# Chapter 2. Probability and Statistics in PHM

## Prognostics and Health Management (PHM)

**Byeng D. Youn**

System Health & Risk Management Laboratory
Department of Mechanical & Aerospace Engineering
Seoul National University

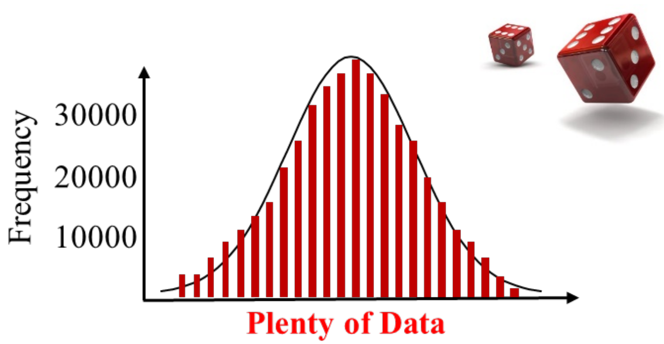
# CONTENTS

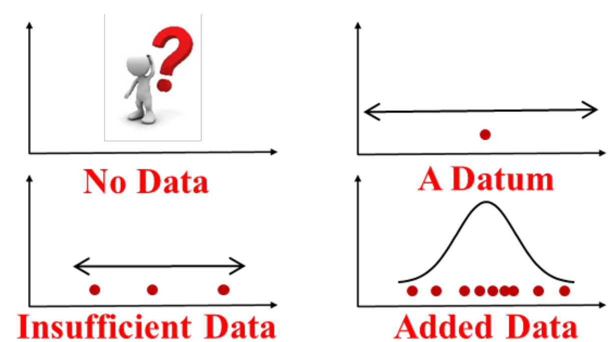
# 1. Uncertainty

Uncertainty Analytics (1)

- Uncertainty Types

| Aleatory Uncertainty | Epistemic Uncertainty |
|---|---|
| - **Inherent randomness** associated with physical systems<br>- Alea: (Latin Word) The rolling of dice<br>- **Irreducible** with the acquisition of additional data<br>- Ex) material properties, product geometry, loading condition, boundary condition, … | - Due to the **lack of knowledge**<br>- **The smaller sample size, the wider confidence interval** in statistical parameter estimation<br>- **Reducible** with additional information<br>- Ex) manufacturing tolerance, material property, expert opinion in case of knowledge absence, .. |

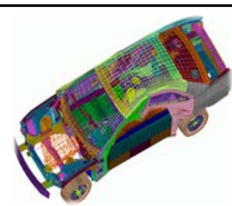


# 1. Uncertainty

Uncertainty Analytics (2)

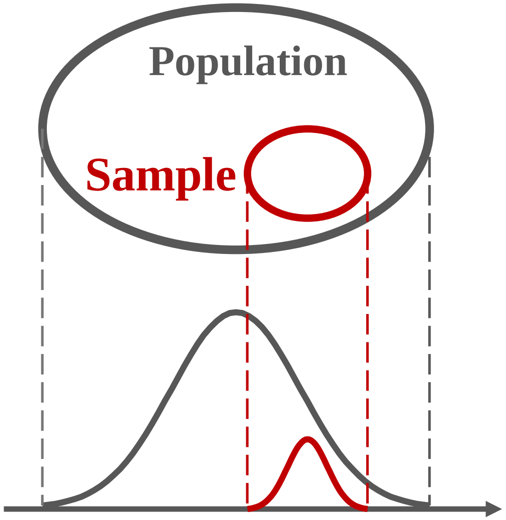
- Uncertainty Sources

| Uncertainty Sources | Meaning |
| --- | --- |
| **Physical Uncertainty**  | - **Inherent variation** in physical quantity<br>- Description by probability distribution<br>- Ex) material property, manufacturing tolerance, loading condition, boundary condition, … |
| **Statistical Uncertainty**   | - **Imprecise statistical estimation** (probability distribution type, parameters, …)<br>- Only depending on the sample size and location<br>- Ex) lack of data, improper sampling |
| **Modeling Uncertainty**  | - Uncertainty from **invalid modeling**<br>- Ex) improper approximation, inaccurate boundary condition, … |

# 2. Characterization of Uncertainty

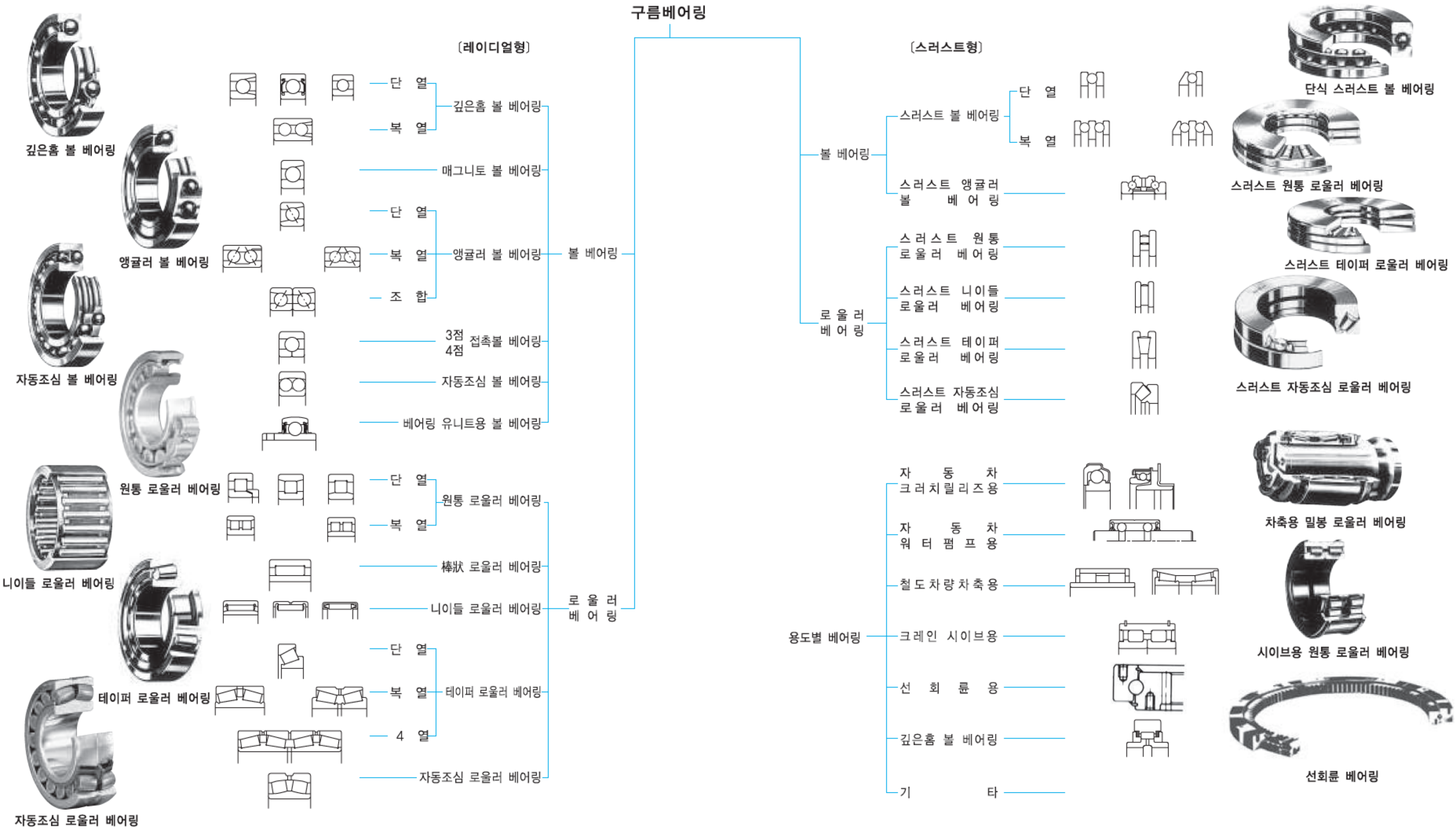
Data Collection

- Population
  - All observation of a random variable
  - Impossible to collect all population
  - Representative sample is collected instead
- Sample
  - Gather information on population
  - A relatively large sample size is always preferable




Data Classification

| Continuous | Discrete |
|---|---|
| - The values can take on any value within a finite or infinite interval | - The values belong to the set are distinct and separate |
| - Measured | - Counted |
| - ex. Failure time | - ex. Number of defective specimen |
|  |  |


