可靠度資料分析
Reliability Data Analysis

許舒涵 (Shu-han Hsu)

成功大學 資訊工程系

Lecture 4 – Common Failure Distribution Models

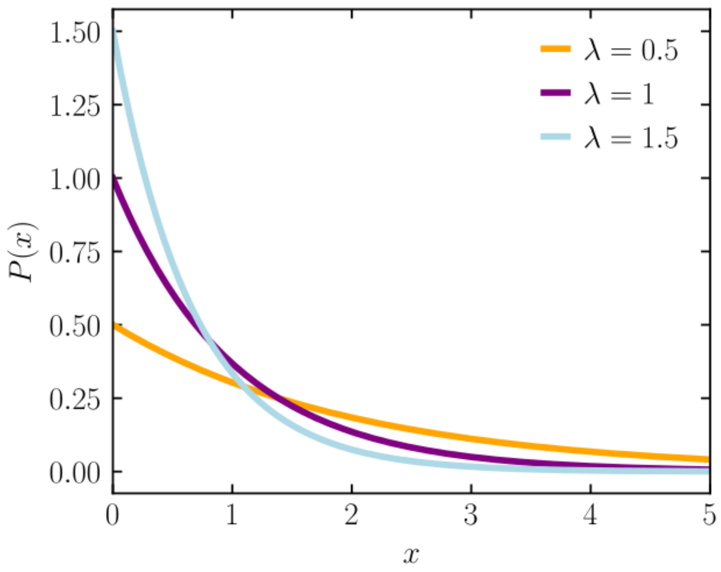
Exponential Distribution

- Constant failure rate (CFR), λ
- Assume that λ(t) = λ, t ≥ 0, and λ > 0.
- Thus, we have the following:

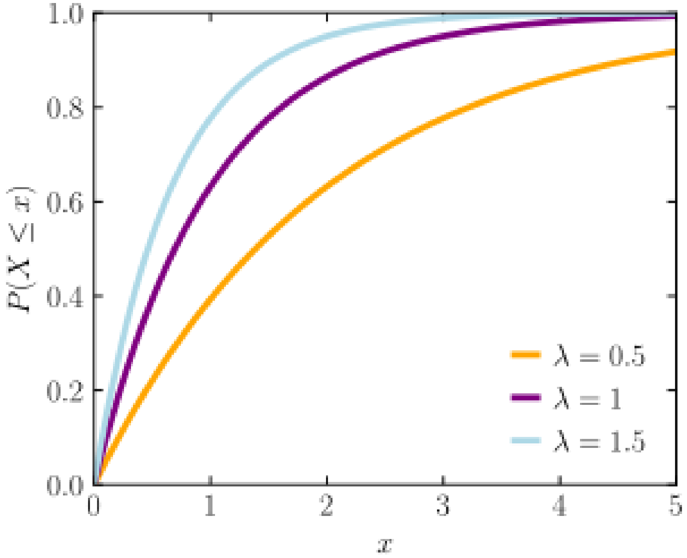
$$f(t) = \lambda e^{-\lambda t}$$
$$F(t) = 1 - e^{-\lambda t}$$
$$R(t) = e^{-\lambda t}, \qquad t \geq 0$$

- $MTTF = \frac{1}{\lambda}$ and $\sigma^2 = \frac{1}{\lambda^2}$

Probability density function



Cumulative distribution function

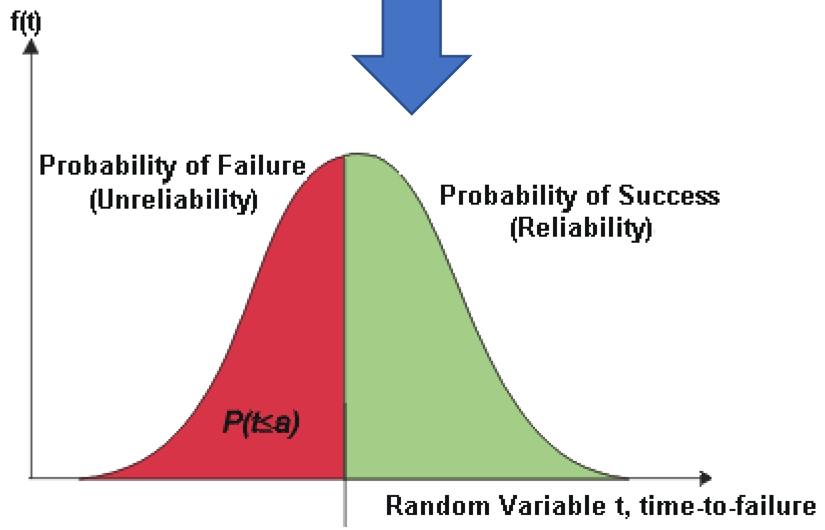


https://en.wikipedia.org/wiki/Exponential_distribution

Quick Note:
Two ways of looking at Reliability Function

- $R(t) = \underbrace{\int_t^{\infty} f(t')dt' = 1 - \int_0^t f(t')dt'}$ $= \underbrace{\exp[-\int_0^t \lambda dt']}$, $t \geq 0$

This is by definition of the PDF curve



This is by looking at hazard rate or instantaneous failure rate point if view

See next slide or “Hazard Rate” from Lecture 3

FYI: Derivation of the Reliability Function, R(t)

$R(t)$ = probability of success
N_S = number successful
N_F = number failed
N_O = total number = $N_S + N_F$ = constant
λ = failure rate

reliability is probability of success

$$\boxed{R(t) = \frac{N_S}{N_O}} = \frac{N_O - N_F}{N_O} = 1 - \frac{N_F}{N_O} \Rightarrow$$

$$\frac{dR}{dt} = \frac{d(1-\frac{N_F}{N_O})}{dt} = -\frac{1}{N_O}\frac{dN_F}{dt} \Rightarrow -N_O\frac{dR}{dt} = \frac{dN_F}{dt} \Rightarrow -\frac{N_O}{N_S}\frac{dR}{dt} = \frac{1}{N_S}\frac{dN_F}{dt} = \lambda \Rightarrow$$

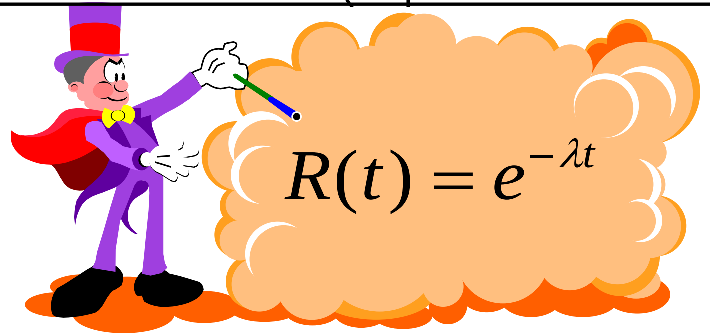
(up to here is derivation of λ)

$$-\frac{1}{R}\frac{dR}{dt} = \lambda \Rightarrow -\frac{dR}{R} = \lambda\ dt \Rightarrow -\int_1^R \frac{dR}{R} = -\ln R = \int_0^t \lambda\ dt \Rightarrow$$

General Case

$$R(t) = e^{-\int_0^t \lambda\ dt} \Rightarrow$$

Constant Failure Rate (Exponential Distribution) Case



Memoryless Property

- Exponential distribution has memoryless property.

- The memoryless property means that the time to failure of a component is not dependent on how long the component has been operating
  - Probability of failure is constant over time, independent of past failures
  - Probability of failure is the same at any point in time

- To see this, consider the following conditional probability:

$$R(t|T_0) = \frac{R(t + T_{0)}}{R(T_0)} = \frac{e^{-\lambda(t+T_0)}}{e^{-\lambda T_0}} = \frac{e^{-\lambda t} \cdot e^{-\lambda T_0}}{e^{-\lambda T_0}} = e^{-\lambda t} = R(t)$$

Exponential Distribution

- Useful for maintenance engineering
- To predict the amount of waiting time until the next event (i.e., success, failure, arrival, etc.)
  - The amount of time until the customer finishes browsing and actually purchases something in your store (success).
  - The amount of time until the hardware fails (failure).
  - The amount of time you need to wait until the bus arrives (arrival).

https://towardsdatascience.com/what-is-exponential-distribution-7bdd08590e2a

Failure Modes

- Complex systems will fail through various means resulting from different physical characteristics.
- These failures can be separated according to the mechanisms or components causing the failures. These categories of failures are referred to as failure modes.
- If we define $R_i(t)$ as the reliability of the ith failure mode, i.e., the probability that the i^{th} failure mode does not occur before time t. Then assuming independence among the failure modes, the system reliability denoted by $R(t)$ can be expressed as follows:

$$R(t) = \prod_{i=1}^{n} R_i(t)$$

- That is, none of the n failure modes occurs before time t.

System Hazard Rate for Exponential Reliability Function

- Let $\lambda_i(t)$ denote the failure rate function for the i^{th} failure mode. Then we have $R_i(t) = \exp[-\int_0^t \lambda_i(t')dt']$, thus the system reliability is

$$R(t) = \prod_{i=1}^{n} \exp\left[-\int_0^t \lambda_i(t')dt'\right]$$

$$= \exp\left[-\int_0^t \sum_{i=1}^{n} \lambda_i(t')dt'\right]$$

$$= \exp\left[-\int_0^t \lambda(t')dt'\right] = \exp[-\lambda t]$$

where $\lambda(t') = \sum_{i=1}^{n} \lambda_i(t')$

- $MTTF = \frac{1}{\lambda} = \frac{1}{\sum_{i=1}^{n} \lambda_i} = \frac{1}{\sum_{i=1}^{n} 1/MTTF_i}$; $MTTF_i = \frac{1}{\lambda_i}$

Example

- An aircraft engine consists of three modules having constant failure rates of $\lambda_1 = 0.002$, $\lambda_2 = 0.015$, and $\lambda_3 = 0.0025$ failures per operating hour. Evaluate the reliability function for the engine and calculate the corresponding MTTF.

