Reliability Analysis

Module 4: Basic Reliability Math: Statistics

Prof. Katrina M. Groth
Mechanical Engineering Department
Center for Risk and Reliability
University of Maryland
kgroth@umd.edu

Reliability Analysis © Katrina Groth, 2023

Reminder: Key concepts from Modules 2, 3

- Module 2:
  - Laws of probability & calculus of probability
  - Boolean algebra for manipulating events
  - Set theory for defining events
  - Joint vs. marginal vs. conditional probability
  - Common parametric probability distributions & their CDFs, PDFs
- Module 3:
  - Definitions of reliability $R(t)$, and hazard rate $h(t)$, MTTF, MRL
  - Bathtub curve and implications for reliability
  - Component reliability and commonly-used distributions

Objectives for Module 4

- Review descriptive statistics
- Introduce empirical distributions (non-parametric pdfs and cdfs)
- Introduce methods for parameter estimation
  - Point estimates & Interval estimates
- Hypothesis testing (aka statistical tests) for Goodness-of-fit
  - ….are these results possible from my chosen model?
    - Why? The model might not be capable of producing those results.

Data, samples, and populations

- Statistics – process of collecting, analyzing, organizing, and interpreting data about a population of interest.
- Data: Sometimes we have data for the full population, but more often we collect *samples* and want to make decisions about *populations*.
- Important terms:
  - **Sample size:** The number of observations in the sample
  - **Independently and identically distributed (i.i.d.):** An assumption made in classical statistics that all the observations are collected under the same conditions and represent independent events.
  - **Realization**: An observed value of the r.v.
  - **Statistical inference**: Using data to make conclusions about populations
    - Deduction: "Given a population, what will a sample look like?"
    - Induction: "Given a sample, what can be inferred about a population?"
  - Samples are imperfect -- we must make various corrections depending on sample characteristics such as sample size, randomness, non-independence, censoring…

Descriptive statistics from data

- Given a sample of size *n*, $\{x_1, x_2, \dots, x_n\}$ from the distribution of r.v. *X*, we can obtain descriptive statistics or sample moments to describe the sample.
- **Sample arithmetic mean:**

$$E(X) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- **Population variance:**

$$Var(X) = s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- **Sample variance:**

$$Var(X) = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

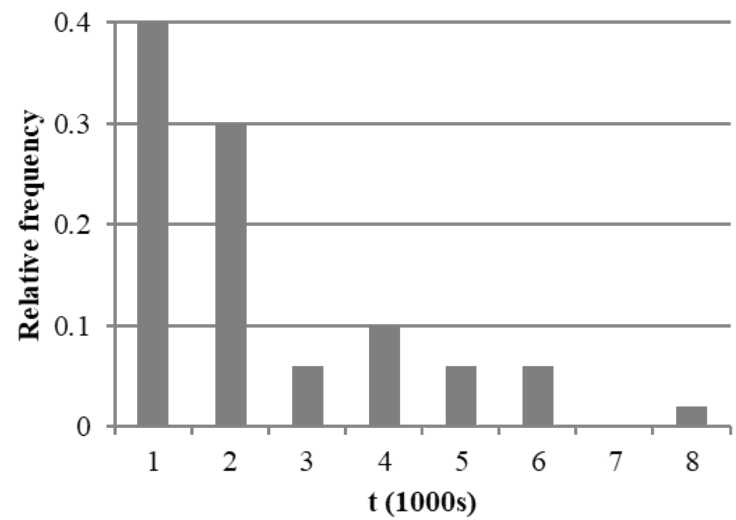
  - Because $\bar{x}$ is estimated from the same sample, the sample variance is biased. This bias is removed by multiplying by $\frac{n}{n-1}$.

Histograms & Empirical distributions

- **Histograms** help visualize data and inform other analysis. They can also display data where no known parametric distribution is a good fit (aka, an **empirical distribution**).
- For example, 50 lightbulb failure times, binned into 8 groups:

Given

t_i (×1000)		Freq.	Relative frequency (n_{t_i}/n)
0	1	20	0.4
1	2	15	0.3
2	3	3	0.06
3	4	5	0.1
4	5	3	0.06
5	6	3	0.06
6	7	0	0
7	8	1	0.02
Total:	**n=**	**50**	**1**



Histogram (Empirical distribution)

100 cables break at the following force, X (lb). What distribution should be used for X? We can use a histogram to visualize the data.

Breaking Strength of Cable (pounds)

Observations

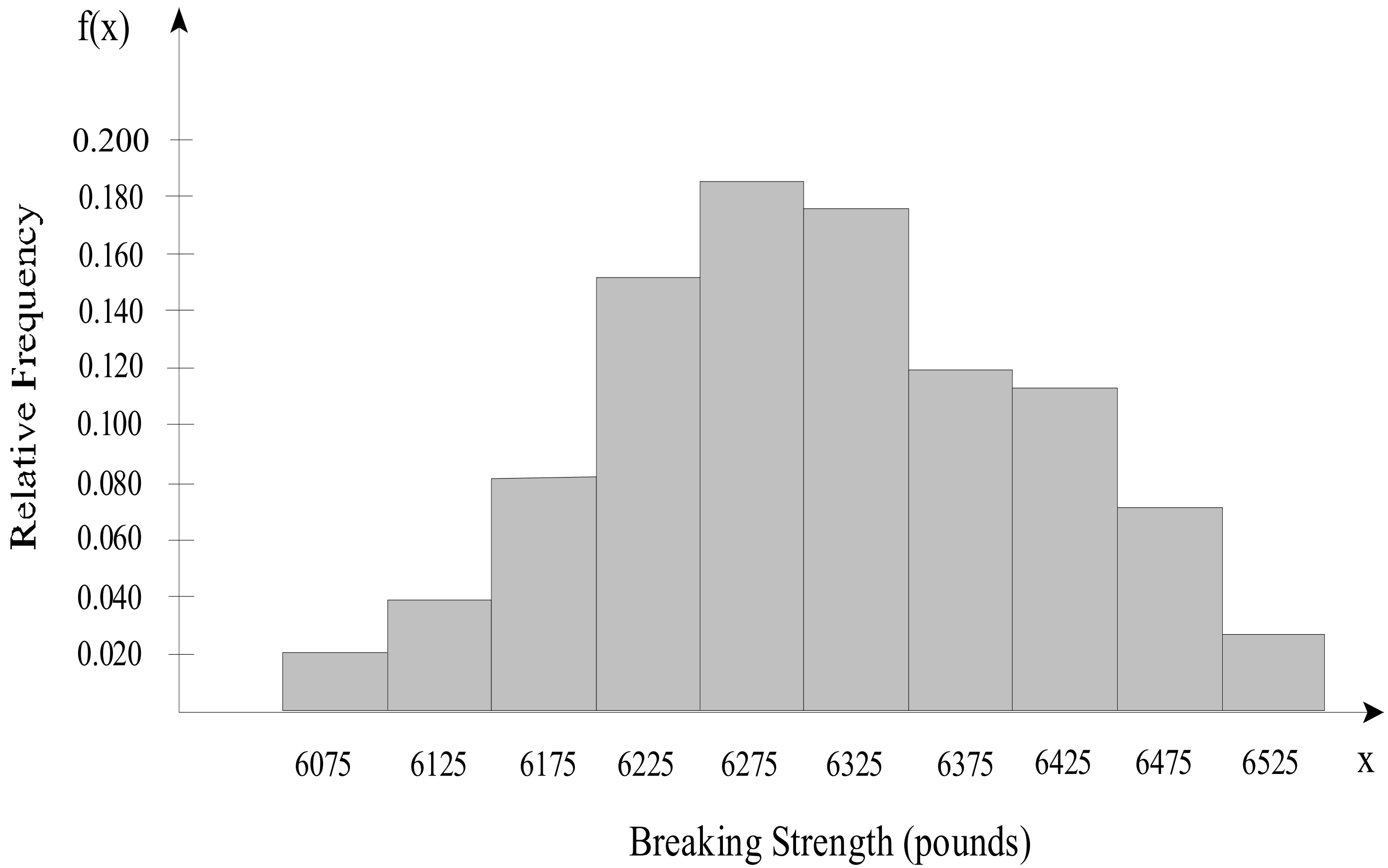
6182	6428	6374	6505	6295	6533	6305	6423	6219	6239
6312	6377	6112	6295	6187	6295	6112	6259	6413	6533
6275	6166	6318	6378	6395	6318	6384	6336	6245	6239
6302	6216	6475	6355	6187	6338	6125	6205	6456	6182
6256	6395	6229	6152	6440	6352	6488	6298	6270	6347
6355	6435	6298	6095	6166	6333	6464	6413	6362	6264
6166	6320	6267	6248	6208	6464	6187	6404	6290	6320
6212	6312	6280	6282	6475	6325	6248	6361	6320	6344
6297	6408	6349	6259	6448	6275	6361	6251	6408	6236
6135	6399	6301	6302	6201	6245	6201	6067	6475	6428