구름베어링
〔레이디얼형〕
〔스러스트형〕
깊은홈 볼 베어링
앵귤러 볼 베어링
자동조심 볼 베어링
원통 로울러 베어링
니이들 로울러 베어링
테이퍼 로울러 베어링
자동조심 로울러 베어링
단 열
복 열
깊은홈 볼 베어링
매그니토 볼 베어링
단 열
복 열
앵귤러 볼 베어링
조 합
3점
4점
접촉볼 베어링
자동조심 볼 베어링
베어링 유니트용 볼 베어링
볼 베어링
단 열
복 열
원통 로울러 베어링
棒狀 로울러 베어링
니이들 로울러 베어링
로 울 러
베 어 링
단 열
복 열
테이퍼 로울러 베어링
4 열
자동조심 로울러 베어링
볼 베어링
스러스트 볼 베어링
단 열
복 열
스러스트 앵귤러
볼 베 어 링
로 울 러
베 어 링
스 러 스 트 원 통
로 울 러 베 어 링
스러스트 니이들
로울러 베어링
스러스트 테이퍼
로울러 베어링
스러스트 자동조심
로 울 러 베 어 링
용도별 베어링
자 동 차
크러치릴리즈용
자 동 차
워 터 펌 프 용
철도차량차축용
크레인 시이브용
선 회 륜 용
깊은홈 볼 베어링
기 타
단식 스러스트 볼 베어링
스러스트 원통 로울러 베어링
스러스트 테이퍼 로울러 베어링
스러스트 자동조심 로울러 베어링
차축용 밀봉 로울러 베어링
시이브용 원통 로울러 베어링
선회륜 베어링

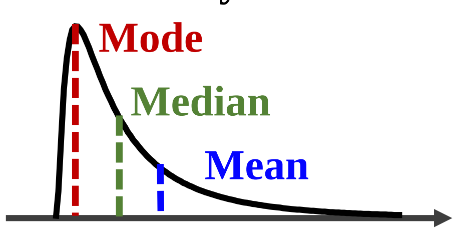
# 2. Characterization of Uncertainty

Characterizing Descriptors

- Measure of Central Tendency

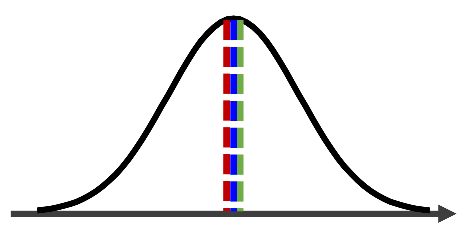
| Mean | Median | Mode |
|---|---|---|
| – Location of centroid<br>– First moment ($\boldsymbol{\mu}$ = E[**X**]) | – Middle value of a set of data | – The most frequentist value in a data set |
| Ex. Data set {1 2 3 4 5 5}<br>$\rightarrow \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1+2+3+4+5+5}{6}$<br>$= \frac{10}{3}$ | Ex. Data set {1 2 3 4 5 5}<br>$\rightarrow \frac{3+4}{2} = 3.5$ | Ex. Data set {1 2 3 4 5 5}<br>$\rightarrow$ 5 |

**< Positively Skewed >**



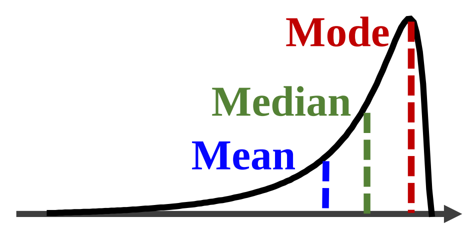

**Mean ≥ Median ≥ Mode**

**< Symmetric >**



**Mean = Median = Mode**

**< Negatively skewed >**



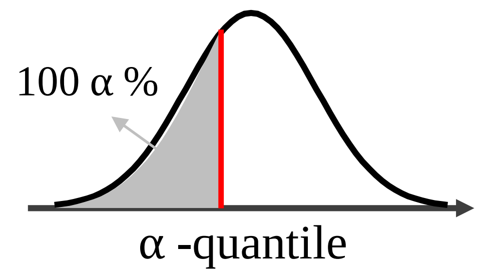

**Mean ≤ Median ≤ Mode**

# 2. Characterization of Uncertainty

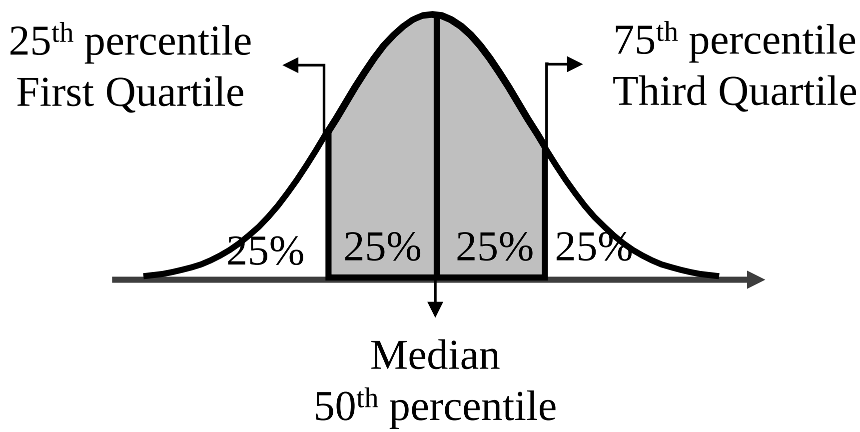
Quantifying Descriptors

- Quantile & Percentile
  - The $x$ value at which the CDF takes a value α is called the α -quantile for 100α -percentile.

$$F(x) = P(x \le x_\alpha) = \alpha$$




- Quartile
  - Data are grouped into four equal parts. Each quartile includes 25% of the data



# 2. Characterization of Uncertainty

Quantifying Descriptors

- Measure of Dispersion

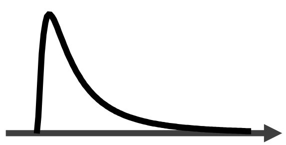
| Variance | Standard Deviation | Coefficient of Variation |
|---|---|---|
| – Mean of squared deviation $$\mathrm{Var}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_{x_i})^2$$ – The second Moment $E[(\mathbf{X}\text{-}\boldsymbol{\mu})^2] = E(\mathbf{X}^2) - E(\mathbf{X})^2$ | – Degree of dispersion of same unit with data $$\boldsymbol{\sigma}_{\mathbf{X}} = \sqrt{\mathrm{Var}(\mathbf{X})}$$ | – Non-dimensional term of standard deviation – COV(X)=0 : deterministic variable $$\boldsymbol{\delta}_{\mathbf{X}} = \frac{\boldsymbol{\sigma}_{\mathbf{X}}}{\boldsymbol{\mu}_{\mathbf{X}}}$$ |
| Ex. Data set {1 2 3 4 5 5} → $\frac{1}{6}\{(3.333 - 1)^2 + \cdots +$ | Ex. Data set {1 2 3 4 5 5} → $\sqrt{2.6667} = 1.633$ | Ex. Data set {1 2 3 4 5 5} → $\frac{1.633}{3.333}$ =0.4899 |

# 2. Characterization of Uncertainty

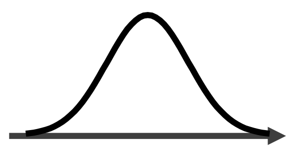
Quantifying Descriptors

- Measure of Asymmetry
  - Skewness = $\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu_{x_i}\right)^{\mathbf{3}}$
  - The third moment (used as a **PHM feature**)
  - Skewness Coefficient = $\boldsymbol{\theta}_{\mathbf{X}} = \frac{\mathbf{skewness}}{\boldsymbol{\sigma}_{\mathbf{X}}^{\mathbf{3}}}$

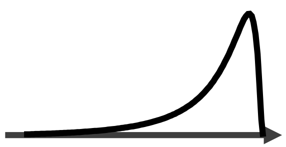
**< Positively Skewed >** **< Symmetric >** **< Negatively skewed >**