Weibull Distribution

- Very flexible distribution
- Applicable to all three phases of life cycle
- Exponential is a special case of Weibull
- Shape parameter: β
- Characteristic life or scale parameter
    - α
    - η

Weibull Distribution

β: shape parameter

$$f(t) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} exp^{\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]}$$

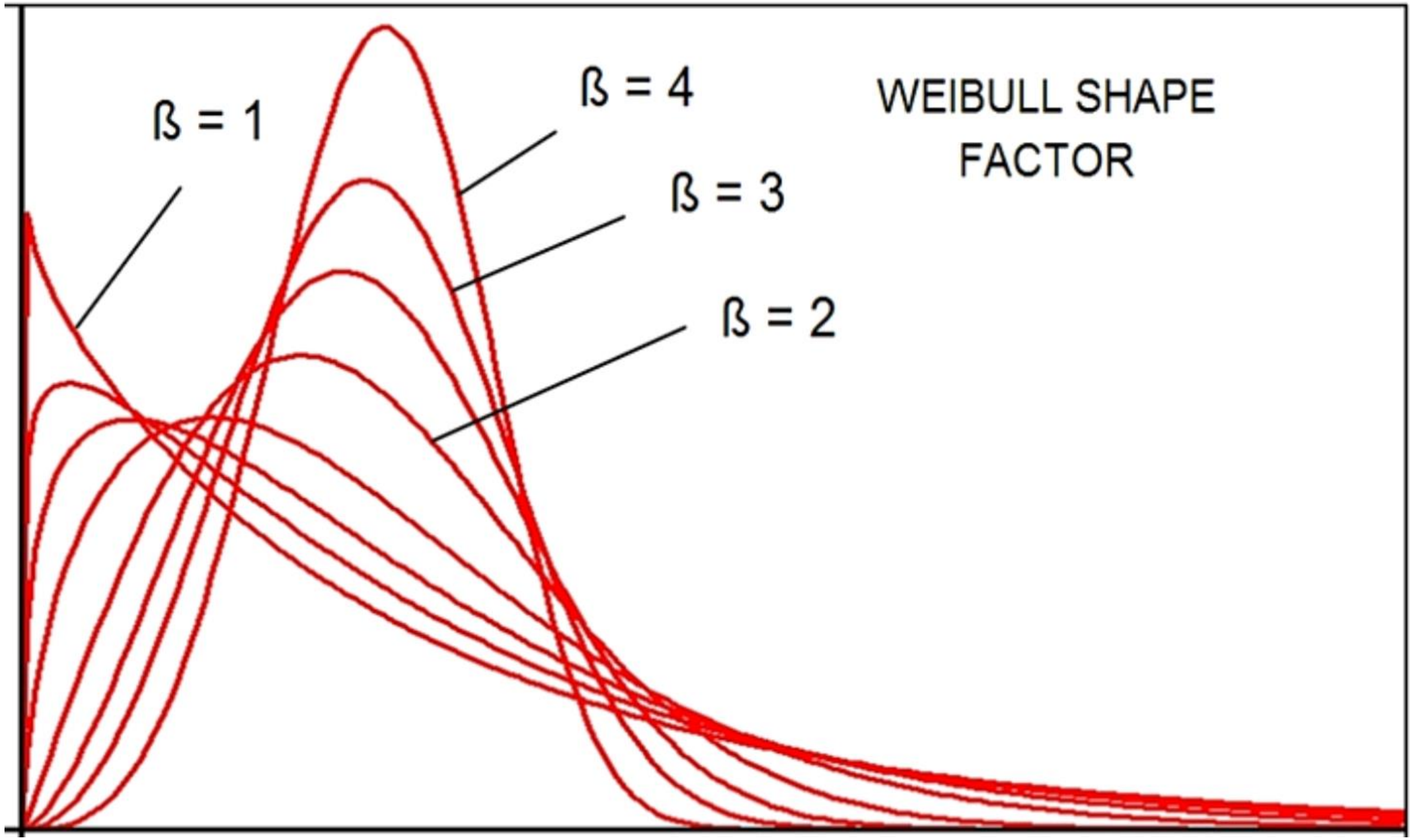
α or η: characteristic life (t_{63})

Weibull PDF and CDF

$$f(t) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} exp^{\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]}$$

$$F(t) = \int_0^t f(t)dt = 1 - exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]$$

Effect of Shape Parameter



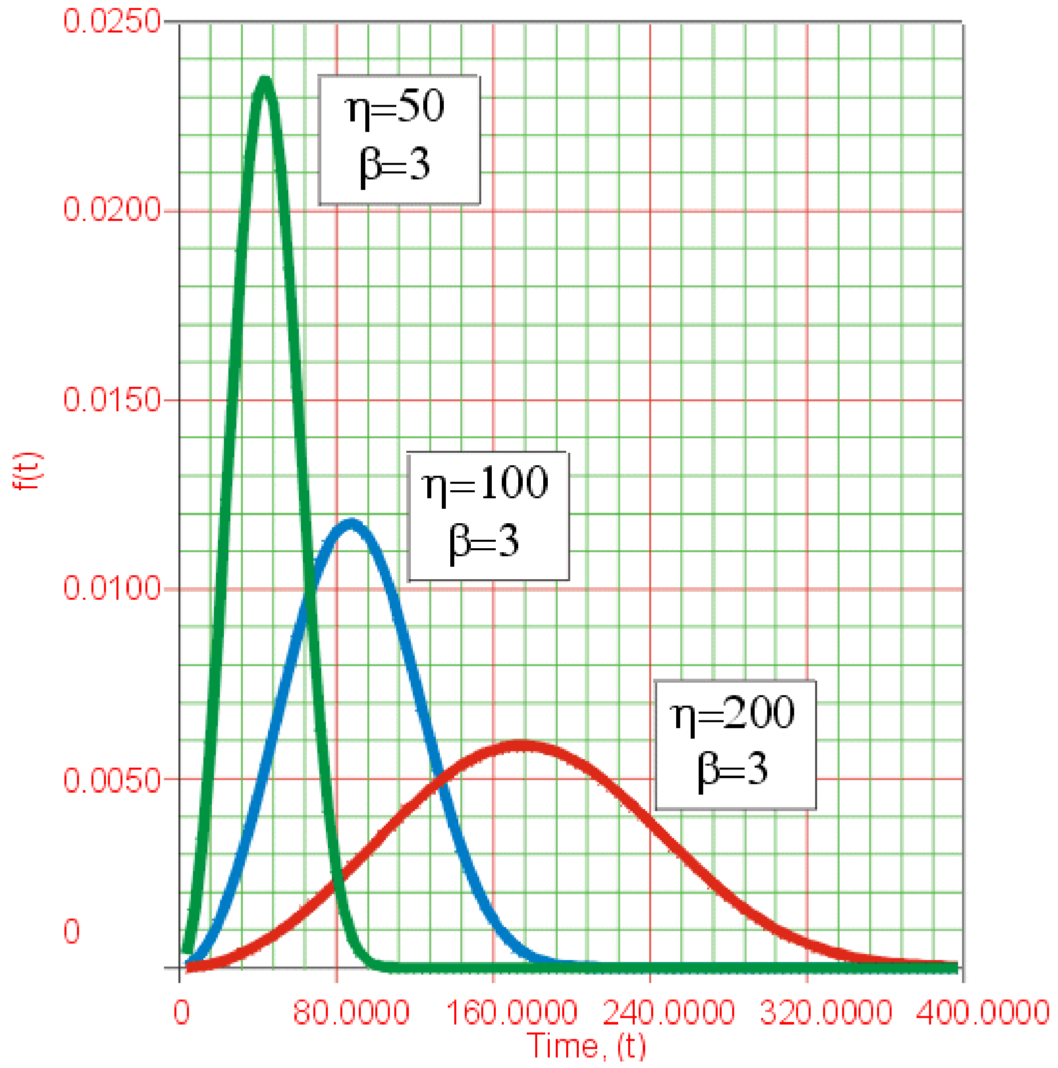
Shape

- β is referred to as the shape parameter
  - It affects the shape of the distribution in the sense that for $\beta < 1$ the distribution looks like an exponential
  - For $\beta = 1$ the distribution is exponential
  - For $\beta > 3$, the distribution is close to symmetrical
  - For $1 < \beta < 3$ it is skewed
    - $\beta = 2.0$: identical to the Rayleigh distribution
    - $\beta = 2.5$: approximates the lognormal distribution (see later slides)

Weibull Distribution PDF

$$f(t) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} exp^{\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]}$$

Effect of Characteristic Life



Scale Parameter

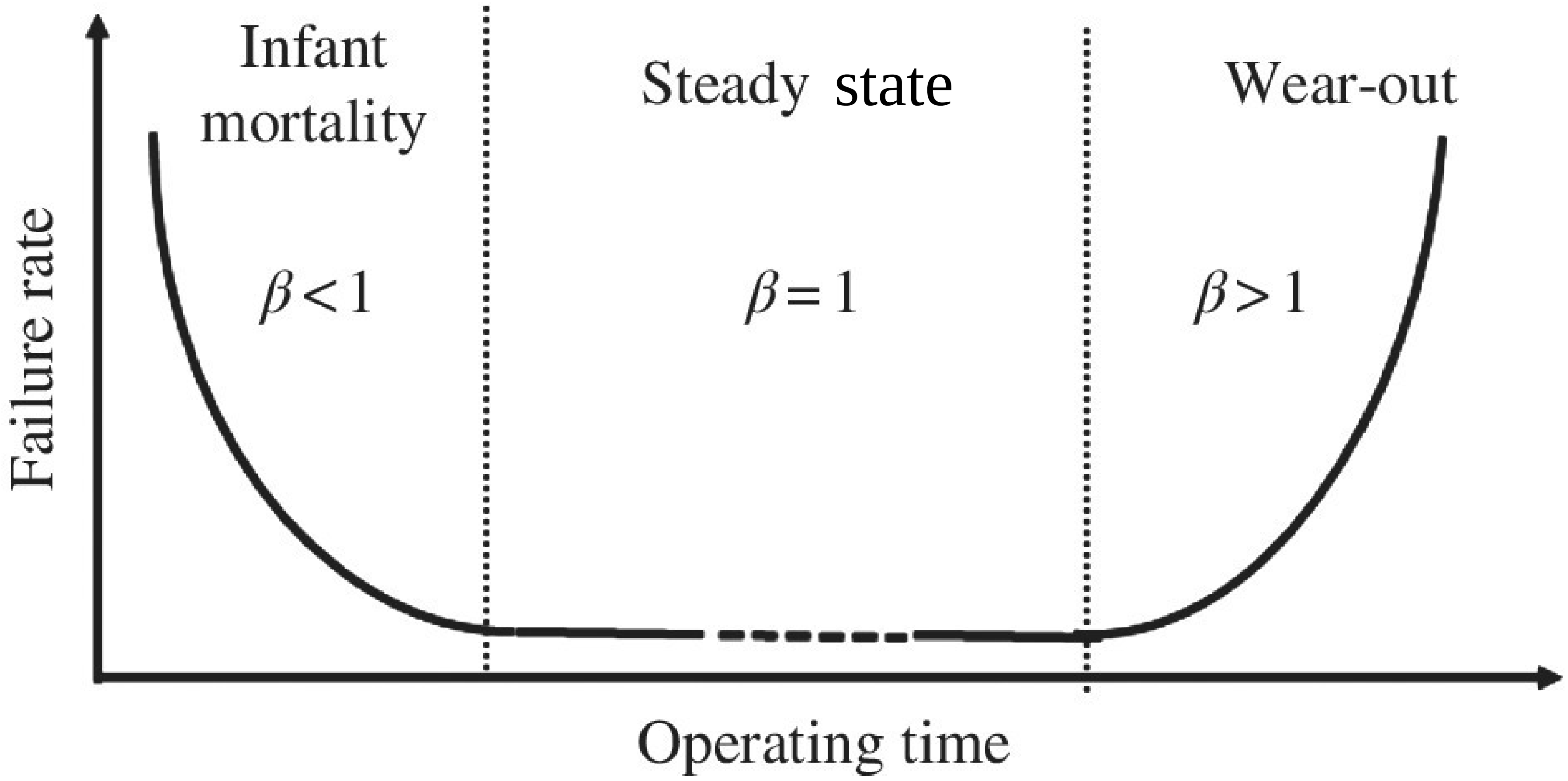
- α or η is the scale parameter
  - Influences both the mean and the spread of the distribution
  - Defines the age by which 63.2% of the units will fail, called characteristic lifetime

