# Statistics: Parameter Estimation

## Parametric Statistics

- Estimation of the Mean
- Estimation of the Variance
- Confidence interval on the Mean of Normal RV with known $\sigma$ (Normal Statistics)
- Confidence interval on the Mean of Normal RV with unknown $\sigma$ ($T_{n-1}$ Statistics)
- Confidence interval on the Variance of Normal RV with known/unknown $\mu$ ($\chi^2$ Statistics)
- Estimation of the Standard Deviation and the Covariance
- Estimation of Non-Gaussian Parameters from Large Samples
- Maximum Likelihood Estimator

# Statistics: Parameter Estimation

**Non-parametric Statistics**

- Sample Median Estimator
- Ordered RVs and Area RVs
- Estimating Percentile Points
- Confidence Interval for the Percentile Point
- Confidence Interval for the Median When $n$ is Large
- Estimation of Vector Means and Covariance Matrices
- Least-Squares Estimator

- Parametric statistics: We know/assume pdf, PMF, or CDF and use it in computing probabilities, estimating parameters, and making decisions

- Non-Parametric (or distribution-free) statistics: Estimation of the properties and parameters of a population without any assumption on the form or knowledge of the population distribution

# Definitions

- An estimator $\hat{\Theta}$ is a function of the observation vector $\boldsymbol{X} \triangleq [X_1, X_2, \ldots, X_n]^T$ that estimates $\theta$ but is not a function of $\theta$.
- An estimator $\hat{\Theta}$ for $\theta$ is unbiased if and only if $E[\hat{\Theta}] = \theta$. The bias is $E[\hat{\Theta}] - \theta$.
- An estimator $\hat{\Theta}$ is a linear estimator of $\theta$ if $\hat{\Theta} = \boldsymbol{b}^T \boldsymbol{X}$ where $\boldsymbol{b}$ is an $n \times 1$ vector that does not depend on $\boldsymbol{X}$.
- Let $\hat{\Theta}_n$ be an estimator computed from $X_1, \ldots, X_n$. Then $\hat{\Theta}_n$ is said to be consistent if

$$\lim_{n \to \infty} P[|\hat{\Theta}_n - \theta| > \epsilon] = 0 \text{ for every } \epsilon > 0. \text{ (convergence in prob.)}$$

- $\hat{\Theta}$ is a minimum-variance unbiased (MVU) estimator if $E[|\hat{\Theta} - \theta|^2] \leq E[|\hat{\Theta}' - \theta|^2]$ where $\hat{\Theta}'$ is any other estimator and $E[\hat{\Theta}] = E[\hat{\Theta}'] = \theta$.
- $\hat{\Theta}$ is a minimum mean-square error (MMSE) estimator if $E[|\hat{\Theta} - \theta|^2] \leq E[|\hat{\Theta}' - \theta|^2]$ where $\hat{\Theta}'$ is any other estimator.

# Estimation of the Mean

- Mean-Estimator Function (MEF): For i.i.d. observations $\{X_i : i = 1, \ldots, n\}$ of a random variable $X$ with mean $\mu_X$, the MEF (a.k.a the sample mean estimator) is

$$\boxed{\hat{\mu}_X(n) \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i}$$

- Properties of the MEF:

$$E[\hat{\mu}_X(n)] = \mu_X, \quad \text{(unbiased)}$$

$$\sigma^2_{\hat{\mu}}(n) = \sigma^2_X / n$$

$$P[|\hat{\mu}_X(n) - \mu_X| \geq \delta] \leq \frac{\sigma^2_{\hat{\mu}}(n)}{\delta^2} = \frac{\sigma^2_X}{n\delta^2} \xrightarrow{n\to\infty} 0. \quad \text{(consistent)}$$

# Variance-Estimator Function (VEF)

- Two VEF (a.k.a. sample variance estimator) definitions are

$$\boxed{\begin{aligned}\hat{\sigma}_X^2(n) &\triangleq \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2, \quad (\text{unbiased} : E[\hat{\sigma}_X^2(n)] = \sigma_X^2)\\ \hat{\sigma}_X^2(n) &\triangleq \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2, \quad (\text{biased} : E[\hat{\sigma}_X^2(n)] \neq \sigma_X^2)\end{aligned}}$$

- Consistency of unbiased VEF:

$$\mathrm{Var}[\hat{\sigma}_X^2(n)] = E[(\hat{\sigma}_X^2(n) - \sigma_X^2)^2] \approx \frac{E[(X_i - \hat{\mu})^4]}{n}$$

$$P[|\hat{\sigma}_X^2(n) - \sigma_X^2| > \varepsilon] \leq \frac{\mathrm{Var}[\hat{\sigma}_X^2(n)]}{\varepsilon^2} \approx \frac{E[(X_i - \hat{\mu})^4]}{n\varepsilon^2} \xrightarrow{n\to\infty} 0.$$

# Confidence Interval

- $\delta$-confidence interval on $\mu_x$ is defined by

$$P[-\gamma_{1,\delta} \leq \hat{\mu}_X(n) - \mu_X \leq \gamma_{2,\delta}] = \delta, \quad (e.g., \delta = 0.95)$$

  With a $100\delta\%$ confidence level, $\mu_X$ will be in the interval $[\hat{\mu}_X[n] - \gamma_{2,\delta},\ \hat{\mu}_X[n] + \gamma_{1,\delta}]$.