$\boldsymbol{\theta}_{\mathbf{X}} > 0$

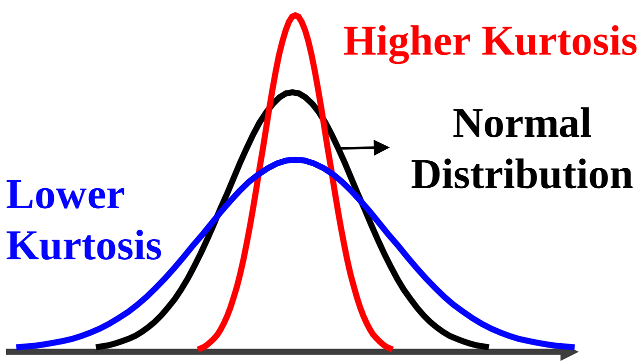


$\boldsymbol{\theta}_{\mathbf{X}} = 0$



$\boldsymbol{\theta}_{\mathbf{X}} < 0$

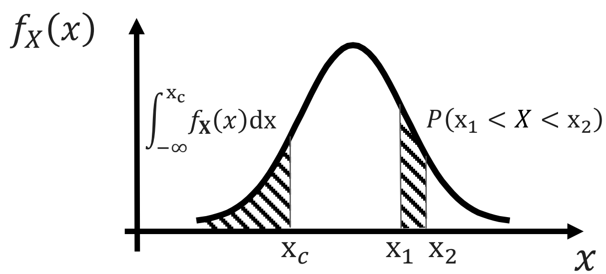
- Measure of Flatness (Peakedness)
  - Kurtosis = $\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu_{x_i}\right)^{\mathbf{4}}$
  - Kurtosis Coefficient = $\frac{\boldsymbol{Kurtosis}}{\boldsymbol{\sigma}_{\mathbf{X}}^{\mathbf{4}}}$
  - The fourth moment
  - Used as a PHM feature

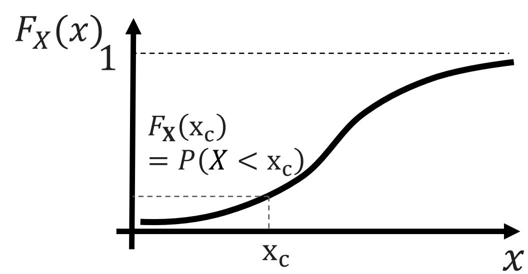


# 2. Characterization of Uncertainty

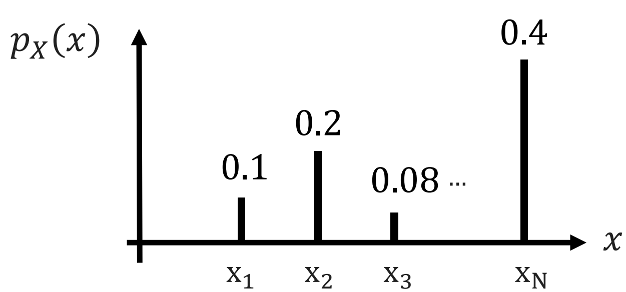
Probability Distribution

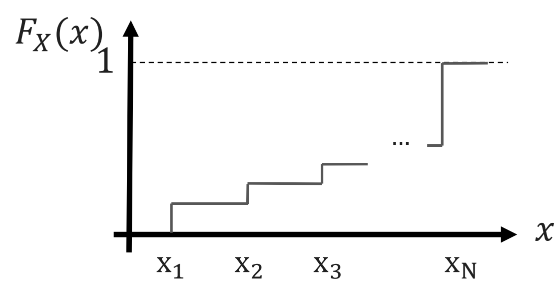
- Continuous
  - Probability Density Function (PDF) : $P(\mathrm{x}_1 < X < \mathrm{x}_2) = \int_{\mathrm{x}_1}^{\mathrm{x}_2} f_X(x)\mathrm{dx}$
  - Cumulative Density Function (CDF) : $P(X < \mathrm{x}) = F_{\mathbf{X}}(x) = \int_{-\infty}^{\mathrm{x}} f_X(x)\mathrm{dx}$
  - $\frac{P(\mathrm{x}<X<\mathrm{x}+\Delta\mathrm{x})}{\Delta\mathrm{x}} = \frac{F_X(x+\Delta x)-F_X(x)}{\Delta\mathrm{x}} = \frac{dF_X(x)}{\mathrm{dx}} = \frac{f_X(x)\Delta\mathrm{x}}{\Delta\mathrm{x}} = f_X(x)$




- Discrete
  - Probability Mass Function (PMF) :$P(X = \mathrm{x}) = p_X(x)$
  - Cumulative Density Function (CDF) : $P(X < \mathrm{x}) = \sum_{x_i<\mathrm{x}} p_X(x_i)$

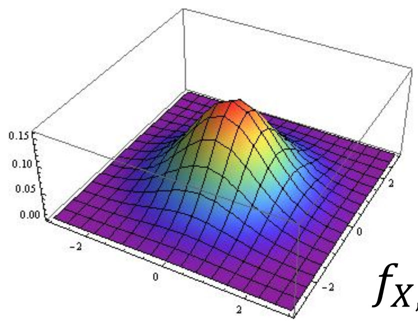
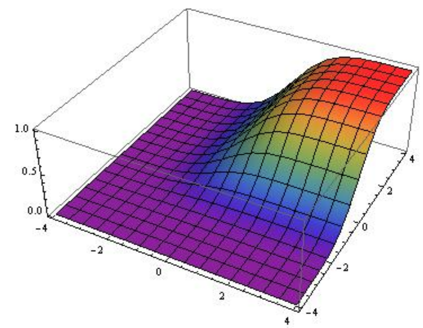
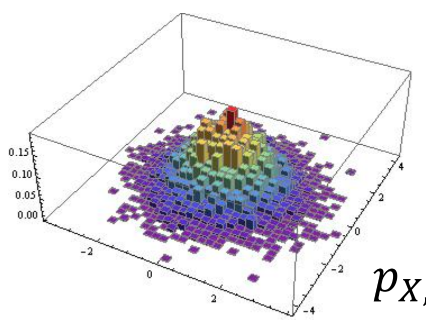
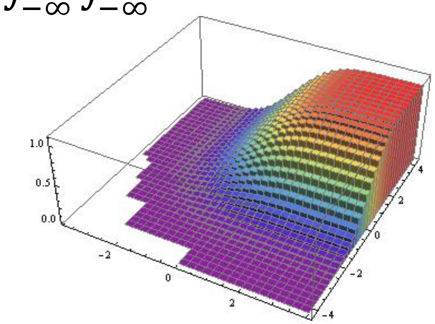


# 2. Characterization of Uncertainty

Multivariate Distribution

- Joint Distribution

| | Probability Density Function | Cumulative Density Function |
|---|---|---|
| Continuous |  $f_{X,Y}(x,y)$ |  |
| Discrete |  $p_{X,Y}(x_i,y_i)$ |  |

$$F_{X,Y}(x,y) = P(X \le \mathrm{x}, Y \le \mathrm{y}) = \int_{-\infty}^{\mathrm{x}} \int_{-\infty}^{\mathrm{y}} f_{X,Y}(u,v)\mathrm{dudv}$$

$$F_{X,Y}(x,y) = \sum_{\mathrm{x}_i \le \mathrm{x}} \sum_{\mathrm{y}_i \le \mathrm{y}} p_{X,Y}(x_i,y_i)$$

# 2. Characterization of Uncertainty

Multivariate Distribution

- Marginal PDF & PMF

The marginal distribution of a random variable X is obtained from the joint probability distribution of two random variables X and T by summing or integrating over the values of the random variable Y.

– Marginal PDF

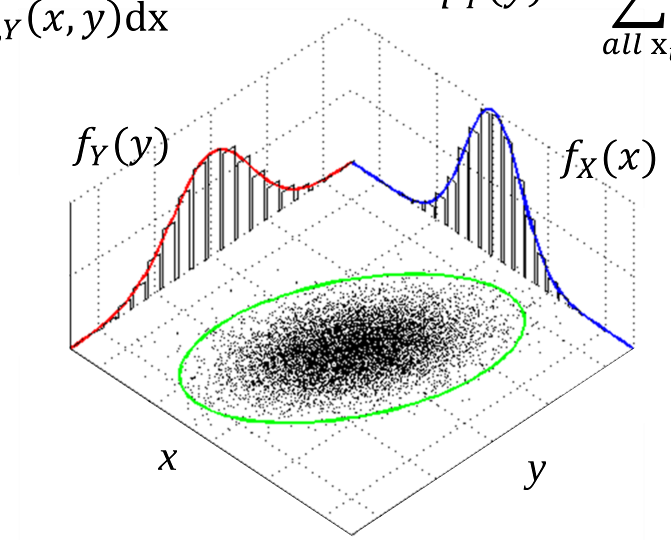
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{dy}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{dx}$$