Weibull Distribution PDF

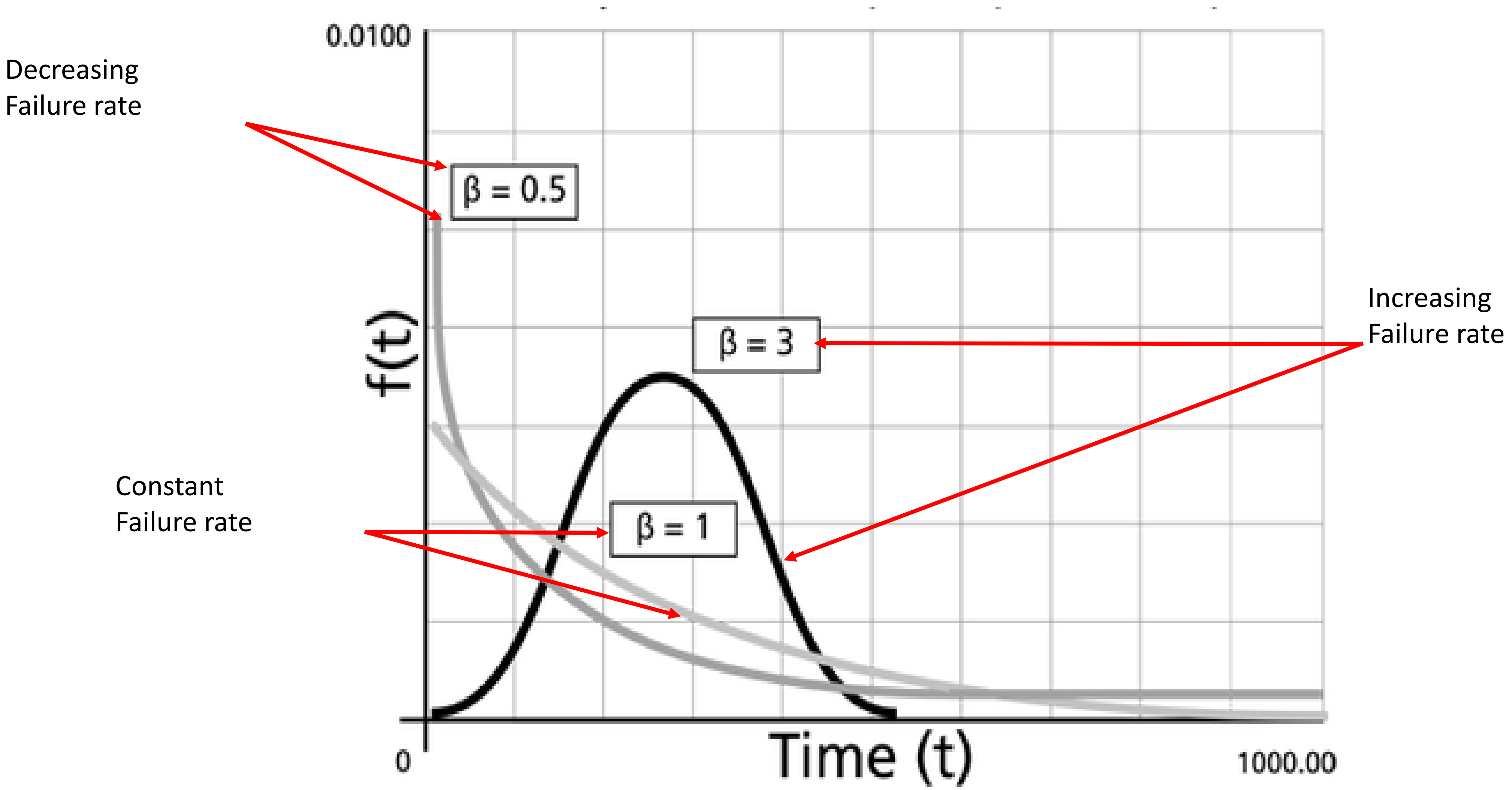
$$f(t) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} exp^{\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]}$$

CDF at characteristic lifetime: $F(\eta) = 1 - e^{\left(-\left(\frac{\eta}{\eta}\right)^{\beta}\right)} = 1 - e^{\left(-(1)^{\beta}\right)} = 1 - \frac{1}{e} \cong 0.632$

Bathtub Curve & Weibull Distribution



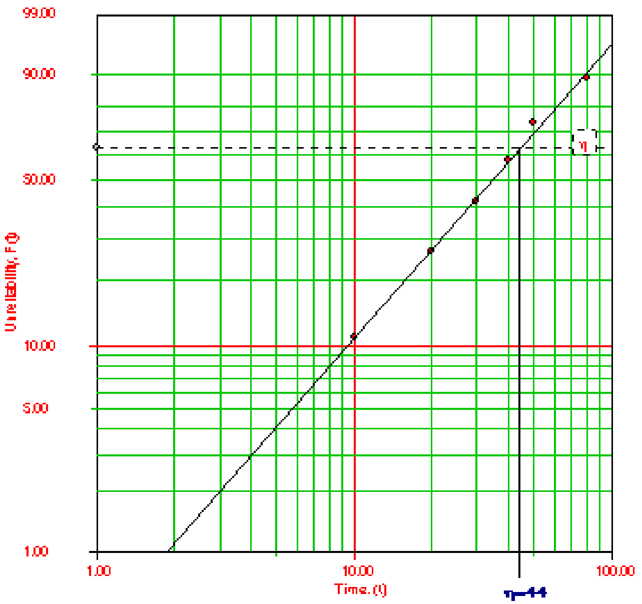
https://www.oreilly.com/library/view/thermodynamic-degradation-science/9781119276227/c09.xhtml
http://www.applied-statistics.org/Glossary/BathTubCurve.html


Decreasing
Failure rate
Constant
Failure rate
Increasing
Failure rate
0.0100
f(t)
β = 0.5
β = 3
β = 1
0
Time (t)
1000.00

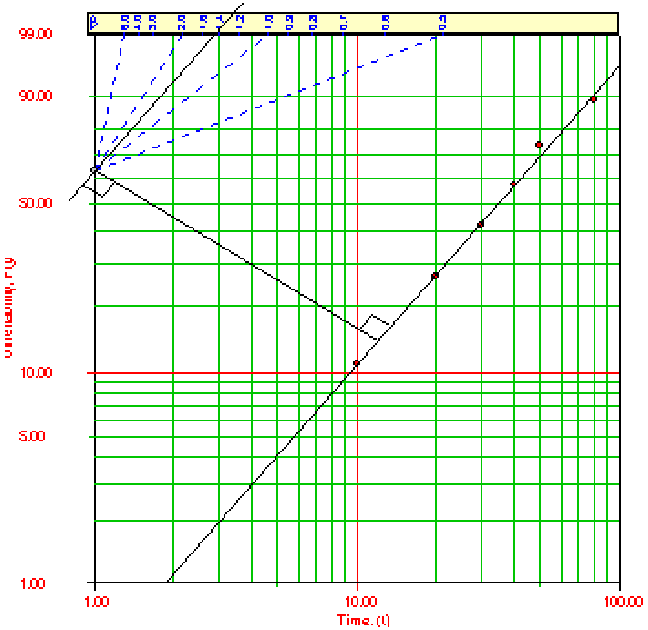
Weibull Plotting

Failure Time (hours)	Unreliability Estimate
10	10.9%
20	26.6%
30	42.2%
40	57.8%
50	73.4%
80	89.1%