- For $\hat{\mu}_X$ with pdf $\mathcal{N}(\mu_X, \sigma_X^2/n)$, $Y \triangleq \frac{\hat{\mu}_X(n) - \mu_X}{\sigma_X/\sqrt{n}}$ has a standard Normal pdf $\mathcal{N}(0,1)$ and its CDF is denoted by $F_{\mathrm{SN}}(z_{[u]}) = u$ where $z_{[u]}$ is called the $100u$-percentile of the standard Normal RV. Due to symmetry of the pdf, $\gamma_{1,\delta} = \gamma_{2,\delta} = \gamma_\delta$. Then,

$$P[-\gamma_\delta\sqrt{n}/\sigma_X \leq Y \leq \gamma_\delta\sqrt{n}/\sigma_X] = 2F_{\mathrm{SN}}(\gamma_\delta\sqrt{n}/\sigma_X) - 1 = \delta$$

$$F_{\mathrm{SN}}(\gamma_\delta\sqrt{n}/\sigma_X) = (1+\delta)/2$$

$$\Rightarrow \quad \gamma_\delta\sqrt{n}/\sigma_X = z_{[(1+\delta)/2]} \quad \text{and} \quad \gamma_\delta = z_{[(1+\delta)/2]}\frac{\sigma_X}{\sqrt{n}}$$

# $\delta$-confidence interval on the Mean of Normal RV $X$ with known $\sigma_X$

- The sample average (mean estimate): $\boxed{\hat{\mu}_X(n) = \frac{1}{n}\sum_{i=1}^{n} X_i}$.
- From the standard Normal CDF table, $\boxed{\text{find } z_{[(1+\delta)/2]} \text{ such that } F_{\mathrm{SN}}(z_{[(1+\delta)/2]}) = (1+\delta)/2}$.
- Then, the interval is $\boxed{\left[\hat{\mu}_X(n) - z_{[(1+\delta)/2]}\frac{\sigma_X}{\sqrt{n}},\ \ \hat{\mu}_X(n) + z_{[(1+\delta)/2]}\frac{\sigma_X}{\sqrt{n}}\right]}$.

- Note: Given an RV $X$ with CDF $F_X(x)$, the $100u$-percentile of $X$ is the number $x_{[u]}$ such that $\boxed{F_X(x_{[u]}) = u. \ \text{(i.e., } x_{[u]} = F_X^{-1}(u)).}$
- Note: The $100u$-percentile of a standard Normal RV is $z_{[u]} = F_{\mathrm{SN}}^{-1}(u)$. The $100u$-percentile of an RV $X$ with $f_X(x) = \mathcal{N}(\mu, \sigma^2)$ is $\boxed{x_{[u]} = \mu + \sigma z_{[u]}}$.

- Example 6.3-1
  Compute $P[|\hat{\mu}_X(n) - \mu_X| \leq 0.1]$ when $X$ is Normal with $\sigma_X = 3$ for two sample size: $n = 64$ and $n = 3600$.

  **Solution:** We have $Y \triangleq (\hat{\mu}_X - \mu_X)/(\sigma_X/\sqrt{n}) \sim N(0,1)$. Hence,

$$\begin{aligned} P[-0.1 < \hat{\mu}_X(n) - \mu_X < 0.1] &= P[-0.1\sqrt{n}/\sigma_X < Y < 0.1\sqrt{n}/\sigma_X] \\ &= 2F_{\mathrm{SN}}\left(\frac{0.1\sqrt{n}}{\sigma_X}\right) - 1 \\ &= 2F_{\mathrm{SN}}\left(0.0333\sqrt{n}\right) - 1 \end{aligned}$$

  When $n = 64$, $P[|\hat{\mu}_X(n) - \mu_X| \leq 0.1] \approx 0.2$.

  When $n = 3600$, $P[|\hat{\mu}_X(n) - \mu_X| \leq 0.1] \approx 0.95$.

  In a single trial, the event $\{|\hat{\mu}_X(n) - \mu_X| \leq 1\}$ will almost certainly happen when $n = 3600$.

- Example: How many samples needed to get a $\delta$-confidence interval on $\mu_X$ of a Normal RV $X$ with known $\sigma_X^2$?