– Marginal PMF

$$p_X(x) = \sum_{all\ \mathrm{y}_i} p_{X,Y}(x, y_i)\mathrm{dy}$$

$$p_Y(y) = \sum_{all\ \mathrm{x}_i} p_{X,Y}(x_i, y)\mathrm{dy}$$



# 2. Characterization of Uncertainty

Covariance and Correlation

- Covariance
  - Degree of linear relationship between two random variables

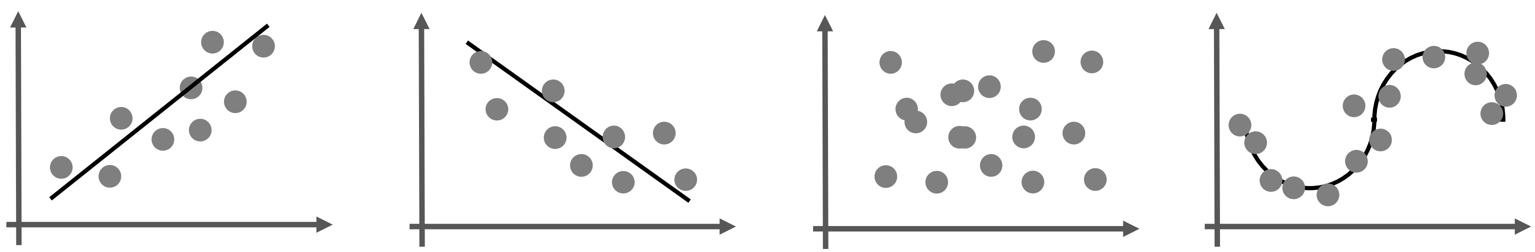
$$Cov(X,Y) = \mathrm{E}\left[(X-\mu_x)\left(Y-\mu_y\right)\right] = \mathrm{E}(XY) - \mathrm{E}(X)E(Y)$$

  - Cov$(X,Y)$=0 when (linearly) independent

- Correlation Coefficient
  - Nondimensionalizing the covariance

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

  - Range between -1 and +1

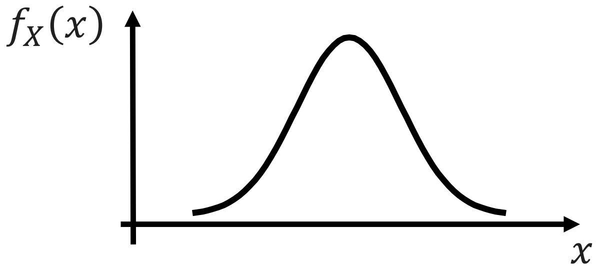


$0 < \rho_{X,Y} < 1$ $\quad -1 < \rho_{X,Y} < 0$ $\quad \rho_{X,Y} = 0$ $\quad \rho_{X,Y} = 0$

# 3. Types of Probability Distribution

Normal (Gaussian) Distribution

- Probability Density Function




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

where μ : Mean of $X$
σ : Standard deviation of $X$

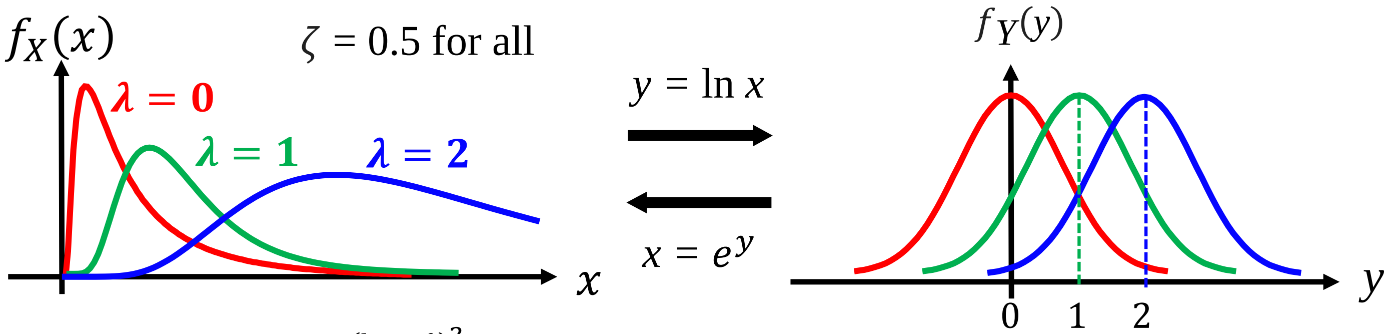
  – The PDF of normal distribution is symmetry
  – Two parameters which are mean (μ) and standard deviation (σ) defines the PDF

- Properties ($\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$, $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$)
  – $Z = \alpha\mathbf{X} + \beta \rightarrow Z \sim N(\alpha\boldsymbol{\mu} + \beta, \alpha^2\boldsymbol{\sigma}^2)$
  – $Z = X_1 + X_2 \rightarrow Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2Cov(X_1, X_2))$
  – $Z = \frac{\mathbf{X}-\boldsymbol{\mu}}{\boldsymbol{\sigma}} \rightarrow Z \sim N(0,1)$
  – $Z \sim N(0,1)$, Z is called standard Gaussian and $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

# 3. Types of Probability Distribution

Lognormal Distribution

- Probability Density Function




  – $f_X(x) = \frac{1}{\zeta\sqrt{2\pi}} e^{-\frac{(\ln x-\lambda)^2}{2\zeta}}$, where $\lambda$: Mean of ln $X$ , $\zeta$: Variance of ln$X$
  – Closely related to the normal distribution
  – Defined for positive values only

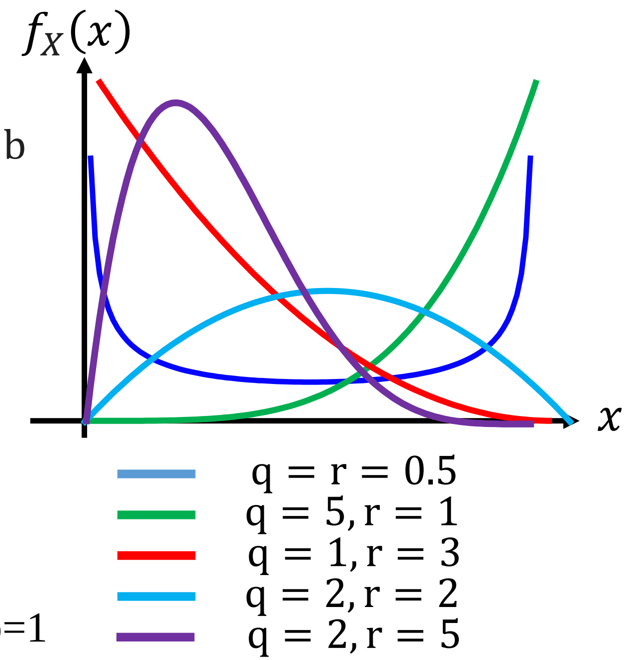
- Properties
  – λ: Mean of ln$X$, ζ: Variance of ln $X$
  – μ: Mean of $X$, σ: Standard deviation of $X$ , $\delta = \sigma/\mu$ : Coefficient of Variation of $X$
  – $\zeta = \sqrt{\ln(1+\delta^2)}$, $\lambda = \ln \mu - 0.5\ln(1+\delta^2)$
  – $\mu = \exp(\lambda + 0.5\zeta^2)$, $\delta = \sigma/\mu = \sqrt{\exp(\zeta^2) - 1}$

# 3. Types of Probability Distribution

Beta Distribution

- Probability Density Function

$$f_X(x) = \frac{1}{B(\mathrm{q},\mathrm{r})}\frac{(x-\mathrm{a})^{\mathrm{q}-1}(\mathrm{b}-x)^{\mathrm{r}-1}}{(\mathrm{b}-\mathrm{a})^{\mathrm{q}+\mathrm{r}-1}}, \mathrm{a} < x < \mathrm{b}$$




- Properties
  - Variable is bounded → a < $x$ < b
  - Parameter relationship

$$\mathrm{E}(X) = \mathrm{a} + \frac{\mathrm{q}}{\mathrm{q}+\mathrm{r}}(\mathrm{b}-\mathrm{a})$$

$$\mathrm{Var}(X) = \frac{\mathrm{qr}}{(\mathrm{q}+\mathrm{r})^2(\mathrm{q}+\mathrm{r}+1)}(\mathrm{b}-\mathrm{a})^2$$

  - Standard beta distribution → substitute a=0, b=1

$$f_X(x) = \frac{1}{B(\mathrm{q},\mathrm{r})}x^{\mathrm{q}-1}(1-x)^{\mathrm{r}-1}$$

$$\mathrm{q}, \mathrm{r} = \text{shape parameter}$$

  - Beta function

$$\mathrm{B}(\mathrm{q},\mathrm{r}) = \int_0^1 x^{\mathrm{q}-1}(1-x)^{\mathrm{r}-1}\mathrm{dx} = \frac{\Gamma(\mathrm{q})\Gamma(\mathrm{r})}{\Gamma(\mathrm{q}+\mathrm{r})}$$