Creating a histogram & empirical distribution

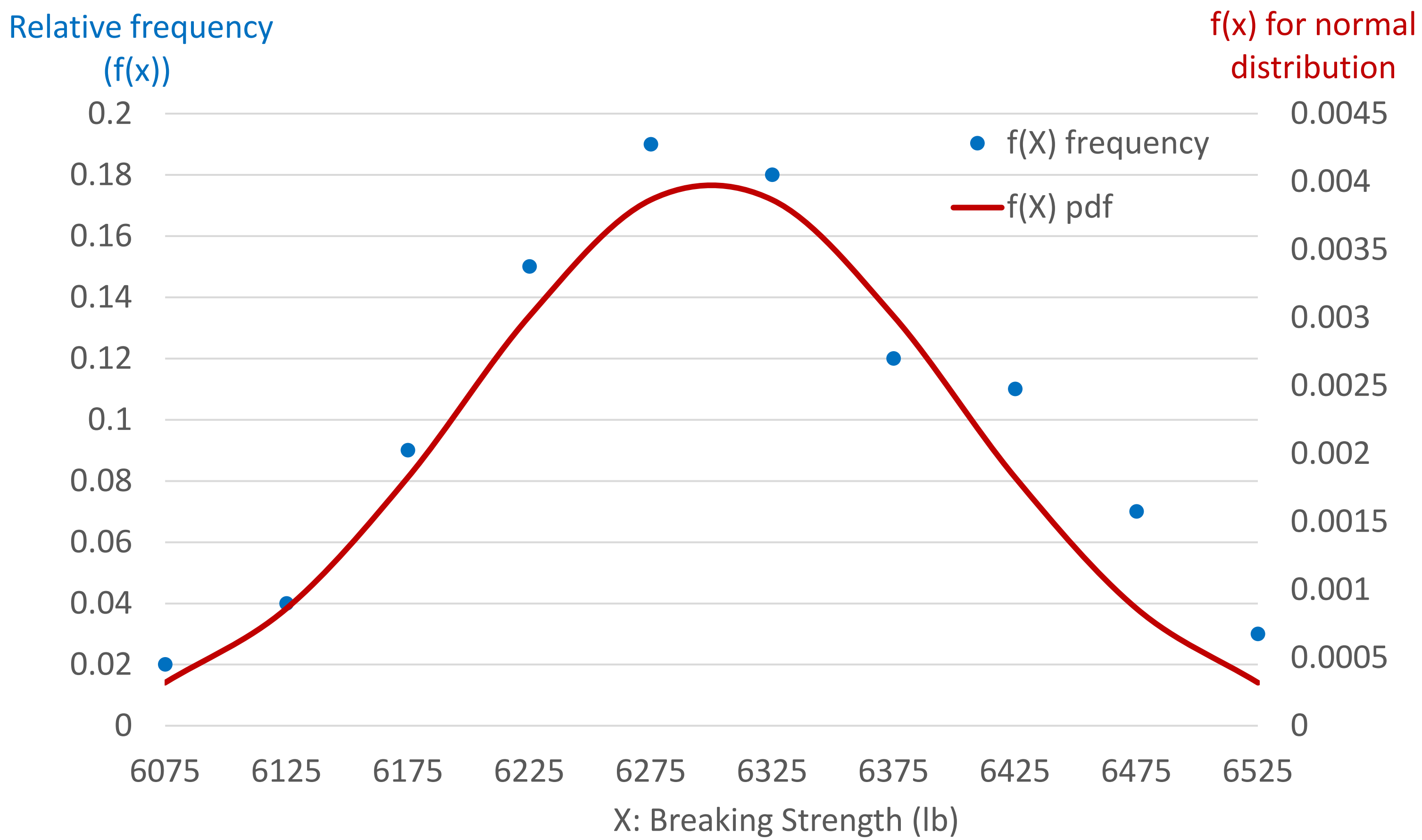
- First, bin the data into equal sized intervals. Then, count the frequencies in each bin and normalize it.

Interval	Midpoint X	Empirical frequency n_e	Relative frequency $f(x) = \frac{n_e}{n}$
$6050-6099$	6075	2	0.020
$6100-6149$	6125	4	0.040
$6150-6199$	6175	9	0.090
$6200-6249$	6225	15	0.150
$6250-6299$	6275	19	0.190
$6300-6349$	6325	18	0.180
$6350-6399$	6375	12	0.120
$6400-6449$	6425	11	0.110
$6450-6499$	6475	7	0.070
$6500-6549$	6525	3	0.030
Totals		100	1.000

Histogram of breaking strengths



Normal curve fit to data



PDF curve for $X \sim norm(\mu = 6300, \sigma = 100)$

Statistical inference

- **Statistical inference:** using sample data to answer questions about the distribution of an r.v.
- For example, how we can use a sample of failure times of an item, t_1, t_2, …, $t_{n,}$ to estimate λ. Therefore, we are inferring from a **specific** sample to a **general** distribution, i.e.,

  Part (sample) $\rightarrow$ Whole (population, model, distribution,)

- This cannot be achieved with **certainty**. Thus, two important aspects of inference are:

  (1) Estimating of distribution (model) parameters & characteristics
  (2) Statistical hypothesis tests

Parameter estimation

- We want to use a set of observations (data) to estimate the parameter(s) of a distribution.
- Desired properties of an estimator: unbiased, consistency, efficient, sufficient
- Let $f(x|\theta)$ denote a distribution (pdf) of r.v. X where θ represents an unknown parameter. Let $x_1, x_2, \ldots, x_n$ denote a sample from $f(x|\theta)$.
  - In frequentist statistics, the parameter set θ is *fixed* (but unknown) – we want to use our *random* sample of data to estimate it.
  - In Bayesian statistics, the parameter set is θ a *random* variable, and we use our *fixed* data (evidence) to update our knowledge of the parameter.

Parameter estimation: Method of moments

- Matching sample moments (e.g., sample mean ($\bar{x}$) and sample variance (s^2) to the distribution moments (e.g., mean ($\hat{\mu}_X$) and variance ($\hat{\sigma}_X^2$)):

$$\bar{x} \Longrightarrow \hat{\mu}_X$$

$$s^2 \Longrightarrow \hat{\sigma}_X^2$$

Example: Sample descriptive stats, method of moments

- **Example**: We want to verify the accuracy of cutting machine. To do this, five samples were cut to different lengths. We recorded the lengths, *X*, and room temperatures, *T*, as:

$$\mathrm{X} = 20.1, 22.13, 19.3, 18.5, 24.3\ \mathrm{mm}$$
$$\mathrm{T} = 20.13, 25.3, 22.4, 19.3, 24.5\ ^{\circ}\mathrm{C}$$

a) Find the sample mean and variance to estimate the parameters of two normal distributions representing *X* and *T*.

b) (If time) Use the distributions you found to calculate $Pr(X<19mm)$ and $Pr(T<20.5\ ^{\circ}\mathrm{C})$

c) (If you finish early, also find the correlation between *X* and *T*).