  The shortest $\delta$-confidence interval on $\mu_X$ is

$$\left[-z_{[(1+\delta)/2]}\frac{\sigma_X}{\sqrt{n}} + \hat{\mu}_X(n),\ z_{[(1+\delta)/2]}\frac{\sigma_X}{\sqrt{n}} + \hat{\mu}_X(n)\right]$$

  where $z_{[u]}$ is the $100u$-percentile of the standard Normal. The width of the confidence interval is

$$W_\delta = 2\ z_{[(1+\delta)/2]}\ \sigma_X/\sqrt{n}$$

  For a desired width of the $\delta$-confidence interval $W_\delta$, the number of samples required is

$$n = \left\lceil \left(\frac{2\ \sigma_X\ z_{[(1+\delta)/2]}}{W_\delta}\right)^2 \right\rceil$$

# $\delta$-confidence interval on the Mean of Normal RV $X$ with unknown $\sigma_X$

- Replace $\sigma_X$ with $\hat{\sigma}_X$ (from the unbiased VEF)

$$\hat{\sigma}_X(n) = \left( \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu}_X(n))^2 \right)^{1/2}$$

in $Y$ to get a new RV $T_{n-1}$ as

$$T_{n-1} \triangleq \frac{\hat{\mu}_X(n) - \mu_X}{\hat{\sigma}_X/\sqrt{n}}.$$

- $T_{n-1}$ has a $t$-distribution with $n-1$ degrees of freedom (DoF) for $n = 2, 3, \ldots$ and its pdf and CDF are denoted by $f_T(t; n-1)$ and $F_T(t; n-1)$. Then, we have

$$P[-\gamma_\delta \leq \hat{\mu}_X(n) - \mu_X \leq \gamma_\delta] = P[-\frac{\gamma_\delta}{\hat{\sigma}_X/\sqrt{n}} \leq T_{n-1} \leq \frac{\gamma_\delta}{\hat{\sigma}_X/\sqrt{n}}] = \delta,$$

$$2F_T(\frac{\gamma_\delta}{\hat{\sigma}_X/\sqrt{n}}; n-1) - 1 = \delta \quad \Rightarrow \quad F_T(\frac{\gamma_\delta}{\hat{\sigma}_X/\sqrt{n}}; n-1) = (1+\delta)/2$$

Then, $\frac{\gamma_\delta}{\hat{\sigma}_X/\sqrt{n}} = t_{[(1+\delta)/2]}$ or $\boxed{\gamma_\delta = t_{[(1+\delta)/2]}\hat{\sigma}_X/\sqrt{n}}$.

# $\delta$-confidence interval on the Mean of Normal RV $X$ with unknown $\sigma_X$

Procedure:

- The mean estimate: $\boxed{\hat{\mu}_X(n) = \frac{1}{n}\sum_{i=1}^{n} X_i}$
- The $\sigma$ estimate:

$$\boxed{\hat{\sigma}_X(n) = \left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2\right)^{1/2}}$$

- From the student-$t$ CDF table,

$$\boxed{\text{find } t_{[(1+\delta)/2]} \text{ such that } F_T(t_{[(1+\delta)/2]}; n-1) = (1+\delta)/2}.$$

- Then, the $\delta$-confidence interval is

$$\boxed{\left[\hat{\mu}_X(n) - t_{[(1+\delta)/2]}\frac{\hat{\sigma}_X}{\sqrt{n}},\quad \hat{\mu}_X(n) + t_{[(1+\delta)/2]}\frac{\hat{\sigma}_X}{\sqrt{n}}\right]}$$

- Example 6.3-3
  $n = 21$ i.i.d. observations are made on a Gaussian RV $X$, denoting as $X_1, X_2, \ldots, X_{21}$. Based on the data, $\hat{\mu}_X(n) = 3.5$, $\hat{\sigma}_X(n)/\sqrt{n} = 0.45$. Find the 90 percent confidence interval on $\hat{\mu}_X(n)$.

  **Solution**:
  $P[-t_{[0.95]} \leq T_{20} \leq t_{[0.95]}] = 0.9 \Rightarrow$

  $$F_T(t_{[0.95]}, 20) = 0.5(1 + 0.9) = 0.95.$$

  From student-t table, we obtain $t_{[0.95]} = 1.725$.

  The corresponding interval is

  $$[\hat{\mu}_X(n) - t_{[(1+\delta)/2]}\frac{\hat{\sigma}_X}{\sqrt{n}}, \quad \hat{\mu}_X(n) + t_{[(1+\delta)/2]}\frac{\hat{\sigma}_X}{\sqrt{n}}]$$
  $$= [3.5 - 1.725 \times 0.45, 3.5 + 1.725 \times 0.45] = [2.72, 4.28]$$

  The width of the interval is $W_\sigma \approx 2 \times 1.725 \times 0.45 = 1.55$.