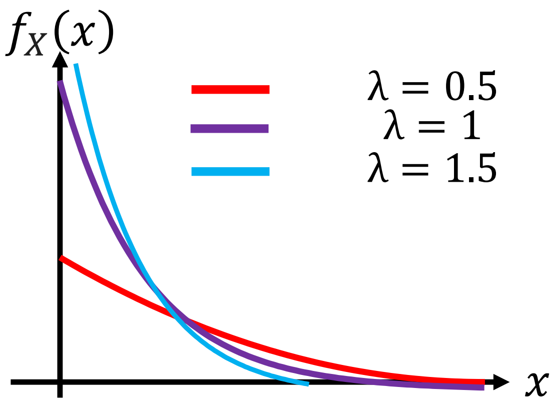
System Health & Risk Management

# 3. Types of Probability Distribution

Exponential Distribution

- Probability Density Function

$$f_X(x) = \lambda e^{-\lambda x}$$




- Properties
  - Parameter and Variable is bounded → $0 < \lambda$, $0 < x$
  - Parameter relationship

$$\mathrm{E}(X) = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

  - Memorylessness

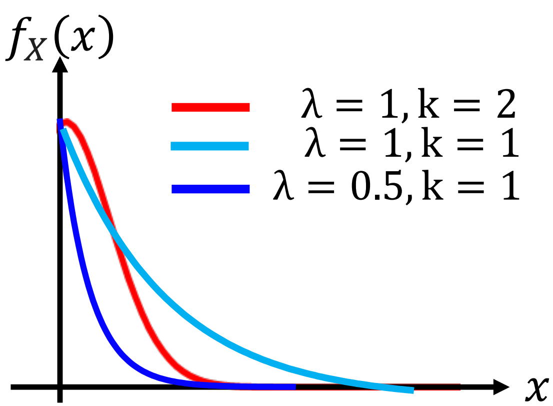
$$\mathrm{P}(X > s + t | X > s) = \mathrm{P}(X > t)$$

Weibull Distribution

- Probability Density Function

$$f_X(x) = \frac{\mathrm{k}}{\lambda}\left(\frac{x}{\lambda}\right)^{\mathrm{k}-1} e^{-\left(\frac{x}{\lambda}\right)^{\mathrm{k}}}$$

  - Generalization of the exponential distribution




- Properties
  - Widely used to model the TTF distribution

$$MTTF = \frac{\Gamma(1 + 1/\mathrm{k})}{\lambda}$$

# 4. Estimation of Parameter and Distribution

Parameter (**θ**)

- A property of an unknown probability distribution.
- For example, mean, variance, or a particular quantile
- One of the goals of statistical inference is to estimate them.
- Examples
  - Mean of Normal Distribution: $\mu$
  - Standard Deviation of Normal Distribution: $\sigma^2$

Statistics

- To denote a quantity that is a property of a sample.
- For example, sample mean, a sample variance, or a particular sample quantile.
- Statistics can be used to estimate unknown parameters.
- Examples
  - Sample mean : $\bar{x} = \frac{x_1 + \cdots\cdots + x_n}{n}$
  - Sample variance : $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

# 4. Estimation of Parameter and Distribution

Point Estimation

- Minimum Mean Square Error Estimation
- Maximum Likelihood Estimation
- Probability Distribution Estimation
  - Method of Moments
  - Goodness of Fit (Chi-square, K-S test)

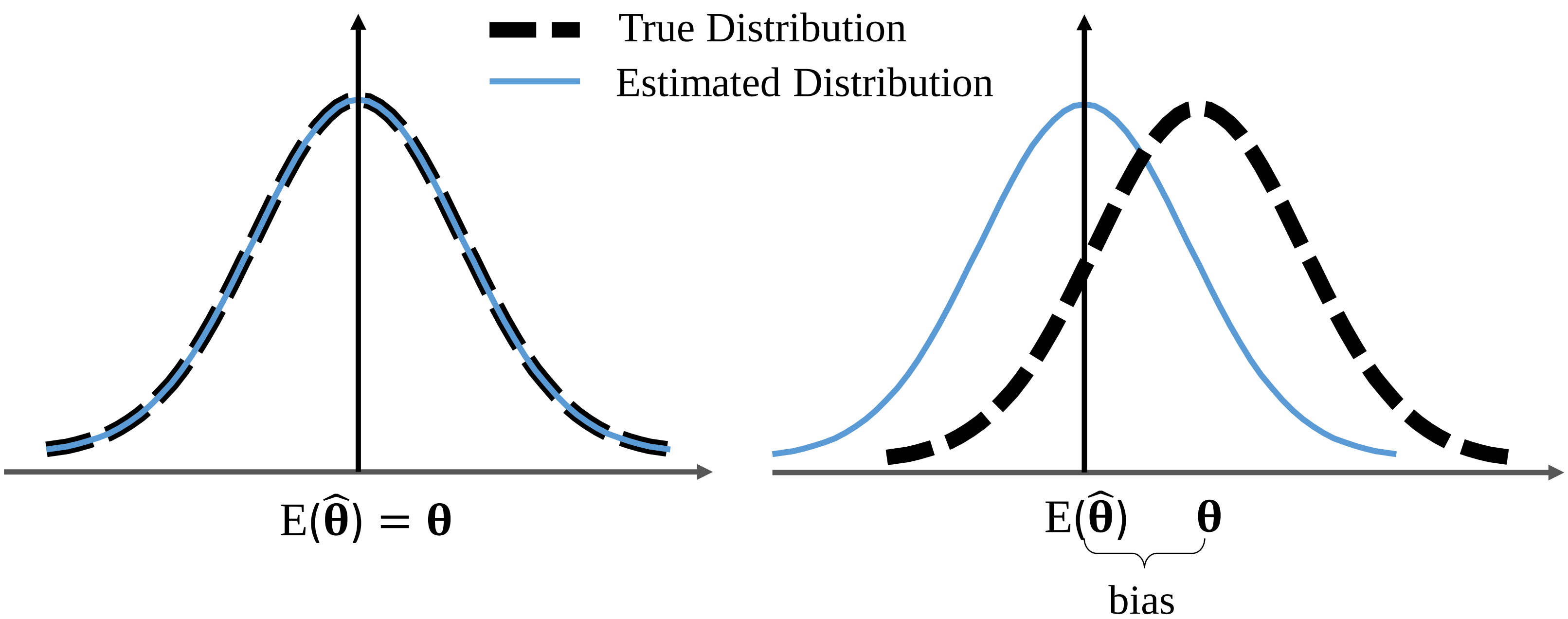
Interval Estimation

Hypothesis Testing

# 4. Estimation of Parameter and Distribution

Point Estimation of Parameters

- Unbiased and Biased Point Estimates
  - $\boldsymbol{\theta}$ : statistical parameter (fixed constant)
  - $\widehat{\boldsymbol{\theta}}$ : a statistics which serves as an estimator of $\boldsymbol{\theta}$
  - Unbiased if $E(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$
  - Not unbiased, bias$= E(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\theta}$
  - To make $E(\widehat{\boldsymbol{\theta}})$ with $\boldsymbol{\theta}$ consistent for eliminating the bias which expresses systematic error.



# 4. Estimation of Parameter and Distribution

- Point Estimate of a Population Mean
  - If $X_1, \ldots, X_n$ is a sample of observations from a probability distribution with a mean μ, then the sample mean $\hat{\mu}=\overline{X}$ is an unbiased point estimate of the population mean μ.
- Point Estimate of a Population Variance
  - If $X_1, \ldots, X_n$ is a sample of observations from a probability distribution with a variance $\sigma^2$, then the sample variance $\widehat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$ is an unbiased point estimate of the population variance $\sigma^2$.