Example: Sample descriptive stats, method of moments

- **Solution a):** Find the mean

$$\hat{\mu}_x = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(20.1 + 22.13 + 19.3 + 18.5 + 24.3) = \mathbf{20.87}\boldsymbol{mm}$$

$$\hat{\mu}_T = \bar{T} = \frac{1}{n}\sum_{i=1}^{n} T_i = \frac{1}{5}(20.13 + 25.3 + 22.4 + 19.3 + 24.5) = \mathbf{22.33}℃$$

Example: Sample descriptive stats, method of moments

- **Solution a) (cont.):** Find the Variance:

$$\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{5-1}[(20.1 - 20.9)^2 + (22.13 - 20.9)^2 + (19.3 - 20.9)^2 + (18.5 - 20.9)^2 + (24.3 - 20.9)^2] = \mathbf{5.51}\boldsymbol{mm}$$

$$\hat{\sigma}_T^2 = s_T^2 = \frac{1}{n-1}\sum_{i=1}^{n}(t_i - \bar{T})^2$$

$$= \frac{1}{5-1}[(20.13 - 22.3)^2 + (25.3 - 22.3)^2 + (22.4 - 22.3)^2 + (19.3 - 22.3)^2 + (24.5 - 22.3)^2] = \mathbf{6.89}\ ℃$$

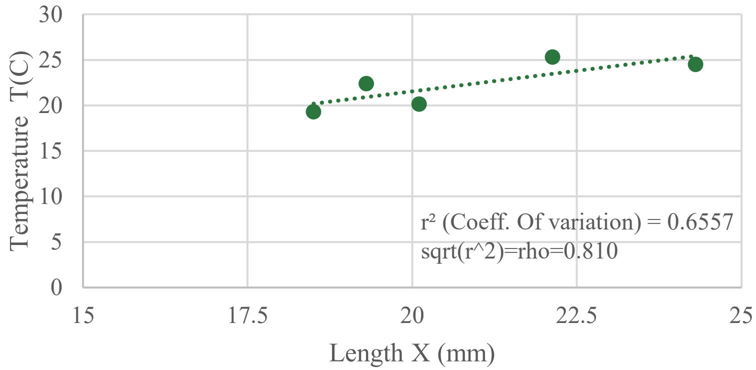
Example: Sample descriptive stats, method of moments

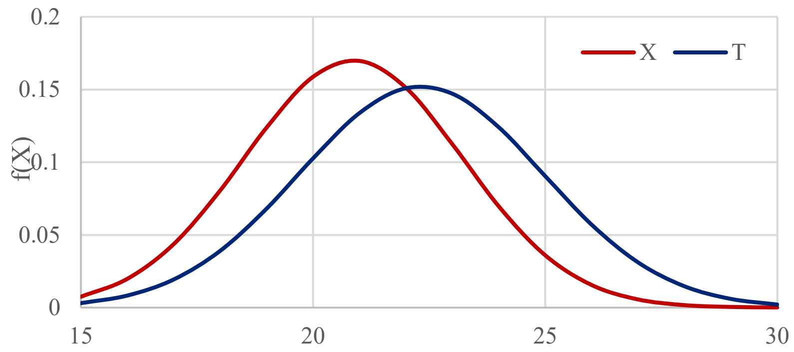
- **Solution b):** Find *Pr(X<19mm)* and *Pr(T<20.5°C)*

$$\mathbf{Pr}(\boldsymbol{X} < \mathbf{19}\boldsymbol{mm}) = Norm.dist\left(20.87, \sqrt{5.51}, true\right) = \mathbf{0.213}$$

$$\mathbf{Pr}(\boldsymbol{T} < \mathbf{20.5}°\mathrm{C}) = Norm.dist\left(22.33, \sqrt{6.89}, true\right) = \mathbf{0.243}$$

- **Solution c):** After class, you can also calculate correlation between these two sets of data.
  - The correlation is 0.810, which indicates a strong positive relationship between sample length and room temperature.





Parameter estimation: Maximum likelihood estimation

- The **likelihood function** of a parameter θ, given the evidence E, (e.g., n sample data points), is given by the probability of the joint occurrence of the data.

$$L(\theta|E) = \prod_{i=1}^{n} f(E_i|\theta)$$

Note: when $f(x|\theta)$ is viewed as a function of (variable) x with θ fixed, it's called a pdf. When it's viewed as a function of (variable) θ with x fixed, it's called a likelihood function, $L(\theta|x)$

  - Likelihood Function L and log-likelihood Λ for complete data:

$$L(\theta|x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$\Lambda(\theta|E) = \ln[L(\theta|E)] = \sum_{i=1}^{n} ln[f(x_i|\theta)]$$

Maximum Likelihood Estimate: Definition

- **Maximum Likelihood Estimate (MLE)** of θ is the value of $\hat{\theta}$ such that

$$L(\hat{\theta}|x_1, x_2, \dots, x_n) \geq L(\theta|x_1, x_2, \dots, x_n),$$

for every value of θ.

- The statistic $\hat{\theta}$ is a r.v. called the ML estimator (MLE) of θ.

- We solve for θ by maximizing the likelihood function, i.e.:

$$\left.\frac{\partial L}{\partial \theta}\right|_{\theta=\hat{\theta}} = \left.\frac{\partial lnL}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0$$

Why MLE?

- Useful for comparing distribution fits to the data: distribution with higher likelihood L is a better fit than one with a lower L.

- The uncertainty or confidence intervals of the parameters can be obtained directly.

- One solution that is not affected by the choice of plotting positions (used in linear regression).

Example: MLE

- **a)** For a r.v. *X*, representing failure times of a widget, you have a set of data of *n* failure times $(x_1, x_2, \ldots x_n)$. Write the likelihood function of this data. Assume that the likelihood of each data point is given by a normal distribution.

- **b)** Compute the point estimate of the normal parameters (μ, σ) using MLE.

Example: MLE

- **a) Solution:** Write the likelihood function of *n* failure times, where the likelihood of each data point is given by a normal distribution.

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

$$L(x_1, x_2, \dots, x_n|\mu,\sigma) = \prod_{i=1}^{n} f(x_i|\mu,\sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]}$$

$$\boldsymbol{L(x_1, x_2, \dots, x_n|\mu,\sigma) = \frac{1}{\sigma^n(2\pi)^{n/2}} \prod_{i=1}^{n} e^{\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]}}$$

Example: MLE

b) Solution: First, compute the log-likelihood by taking $\ln(L)$ and simplifying it:

$$\ln(L) = -n \cdot \ln\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

- Then differentiate the log-likelihood wrt to μ and σ:

$$\frac{\partial \ln(L)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{\mu n}{\sigma^2} = 0$$

$$\frac{\mu n}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i$$

$$\mu n = \sum_{i=1}^{n} x_i$$

$$\boldsymbol{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

$$\frac{\partial \ln(L)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

$$\frac{n}{\sigma} = \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$n = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_{i-}\hat{\mu})^2$$

$$\boldsymbol{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}$$

- In most cases however, unbiased estimators are used so we multiply the variance estimate by $n/(n-1)$ to get the familiar form with 1/(n-1).

MLE Parameters of normal and lognormal distributions

Continuous, for n data points representing t_i failure times:

Normal Distribution (biased correction)

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

$$\boldsymbol{\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(t_i - \hat{\mu})^2}$$

Lognormal Distribution (bias corrected)

$$\boldsymbol{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} lnt_i}$$

$$\boldsymbol{\hat{\sigma}^2 = \frac{1}{\mathrm{n}-1}\sum_{\mathrm{i}=1}^{n}(\mathrm{ln}t_i - \hat{\mu})^2}$$

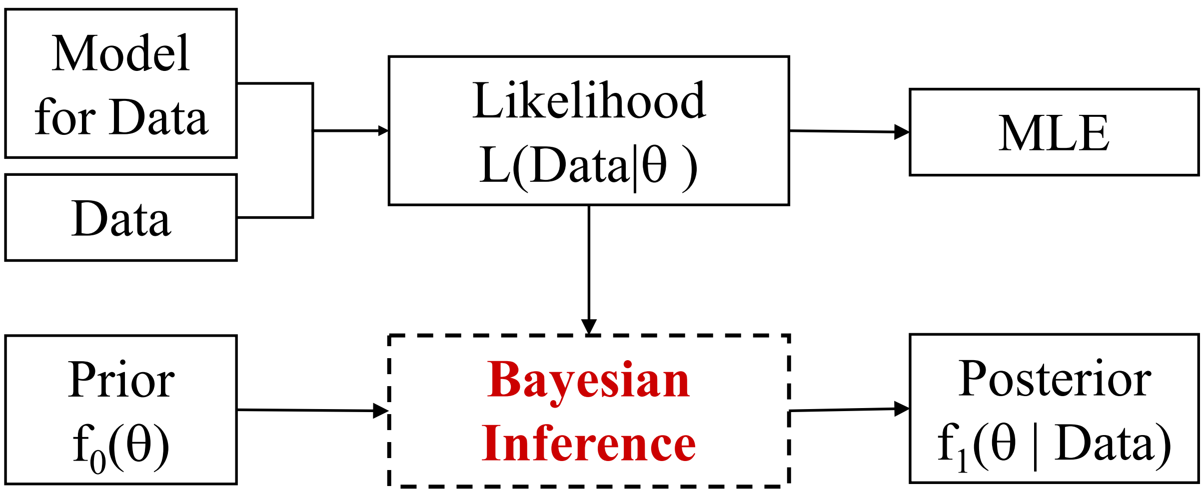
More to be discussed in Ch 5.