# Estimation of the Variance of a Normal RV $X$

- If $\mu_X$ is known, the unbiased VEF is

$$\boxed{\hat{\sigma}_X^2(n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)^2}.$$

- If $\mu_X$ is unknown, the unbiased VEF is

$$\boxed{\hat{\sigma}_X^2(n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2}.$$

- Note: $U_i \triangleq (X_i - \mu_X)/\sigma_X$ is $\mathcal{N}(0,1)$ and

  $Z_n \triangleq \sum_{i=1}^{n} U_i^2$ has a Chi-square ( $\chi^2$ ) pdf $f_{\chi^2}(z;n)$ with DoF of $n$.
- Note: For $V_i \triangleq (X_i - \hat{\mu}_X(n))/\sigma_X$,

  $Z_{n-1} \triangleq \sum_{i=1}^{n} V_i^2$ has a $\chi^2$ pdf $f_{\chi^2}(z;n-1)$ with DoF of $n-1$.

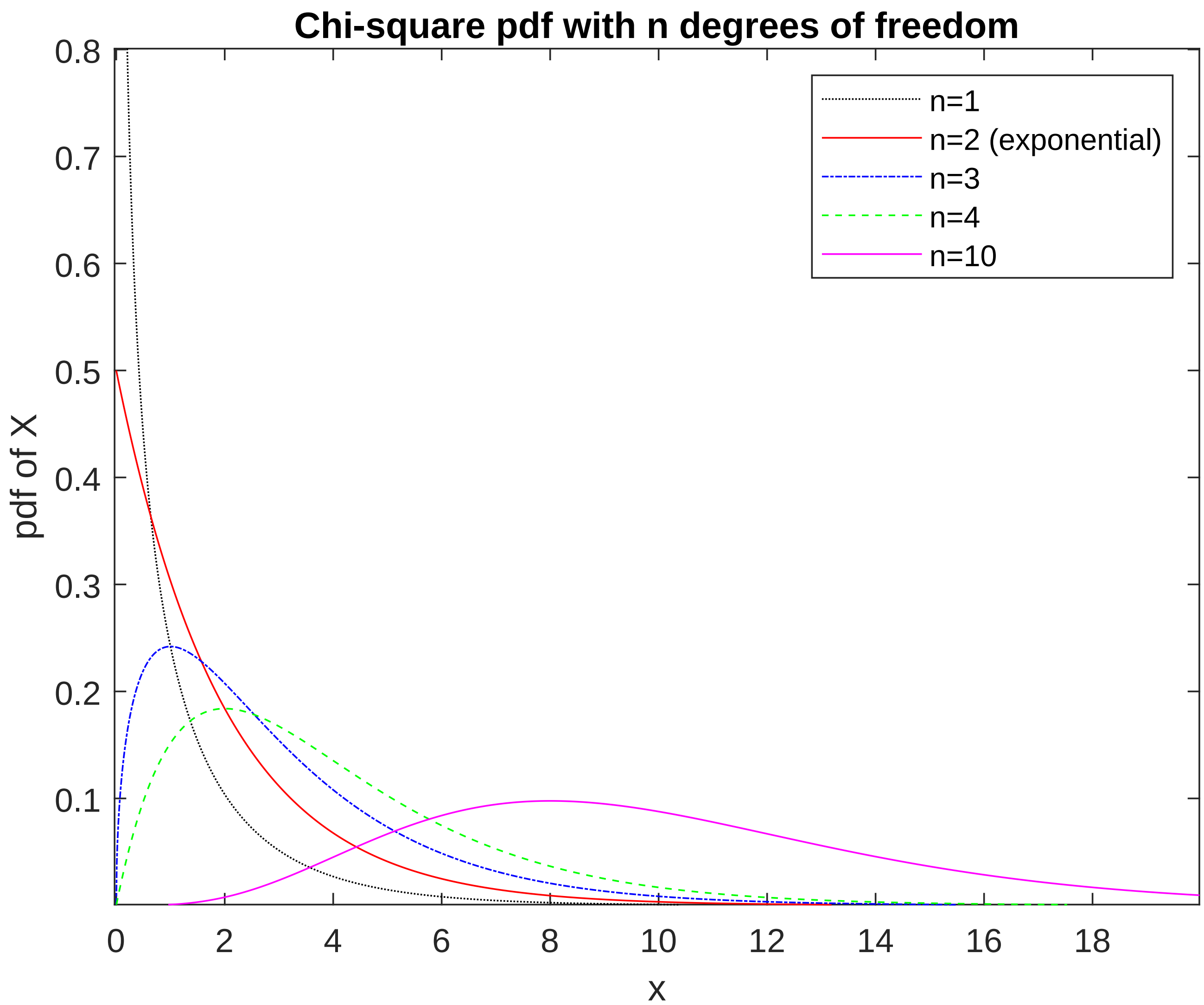



Figure: The central Chi-square pdf with $n$ degrees of freedom

# $\delta$-confidence interval on the Variance of a Normal RV $X$

- Let $Z \triangleq \sum_{i=1}^{n} U_i^2$ for known $\mu_X$ and $Z \triangleq \sum_{i=1}^{n} V_i^2$ for unknown $\mu_X$. Then, the pdf of $Z$ is $f_{\chi^2}(z;k)$ with the $\boxed{\text{DoF } k = n \text{ for known } \mu_X \text{ and } k = n-1 \text{ for unknown } \mu_X}$.
- Note that $Z = \hat{\sigma}_X^2 k/\sigma_X^2$ and hence