Proof.

$$
\begin{aligned}
E(S^2) &= \frac{1}{n-1} E\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right) \\
&= \frac{1}{n-1} E\left(\sum_{i=1}^{n}(x_i-\mu)-(\bar{x}-\mu)\right)^2) \\
&= \frac{1}{n-1} E\left(\sum_{i=1}^{n}(x_i-\mu)^2 - 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu) + n(\bar{x}-\mu)^2\right) \\
&= \frac{1}{n-1} E\left(\sum_{i=1}^{n}(x_i-\mu)^2 - n(\bar{x}-\mu)^2\right) \\
&= \frac{1}{n-1} E\left(\sum_{i=1}^{n}(x_i-\mu)^2\right) - n\,E((\bar{x}-\mu)^2) \\
&= \frac{1}{n-1}\left(n\sigma^2 - n\left(\frac{\sigma^2}{n}\right)\right) = \sigma^2
\end{aligned}
$$

# 4. Estimation of Parameter and Distribution

Minimum Mean Square Error Estimation (MMSE)

- The average of the square of the errors between the estimator and what is estimated.
- Because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate.

$$
\begin{aligned}
\mathrm{MSE}(\hat{\boldsymbol{\theta}}) &= E((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2) \\
&= E[(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}) + E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})]^2 \\
&= E[(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))^2 + 2(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))(E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}) + (E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})^2] \\
&= E[(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))^2] + 2(E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})E(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}})) + (E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})^2] \\
&= E((\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))^2 + (E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})^2 \\
&= \mathrm{Var}(\hat{\boldsymbol{\theta}}) + \mathrm{bias}^2
\end{aligned}
$$

- Example

When $\widehat{\theta_1} \sim N(1.2\theta, 0.02\theta^2)$, $\widehat{\theta_2} \sim N(0.9\theta, 0.04\theta^2)$, Mean Square Error of each estimator

$\rightarrow \mathrm{Var}(\widehat{\theta_1}) < \mathrm{Var}(\widehat{\theta_2})$

$\mathrm{bias}_1 = \mathrm{E}(\widehat{\theta_1}) - \theta = 0.2\,\theta$, $\mathrm{bias}_2 = \mathrm{E}(\widehat{\theta_2}) - \theta = -\,0.1\,\theta$

$\mathrm{MSE}(\widehat{\theta_1}) = \mathrm{Var}(\widehat{\theta_1}) + (\mathrm{bias}_1)^2 = 0.06\theta^2$, $\mathrm{MSE}(\widehat{\theta_2}) = \mathrm{Var}(\widehat{\theta_2}) + (\mathrm{bias}_2)^2 = 0.05\theta^2$

$\rightarrow \mathrm{MSE}(\widehat{\theta_1}) > \mathrm{MSE}(\widehat{\theta_2})$

# 4. Estimation of Parameter and Distribution

Maximum Likelihood Estimation (MLE)

- Likelihood function can be defined as

$$L(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n} | \boldsymbol{\theta}) = f_X(\mathrm{x}_1 | \boldsymbol{\theta}) f_X(\mathrm{x}_2 | \boldsymbol{\theta}) \ldots, f_X(\mathrm{x}_n | \boldsymbol{\theta}).$$

$f_X(x_i|\boldsymbol{\theta})$ = The PDF values of random variable $X$ at $x_i$
when statistical parameter is gievn as $\boldsymbol{\theta}$

- Maximum Likelihood Estimation (MLE) is to find the $\boldsymbol{\theta}^*$ that maximizes the likelihood function, so the following equation is satisfied.

$$\frac{\partial L}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = 0$$

Ex) There's a big box with some black and white balls. But we don't know the number of the balls. Someone picked up a ball 10 times by sampling with replacement. And he picked up black balls 1 time, and white balls 9 times. Likelihood of black ball $p$?

Sol.) $L = p(1-p)^9$

$$\frac{\partial L}{\partial p} = (1-p)^9 - 9p(1-p)^8 = 0 \qquad \therefore p = \frac{1}{10}$$

# 4. Estimation of Parameter and Distribution

Method of Moments

- Basic concept : All the parameters of a distribution can be estimated using the information on its moment.
- Parameters of a distribution have a definite relation with the moments of the random variable.

| Distribution | Relation to mean and variance | Inverse Relation |
|---|---|---|
| Normal | $\mathrm{E}(X) = \mu_x \,, \mathrm{Var}(X) = \sigma_x^2$ | $\mu_X = \mathrm{E}(X), \sigma_X = \sqrt{\mathrm{Var}(X)}$ |
| Lognormal | $\mathrm{E}(X) = \exp(\lambda + \frac{1}{2}\zeta^2)$<br>$\mathrm{Var}(X) = \mathrm{E}^2(X)[e^{\zeta^2} - 1]$ | $\lambda = \ln \mathrm{E}(X) - 0.5 \ln(1 + \delta^2)$<br>$\zeta = \sqrt{\ln(1 + \delta^2)}$<br>$\delta = \sqrt{\mathrm{Var}(X)}/\mathrm{E}(X)$ |
| Weibull | $\mathrm{E}(X) = \lambda\, \Gamma(1 + \frac{1}{\mathrm{k}})$<br>$\mathrm{Var}(X) = \lambda^2 \left[ \Gamma\left(1 + \frac{2}{\mathrm{k}}\right) - \left(\Gamma\left(1 + \frac{1}{\mathrm{k}}\right)\right)^2 \right]$ | $\lambda = \frac{\mathrm{E}(X)}{\Gamma(1 + \frac{1}{\mathrm{k}})}, \frac{\Gamma\left(1 + \frac{2}{\mathrm{k}}\right)}{\left(\Gamma\left(1 + \frac{1}{\mathrm{k}}\right)\right)^2}$<br>$= \frac{\mathrm{Var}(X)}{\mathrm{E}(X)^2} - 1 \rightarrow$ implicit<br>Approximation : $\lambda = \frac{\mathrm{E}(X)}{\Gamma(1+\frac{1}{\mathrm{k}})},$<br>$\mathrm{k} = \left(\sqrt{\mathrm{Var}(X)}/\mathrm{E}(X)\right)^{-1.086}$ |

# 4. Estimation of Parameter and Distribution

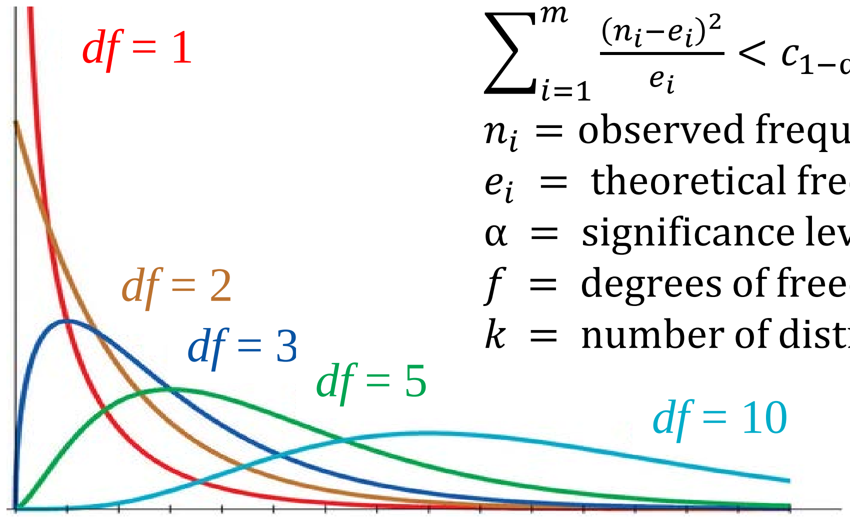
| Distribution | Relation to mean and variance | Inverse Relation |
| --- | --- | --- |
| Rayleigh | $\mathrm{E}(X) = \sqrt{\frac{\pi}{2}}\alpha$<br>$\mathrm{Var}(X) = (2 - \frac{\pi}{2})\alpha^2$ | $\alpha = \sqrt{\frac{2}{\pi}\mathrm{E}(X)}$ *or* $\alpha = \sqrt{\frac{2\,\mathrm{Var}(X)}{4 - \pi}}$ |
| Exponential | $\mathrm{E}(X) = \frac{1}{\mathrm{v}}$, $\mathrm{Var}(X) = \frac{1}{v^2}$ | $v = \frac{1}{\mathrm{E}(X)}$, $v = \frac{1}{\sqrt{\mathrm{Var}(X)}}$ |

# 4. Estimation of Parameter and Distribution

Goodness of Fit

- Quantitative method
- Based on the error between observed data and an assumed PDF
- Assume a distribution will be acceptable if an error between the observed data and the assumed PDF is less than a critical value.
- Examples
  - Chi-Square test
  - Kolmogorov-Smirnov (KS) test

