# Statistical Estimation of Failure Distributions

Prof. Nagi Gebraeel
Industrial and Systems Engineering
Georgia Tech



## Statistical Analysis and Estimation

- Identifying Failure Distributions
  - Least-Squares Estimation
  - Maximum Likelihood Estimation
- Interactive Examples
- Censoring
  - Types of Censored Data
  - Estimation using Censored Data
- Empirical Methods

## Identifying Failure Distributions

- One of our key objectives is to outline procedures for identifying and specifying parametric distributions that best describe the probability characteristics of failure time using a random sample of failure data.
- This is accomplished by performing a statistical **hypothesis test** to evaluate whether the observed sample of data comes from a pre-specified distribution.
  - These tests are typically known as **goodness-of-fit tests**.

## Identifying Failure Distributions

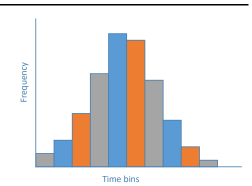
- Our goal is to **fit** a theoretical distribution to a random sample of failure data.
- Why is this useful?
  - Properties of the theoretical distribution can be used to provide strong results and reliability measures, e.g., mean, variance,
  - Theoretical distributions provide the means to easily evaluate (conditional) reliability at any future point in time.
  - Provides the foundation for deriving theoretical/analytical reliability models that can be used to perform complex analyses

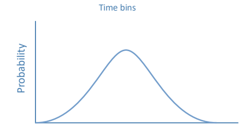
## Identifying Failure Distributions

- Fitting a distribution to a sample of failure data may be viewed as a 3-step process.
  1. Identify candidate distribution (hypothesis).
  2. Estimate its parameters.
  3. Perform a goodness-of-fit test.
- This can be accomplished by one or more of the following
  - Preliminary statistical analysis.
  - Leverage understanding of the failure process.
  - Leverage knowledge of the theoretical distribution.

## Identifying Failure Distributions

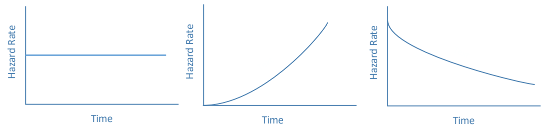
- A comprehensive framework involves the following steps:
  1. Construct histogram of failure times
     - Histogram involves grouping failure times into classes and plotting the frequency of the number of observations in each class.
  2. Compute descriptive statistics
     - If mean & median are close, then distribution is likely symmetric, e.g. Normal
     - If mean and standard deviation are equal, then the distribution is likely Exponential.



## Identifying Failure Distributions

3. Analyze failure rate
   - Graphing hazard rate indicates whether $\lambda(t)$ is constant, increasing, or decreasing.




4. Use prior knowledge of failure process.
5. Use properties of theoretical distributions.
6. Construct a probability plot.

## Probability Plots & Least Squares Method

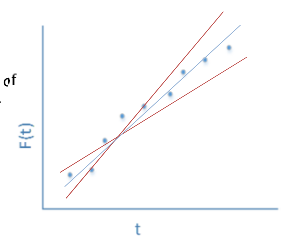
- Probability plots are one of the methods used for evaluating the fit of a sample of data to a distribution.
- Probability plots involve plotting the points $(t_i, \hat{F}_i(t))$ on appropriate graph paper that depends on the type of distribution being considered.
  - $\hat{F}_i(t)$ is often a transformation of the cdf of the hypothesized distribution.
- Data that is considered to be a proper fit to the distribution would graph as an approximate straight line.