$$P[a \le Z \le b] = P[a \le \hat{\sigma}_X^2 k/\sigma_X^2 \le b] = P[(1/b) \le \sigma_X^2/(k\hat{\sigma}_X^2) \le (1/a)]$$

$$= P[\frac{k\hat{\sigma}_X^2}{b} \le \sigma_X^2 \le \frac{k\hat{\sigma}_X^2}{a}] = F_{\chi^2}(b;k) - F_{\chi^2}(a;k) = \delta$$

- A simple equal error probability approach for finding $a$ and $b$ with near-shortest interval:

$$P[Z < a] = \quad \boxed{F_{\chi^2}(a;k) = (1-\delta)/2 \quad \Rightarrow \quad a = x_{[(1-\delta)/2]}}$$

$$P[Z > b] = 1 - F_{\chi^2}(b;k) = (1-\delta)/2$$

$$\Rightarrow \quad \boxed{F_{\chi^2}(b;k) = (1+\delta)/2 \quad \Rightarrow \quad b = x_{[(1+\delta)/2]}}$$

- The $\delta$-confidence interval for $\sigma_X^2$ is $\boxed{\left[\frac{k\hat{\sigma}_X^2}{x_{[(1+\delta)/2]}}, \quad \frac{k\hat{\sigma}_X^2}{x_{[(1-\delta)/2]}}\right]}$.

# $\delta$-confidence interval on the Variance of Normal RV $X$ with known $\mu_X$

Procedure for the equal error probability approach:

- The statistic is $\chi^2$ with $n$ DoF and the steps are:

$$\boxed{\hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)^2}$$

$$\boxed{F_{\chi^2}(a; n) = (1-\delta)/2 \quad \Rightarrow \quad a = x_{[(1-\delta)/2]}}$$

$$\boxed{F_{\chi^2}(b; n) = (1+\delta)/2 \quad \Rightarrow \quad b = x_{[(1+\delta)/2]}}$$

The confidence interval is

$$\boxed{\left[\frac{n\ \hat{\sigma}_X^2}{b}, \quad \frac{n\ \hat{\sigma}_X^2}{a}\right]}$$

# $\delta$-confidence interval on Variance of Normal RV $X$ with unknown $\mu_X$

Procedure for the equal error probability approach:

- The statistic is $\chi^2$ with $(n-1)$ DoF and the steps are:

$$\boxed{\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

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

$$\boxed{F_{\chi^2}(a; n-1) = (1-\delta)/2 \quad \Rightarrow \quad a = x_{[(1-\delta)/2]}}$$

$$\boxed{F_{\chi^2}(b; n-1) = (1+\delta)/2 \quad \Rightarrow \quad b = x_{[(1+\delta)/2]}}$$

The confidence interval is

$$\boxed{\left[\frac{(n-1)\ \hat{\sigma}_X^2}{b}, \quad \frac{(n-1)\ \hat{\sigma}_X^2}{a}\right]}$$

- Example 6.4-3 (extended) 16 i.i.d. observations are made on $X : N(\mu_X, \sigma_X^2)$. Find the numbers $a$, $b$ that will give a near-shortest 95 percent confidence interval on $\sigma_X^2$ using the "equal error probability" rule. Given the observations $X_1, \ldots, X_{16}$, find the above confidence interval.

  **Solution**: Statistics for confidence interval on $\sigma_X^2$ of Gaussian RV with unknown mean: Chi-square with $n-1$ DoF

  $$F_{\chi^2}(a;k) = (1-\delta)/2 \quad \Rightarrow \quad F_{\chi^2}(x_{[0.025]};15) = 0.025$$

  $$F_{\chi^2}(b;k) = (1+\delta)/2 \quad \Rightarrow \quad F_{\chi^2}(x_{[0.975]};15) = 0.975$$

  From the table of the Chi-square distribution, we find

  $$a = x_{[0.025]} = 6.26 \qquad \text{and} \qquad b = x_{[0.925]} = 27.5.$$

  As $\mu_X$ is unknown, the unbiased variance estimate is

  $$\hat{\sigma}_X^2 = \frac{1}{15}\sum_{i=1}^{16}(X_i - \hat{\mu}_X)^2 \quad \text{where} \quad \hat{\mu}_X = \frac{1}{16}\sum_{i=1}^{16} X_i$$

  The corresponding confidence interval is $\left[\frac{15\ \hat{\sigma}_X^2}{b}, \quad \frac{15\ \hat{\sigma}_X^2}{a}\right]$.