Chi-Square Test




$$\sum_{i=1}^{m} \frac{(n_i - e_i)^2}{e_i} < c_{1-\alpha,f}\text{ , where}$$

$n_i$ = observed frequency at $i^{th}$ interval

$e_i$ = theoretical frequency of an assumed distribution

$\alpha$ = significance level

$f$ = degrees of freedom $(= m - 1 - k)$

$k$ = number of distribution parameter

# 4. Estimation of Parameter and Distribution

Kolmogorov-Smirnov (KS) Test

$D_n = \max|F_X(x_i) - S_n(x_i)|$

$D_n$: Maximum difference between CDFs

$F_X(x_i)$: CDF of the theoretical CDF

$S_n(x_i)$: CDF of the observed data

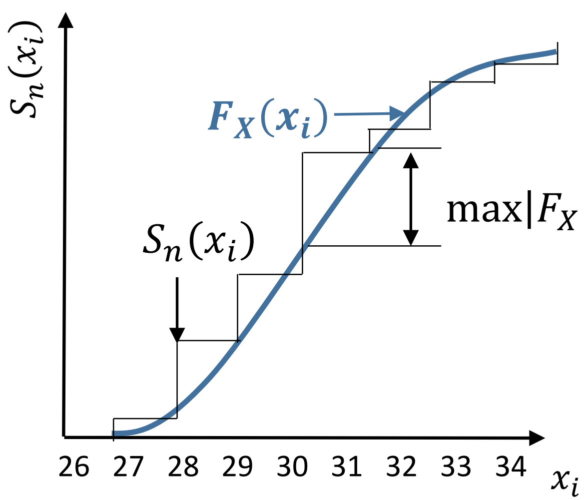
$P(D_n \leq D_n^\alpha) = 1 - \alpha$

$D_n^\alpha$: Values at significance level $\alpha$

$$S_n(x_i) = \begin{cases} 0 & x < x_1 \\ m/n & x_m \leq x \leq x_{m+1} \\ 1 & x \geq x_n \end{cases}$$

$S_n(x_i)$: Emperical CDF of the observed data

$n$: Sample size



# 5. Bayesian

Approaches to determine a probability

- **Frequentist's approach**
  - Postulate its probability based on the number of times the event occurs in a large number of samples

$$P(A) = \lim_{n \to \infty} \frac{k}{n}$$

- **Bayesian approach**
  - Employs a degree-of-belief, which is subjective information (e.g. previous experience, expert's opinion, data from handbook)
  - Express in the form of probability density function (PDF) and observations are used to change or update the PDF

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \,(P(B) \neq 0)$$

  - $P(A)$ is initial degree-of-belief in event $A$ or called the *prior.*
  - $P(A|B)$ is the degree-of-belief after accounting for evidence $B$ or called the *posterior.*
  - The Bayes' theorem is modifying or updating the prior probability $P(A)$ to the posterior probability $P(A|B)$ after accounting for evidence.

# 5. Bayesian

Bayesian Theorem in Probability Density Form (PDF)

- $f_X$ be a PDF of uncertainty variable X and test measure a value Y, random variable, whose PDF denoted by $f_Y$

$$f_{XY}(x,y) = f_X(x|Y=y)f_Y(y) = f_Y(y|X=x)f_X(x)$$

- **Ex**. Fatigue life of *X* has epistemic uncertainty in the form of $f_X$. After measuring a fatigue life *y* of a specimen, our knowledge on fatigue life of *X* can be changed to $f_X(x|Y=y)$.

$$f_X(x|Y=y) = \frac{f_Y(y|X=x)f_X(x)}{f_Y(y)} \quad \left(f_Y(y) = \int_{-\infty}^{\infty} f_Y(y|X=\varepsilon)f_X(\varepsilon)d\varepsilon\right)$$

- Analytical calculation is possible when prior distribution is as $f_X(x) = N(\mu_0, \sigma_0^2)$ and likelihood is normal distribution as $f_Y(y|X=x) = N(y, \sigma_y^2)$.

$$f_X(x|Y=y) = \frac{f_Y(y|X=x)f_X(x)}{f_Y(y)} \sim \exp\left[-\frac{(y-x)^2}{2\sigma_y^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right]$$

System Health & Risk Management

# 5. Bayesian

**Example 3.1**

There are three doors and behind two of the doors are goats and behind the third door is a new car with each door equally likely to provide the car. Thus the probability of selecting the car for each door at the beginning of the game is simply 1/3. After you have picked a door, say A, before showing you what is behind that door, Monty opens another door, say B, revealing a goat. At this point, Monty gives you the opportunity to switch doors from A to C if you want to. What should you do? (Given that Monty is trying to let you get a goat.)

**Solution**

The question is whether the probability is 0.5 to get the car since only two doors left, or mathematically, $P(A|B_{Monty}) = P(C|B_{Monty}) = 0.5$. Basically we need to determine the probabilities of two event $E_1 = \{A|B_{Monty}\}$, $E_2 = \{C|B_{Monty}\}$. We elaborate the computation in the following steps:

**1.** The prior probabilities read $P(A) = P(B) = P(C) = 1/3$.

**2.** We also have some useful conditional probabilities $P(B_{Monty}|A) = ½$, $P(B_{Monty}|B) = 0$, and $P(B_{Monty}|C) = 1$.

**3.** We can compute the probabilities of joint events as $P(B_{Monty}, A) = ½\times 1/3 = 1/6$, $P(B_{Monty}, B) = 0$, and $P(B_{Monty}, C) = 1\times 1/3 = 1/3$.

**4.** Finally, with the denominator computed as $P(B_{Monty}) = 1/6 + 0 + 1/3 = ½$, we then get $P(A|B_{Monty}) = 1/3$, $P(C|B_{Monty}) = 2/3$. Thus, it is better to switch to C.

# 5. Bayesian

Bayesian Updating

- **Overall Bayesian Update**

$$f_X(x|Y=y) = \frac{1}{K}\prod_{i=1}^{N}[f_Y(y_i|X=x)]f_X(x)$$

  – Likelihood functions of individual tests are multiplied together to build the total likelihood function.

  – $K$ is a normalizing constant.

- **Recursive Bayesian Update**

$$f_X^{(i)}(x|Y=y_i) = \frac{1}{K_i}f_Y(y_i|X=x)f_X^{(i-1)}(x), \qquad i=1,\cdots,N$$

  – $K_i$ is a normalizing constant at $i$-th update and $f_X^{(i-1)}(x)$is the PDF of X, updated using up to $(i-1)$th tests.

# 5. Bayesian

## Bayesian Parameter Estimation

- Bayes theorem's main purpose is parameter estimation and calibration of model parameters.
- Vector of unknown model parameters is denoted as **θ**, while the vector of measured data is denoted as **y**.

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{\theta})}{f(\mathbf{y})}$$

- Denominator in the above equation is independent of unknown parameters and a normalizing constant to make the one.

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{\theta})$$

  – $f(\mathbf{y}|\boldsymbol{\theta})$ is a likelihood function that is the PDF value at **y** conditional on given **θ**.
  – $f(\boldsymbol{\theta})$ is the prior PDF of **θ**, which is updated to $f(\boldsymbol{\theta}|\mathbf{y})$, the posterior PDF of **θ** conditional on given **θ**.

System Health & Risk Management

# 5. Bayesian

**Example 3.2**: Suppose that we have a set of random samples $\mathbf{x} = \{x_1, x_2, \ldots, x_M\}$ from a normal PDF $f_X(x; \mu, \sigma)$ of a random variable $X$, where $\mu$ is unknown and $\sigma$ is known. Assume that the prior distribution of $\mu$, $f_M(\mu)$, is a normal distribution with its mean, $u$, and variance, $\tau^2$. Determine the posterior distribution of $\mu$, $f_{M|X}(\mu|\mathbf{x})$.