## Probability Plots & Least Squares

- Fitting is accomplished using the method of **Least Squares**.
- This approach involves fitting a linear regression of the form $(y = bx + a)$ to a set of transformed data, using the method of least-squares.
- The least square equations for estimating $a$ and $b$ are given below:

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \quad \text{where } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\hat{a} = \bar{y} - b\bar{x} \quad \text{where } \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$




## Example: Least Squares for Exponential Distribution

- We know that the cdf of the exponential distribution is $F(t) = 1 - e^{-\lambda t}$.
- Rewriting the equation $1 - F(t) = e^{-\lambda t}$ and take the natural log of both sides, we arrive at $ln[1 - F(t)] = -\lambda t$. Thus:

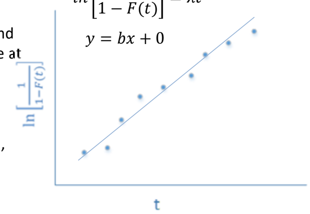
$- \ln(x) = \ln(x)^{-1}$

$$-ln[1 - F(t)] = ln\left[\frac{1}{1 - F(t)}\right] = \lambda t$$

- Given a sample of failure times $t_1, t_2, \dots, t_n$, and plotting the points $(t_i, \hat{F}_i(t))$ where $\hat{F}(t) = ln\left[\frac{1}{1-F(t)}\right]$,
- **It is clear that that the slope of the line can be used as an estimate of $\lambda$.**

$$ln\left[\frac{1}{1 - F(t)}\right] = \lambda t$$

$$y = bx + 0$$




## Probability Plots & Least Squares

- Probability plots can be performed for other known probability distributions such as the Weibull, Normal, Lognormal, etc.
- Least squares can also be used to estimate the respective parameters of these distributions.
- All these plots and computations can be performed using commercially available software as well as open source (free) statistical software packages.

## Maximum Likelihood Estimation

- Probability plots provided a way to identify the best distribution that fits the data.
- The least squares method provided a way to estimate the parameter of the distribution.
  - Recall the exponential case where the slope of the probability plot was used as an estimate for the disituirbution parameter $\lambda$.
- Although least squares provide estimates of the distribution parameters, they are often not necessarily the best estimates.

## Maximum-Likelihood Estimation MLE

- Maximum-likelihood estimation is a method for estimating the parameters of a distribution by selecting the values of the distribution parameter that maximizes a **likelihood function**.
- The likelihood function is a function is derived from the joint PDFs of the data.
  - If we represent observed failure time by $t_1, t_2, \dots, t_n$ where $t_i$ is the failure time of the $i^{th}$ unit.
  - Since failures are independent and from the same population, then the data is an **independent and identically distributed** sample of failure times.
  - If $f(t)$ is the pdf of the underlying population, and $f(t_i)$ is the pdf of the $i^{th}$ observation, the joint pdf of the sample is expressed as follows;

$$f(t_1, t_2, \dots, t_n) = f(t_1)f(t_2)\dots f(t_n) \longleftarrow \text{Likelihood Function}$$

## Maximum-Likelihood Estimation MLE

- Generally, the likelihood function is given by, $\mathcal{L}(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^{n} f(t_i)$ where $\theta_1, \theta_2, \dots, \theta_k$ are distribution parameters.
  - For example, for a single-parameter distribution like the exponential,

$$\mathcal{L}(\theta) = \mathcal{L}(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i}$$

- Ultimately, the objective of the MLE method is to maximize the likelihood function $\mathcal{L}(\theta_1, \theta_2, \dots, \theta_k)$ over the distributions parameters $\theta_1, \theta_2, \dots, \theta_k$.
- This is done by taking partial derivatives with respect to each parameter and equating to 0.

$$\frac{\partial \mathcal{L}(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_i} = 0$$

## Maximum-Likelihood Estimation MLE

- Recall that the likelihood function, $\mathcal{L}(\theta)$, is a product of the pdf, and that our goal is to maximize this function by differentiating it.
- Rather than taking the derivate of the original $\mathcal{L}(\theta)$, it is often easier to take the derivative of the $ln[\mathcal{L}(\theta)]$.
  - Taking the natural logarithm does not alter the result, but it simplifies the derivation.
  - Differentiating $\mathcal{L}(\theta)$ requires invoking the (differentiation) chain rule, which complicates the derivation.
  - Taking the natural-log changes the product to a summation.
- Let us look at an example.