Why Bayesian parameter estimation methods?

- To incorporate prior knowledge of the phenomena
  - Combinations of extensive past experience, physical/chemical theory can provide prior information to form a framework for inference and decision making. **--** *it would be shortsighted to ignore this prior information* – even if we do have "enough" data.
- For problems that have insufficient data for (or violate other assumptions inherent in use of) frequentist methods.
- To directly quantify uncertainty

Bayesian parameter estimation

- Bayesian parameter estimation methods are closely related to maximum likelihood methods
- The basic process of Bayesian parameter estimation is to update our prior knowledge of an *unknown* parameter of interest using new data



- Notice use of $L(\theta|x_1, x_2, \dots, x_n)$ in MLE but $L(x_1, x_2, \dots, x_n|\theta)$ in Bayesian estimation; MLE treats $L(\theta|E)$ and $L(E|\theta)$ as equivalent, Bayesian est. does not.

Bayesian parameter estimation

- The basic process of Bayesian parameter estimation is to update our prior knowledge of an *unknown* parameter of interest using new data (evidence).
- Using the continuous form of Bayes' theorem w/ evidence E:

$$f_1(\theta|E) = \frac{f_0(\theta)L(E|\theta)}{\int_{-\infty}^{\infty} f_0(\theta)L(E|\theta)d\theta}$$

  - **Prior distribution:** $f_0(\theta) or\ \pi_0(\theta)$ pdf representing the distribution of parameter θ estimated from prior information, or the degree of belief of expert judgment, etc.
  - **Likelihood function:** $L(E|\theta)$ or $\Pr(E|\theta)$ representing actual data (*evidence*) about the r.v.
  - **Posterior distribution:** $f_1(\theta|E)$ or $\pi_1(\theta|E)$ or $f_{\Theta|X_1,X_2,X_3,..}(\theta|x_i, x_2, x_x, \dots)$

Bayesian parameter estimation (cont.)

- **Discrete Form**
  - If available data are represented best by a discrete probability function $\Pr(data|\theta)$, then,

$$\Pr(\theta|E) = \frac{\Pr(E|\theta)\Pr(\theta)}{\sum_{i=1}^{\infty}\Pr(E|\theta_i)\Pr(\theta_i)}$$

Prior distribution and likelihood function are both discrete

- **Mixture Form**

$$f_1(\theta|E) = \frac{\Pr(E|\theta)\, f_0(\theta)}{\int_{-\infty}^{\infty}\Pr(E|\theta)\, f_0(\theta)d\theta}$$

Prior distribution is continuous likelihood function is discrete

More to be discussed in Ch 5.

Interval estimation

- Why? Different amounts of data influence our (un)certainty about the parameter
- An attempt to answer: How 'far off' could we be?
- Frequentist approach: Confidence intervals
  - Quantify uncertainty due to sampling error (i.e., limited number of samples)
- Bayesian approach: Credible intervals (aka Bayesian probability interval)
  - Statements about the probability of a parameter based on the evidence

- **Not a characterization of uncertainty due to incorrect model selection or assumptions!**

Parameter estimation: Confidence intervals

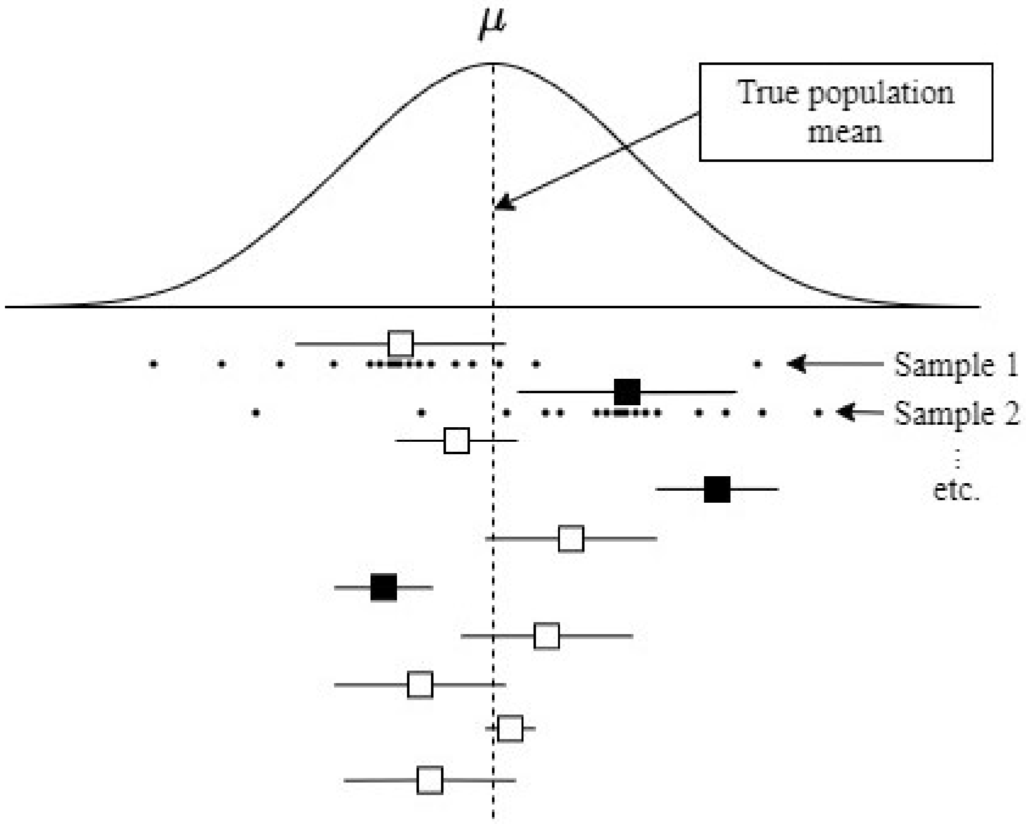
- **Confidence intervals** are a frequentist expression of uncertainty about estimated parameter values.
  - The main purpose is to find an interval with a high probability of containing the true (but unknown) value of a parameter θ
  - Often misinterpreted as a probability that θ is in this interval.
- Consider r.v.s $\theta_l(x_1, \dots, x_n)$ and $\theta_u(x_1, \dots, x_n)$, such that the probability that interval $[\theta_l, \theta_u]$ contains the true value of θ is:

$$\Pr[\theta_l(x_1, \dots, x_n) < \theta < \theta_u(x_1, \dots, x_n)] = 1 - \gamma$$

- Interval $[\theta_l, \theta_u]$ is the **(two-sided) k% confidence interval** for parameter θ with **confidence level** (**confidence coefficient**): $k\% = 100(1-\gamma)\%$.
- **Ex:** For a 90% confidence interval around $R(t)$, $\gamma = 0.1$. Thus, $\Pr(R_L \le R(t) \le R_U) = 1 - \gamma = 0.9$.

Illustration of how confidence intervals are generated

- See a known distribution with mean parameter μ, and 10 rows of samples from the distribution.
- Whiskers represent 70% confidence intervals.
- What percent of the generated intervals contain the true mean? (Hint: box colors)
- Interpretation:
  - In $100(1-\gamma)\%$ of repetitions, μ falls between μ_L and μ_U.
  - Not: the probability that μ is in the given interval.