**Solution**

Firstly, we compute the conditional probability of obtaining $\mathbf{x}$ given $\mu$ as

$$\begin{aligned} f_{X|M}(\mathbf{x}\mid\mu) &= \prod_{i=1}^{M}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right] \\ &= \left(2\pi\sigma^2\right)^{-M/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{M}(x_i-\mu)^2\right] \end{aligned} \tag{28}$$

Next, we compute the joint probability of $\mathbf{x}$ and $\mu$ as

$$\begin{aligned} f_{X,M}(\mathbf{x},\mu) &= f_{X|M}(\mathbf{x}\mid\mu)\,f_M(\mu) \\ &= \left(2\pi\sigma^2\right)^{-M/2}\left(2\pi\tau^2\right)^{-1/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{M}(x_i-\mu)^2-\frac{1}{2\tau^2}(\mu-u)^2\right] \\ &= K_1(x_1,\ldots,x_M,\sigma,u,\tau)\exp\left[-\left(\frac{M}{2\sigma^2}+\frac{1}{2\tau^2}\right)\mu^2+\left(\frac{M\bar{x}}{\sigma^2}+\frac{u}{\tau^2}\right)\mu\right] \end{aligned}$$

# 5. Bayesian

We then set up a square with $\mu$ in the exponent as

$$f_{X,M}(\mathbf{x},\mu) = K_2(x_1,\ldots,x_M,\sigma,u,\tau)\exp\left[-\frac{1}{2}\left(\frac{M}{\sigma^2}+\frac{1}{\tau^2}\right)\left(\mu-\frac{\frac{M\bar{x}}{\sigma^2}+\frac{u}{\tau^2}}{\frac{M}{\sigma^2}+\frac{1}{\tau^2}}\right)^2\right]$$

$$= K_2(x_1,\ldots,x_M,\sigma,u,\tau)\exp\left[-\frac{1}{2}\left(\frac{M}{\sigma^2}+\frac{1}{\tau^2}\right)\left(\mu-\frac{M\tau^2\bar{x}+\sigma^2 u}{M\tau^2+\sigma^2}\right)^2\right]$$

Since the denominator $f_X(x_1, x_2,\ldots, x_M)$ does not depend on $\mu$, we then derive the posterior distribution of $\mu$ as

$$f_{M|X}(\mu\,|\,\mathbf{x}) = K_3(x_1,\ldots,x_M,\sigma,u,\tau)\exp\left[-\frac{1}{2}\left(\frac{M}{\sigma^2}+\frac{1}{\tau^2}\right)\left(\mu-\frac{M\tau^2\bar{x}+\sigma^2 u}{M\tau^2+\sigma^2}\right)^2\right]$$

Clearly, this is a normal distribution with the mean and variance as

$$\hat{u} = \frac{M\tau^2\bar{x}+\sigma^2 u}{M\tau^2+\sigma^2}, \quad \hat{\tau} = \left(\frac{M}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1} = \frac{\sigma^2\tau^2}{M\tau^2+\sigma^2} \qquad (29)$$

Therefore, the Bayes estimate of $\mu$ is essentially a weighted-sum of the sample mean and the prior mean. In contrast, the maximum likelihood estimator is only the sample mean. As the number of samples $M$ approaches the infinity, the Bayes estimate becomes equal to the maximum likelihood estimator since the sample data tend to have a predominant influence over the prior information. However, for the case of a small sample size, the prior information often plays an important role, especially when the prior variance $\tau^2$ is small (or we have very specific prior information).

# THANK YOU
# FOR LISTENING

Reference

[1] Achintya Haldar, Sankaran Mahadevan, Probability, Reliability and Statistical Methods in Engineering Design, John Wiley, 2000.

[2] Anthony Hayter, Probability and Statistics For Engineers and Scientists, Duxbury Resource Center, 2012.

# 4. Estimation of Parameter and Distribution

Interval Estimation of Parameters

- An interval that contains a set of plausible value of the parameter.
  - The confidence level : $1-\alpha$

  ex) confidence interval for $\mu$

$$P\left(\bar{X}-\frac{t_{\alpha/2,n-1}S}{\sqrt{n}}\le\mu\le\bar{X}+\frac{t_{\alpha/2,n-1}S}{\sqrt{n}}\right)=1-\alpha$$

  - Confidence interval length
  - $L=\frac{2t_{\alpha/2,\,n-1}\times S}{\sqrt{n}}\ \propto\ \frac{1}{\sqrt{n}}$
  - Higher confidence levels require longer confidence intervals. $(\alpha_2>\alpha_1)$

- t-Interval

$$\mu\in\left(\bar{x}-\frac{t_{\alpha/2,n-1}s}{\sqrt{n}},\bar{x}+\frac{t_{\alpha/2,n-1}s}{\sqrt{n}}\right)$$

  - with **unknown** population variance
  - small sample sizes when the data are taken to be normally distributed.
  - not normally distributed small sample data (nonparametric techniques)

# 4. Estimation of Parameter and Distribution

- z-Interval

$$\mu \in \left( \bar{x} - \frac{z_{\alpha/2,n-1}\sigma}{\sqrt{n}}, \bar{x} + \frac{z_{\alpha/2,n-1}\sigma}{\sqrt{n}} \right)$$

  – with **known** population standard-deviation($\sigma$)
  – observations : $x_1, x_2, \cdots x_n$
  independent RV : $X_1, X_2, \cdots X_n$
  sample mean is itself a RV
  $(\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i)$

- One-sided t-Interval

$$\mu \in \left( -\infty, \bar{x} + \frac{t_{\alpha,n-1}s}{\sqrt{n}} \right) \ and\ \mu \in \left( \bar{x} - \frac{t_{\alpha,n-1}s}{\sqrt{n}}, \infty \right)$$

- One-sided z-Interval

$$\mu \in \left( -\infty, \bar{x} + \frac{z_{\alpha,n-1}\sigma}{\sqrt{n}} \right) \ and\ \mu \in \left( \bar{x} - \frac{z_{\alpha,n-1}\sigma}{\sqrt{n}}, \infty \right)$$

# 4. Estimation of Parameter and Distribution

Hypothesis Testing

- Deciding the rejection yes or no of ‘Null hypothesis’ by providing the intensity of it’s counterevidence.

| | Two sided | One sided | |
|---|---|---|---|
| Null hypothesis($H_o$) | $\mu = \mu_o$ | $\mu \leq \mu_o$ | $\mu \geq \mu_o$ |
| Alternative hypothesis($H_A$) | $\mu \neq \mu_o$ | $\mu > \mu_o$ | $\mu < \mu_o$ |

- Ex) The machine that produces metal cylinders is set to make cylinders with a diameter 50mm. Is it calibrated correctly?

$$H_o : \mu=50 \quad vs \quad H_A : \mu\neq 50$$

- p-Value(significance probability) : the probability of obtaining the worse data set when the null hypothesis is true. (usually 0.01)
  - The smaller the p-value, the less plausible is the null hypothesis.
  - $H_A$ cannot be proven to be true; $H_o$ can only be shown to be implausible.

# 4. Estimation of Parameter and Distribution

- Two-sided problem

$$H_o\colon \mu = \mu_o \quad vs \quad H_A : \mu \neq \mu_o$$

$$\text{Test statistic: } t = \frac{\sqrt{n}(\bar{x}-\mu_o)}{s}$$

- One-sided problem

$$H_o\colon \mu \leq \mu_o \quad vs \quad H_A : \mu > \mu_o$$

$$H_o\colon \mu \geq \mu_o \quad vs \quad H_A : \mu < \mu_o$$

- Rejection region
  - The set of values for the test statistic that leads to rejection of $H_O$.
  - If the value falls inside the rejection region, you reject the null hypothesis.
  - If you choose the alpha level 5%, that level is the rejection region.

| $H_A$ | P-value(reject) , $X \sim \mathrm{t}(\mathrm{n}-1)$ | Rejection region |
|---|---|---|
| $\mu \neq \mu_o$ | $P\{\lvert X\rvert \geq \lvert t\rvert\} < \alpha$ | $\lvert t\rvert > t_{\frac{\alpha}{2},n-1}$ |
| $\mu > \mu_o$ | $P\{X \geq t\} < \alpha$ | $\lvert t\rvert > t_{\alpha,n-1}$ |
| $\mu < \mu_o$ | $P\{X \leq t\} < \alpha$ | $\lvert t\rvert < -t_{\alpha,n-1}$ |

# 4. Estimation of Parameter and Distribution

- Ex) The data : the times in minutes taken to remove paint.

  Question : Is the average blast time is less than 10 min?

Data : 10.3, 9.3, 11.2, 8.8, 9.5, 9.0

1. Data summary
   $n = 6, \bar{x} = 9.683$, s=0.906

2. Determination of suitable hypothesis
   $H_o: \mu \geq 10 \quad vs \quad H_A : \mu < 10$

3. Calculation of the test statistic
   $$t = \frac{\sqrt{n}(\bar{x}-\mu_o)}{s} = \frac{\sqrt{6}(9.683-10)}{0.906} = -0.857$$

4. Expression for the $p - value$
   $p - value = P(X \leq -0.857), X \sim t(5)$

5. Evaluation of the $p - value$
   set $\alpha$=0.1, P$(X \leq -0.857) >$ 0.1 or $t = -0.857 > -t_{0.1,5} = -1.476$

6. Decision
   $H_o$ is accepted.

7. Conclusion
   The data can't provide sufficient evidence that the average blast time is less than 10 min.

# 4. Estimation of Parameter and Distribution

- Type of errors

| | | Real | |
|---|---|---|---|
| | | $H_o$ true | $H_A$ true |
| Result of test | $select \quad H_o$ | OK | Type 2 error($\boldsymbol{\beta}$) |
| | $select \quad H_A$ | Type 1 error($\boldsymbol{\alpha}$) | OK |