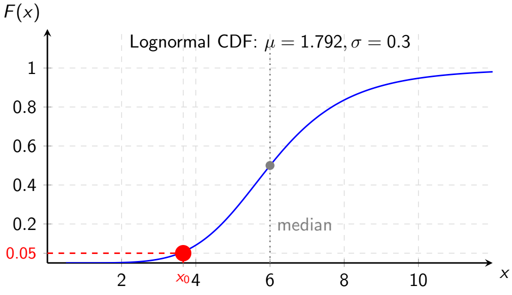


Figure: Lognormal CDF showing the 5th percentile at $x = x_0 = 3.66$

Example 3: Equipment breakdown (cont.)

Solution (using Python)

$P(X \leq x_0) = 0.05$ implies:

$$x_0 = F^{-1}(0.05)$$

Code:

```
from scipy import stats
import numpy as np
x0 = stats.lognorm.ppf(q=0.05, s=0.30, scale=6)
```

Note the following about the *arguments*`stats.lognorm.ppf` function:

- The first, `q`, is the cumulative probability or quantile (0.05)
- The second, `s`, is the shape parameter σ (0.30)
- The third, `scale`, is e^{μ} (median) (6)

This returns $x_0 = 3.66$ months.

Example 3: Equipment breakdown (cont.)

Solution (using tables)

Thus:

$$\Phi\left(\frac{\ln(x_0) - 1.792}{0.30}\right) = 0.05$$

$$\begin{aligned}
\frac{\ln x_0 - 1.792}{0.30} &= \Phi^{-1}(0.05) \\
\ln x_0 - 1.792 &= 0.30[-\Phi^{-1}(0.95)] \\
\ln x_0 &= 1.792 + 0.30(-1.65) \\
\ln x_0 &= 1.792 - 0.495 = 1.297
\end{aligned}$$

Therefore, the required inspection interval is:

$$x_0 = e^{1.297} = 3.66 \text{ months}$$

Modeling probabilities of elapsed times

Consider the random variable X which represents the *number of arrivals* at a restaurant within a given time interval.

- The probability of X in t time units can be modeled by the Poisson distribution with a rate parameter λt

Now consider the variable Y representing the **elapsed time** between successive arrivals.

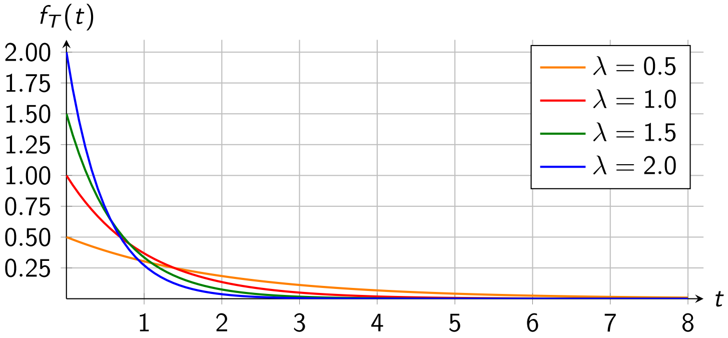
- What is the probability the time between the third and fourth arrivals is less than y minutes, for instance?
- This is modeled by the **exponential distribution** with parameter λ.

Exponential distribution

Definition

A random variable X that is exponentially distributed with parameter λ has the PDF:

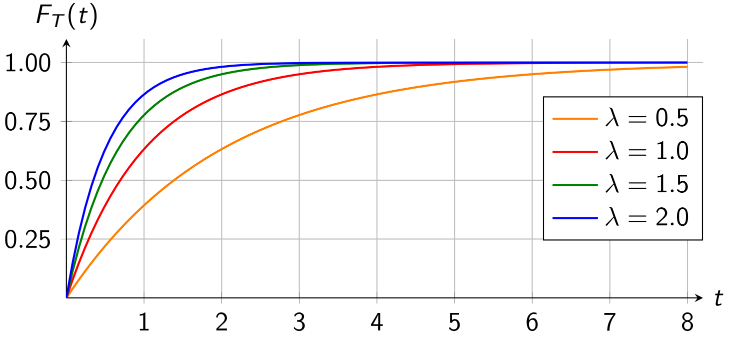
$$f_X(x) = \lambda e^{-\lambda x} \qquad x > 0 \tag{12}$$



CDF of the exponential distribution

The CDF of the exponential distribution is derived as:

$$F_X(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt$$
$$F_X(x) = 1 - e^{-\lambda x}$$



Note that $P(X \leq x) = 1 - e^{-\lambda x}$, while $P(X > x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$

Mean and variance of the exponential distribution

Let $X \sim \text{Exponential}(\lambda)$.

Mean

The mean of X is given by:

$$\mathbb{E}(X) = \frac{1}{\lambda} \tag{13}$$

Variance

The variance of X is given by:

$$\mathbb{V}(X) = \frac{1}{\lambda^2} \tag{14}$$

Example 3: Waiting for a flight

The delay time T of a flight is exponentially distributed wtih $\lambda = 2$ (delays per hour). Answer the following questions:

(a) What is the mean delay (waiting) time, $\mathbb{E}(T)$?