It is much easier to differentiate a sum than it is a product

## MLE: Exponential Case

- Let $T_i$ be a random variable representing failure times of a specific machine, and assume that $T_i$ follows an exponential distribution.

$$P(T_i = t_i) = \lambda e^{-\lambda t_i}$$

- For the exponential case, $(\theta = \lambda)$, the above expression can be rewritten as :

$$\mathcal{L}(\theta) = \mathcal{L}(\lambda) = \left(\lambda e^{-\lambda t_1}\right) \times \left(\lambda e^{-\lambda t_2}\right) \times \cdots \times \left(\lambda e^{-\lambda t_n}\right) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i}$$

- The goal now is to maximize the above likelihood function by taking the derivative and equating to zero.
- The resulting expression is the maximum likelihood estimator of $\lambda$.

## MLE: Exponential Case

- Let us see how.

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i}$$

$$\mathcal{L}(\lambda) = \lambda^n \times e^{-\sum_{i=1}^{n} \lambda t_i}$$
$$= \lambda^n \times e^{-\lambda \sum_{i=1}^{n} t_i}$$

- Taking the natural-log of both sides yields,

$$ln[\mathcal{L}(\lambda)] = ln\,\lambda^n + ln\left[e^{-\lambda \sum_{i=1}^{n} t_i}\right]$$
$$= n \ln \lambda - \lambda \sum_{i=1}^{n} t_i \times ln[e]$$
$$= n \ln \lambda - \lambda \sum_{i=1}^{n} t_i$$

## MLE: Exponential Case

- Let us now, differentiate the **'log'-likelihood function** and equate to "0".

$$\ln[\mathcal{L}(\lambda)] = n \ln \lambda - \lambda \sum_{i=1}^{n} t_i$$

$$\frac{\mathrm{d}(\ln[\mathcal{L}(\lambda)])}{\mathrm{d}\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} t_i = 0$$

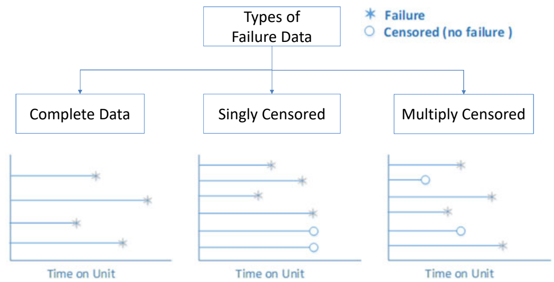
- The MLE of $\lambda$ can therefore be written as,

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}$$

## Censoring

- Up till now, we have assumed that a **complete** set of failure data is available to identify failure distributions and estimate their parameters.
- This is not necessarily always the case. Failure data can sometimes be **incomplete**, a phenomenon known as **Censoring**.
- Censoring occurs when the data is incomplete because,
  - Reliability tests are completed/stopped prior to all the units failing
  - Units are removed from reliability tests prior to their failure.

## Types of Censoring

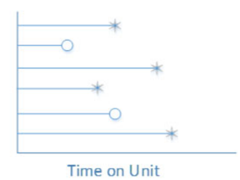



## Singly Censored Data

- **Singly Censored Data:** Arise when all units have the same (failure) test, and the test is concluded before all the units fail. <u>There are 3 main kinds.</u>
  - **Left Censoring:** Failure times of units are known to be *before some specified time* but it is unknown by how much.
  - **Interval Censoring:** Failure times of units are *between two values*.
  - **Right Censoring:** Failure times of units are known to be *after some specified time* but it is unknown by how much. <u>There are 2 types.</u>
    - **Type I Censoring:** Failure testing is terminated after a fixed time period $t^*$ has been reached. Thus, the remaining units are right-censored.
    - **Type II Censoring:** Failure testing is terminated after a fixed number of failures has occurred $r$ has occurred. Test time is given as $t_r$, time of the $r^{th}$ failure. The remaining units are right-censored.