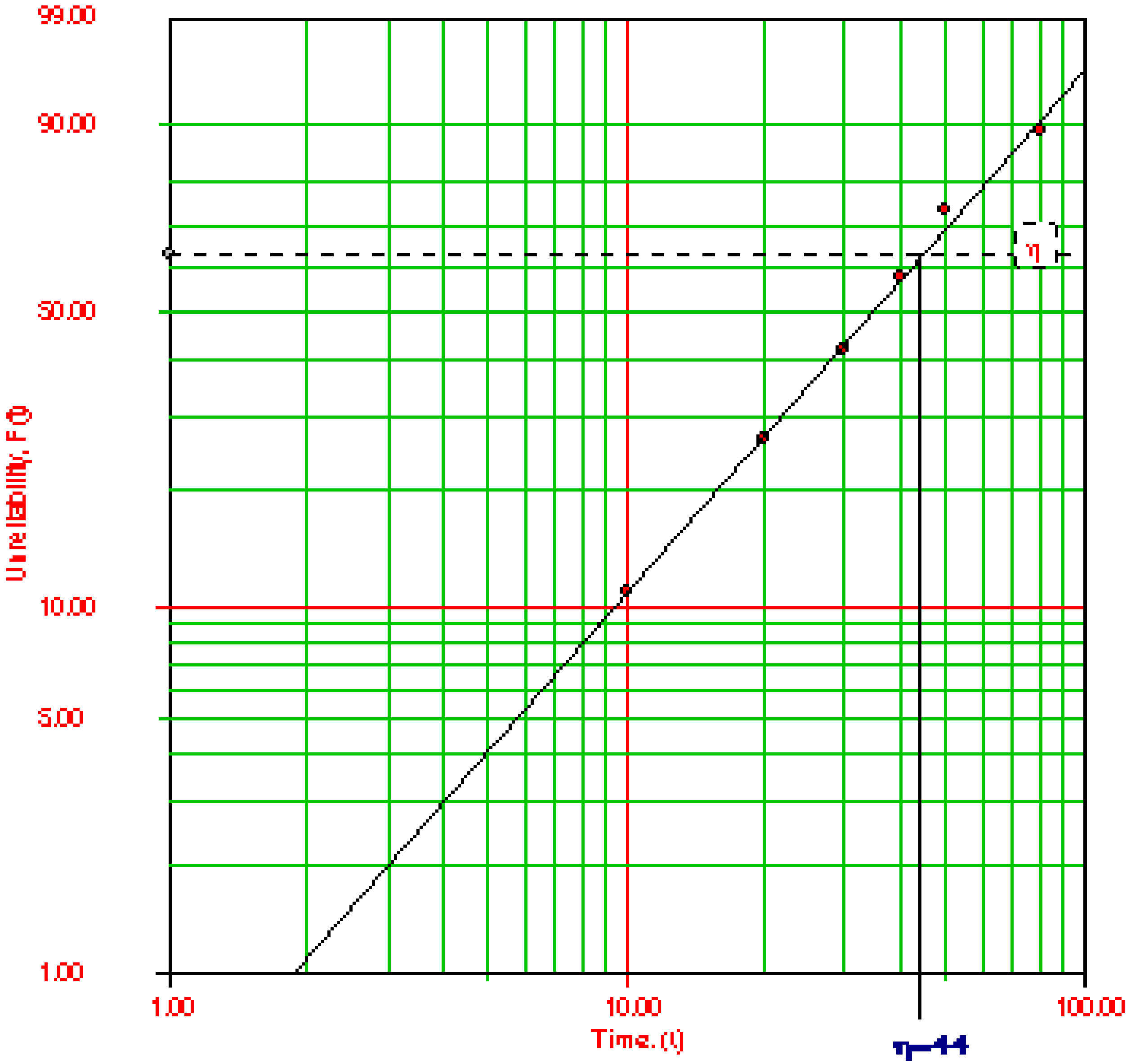
α or η is when 63% failed



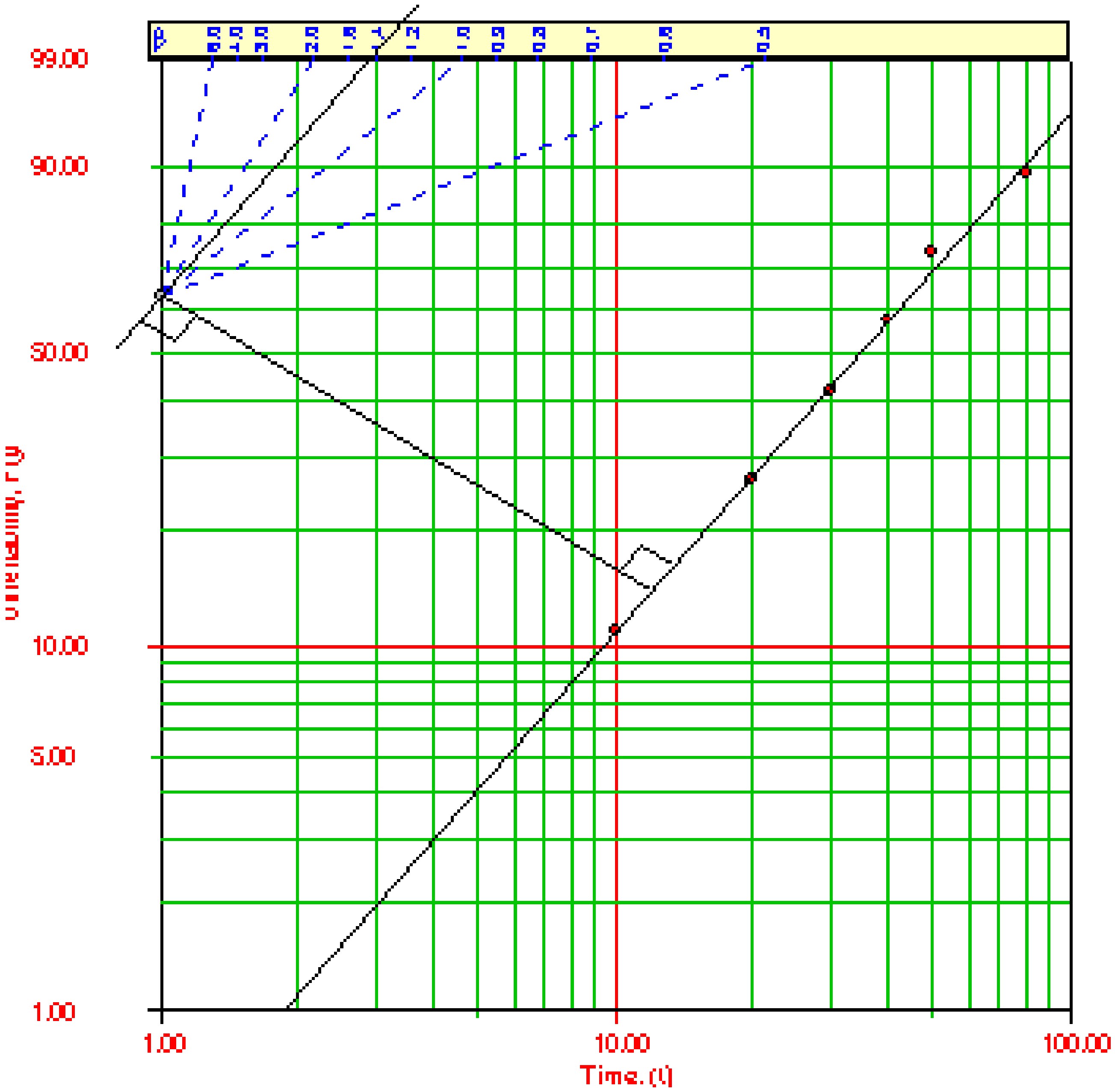
β: is the slope



α or η is when 63% failed



β: is the slope



Example

- An item is randomly drawn from a two - parameter Weibull population having a shape parameter $\beta = 1.5$ and a scale parameter $\eta = 100.0$ hours. What is the probability that the item fails before achieving a life of $x = 25$ hours?

Why the Lognormal Distribution?

- Normal distribution
  - Most common
  - Represents addition of random variables
  - Application: dimension of a manufactured product
  - Used by quality engineers to reduce variability
- Lognormal distribution
  - Significant applications in science, finance, etc.
  - Represents multiplication of random variables

Background of Lognormal

So lets say that Y is the product of lots of independent, positive random variables ...

$$Y = X_1 \times X_2 \times X_3 \times \ldots \times X_i \times \ldots$$

independent random variables

Multiplication to Addition

$$Y = X_1 \times X_2 \times X_3 \times \ldots \times X_i \times \ldots$$

how do we model Y?

we can then apply the 'natural logarithm' to restate the above equation ...

$$\ln Y = \ln X_1 + \ln X_2 + \ln X_3 + \ldots + \ln X_i + \ldots$$

Addition of Random Variables

each one of these logarithms
is a random variable

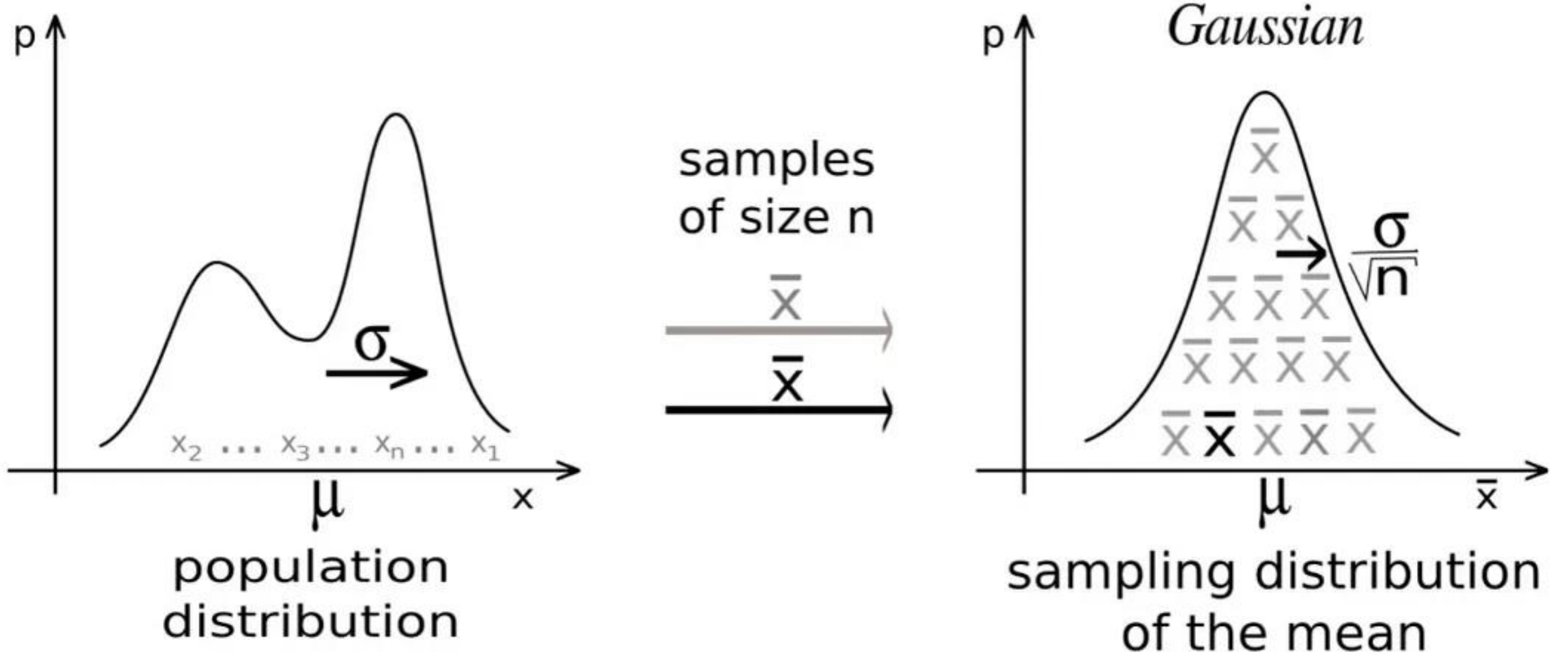
$$\ln Y = \ln X_1 + \ln X_2 + \ln X_3 + \ldots + \ln X_i + \ldots$$

which means that $\ln Y$ is a ***normally distributed*** *random variable*
... due to the ***Central Limit Theorem***

Note: Central Limit Theorem

No matter the shape of the population of a distribution

If sample batches of data from that distribution (with replacement) and take the mean of each batch, the mean values from all those batches will be normally distributed.