# Estimation of the Standard Deviation

- Standard Deviation Estimating Function (SDEF) based on VEF:

$$\boxed{\hat{\sigma}_X(n) = \left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2\right)^{1/2}}$$

- A Pair-wise based SDEF (with an even $n$):

$$\boxed{\hat{\sigma}_X(n) = \frac{2}{n}\sum_{i=1}^{n/2}\sqrt{\pi}\,(\max(X_{2i-1}, X_{2i}) - 0.5(X_{2i-1} + X_{2i}))}$$

and its variance is $\mathrm{Var}(\hat{\sigma}_X(n)) = \frac{\pi-2}{n}\,\sigma_X^2$.

# Estimation of the Covariance $c_{11} \triangleq \text{Cov}[X, Y]$

- $\text{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$
- Covariance estimating function (CEF):

$$\boxed{\hat{c}_{11} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu}_X(n))(Y_i - \hat{\mu}_Y(n))}$$

- The correlation coefficient $\rho_{XY} \triangleq c_{11} / \sqrt{\sigma_X^2 \sigma_Y^2}$ can be estimated as

$$\hat{\rho}_{XY} \triangleq \hat{c}_{11} / \sqrt{\hat{\sigma}_X^2 \hat{\sigma}_Y^2}$$

# Estimation of Non-Gaussian Parameters from Large Samples

- Based on $n$ i.i.d. observations $\{X_i\}$ of an RV $X$ with mean $\mu$ and finite variance $\sigma^2$, the sample mean estimator $\hat{\mu}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ has approximately Normal distribution as $\mathcal{N}(\mu, \sigma^2/n)$ for large $n$, due to the Central Limit Theorem.

- The $\delta$-confidence interval can be obtained from

$$P[-a \leq \frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}} \leq a] \approx 2F_{\mathrm{SN}}(a) - 1 = \delta$$

$$P[\hat{\mu} - (a\sigma/\sqrt{n}) \leq \mu \leq \hat{\mu} + (a\sigma/\sqrt{n})] = \delta$$

- $F_{\mathrm{SN}}(a) = (1+\delta)/2 \Rightarrow a = z_{[(1+\delta)/2]}$ which yields the confidence interval as

$$[\hat{\mu} - (z_{[(1+\delta)/2]}\ \sigma/\sqrt{n}), \quad \hat{\mu} + (z_{[(1+\delta)/2]}\ \sigma/\sqrt{n})]$$

- Example 6.6-1
  Confidence interval for $\lambda$ in the exponential pdf $f_X(x) = \lambda e^{-\lambda x} u(x)$.

  **Solution**: We have $\mu \triangleq E[X] = \lambda^{-1}$ and $\sigma^2 = \lambda^{-2}$.

  Inserting these results into $P[-a \le \frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}} \le a] = \delta$ yields

$$P\left[\frac{(-a/\sqrt{n})+1}{\hat{\mu}} \le \lambda \le \frac{(a/\sqrt{n})+1}{\hat{\mu}}\right] = \delta.$$

  We can obtain $a$ by approximating $Z \triangleq (\hat{\mu}-\mu)\sqrt{n}/\sigma$ as $N(0,1)$ RV.

  This yields $a = z_{[(1+\delta)/2]}$.

  Thus, a $100 \times \delta$ percent confidence interval for $\lambda$ is

$$\left[\frac{(-z_{[(1+\delta)/2]}/\sqrt{n})+1}{\hat{\mu}}, \; \frac{(z_{[(1+\delta)/2]}/\sqrt{n})+1}{\hat{\mu}}\right]$$

  and its width is

$$W_\delta = 2z_{[(1+\delta)/2]}/(\hat{\mu}\sqrt{n})$$

- Example 6.6-2
  Find a 95 percent confidence interval on the parameter $\lambda$ of the exponential distribution from 64 i.i.d. observations on an exponential RV $X$. The estimate is $\hat{\mu}_X = 3.5$.

  **Solution**:
  From $P[-a \leq \frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}} \leq a] = \delta$, we obtain

  $$F_{\mathrm{SN}}(z_{[(1+\delta)/2]}) = (1+\delta)/2 = 0.975 \Rightarrow a = z_{[0.975]} = 1.96.$$

  Then, from

  $$P\left[\frac{(-a/\sqrt{n})+1}{\hat{\mu}} \leq \lambda \leq \frac{(a/\sqrt{n})+1}{\hat{\mu}}\right] = \delta \quad \text{and} \quad W_\delta = 2z_{[(1+\delta)/2]}/(\hat{\mu}\sqrt{n}),$$

  we compute that the 95 percent confidence interval for $\lambda$ is

  $$\left[\frac{(-a/\sqrt{n})+1}{\hat{\mu}},\ \frac{(a/\sqrt{n})+1}{\hat{\mu}}\right] = [0.22,\ 0.36]$$

  and it has an approximate width of $W_\delta = 0.14$.