Confidence intervals express uncertainty due to sample size

- **Example**: If 100 units are tested, consider two situations for exponential parameter estimation:
  - **Case 1: For r = 1** $\boldsymbol{failure, t_0 = 10}$ **hrs**
    - $T = 10 \cdot 100 = 1000$ test hours
    - $\hat{\lambda} = \frac{r}{T} = \frac{1}{1000} = 10^{-3} hr^{-1}$
  - **Case 2: For r = 10 failures,** $\boldsymbol{t_0 = 100}$ **hrs**
  - $T = 100 \cdot 100 = 10{,}000$ test hours
  - $\hat{\lambda} = \frac{r}{T} = \frac{10}{10{,}000} = 10^{-3} hr^{-1}$
  - Both gives you the same $\hat{\boldsymbol{\lambda}}$ estimate, but one has more data. The MLE parameter is the same, but the confidence interval is different for these two datasets.

Normal distribution confidence interval

- **If σ is known:**
  - Construct a confidence interval for μ based on a random sample $\{x_1, \ldots, x_n\}$ from a normal distribution: $X \sim norm(\mu, \sigma)$
  - Sample mean is normally distributed: $\bar{x} \sim norm\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Then use the Z transformation

$$\Pr\left(z_{\left(\frac{\gamma}{2}\right)} \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\left(1-\frac{\gamma}{2}\right)}\right) = 1-\gamma$$

  - Thus:

$$\boxed{\Pr\left(\bar{x} - z_{\left(1-\frac{\gamma}{2}\right)} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\left(1-\frac{\gamma}{2}\right)} \frac{\sigma}{\sqrt{n}}\right) = 1-\gamma}$$

Confidence interval example

- **Example:** Consider the following sample, taken from a normal distribution with known $\sigma = 3.1$. Find the 90% confidence interval around μ.

5.19, 11.84, 11.54, 8.44, 6.87, 12.96, 13.69, 14.39, 7.83, 10.62

Confidence interval example

Solution:

- $x_i = \{5.19, 11.84, 11.54, 8.44, 6.87, 12.96, 13.69, 14.39, 7.83, 10.62\}$
- $n = 10$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10} \cdot (103.37) = 10.34$$

- $\gamma = 0.1 \Longrightarrow \frac{\gamma}{2} = 0.05$
- Using Appendix A, $z_{\left(1-\frac{\gamma}{2}\right)} = 1.65$, so:

$$\Pr\left(\bar{x} - z_{\left(1-\frac{\gamma}{2}\right)}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\left(1-\frac{\gamma}{2}\right)}\frac{\sigma}{\sqrt{n}}\right) = 1-\gamma = 0.9$$

$$\Pr\left(10.34 - 1.65 \cdot \left(\frac{3.1}{\sqrt{10}}\right) \le \mu \le 10.34 + 1.65 \cdot \left(\frac{3.1}{\sqrt{10}}\right)\right) = 0.9$$

$$\Pr(8.72 \le \mu \le 11.96) = 0.9 \Longrightarrow \boldsymbol{\mu} \in [\mathbf{8.72}, \mathbf{11.96}]$$

Normal distribution confidence interval

- **If $\boldsymbol{\sigma}$ is unknown:**
  - If σ is estimated from the sample as $\hat{\sigma} = s$, or if sample is small (e.g., $n < 30$), then
  - The r.v. $\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ follows the **Student's *t*-distribution** with *n-1* degrees of freedom, and the **confidence interval around $\boldsymbol{\mu}$** is given:

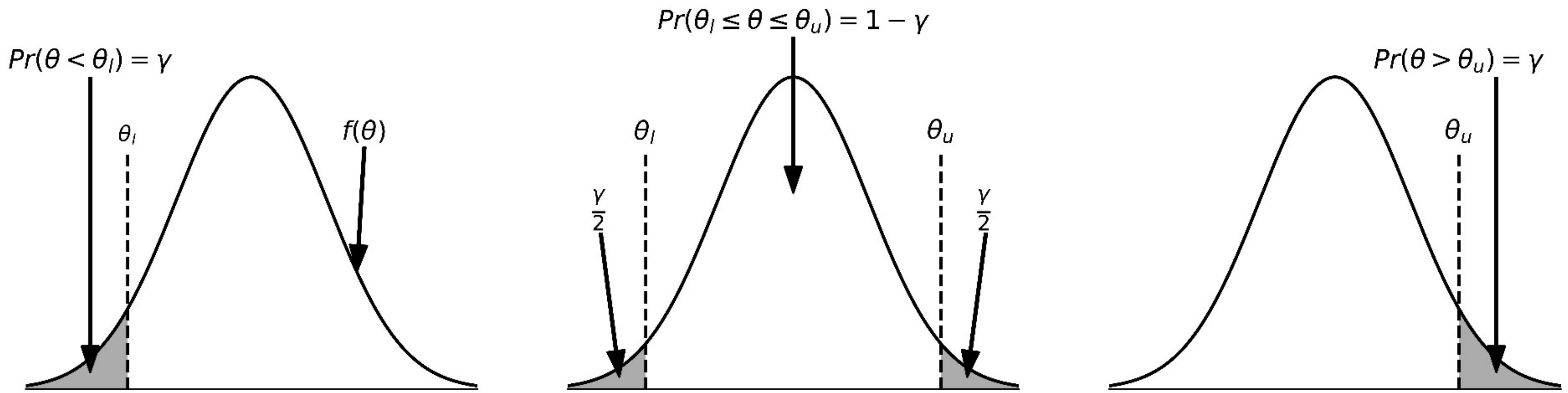
$$\Pr\left(\bar{x} - t_{\left(1-\frac{\gamma}{2}\right)}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\left(1-\frac{\gamma}{2}\right)}\frac{s}{\sqrt{n}}\right) = 1 - \gamma$$

  - Additionally, the **confidence interval around $\boldsymbol{\sigma^2}$** is given:

$$\Pr\left(\frac{(n-1)s^2}{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[n-1]} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\left(\frac{\gamma}{2}\right)}[n-1]}\right) = 1 - \gamma$$

Bayesian credible intervals

- The **credible interval** (Bayesian probability interval) is the Bayesian analog to the confidence interval.
- Using the posterior distribution of the parameter θ:
$$\Pr(\theta_l \le \theta \le \theta_u) = 1 - \gamma$$
- The interval $[\theta_l, \theta_u]$ is the $100(1-\gamma)\%$ credible interval for θ.
- Interpretation:
  - There is a $100(1-\gamma)\%$ probability that the true value of θ is contained within the interval.



(a) One-tailed lower interval (b) Two-tailed interval (c) One-tailed upper interval

Credible interval vs. confidence interval

- Recall that the **confidence interval** is interpreted:
  - $k\% = 100(1-\gamma)\%$ confidence that any generated interval $[\theta_l, \theta_u]$ contains the true value of θ
  - Confidence about the **interval** $[\boldsymbol{\theta_l}, \boldsymbol{\theta_u}]$
- While the **credible interval** is interpreted:
  - Directly gives the probability that the true value of θ falls in the interval $[\theta_l, \theta_u]$
  - The interval has a $k\% = 100(1-\gamma)\%$ probability of containing θ
  - Confidence about the **value of $\boldsymbol{\theta}$**

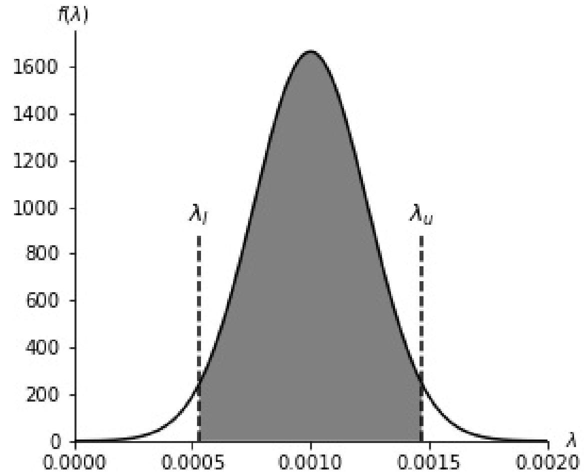
Example: credible intervals

- **Example**: A capacitor has a time to failure, *T*, that can be represented by the exponential distribution with a failure rate λ. The mean value and uncertainty about λ are modeled through a Bayesian inference with the posterior pdf for λ expressed as a normal distribution with a mean of 1×10^{-3} and standard deviation of 2.4×10^{-4}. Find the 95% Bayesian credible interval for the failure rate.