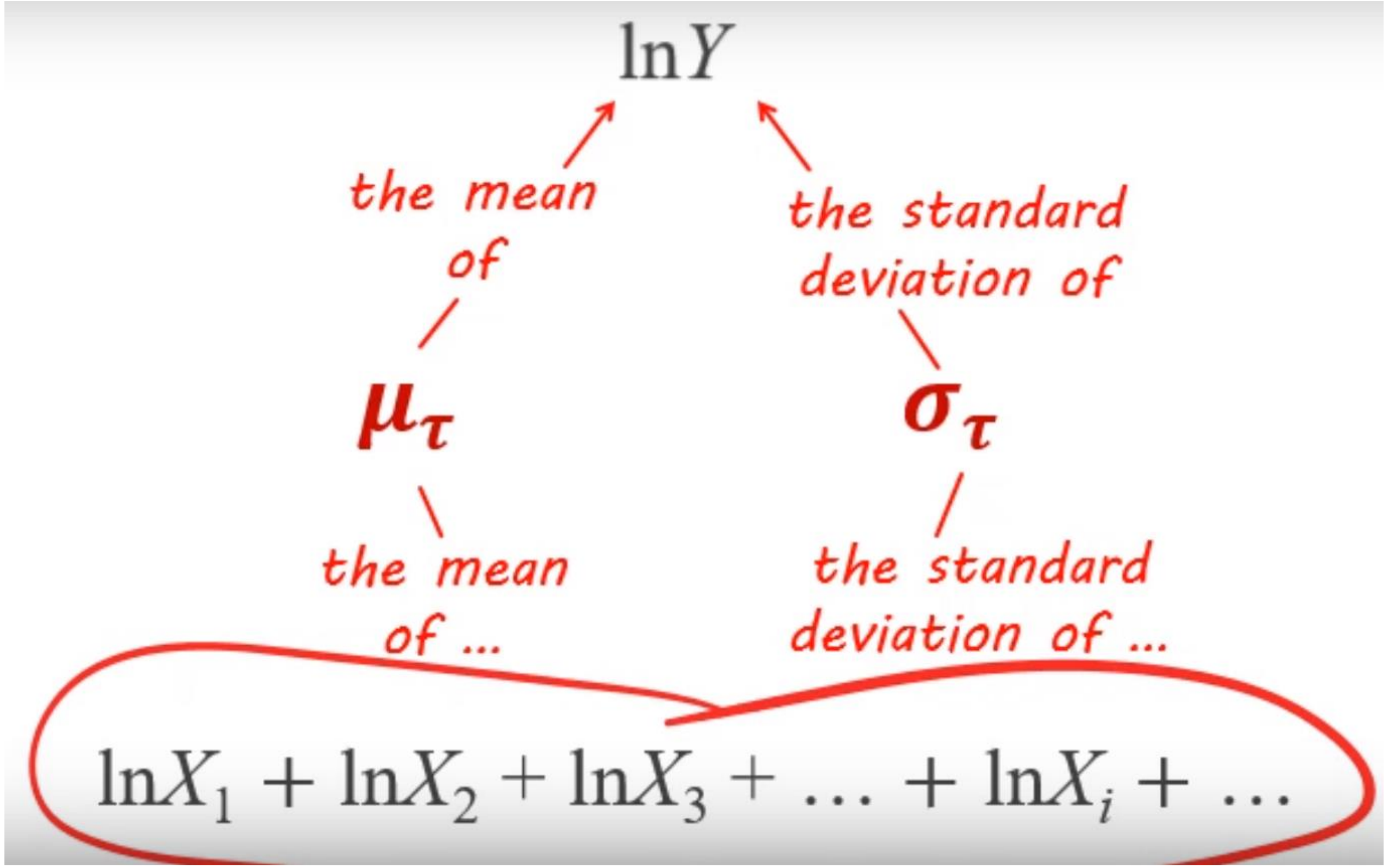


https://ai.plainenglish.io/central-limit-theorem-a-complete-guide-for-beginners-c17a6bd7417
https://medium.com/@seema.singh/central-limit-theorem-simplified-46ddefeb13f3

Note: Central Limit Theorem Demo Explanation

Central Limit Theorem Premises	**Candy Jar Demo**
• Independent	• Each trial (or individual estimate) does not know about other trials or estimates
• Identically distributed	• Looking at the same underlying problem
• Finite variance	• Estimates are not so unreasonable that a single estimate will throw off the average

Attributes of Ln(Y)

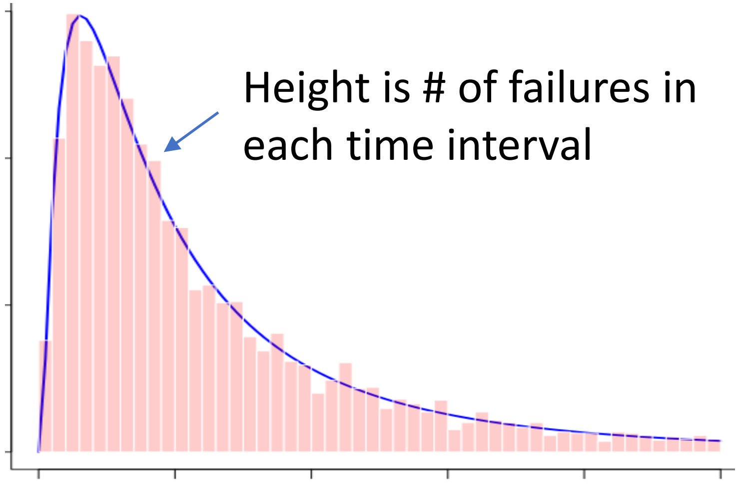


τ: *referring to the X_i or normal domain* (see following slides on "Conversion")

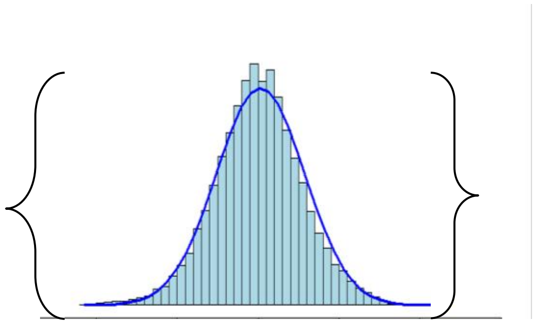
Visual Interpretation from Histogram Perspective