## Multiply Censored Data

- **Multiply Censored Data:** Arise when test times (or operating times) differ among censored units (i.e. some units are removed while still operational).
  - Censored units are removed at various times from the sample
  - Units have gone into service at different times.
- For example, the figure shows two units being removed at various times.
- **Remark:** Recording failure data by failure mode will result in multiply censored data since units will be removed from a particular sample depending on the nature of their failure.




## Censored Data

- Censoring introduces additional challenges in the statistical analysis of failure times.
- However, they cannot be ignored. Ignoring censored data eliminates valuable information.
- For example:
  - If the remaining operating units from a Type I testing procedure were ignored, then only the weak units having the earliest failure times would be considered in during reliability analysis.
  - Thus, the reliability of the component would be underestimated.

## Estimation and Censored Data

- Recall that the Maximum-Likelihood Estimation (MLE) was a procedure used to estimate the parameters of a theoretical distribution used to describe the probability distribution of failure times.
- The likelihood function was derived as the product of the PDFs of the failure times.
- The key question that we need to ask is:
  - **How can we translate the testing/operating time of a censored unit into a probability statement?**

## Estimation and Censored Data

- Let's consider a case where the failure time of a censored unit means that its failure time is greater than a given time—often the time at which the unit was removed from testing (before its failure).
  - $t^*$ in the case of Single-Type I censoring.
  - $t_r$ in the case of Single-Type II Censoring.
  - $t_i^+$ in the case of Multiple Censoring.

$t^*$ is the test completion time

$t_r$ time of $r^{th}$ failure

$t_i^+$ implies that the failure time of the $i^{th}$ unit is greater than $t_i$, the time at which the test unit was removed.

## Estimation and Censored Data

- Formally, the failure time of a censored unit $i$ can be defined mathematically as $\mathbb{P}(T_i > t_C)$, where $t_C$ is the censoring time and is defined as,
  - $t_C = t^*$ in the case of Single-Type I censoring, $\mathbb{P}(T_i > t^*) = R(t^*)$.
  - $t_C = t_r$ in the case of Single-Type II Censoring, $\mathbb{P}(T_i > t_r) = R(t_r)$.
  - $t_C = t_i^+$ in the case of Multiple Censoring, $\mathbb{P}(T_i > t_i^+) = R(t_i^+)$.

## Estimation and Censored Data

- In the presence of censored data, the likelihood function must reflect the fact that no failures occurred at the censored times.
- This can be accomplished by modifying the likelihood function as follows:

$$\underbrace{\mathcal{L}(\theta) = \prod_{i\in F} f(t_i)}_{\text{All test units failed}} \longrightarrow \mathcal{L}(\theta) = \underbrace{\prod_{i\in F} f(t_i)}_{\text{Some test units failed}} \underbrace{\prod_{i\in C} R(t_i)}_{\text{Some test units censored}}$$

$R(t^*)$ Single-Type I

$R(t_r)$ Single-Type II

$R(t_i^+)$ Multiple

Where, $i \in F$ means that unit $i$ belongs to the set of failed units, and $i \in C$ means that unit $i$ belongs to the set of censored units.

## MLE and Censored Data: Exponential Case

- Recall the likelihood function for a sample of $n$ failed units from an exponentially distributed population with unknown parameter $\lambda$ is given as

$$\mathcal{L}(\theta) = \prod_{i\in F} f(t_i) \longrightarrow \mathcal{L}(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i}$$

- Now, let us consider a case with singly censored to right (Type I) data with $r$ units failed and the remaining $n-r$ components censored at time $t^*$ .