- Which means:
  - T~exp(λ) where $\lambda \sim norm(\mu = 1 \times 10^{-3}, \sigma = 2.4 \times 10^{-4})$

Example: credible intervals

- **Solution:**
- To obtain the 95% credible interval for the normal distribution on λ, we find the 2.5th and 97.5th percents of the normal distribution representing uncertainty on the parameter λ.
- $\lambda_l = \Phi^{-1}(0.025, 1 \times 10^{-3}, 2.4 \times 10^{-4}) = 5.30 \times 10^{-4}$,
- $\lambda_h = \Phi^{-1}(0.975, 1 \times 10^{-3}, 2.4 \times 10^{-4}) = 1.47 \times 10^{-3}$.
- Therefore, $\mathbf{5.30 \times 10^{-4} < \hat{\lambda} < 1.47 \times 10^{-3}}$

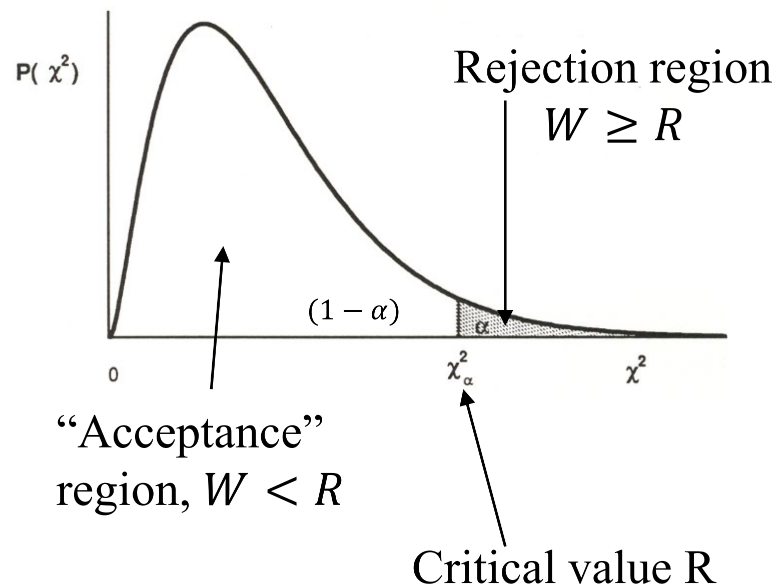


Hypothesis testing and goodness of fit

- *Question*: is observed data consistent with the assumption that it comes from a specific theoretical distribution (i.e., with known form and parameter)?
  - It is necessary to choose properly which distribution appropriately fits a set of data.
  - Even with the obvious choices (e.g., Weibull distribution for failure times), you may need to check for appropriateness of the distribution.
  - Test whether the **hypothesis** that the data originate from a known distribution is true.
- For this, reliability engineers rely on several methods to see which distribution is the best fit:
  1) Chi-Square Goodness-of-fit test
  2) Kolmogorov-Smirnov (K-S) test
  3) Linear Regression Analysis (Least Squares estimation)
- One or more of these methods may be applied to support your choice of distribution.

Generic hypothesis estimation procedure

1. Set **Null Hypothesis** (H_0: $\theta = \theta_0$)
   - Define an appropriate alternative hypothesis (H_1: $\boldsymbol{\theta} \neq \boldsymbol{\theta_0}, \theta = \theta_1, \theta < \theta_0, \theta > \theta_0$, etc.)
2. Define a **significance level**, γ
3. Calculate a **test statistic** (W) using sample data.
4. Calculate **critical value** of the test statistic (R) (i.e., boundary of rejection region)
5. Decision: If the test statistic from the sampled data is in the rejection region, reject H_0; otherwise, don't reject H_0.



Note: we can only *fail to support* the null hypothesis, so we can only reject it rather than positively affirm it as truth. Therefore, the hypothesis is either *rejected* or *not rejected* rather than *accepted* or *not accepted*.

Chi-squared test in frequentist statistics

- Based on a statistic that has an approximate χ^2 distribution.
- Compares observed frequencies (o_i) of data to the expected frequencies (e_i) generated from the hypothesized distributions.
  - Suitable for data sets which are frequencies or counts (not probabilities).
  - Data must be in **mutually exclusive intervals**.
- Checks:
  - The value of the cell "expected" should be ≥ 5 in at least 80% of the cells, and no cell should have an "expected" of less than 1.
    - If not: group some intervals. Only adjacent intervals may be grouped together.
    - This assumption is most likely to be met if the sample size $\geq \#intervals \cdot 5$
- Test statistics χ^2 is a value of an r.v. whose sampling distribution is approximated very closely by the chi-squared distribution with:
  - $v = k - m - 1$ degrees of freedom.
  - (k = # of intervals, m = # of parameters estimated from the data)

Chi-square goodness-of-fit test

- The steps in the **Chi-Square Goodness-of-Fit Test** are:
  a) Choose the hypothesized distribution, H_0 (e.g., an exponential distribution with $\lambda = 2.1$)
  b) Select a significance level (γ)
  c) Calculate the test statistic, W:

$$W = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}; \text{where} \quad \begin{array}{l} o_i = \text{observed frequency at } i \\ e_i = \text{expected frequency at } i \end{array}$$

  e) Establish a critical value (edge of rejection region) $R = \chi^2_{(1-\gamma)}[df]$
     - df=k-m-1: k = # of intervals, m = # of parameters estimated from the data

       Table in Appendix A
       Excel: chisq.inv(1-γ, df)
       Matlab: chi2inv(1-γ, df)

  e) If $W > R$ reject H_0; otherwise do not reject.

Chi-square goodness-of-fit test

- **Example**: Consider the data below for X, the number of replacement parts which need to be ordered each week by a repair facility. Test the hypothesis that these come from a Poisson distribution. Use a significance level of $\gamma = 10\%$.

Number of Failed Parts/Week (x)	Observed Frequency (o_i)
0	18
1	18
2	8
3	5
4	2
5	1

Hints:

- Calculate the Poisson parameter first
- Don't get the total number of weeks confused with the total number of parts ordered.

Chi-square goodness-of-fit test

- **Solution:** We are told to test the fit to a Poisson distribution, but we're not given a parameter, so first the parameter μ has to be estimated:

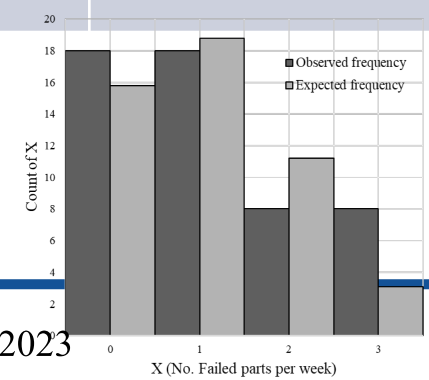
$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n\ weeks} \cdot \sum_{i=1}^{n} x_i\ parts = \frac{62}{52} = 1.192\,\frac{\text{parts}}{\text{week}}$$

- For this test, use k (number of intervals) = 4 [because the last three rows should be combined].
  - The grouping is a good idea when e_i has a very low value or differs by a large factor (5-10) different from other e_i's calculated.