$$Time\ to\ failure = A_0 e^x$$

Bucketizing failure times in a test



$= A_0 e$

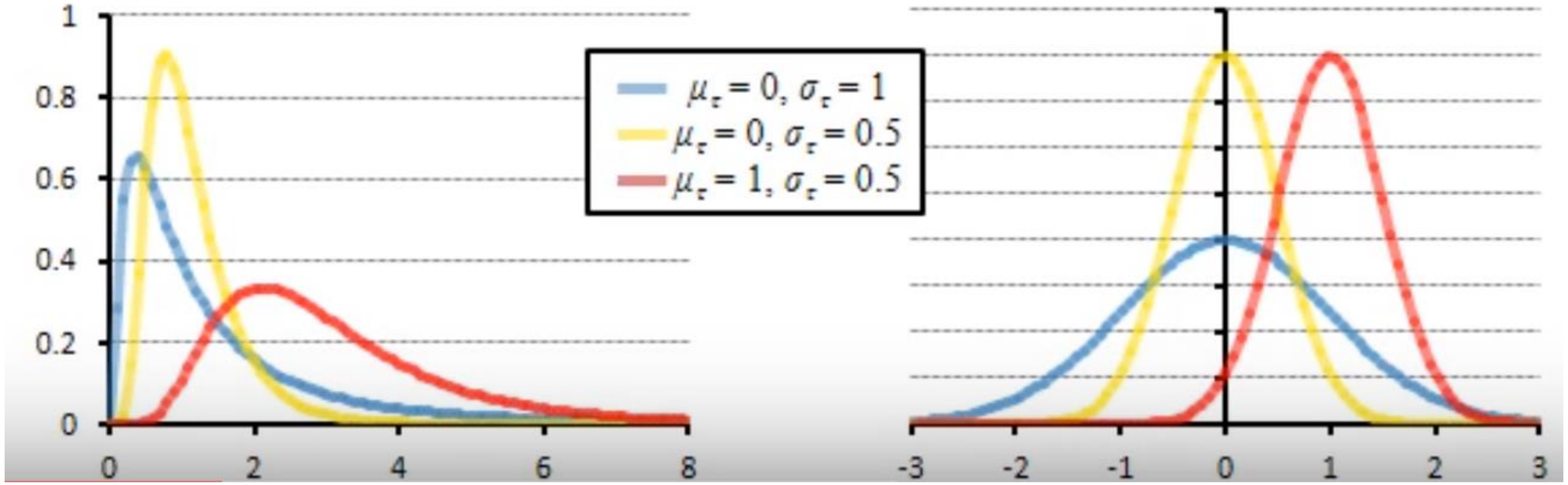


Reality

Underlying factor x is normally distributed that influences overall distribution

Conversion

Reality



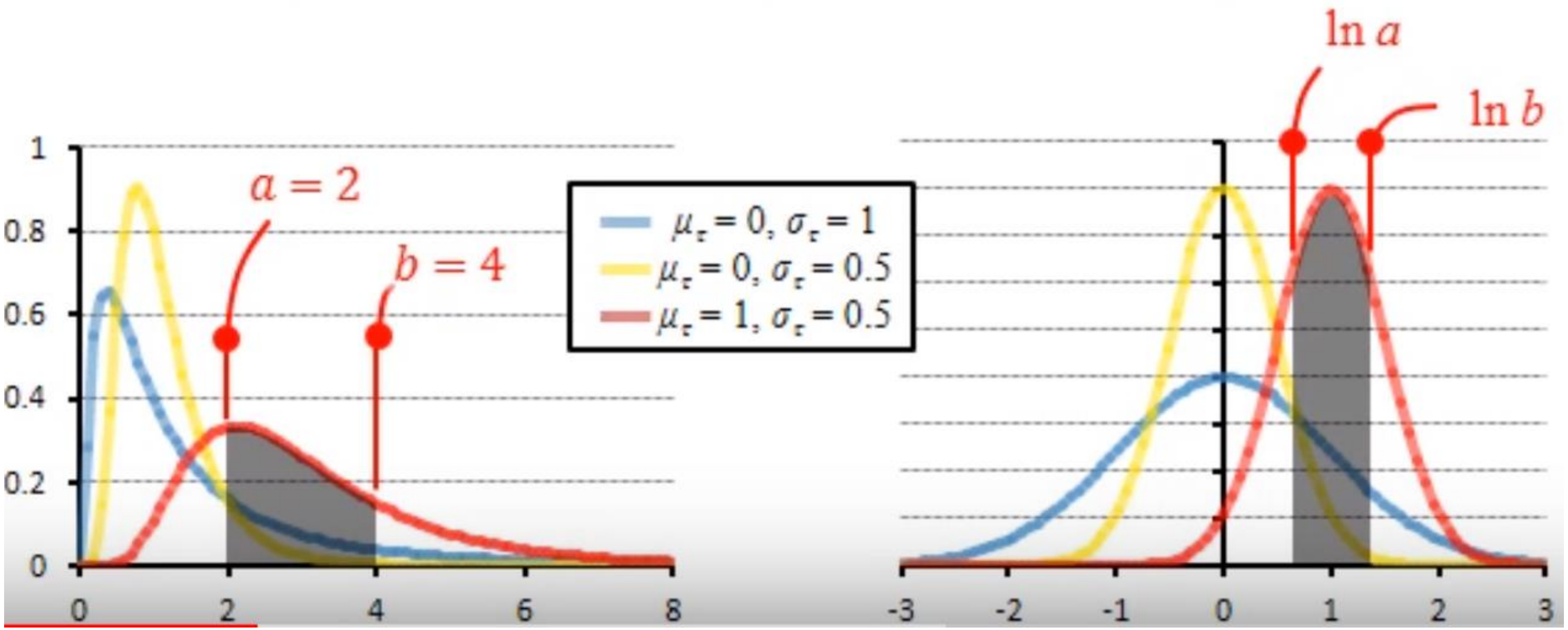
t

τ

$$\tau = \ln(t)$$

Note: we are now using t instead of Y

Equivalent Areas



Comparison to Normal

so the **probability density function** for the **normal distribution** ...

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

... becomes ...

$$f(y) = \frac{1}{y\sigma_\tau\sqrt{2\pi}} e^{-\frac{(\ln y-\mu_\tau)^2}{2\sigma_\tau^2}}$$

... for the **lognormal distribution**

1/y pre-factor is needed to ensure that the integral of f(y) x dy will continue to represent the probability of failure. Due to the fact that dln(y) = (1/y) dy

Lognormal Probability Density Function

$$f(t) = \frac{1}{t\sigma_\tau\sqrt{2\pi}} e^{-\frac{(\ln t - \mu_\tau)^2}{2\sigma_\tau^2}}$$

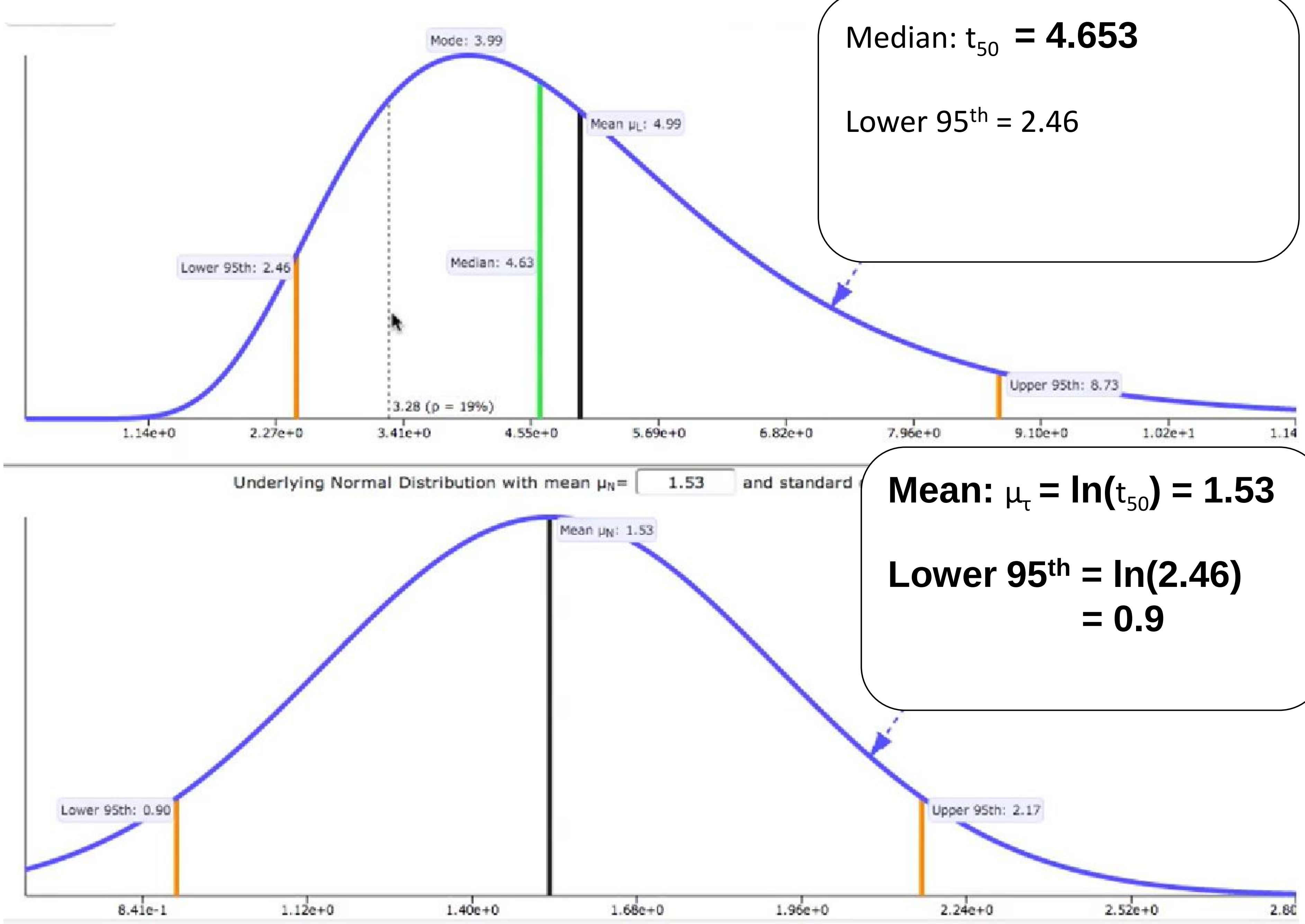
μ_τ also written as $\ln(t_{50})$

The mean is $\mu = e^{\mu_\tau + \frac{\sigma_\tau^2}{2}}$

The median is $t_{50} = e^{\mu_\tau}$

The mode is $\hat{t} = e^{\mu_\tau - \sigma_\tau^2}$

τ: *referring to the* X_i *or normal domain*



Median: t_{50} **= 4.653**

Lower 95th = 2.46

Mean: μ_τ = ln(t_{50}) = 1.53

Lower 95th = ln(2.46)
= 0.9

Lognormal PDF and CDF

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} exp\left\{-\left[\frac{\ln(t) - \ln(t_{50})}{\sigma\sqrt{2}}\right]^2\right\}$$