# Maximum Likelihood Estimator (MLE)

- Likelihood function $L(\theta; \boldsymbol{X})$ or simply $L(\theta)$ is the joint pdf $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \cdots, x_n; \theta)$ considered as a function of the unknown parameter $\theta$.
- MLE of $\theta$ for a given observation $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$:

$$\boxed{\theta^*(\boldsymbol{x}) = \arg_\theta \max L(\theta; \boldsymbol{x})}$$

- In many cases, $\theta^*$ can be obtained by $\boxed{\text{solving } \frac{dL(\theta)}{d\theta} = 0 \text{ or } \frac{d \log_e L(\theta)}{d\theta} = 0}$ where $\log_e L(\theta)$ is called the log-likelihood function.
- MLE Properties:
  - Squared-error consistency
  - Invariance: If $\hat{\theta}$ is MLE for $\theta$, then $h(\hat{\theta})$ is the MLE for $h(\theta)$.
  - No guarantee for unbiasedness.

- Example 6.7-2
  Consider a Bernoulli RV $X \sim P_X(k) = p^k(1-p)^{1-k}$, where $P[X=1] = p$, and $P[X=0] = 1-p$. Find the ML estimation of $p$ using likelihood function, with $n$ i.i.d. observations $X_1, \ldots, X_n$ on $X$.

  **Solution**: The likelihood function is

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i} \times (1-p)^{n-\sum_{i=1}^{n} x_i}.$$

  By setting $dL(p)/dp = 0$, we obtain three roots:

$$p = 0,\ p = 1,\ p = \sum_{i=1}^{n} x_i/n.$$

  The first two roots yield a minimum, while the last root yields a maximum.

  Thus, $p^*(\boldsymbol{x}) = \sum_{i=1}^{n} x_i/n$ and the MLE of $p$ is

$$\hat{p} = p^*(X_1, X_2, \ldots, X_n) = \textstyle\sum_{i=1}^{n} X_i/n.$$

- Example 6.7-3 Assume $X : N(\mu, \sigma)$, where $\sigma$ is known. Compute the MLE of the mean $\mu$ with $n$ realizations of $X$.

**Solution**: The likelihood function is

$$L(\mu) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right).$$

The maximum of $L(\mu)$ is also that of $\log L(\mu)$. Hence

$$\log L(\mu) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Setting $\partial \log L(\mu)/\partial\mu = 0$ yields

$\sum_{i=1}^{n}(X_i - \mu) = 0$ and $\mu^* = \frac{1}{n}\sum_{i=1}^{n} X_i$.

This implies the MLE of $\mu$ should be

$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

- Example 6.7-5. Consider the Normal pdf

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \quad -\infty < x < \infty.$$

Compute the MLE of $\mu$ and $\sigma^2$ with $n$ realizations of $X$.

**Solution**: The log-likelihood function is

$$\bar{L}(\mu,\sigma) \triangleq \log L = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Solving $\partial\bar{L}/\partial\mu = 0,\ \partial\bar{L}/\partial\sigma = 0$ gives

$$\sum_{i=1}^{n}(x_i-\mu) = 0, \quad -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 = 0.$$

which yields

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}X_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\mu})^2.$$

- Example 6.7-6 (Application of Invariance property of MLE)

  Consider $n$ observations on a Normal RV. Assume that it is known that the mean is zero.

  The MLE of the variance is $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

  Hence, by applying the invariance property of MLE, the MLE of the standard deviation is

  $$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}.$$

# Parametric versus Non-Parametric Statistics

- Parametric statistics: We know/assume pdf, PMF, or CDF and use it in computing probabilities, estimating parameters, and making decisions
- Non-Parametric (or distribution-free) statistics: Estimation of the properties and parameters of a population without any assumption on the form or knowledge of the population distribution
- Mean and standard deviation indicate the center and the dispersion of the population in the parametric case while the median and the range play a comparable role in the non-parametric case.
- Median of the population $X$ is the point $\boxed{x_{[0.5]} \text{ such that } F_X(x_{[0.5]}) = 0.5}$.

# Sample Median Estimator

- Order the observations $X_1, X_2, \cdots, X_n$ to get $\boxed{Y_1 < Y_2 < \cdots < Y_n}$.