Chi-square goodness-of-fit test

- **Solution:** Test the hypothesis that these data come from a Poisson distribution

Number of Failed Parts/Week (x_i)	**Observed Frequency (o_i)**	**Hypothesized or Expected Frequency (e_i)**	$W = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}$
0	18	$52 \times \Pr(X = 0) = 15.8$	0.31
1	18	$52 \times \Pr(X = 1) = 18.8$	0.036
2	8	$52 \times \Pr(X = 2) = 11.2$	0.923
3	5	$52 \times \Pr(X = 3) = 4.5$	
4	$2 \;\}\; 8$	$52 \times \Pr(X = 4) = 1.3 \;\}\; 6.1$	0.588
5	1	$52 \times \Pr(X = 5) = 0.3$	
$\sum x_i = 62$ parts	n= 52 weeks	Check: sum = 52 weeks	$\mathbf{W = 1.86}$



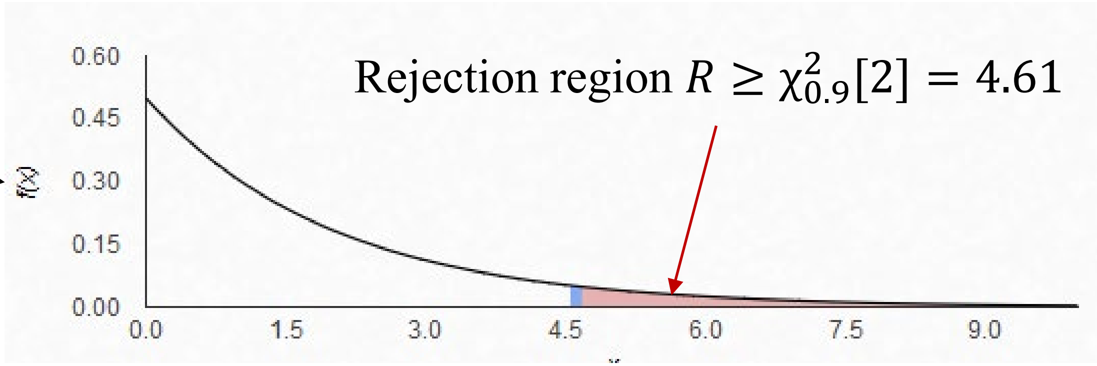
Chi-square goodness-of-fit test

- **Solution:** With k = 4 bins and m=1 (because 1 parameter, $\hat{\mu}$, is estimated from the data).

$$Test\ stat: \chi^2_{0.9}[k-m-1] = \chi^2_{0.9}[4-1-1] = \chi^2_{0.9}[2] = 4.61$$
$$Reject \geq (\chi^2_{0.9}[2] = 4.61)$$

- Our $W = 1.86$ so $W < R$
- So the hypothesis **is not rejected.**

pdf for $\chi^2_{0.9}[2]$



K-S (Kolmogorov-Smirnov) Test

- χ^2 requires substantial amounts of data to provide frequencies
- K-S test uses each individual data point (not frequencies) and is effective for small samples.
- Suppose for an r.v. T, we observe an **ordered sample** of data $t_1 \leq t_2 \leq t_3 \leq \cdots \leq t_n$

- The **sample cdf** $S_n(t)$ is defined for the ordered sample as:

$$S_n(t) = \begin{cases} 0 & -\infty < t < t_{i=1}, \\ \dfrac{i}{n} & t_i \leq t < t_{i+1}, i = 1, 2, \dots, n-1, \\ 1 & t_n \leq t < \infty. \end{cases}$$

The K-S Procedure

- **K-S Test Procedure:**
  1) Choose a hypothesized cumulative distribution $F(t)$ for sample.
  2) Select a significance level γ for the test
  3) Define the rejection region $R > D_n(\gamma)$. Critical values of the test statistic $D_n(\gamma)$ can be obtained from Appendix A.
  4) If $D > D_n(\gamma)$, reject the hypothesized distribution and conclude that $F(t)$ does not fit the data. Otherwise, do not reject.
- The K-S Test is performed similar to the χ^2 test:
  - H_0: T has cdf $F(t)$
    H_1: T does not have cdf $F(t)$
  - The K-S test statistic, W, is defined:

$$W = \max[|F(t_i) - S_n(t_i)|, |F(t_i) - S_n(t_{i-1})|]$$

  - The critical value of test statistic is $R = D_n(\gamma)$
    - n: number of observations.
    - $D_n(\gamma)$: See Appendix A

Example: K-S test

- **Example:** Are the failure times 8, 20, 34, 46, 63, 86, 111, 141, 186 and 266 hrs adequately modeled by an exponential distribution with parameter $\lambda = 0.01/\text{hr}$. Use $\gamma = 0.05$.
  - H_0: The r.v. T has cdf $F(t) = 1 - e^{-\lambda t} = 1 - e^{-0.01t}$
  - H_1: The r.v. T does not have cdf $F(t) = 1 - e^{-0.01t}$
- The calculated rejection region is:

$$R > D_{10}(0.05) = 0.409$$

  - $n = 10$ (number of observations)
  - $\gamma = 0.05$ (chosen significance level)

Example: K-S test solution

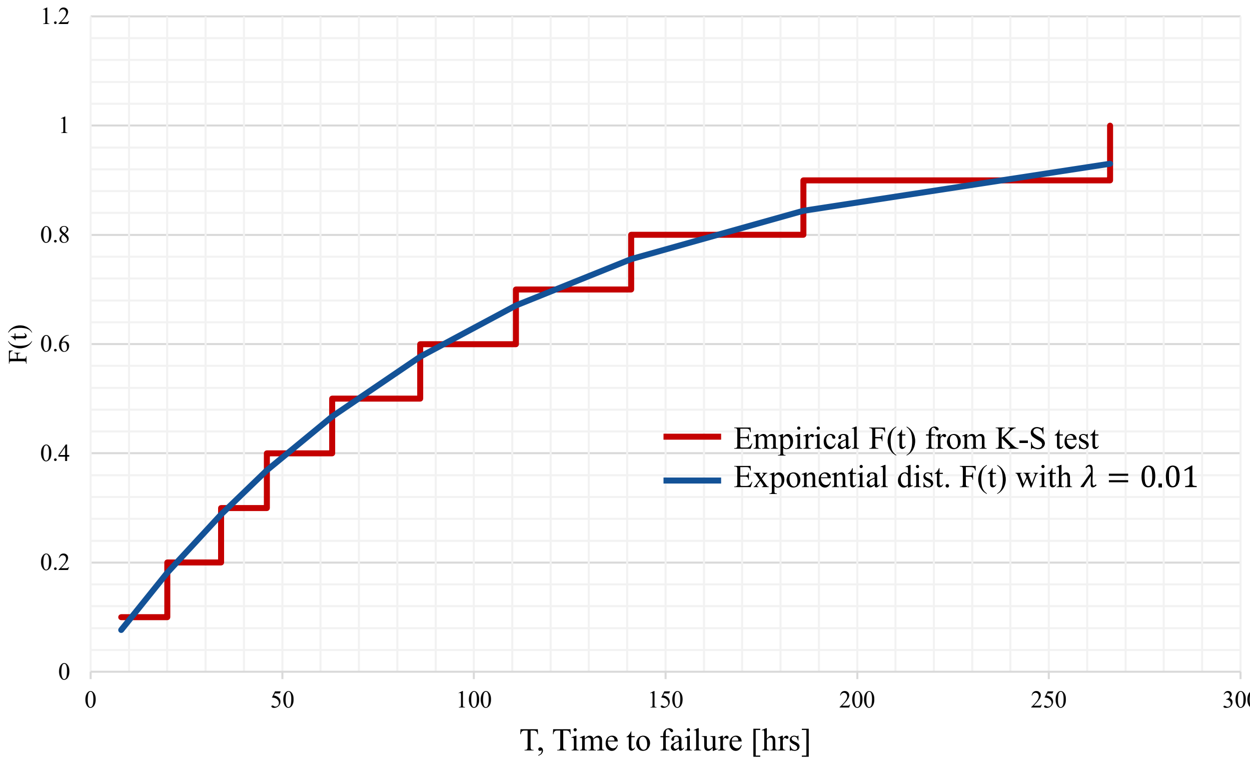
i	t_i	$S_n(t_i)$	$S_n(t_{i-1})$	$F(t_i)$	$F(t_i) - S_n(t_i)$	$F(t_i) - S_n(t_{i-1})$
1	8	0.1	0.0	0.077	0.023	0.077
2	20	0.2	0.1	0.181	0.019	0.081
3	34	0.3	0.2	0.288	0.012	0.088
4	46	0.4	0.3	0.369	0.031	0.069
5	63	0.5	0.4	0.467	0.033	0.067
6	86	0.6	0.5	0.577	0.023	0.077
7	111	0.7	0.6	0.670	0.030	0.070
8	141	0.8	0.7	0.756	0.044	0.056
9	186	0.9	0.8	0.844	0.056	0.044
10	266	1.0	0.9	0.930	0.070	0.030

$$W = \max[|F(t_3) - S_n(t_2)|] = \mathbf{0.088}$$

$W = 0.088 < 0.41$ thus we **don't reject $\mathbf{H_0}$** and conclude that the exponential distribution, T~exp($\lambda = 0.01$), is an acceptable model.

Example: K-S test solution shows good visual fit

- **Example:** By plotting the Empirical and fitted CDF we also see that the solution shows good visual fit to the exponential distribution.