$$F(t) = \int_0^t f(t)dt$$

The integral needs to be numerically evaluated

Lognormal CDF Evaluation

$$F(t) = \int_0^t f(t)dt$$

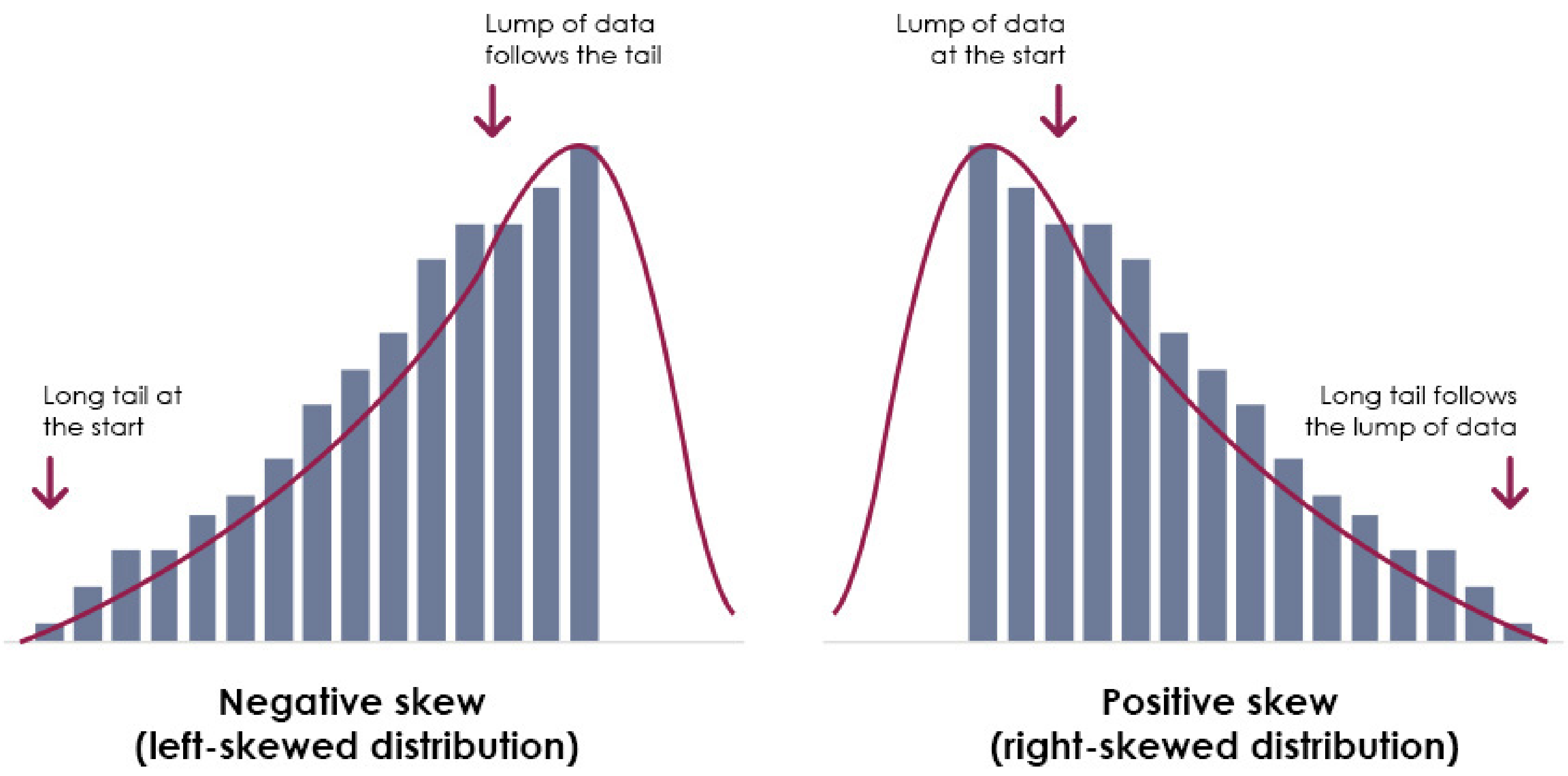
Numerical evaluation of integral by error function complement

$$F(t) = \frac{1}{2} erfc\left(\frac{\ln(t_{50}) - \ln(t)}{\sigma\sqrt{2}}\right) \qquad \text{for t} \leq t_{50}$$

$$F(t) = 1 - \frac{1}{2} erfc\left(\frac{\ln(t) - \ln(t_{50})}{\sigma\sqrt{2}}\right) \qquad \text{for t} \geq t_{50}$$

Skewed Data

- Skewness is a way to describe the symmetry of a distribution

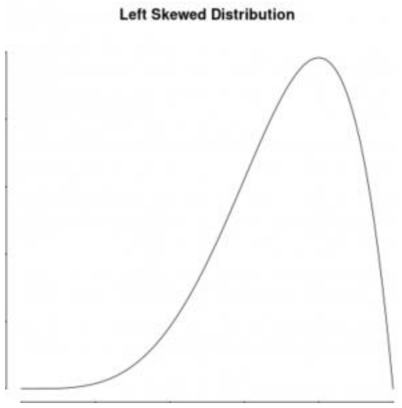


https://www.cambridgemaths.org/blogs/skewed-usage-skewed-distribution/

Skewed Data

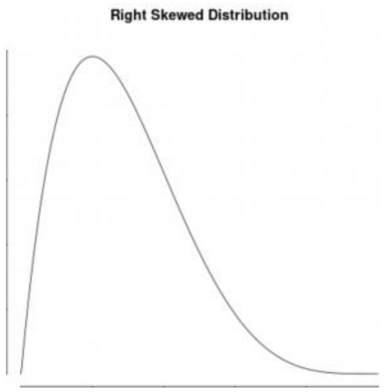
- Skewness is a way to describe the symmetry of a distribution

Left Skewed



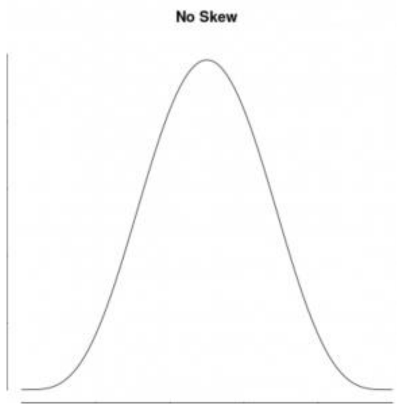
- "tail" on the left side of the distribution
- negatively-skewed

Right Skewed



- "tail" on the right side of the distribution
- positively-skewed

No Skew



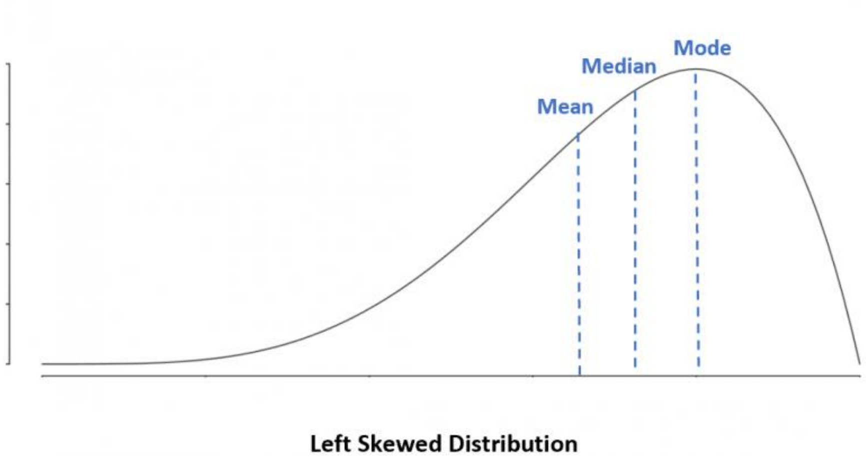
- symmetrical on both sides

https://www.statology.org/left-skewed-vs-right-skewed/

Skewed Data

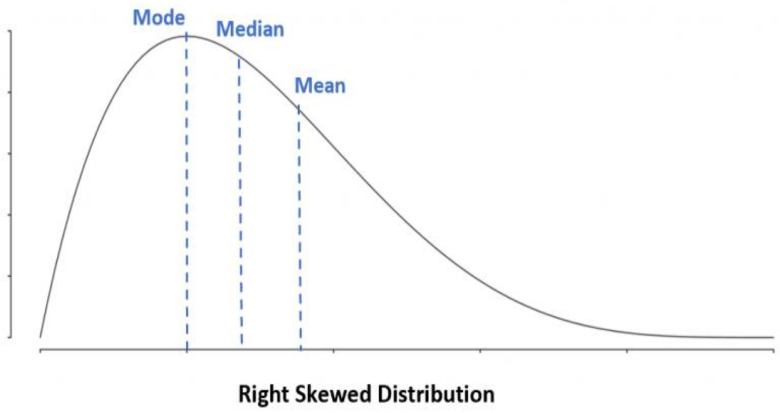
- Skewness is a way to describe the symmetry of a distribution

<u>Left Skewed</u>



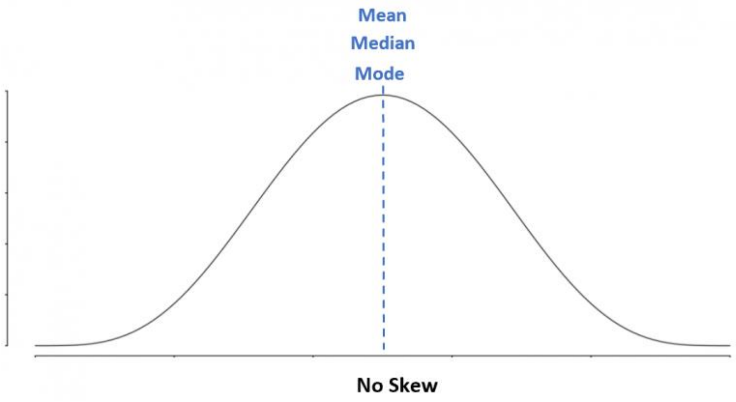
- Mean < Median < Mode
- Most of the data points falls on the **<u>right side</u>** of the mean

<u>Right Skewed</u>



- Mode < Median < Mean
- Most of the data points falls on the **<u>left side</u>** of the mean

<u>No Skew</u>

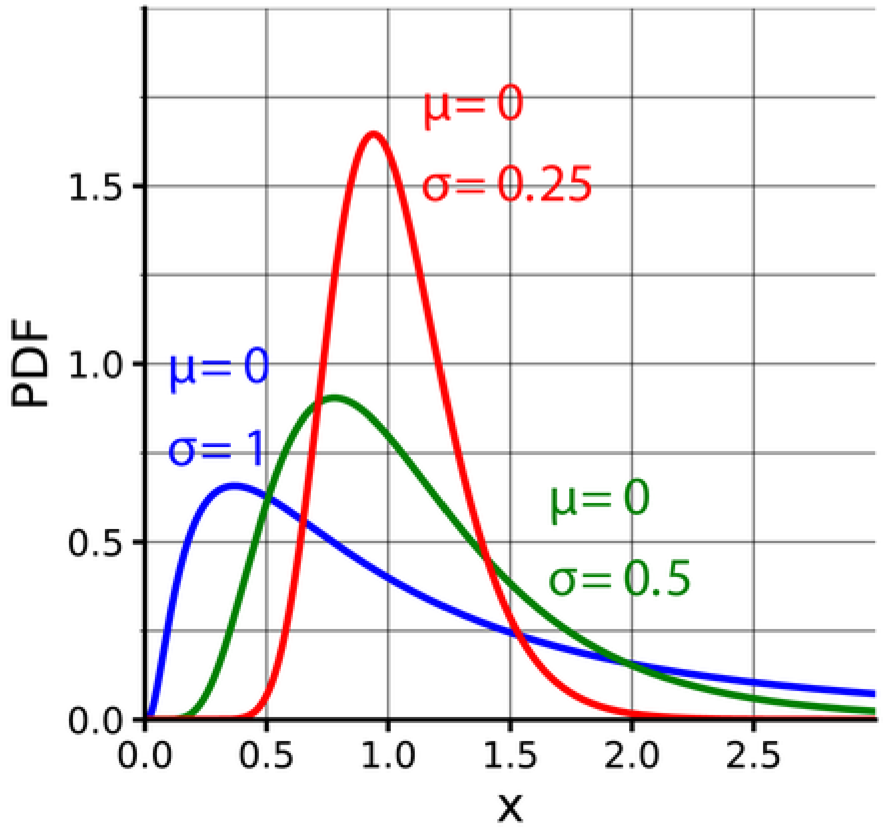


- Mean = Median = Mode

https://www.statology.org/left-skewed-vs-right-skewed/
https://www.quora.com/What-does-it-mean-when-data-has-a-skewed-distribution