(b) What is the variance of the delay time $\mathbb{V}(T)$?

(c) Find the probability that a flight will be delayed by no more than 10 minutes.

(d) Assuming you have been waiting for a flight for an hour, what is the probability that the flight will be delayed for an additional 30 minutes? (i.e. Find $P(T > 1.5 | T > 1)$).

Example 3: Waiting for a flight (cont.)

Solution

(a) The mean delay is given by

$$\mathbb{E}(T) = \frac{1}{\lambda} = \frac{1}{2} = \boxed{0.5\text{hr}}$$

(b) The variance is:

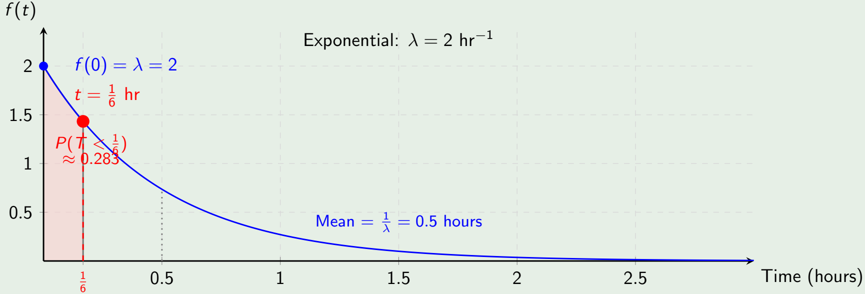
$$\mathbb{V}(T) = \frac{1}{\lambda^2} = \frac{1}{2^2} = \boxed{0.25\text{hr}^2}$$

Example 3: Waiting for a flight (cont.)

Solution

(c) The probability the flight will be delayed by no more than 10 minutes ($\frac{1}{6}$ hr) is given by:

$$
\begin{aligned}
P\left(T \leq \frac{1}{6}\right) &= 1 - \exp\left[-\lambda \cdot \frac{1}{6}\right] = 1 - \exp\left[-2\left(\frac{1}{6}\right)\right] = 1 - \exp\left[-\frac{1}{3}\right] \\
&= \boxed{0.283}
\end{aligned}
$$



Example 3: Waiting for a flight (cont.)

Solution

(c) In python, we can use the `scipy.stats.expon.cdf` function to find $P(T \leq 1/6)$:

```
from scipy import stats
import numpy as np
p = stats.expon.cdf(1/6, scale=1/2)
```

where:

- 1/6 is the value at which we want to evaluate the CDF
- `scale` $= 1/\lambda = 1/2$ (scale parameter/mean)

This returns `p = 0.283`, i.e., about 28.3% probability that the flight will be delayed by no more than 10 minutes.

Example 3: Waiting for a flight (cont.)

Solution

(d) The probability that the flight will be delayed by a further 0.5hr after 1hr of waiting is given by:

$$\begin{aligned} P(T > (0.5+1) | T > 1) &= P(T > 1.5 | T > 1) \\ &= \frac{P((T > 1.5) \cap (T > 1))}{P(T > 1)} \quad \text{(mult. rule)} \\ &= \frac{P(T > 1.5)}{P(T > 1)} \\ &= \frac{e^{-2(1.5)}}{e^{-2(1)}} = e^{-2[1.5-1.0]} \\ &= e^{-2(0.5)} \quad (= P(T > 0.5)) \\ &= e^{-1} = \boxed{0.37} \end{aligned}$$

In Python: `p = 1 - stats.expon.pdf(.5, scale=1/2)` also returns `p = 0.37`.

Memorylessness of the exponential distribution

This leads us to an important property of the exponential distribution

Memoryless property

$$P(T > t + s | T > s) = P(T > t) \tag{15}$$

That is, it does not matter from which time the waiting begins (i.e. conditioning); the probability of an elapsed time remains the same.

Recap

- **Lognormal distribution:** $X \sim \mathcal{LN}(\mu, \sigma^2)$
  CDF: $F_X(x) = P(X \leq x) = \Phi((\ln(x) - \mu)/\sigma)$
  Mean:
$$\mathbb{E}(X) = e^{\left(\mu + \frac{1}{2}\sigma^2\right)} \tag{16}$$
  Variance:
$$\mathbb{V}(X) = (e^{\sigma^2} - 1)e^{(2\mu+\sigma^2)} \tag{17}$$

- **Exponential distribution**: $X \sim \text{Exponential}(\lambda)$

$$\text{PDF:} \quad f_X(x) = \lambda e^{-\lambda x}, \qquad x > 0 \tag{18}$$
$$\text{CDF:} \quad F_X(x) = P(X \leq x) = 1 - e^{-\lambda x}, \qquad x > 0 \tag{19}$$

  Mean:
$$\mathbb{E}(X) = \frac{1}{\lambda} \tag{20}$$
  Variance:
$$\mathbb{V}(X) = \frac{1}{\lambda^2} \tag{21}$$