$$\mathcal{L}(\theta) = \prod_{i\in F} f(t_i) \prod_{i\in C} R(t^*) \longrightarrow \mathcal{L}(\lambda) = \prod_{i=1}^{r} \lambda e^{-\lambda t_i} \times [R(t^*)\,]^{n-r}$$

## MLE and Censored Data: Exponential Case

$$\mathcal{L}(\lambda) = \prod_{i=1}^{r} \lambda e^{-\lambda t_i} \times [R(t^*)\,]^{n-r}$$

$$= \lambda^r \exp\left(-\lambda \sum_{i=1}^{r} t_i\right) \times [\exp(-\lambda t^*)]^{n-r} = \lambda^r \exp\left\{-\lambda \sum_{i=1}^{r} t_i - \lambda(n-r)t^*\right\}$$

- Taking the logarithm and differentiating yields the following

$$\ln \mathcal{L}(\lambda) = r \ln \lambda - \lambda \sum_{i=1}^{r} t_i - \lambda(n-r)t^*$$

$$\frac{\mathrm{d}(\ln \mathcal{L}(\lambda))}{\mathrm{d}\lambda} = \frac{r}{\lambda} - \sum_{i=1}^{r} t_i - (n-r)\,t^* = 0 \longrightarrow \hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_i + (n-r)\,t^*}$$

## MLE and Censored Data: Exponential Case

- A similar derivation approach can be employed for the Type II and the Multiply censored failure data.

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_i + (n-r)\,t^*} \longrightarrow \hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_i + (n-r)\,t_r} \quad \text{Single-Type II}$$

$$\longrightarrow \hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_i + \sum_{i=r+1}^{n} t_i^+} \quad \text{Multiple}$$

- In some cases, you can have 2 different types of censored data within the same testing procedure.
  - 12 units tested and test terminated after 8 failures. In addition 2 units had to be removed prior to their failure and test termination).

## MLE and Censored Data

- Derivations for other types of probability distributions such as the Weibull, Normal, Lognormal, etc. can be performed in a similar manner.
- Some of the derivations may be a little more involved than others, however, the book provides several examples in Chapter 15.
- The appendix of chapter 15 also contains additional examples of MLE derivations for various types of censored data.

## Empirical Methods

- Empirical Methods
  - Empirical Probability & Reliability Plots
  - Empirical Hazard Function & MTTF
- One such package is Minitab.
- Accessible to all Georgia Tech students.

## Empirical Methods

- Generally, there are two approaches for fitting failure distributions. Up till now, we have focused mostly on fitting data to a theoretical probability distribution.
  - Cases where no theoretical distribution can adequately fit/describe the data.
- The second method involves deriving **empirical** distributions, reliability functions, or hazard functions **directly from the data**.
  - Also known as nonparametric methods since they do not require estimating parameters of known distributions.

## Empirical CDF & Reliability Function

- Let $t_1, t_2, \dots, t_n$ be an ordered set of $n$ failure times, where $t_i \le t_{i+1}$.
- If $n - i$ units survived at time $t_i$, then a possible estimate of the reliability function would be.

$$\hat{R}(t_i) = \frac{n-i}{n} = 1 - \frac{i}{n}$$

- This can be used to estimate the CDF of failure.

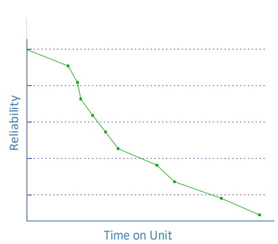
$$\hat{F}(t_i) = 1 - \hat{R}(t_i) = \frac{i}{n}$$

- Probability plots can be obtained by plotting the points $\left(t_i, \hat{F}(t_i)\right)$.