Lognormal: Right-skewed data

- Values are positively skewed
- Better at predicting longer lifetimes
- Gradual degradation over time



https://en.wikipedia.org/wiki/Log-normal_distribution

Normal vs. Lognormal

Distribution	$Y = \ln(X)$	$X = e^Y$
Notation	$N(\mu_y, \sigma_y^2)$	$lognormal\ (\mu_x, \sigma_x^2)$
PDF	$\frac{1}{\sigma_y\sqrt{2\pi}} exp\left\{-\left[\frac{x-\mu_y}{\sigma_y\sqrt{2}}\right]^2\right\}$	$\frac{1}{\sigma_y x\sqrt{2\pi}} exp\left\{-\left[\frac{\ln x-\mu_y}{\sigma_y\sqrt{2}}\right]^2\right\}$
Mean	μ_y	$\mu_x = \exp(\mu_y + \frac{1}{2}\sigma_y^2)$
Variance	σ_y^2	$\sigma_x^2 = \mu_x^2\left(e^{\sigma_y^2} - 1\right)$ $= \exp(2\mu_y + \sigma_y^2)\left(e^{\sigma_y^2} - 1\right)$

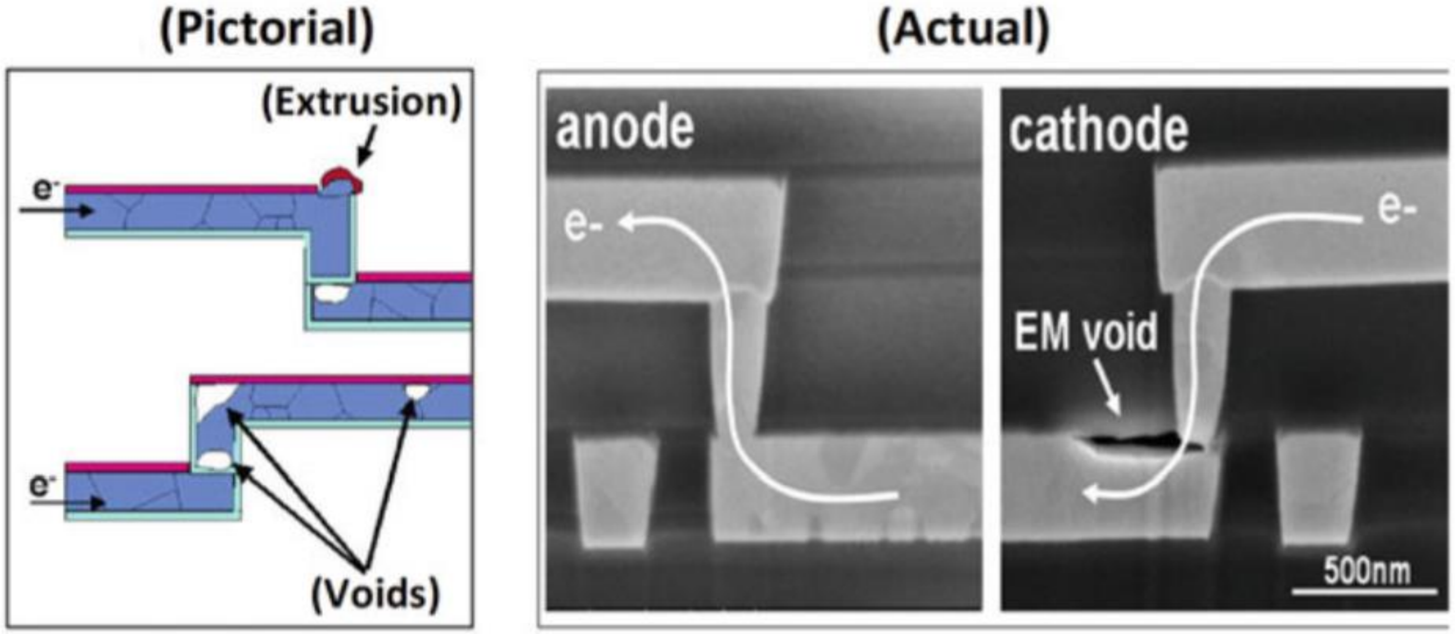
Thomopoulos, Nick T. "Statistical distributions." Applications and Parameter Estimates. Cham, Switzerland: Springer International Publishing (2017).

Example

- In a fracture failure experiment, one of the variables, x, is detected as lognormal with LN(2.5, 1.1^2). Compute the corresponding mean and variance of the x.

Lognormal Applications

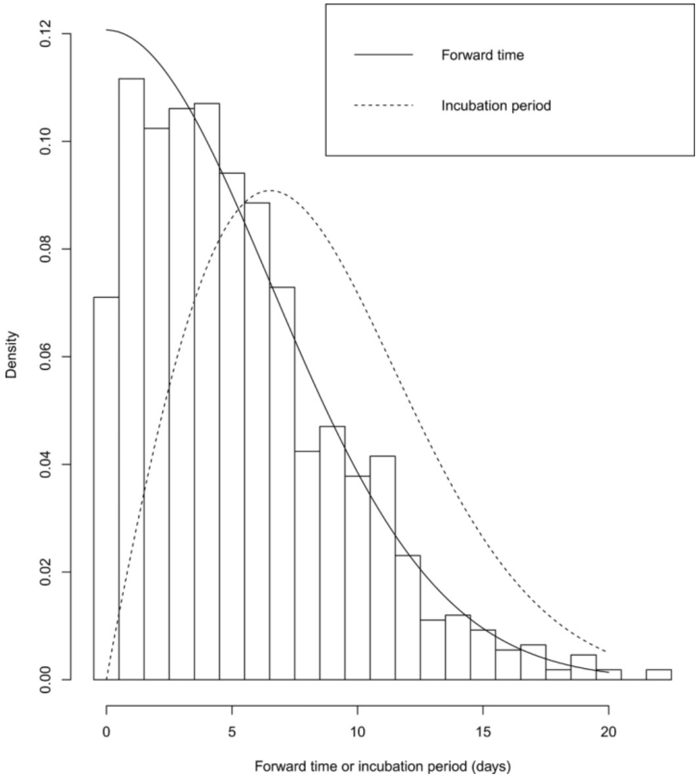
- Electromigration



McPherson, Joe W. Reliability physics and engineering: time-to-failure modeling. Springer, 2018.

Lognormal Applications

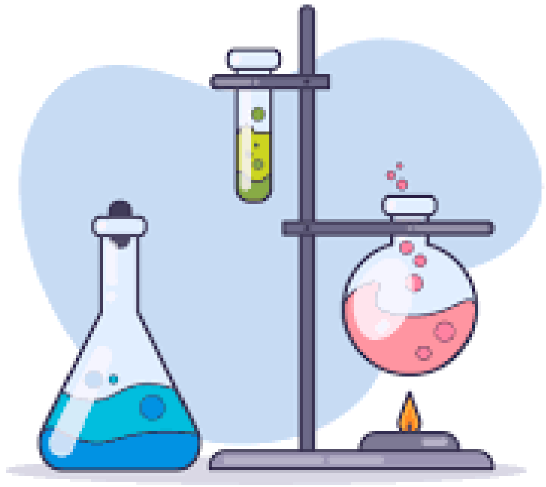
Medical Science



Modeling onset of symptoms from COVID

Qin, Jing, et al. "Estimation of incubation period distribution of COVID-19 using disease onset forward time: a novel cross-sectional and forward follow-up study." Science advances 6.33 (2020)

Chemical Reactions



Modeling chemical kinetics

Schnoerr, David, Guido Sanguinetti, and Ramon Grima. "Comparison of different moment-closure approximations for stochastic chemical kinetics." The Journal of Chemical Physics 143.18 (2015)

Corrosion & Fatigue Cracking

Pic from https://www.thoughtco.com/how-rust-works-608461

Fu, Guoyang, et al. "Prediction of fracture failure of steel pipes with sharp corrosion pits using time-dependent reliability method with lognormal process." Journal of Pressure Vessel Technology 141.3 (2019).

Summary

- Times of failures should lead your selection of appropriate probability distribution

Model	Description	Characteristic	Advantages	Disadvantages
Exponential	The failure rate is constant over time	Memoryless (the age of the item has no effect on its future failure rate)	Simple and easy to understand (only one parameter to estimate)	May not reflect real-life scenarios where the failure rate changes over time
Log-normal	The failure rate changes over time	Model is appropriate when underlying process has a large number of causes of failure and the failure times are spread out over a large range	Can reflect real-life scenarios where the failure rate changes over time	Requires more parameters to estimate
Weibull	The shape parameter determines the shape of the failure rate curve, which can be constant, increasing or decreasing over time	Flexibility in modeling	Can model a wide range of failure rate patterns	Estimation process can be more complex

Summary

	Exponential	Log-normal	Weibull
PDF	$f(t) = \lambda e^{-\lambda t}$	$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} exp\left\{-\left[\frac{\ln(t) - \ln(t_{50})}{\sigma\sqrt{2}}\right]^2\right\}$	$f(t) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} exp^{\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]}$
CDF	$F(t) = 1 - e^{-\lambda t}$	$F(t) = \frac{1}{2} erfc\left(\frac{\ln(t_{50}) - \ln(t)}{\sigma\sqrt{2}}\right)$	$F(t) = 1 - exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]$
Parameter	λ	$\mu = \ln(t_{50}), \sigma$ (*note: these refer to the parameters in normal form, i.e. $\mu = \mu_y$, $\sigma = \sigma_y$ on slide "Normal vs. Lognormal")	$\beta = shape\ parameter$ (affects shape of distribution) $\eta = scale\ parameter$ $or\ characteristic\ lifetime$ (affects spread of distribution)