Regression analysis

- **Definition**: In **Regression Analysis**, a correlation is made between a given dependent variable Y, and j number of independent variables $X_1, X_2, \dots, X_j$ (also called explanatory variables).
  - Explanatory variables $X_1, \dots, X_j$ do not have to be statistically independent
- Such a relation can be expressed as:

$$E[Y|X_1, \dots, X_j] = \beta_0 + \beta_1 x_1 + \ \dots \beta_j x_j$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \varepsilon$$

- This is the regression model where $\beta_1, \beta_2, \dots, \beta_k$ are regression parameters (or coefficients) and ε is the random model error (which can be expressed in any form of pdf distribution)
  - The normal distribution is commonly used to describe ε, with $E(\varepsilon) = 0$ and finite σ^2.
- An example of such a correlation would be the dependency between a crack's length and two factors: the number of cycles elapsed and the tensile force on the material.

Regression: Least squares method

- Consider a very simple example of linear regression case with dependent variable (*Y*) and one explanatory factor (*X*):

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- Take some data containing n pairs of observations:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

- The parameters are then estimated via MLE of the estimation, or minimizing the **sums of squares**:

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x)^2$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 0, \qquad \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 0$$

Regression analysis

- The resulting **least square point estimates** of the regression parameters are:

$$\boldsymbol{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}} \quad and \quad \boldsymbol{\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

- Where:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad and \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- The variance of Y may also be obtained as a result (assuming that random error is a normal distribution of mean zero):

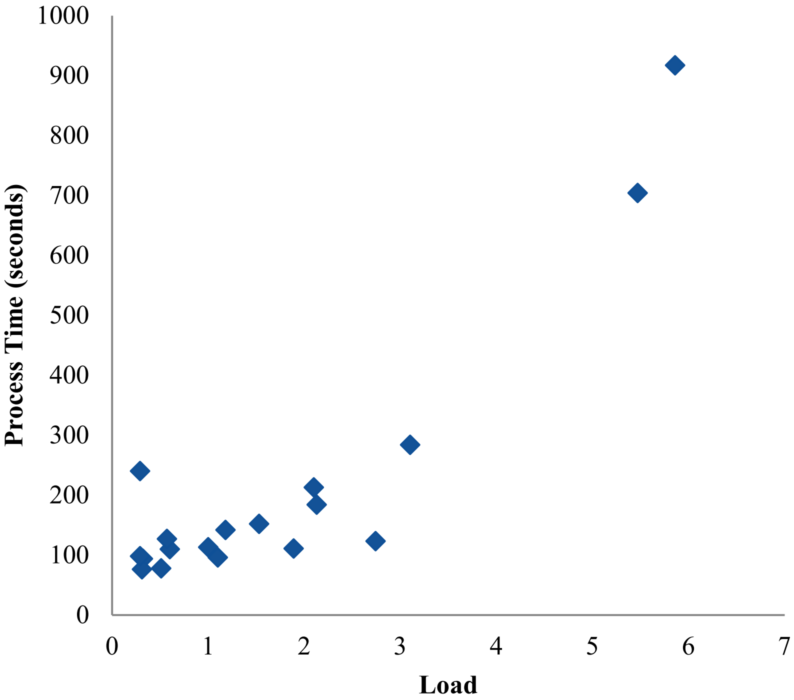
$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} \text{ where } \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Example: Regression analysis

- **Example:** The following table and figure shows how long it takes for a computer to run a certain program as a function of the system load.

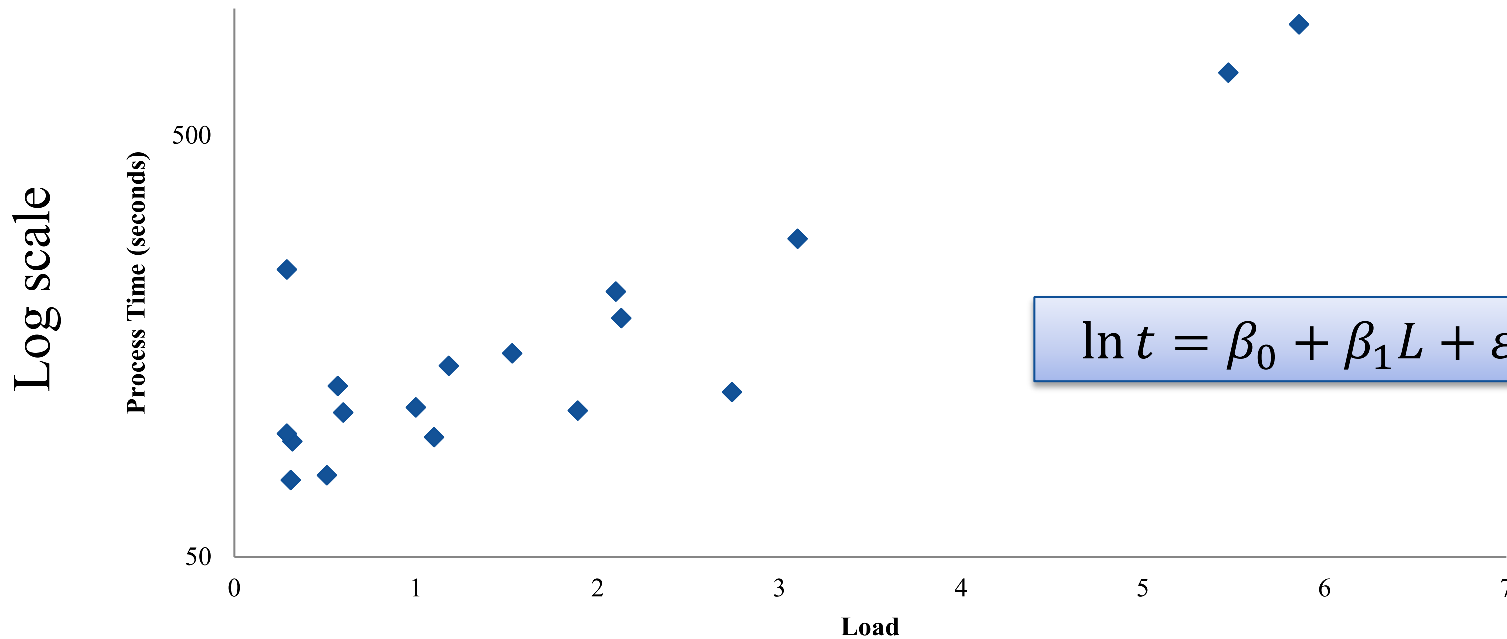
Process time, t_i	Load, L_i	Process time, t_i	Load, L_i
123	2.74	110	0.6
704	5.47	213	2.1
184	2.13	284	3.1
113	1	917	5.86
94	0.32	142	1.18
76	0.31	127	0.57
78	0.51	96	1.1
98	0.29	111	1.89
240	0.29	152	1.53

$n = 18$



Regression analysis

- **Example (cont):** Assume that this can be expressed semilog-linearly:



$$\ln t = \beta_0 + \beta_1 L + \varepsilon$$

- Where L is the system load and t is the processing time in seconds.
- Use linear regression analysis to find the point estimates for β_0, β_1, and σ. Assume the random error follows a normal distribution with a mean of 0.

Regression analysis (cont.)

- **Example (cont.):** Following the equations gets:

$$\ln(t) = \beta_0 + \beta_1 L + \varepsilon \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(L_i - \bar{L})(\ln(t_i) - \ln(\bar{t}))}{\sum_{i=1}^{n}(L_i - \bar{L})^2} = 0.367$$

$$\bar{L} = \frac{1}{n}\sum_{i=1}^{n} L_i = 1.72 \qquad \hat{\beta}_0 = \ln(\bar{t}) - \hat{\beta}_1 \bar{L} = 4.44$$

$$\ln(\bar{t}) = \frac{1}{n}\sum_{i-1}^{n} \ln(t_i) = 5.07$$

$$\ln(\hat{t}_i) = \hat{\beta}_0 + \hat{\beta}_1 L_1 = 4.44 + 0.367 L_i$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(\ln(t_i) - \ln(\hat{t}_i))^2}{n-2} = 0.113, \sigma = 0.336$$

Regression analysis (cont.)

- **Example (cont.):** The final plot