## Empirical CDF & Reliability Function

- If we analyze $\hat{F}(t_i) = \frac{i}{n}$, we would notice that $\hat{F}(t_n) = \frac{n}{n} = 1$ which implies that there is a zero probability of any unit surviving beyond $t_n$.
- Such an assumption may be unreasonable, especially since we are considering failure data from a sample.
- To overcome this technicality, we use the following modified estimate of the CDF,

$$\hat{F}(t_i) = \frac{i}{n+1}, \qquad \hat{R}(t_i) = 1 - \frac{i}{n+1}$$




## Empirical Hazard Function

- An estimate of the PDF may be obtained using the following expression.

$$\hat{f}(t) = \frac{\hat{F}(t_{i+1}) - \hat{F}(t_i)}{t_{i+1} - t_i} = \frac{1}{(t_{i+1} - t_i)(n+1)}$$

- The hazard function can be obtained as follows:

$$\hat{\lambda} = \frac{\hat{f}(t)}{\hat{R}(t)} = \frac{1}{(t_{i+1} - t_i)(n+1-i)}$$

- Recall,

$$\hat{R}(t) = 1 - \frac{i}{n+1} = \frac{n+1-i}{n+1}$$

## Empirical MTTF

- An estimate of the MTTF can be obtained using the following expression.

$$\widehat{MTTF} = \sum_{i=1}^{n} \frac{t_i}{n}$$

- An estimate of the variance of the failure distribution is

$$s^2 = \sum_{i=1}^{n} \frac{\left(t_i - \widehat{MTTF}\right)^2}{n-1} = \frac{\sum_{i=1}^{n} t_i - n\,\widehat{MTTF}^2}{n-1}$$

## Empirical Methods: Kaplan-Meier Estimator

- There are cases where a combination of censoring may occur within the same test. The resulting data is said to be **multicensored**.
- To date there are no known parametric methods for estimating reliability of multi-censored data
- Kaplan-Meier estimator is one of the popular methods for deriving reliability functions for multicensored data.
- This estimator relies on the fact that the probability of a component surviving during the interval $(t_i, t_{i+1})$ is estimated as the ratio between the number of units that did not fail during the interval and the units that were under the reliability test at the beginning of the interval (at time $t_i$).
- The reliability estimate at that interval is obtained as the product of all ratios from time zero until time $t_{i+1}$.

## Empirical Methods: Kaplan-Meier Estimator

- Each term in the equation below provides the conditional probability of surviving past time $t_j$ given that the unit survived to time $t_j$.

$$\hat{R}(t) = \prod_{\{j: t_j \le t\}} \left(1 - \frac{1}{n_j}\right)$$

Recall that $\hat{R}(t_i) = \frac{n-i}{n}$

- The product of these conditional probabilities provide the unconditional probability of surviving past time $t$.
- If two or more failures occur at time $t_j$, then the corresponding term in the above equation is replaced by $\left(1 - \frac{d_j}{n_j}\right)$ where $d_j$ is the number of failure at time $t_j$.

## Empirical Methods: Kaplan-Meier Estimator

- Let $t_i$ is an ordered set of failure times and $n_i$ is the number of units still operational prior to the $i^{th}$ failure.
- The K-P estimator is given by:

$$\hat{R}(t) = \prod_{\{j: t_j \le t\}} \left(1 - \frac{1}{n_j}\right)$$

| $i$ | $t_j$ | $n_j$ | $1 - \frac{1}{n_j}$ | $\hat{R}(t_i)$ |
|---|---|---|---|---|
| 1 | 150 | 10 | 9/10 | R(150)=(9/10)(1.0)=0.9 |
| 2 | $340^{+}$ | | | |
| 3 | 560 | 8 | 7/8 | R(560)=(7/8)(0.9)=0.787 |
| 4 | 800 | 7 | 6/7 | R(800)=(6/7)(0.787)=0.675 |
| . | | | | |
| . | | | | |
| 10 | | | | |

## Section Summary

Identifying Failure Distributions
- Least-Squares Estimation
- Maximum Likelihood Estimation (MLE)

Censoring
- Types of Censored Data
- MLE for Censored Data

Empirical Methods
- Traditional Empirical techniques
- Kaplan Meier Estimator