- The sample median estimator is

$$\boxed{\hat{x}_{[0.5]} = \begin{cases} Y_{k+1}, & \text{if } n = 2k+1 \ \ (\text{odd } n) \\ 0.5(Y_k + Y_{k+1}), & \text{if } n = 2k \ \ (\text{even } n) \end{cases}}$$

- The sample median estimate is not unbiased but becomes nearly so when $n$ is large.

- The dispersion in the nonparametric case is measured from an appropriate range, e.g., $\Delta x_{[0.50]} \triangleq x_{[0.75]} - x_{[0.25]}$ for the 50% range and $\Delta x_{[0.90]} \triangleq x_{[0.95]} - x_{[0.05]}$ for the 90% range.

  These percentile points (e.g., $x_{[0.95]}, x_{[0.05]}$) have to be estimated from the observations (to be discussed in Estimating Percentile Points).

# Ordered RVs and Area RVs

- Let $Y_1,\ Y_2,\ \cdots,\ Y_n$ be the ordered RVs of i.i.d. RVs $X_1,\ X_2,\ \cdots, X_n$ with pdf $f_{X_i}(x) = f_X(x)$ such that $Y_1 < Y_2 < \cdots < Y_n$.
- The joint pdf of $\{Y_i\}$ is

$$f_{Y_1,\cdots,Y_n}(y_1,\cdots,y_n) = \begin{cases} n! \prod_{i=1}^n f_X(y_i), & -\infty < y_1 < \cdots < y_n < \infty \\ 0, & \text{else} \end{cases}$$

- Area RVs are defined as

$$Z_i \triangleq \int_{-\infty}^{Y_i} f_X(x)dx = F_X(Y_i), \quad i = 1, \cdots, n$$

- The joint pdf of $\{Z_i\}$ (where $0 < Z_1 < \cdots < Z_n < 1$) is

$$f_{Z_1,\cdots,Z_n}(z_1,\cdots,z_n) = \begin{cases} n!, & 0 < z_1 < \cdots < z_n < 1 \\ 0, & \text{else} \end{cases}$$

with $E[Z_i] = \frac{i}{n+1}$ and $\sigma_{Z_i}^2 = \frac{i(i+1)}{(n+1)(n+2)} - \frac{i^2}{(n+1)^2} \approx \frac{i}{(n+1)^2}$, $n \gg 1$.

- The pdf of the area $V_{l,m} \triangleq Z_m - Z_l = \int_{Y_l}^{Y_m} f_X(x)dx$ between any ordered RVs with $m > l$ is

$$f_{V_{l,m}}(v) = \begin{cases} \frac{n!}{(m-l-1)!(n-m+l)!} \; v^{m-l-1}(1-v)^{n-m+l}, & 0 < v < 1 \\ 0, & \text{else} \end{cases}$$

- For $l = 1$ and $m = n$, the pdf (a beta pdf with $\alpha = n-2$ and $\beta = 1$) and CDF are

$$f_{V_{1,n}}(v) = \begin{cases} n(n-1) \; v^{n-2}(1-v), & 0 < v < 1, \; n \geq 2 \\ 0, & \text{else} \end{cases}$$

$$F_{V_{1,n}}(v) = \begin{cases} nv^{n-1} - (n-1)v^n, & 0 < v < 1, \\ 1, & v \geq 1 \\ 0, & \text{else} \end{cases}$$

- The pdf of $V_{i,i+1}$ is

$$f_{V_{i,i+1}}(v) = \begin{cases} n(1-v)^{n-1}, & 0 < v < 1, \\ 0, & \text{else} \end{cases}$$

with $E[V_{i,i+1}] = \frac{1}{n+1}$ and $E[V_{i,i+1}^2] = \frac{2}{(n+2)(n+1)}$.

# Estimating Percentile Points

- Recall $Z_i = F_X(Y_i)$, $E[Z_i] = \frac{i}{n+1}$ and $\sigma^2_{Z_i} \approx \frac{i}{(n+1)^2}$ for $n \gg 1$. Thus, for $n \gg 1$, $Z_i \approx E[Z_i]$ and hence $F_X(Y_i) \approx E[Z_i]$.

- Given i.i.d Observations $x_1, \cdots, x_n$ and associated ordered observations $y_1, \cdots, y_n$, for large $n$, we can estimate the percentile points from $\boxed{\hat{F}_X(y_i) = E[Z_i] = \frac{i}{n+1}}$, i.e., (here, $u = i/(n+1)$),

  $\boxed{\text{the } \frac{100i}{n+1} \text{ percentile point } x_{[i/(n+1)]} \text{ is estimated to be } y_i}$.

- For a 100$u$ percentile with $\frac{i}{n+1} < u < \frac{i+1}{n+1}$, the estimate of the percentile point $x_{[u]}$ can be obtained by interpolation as

$$\boxed{\hat{x}_{[u]} = y_i + \frac{(y_{i+1} - y_i)(u - \frac{i}{n+1})}{1/(n+1)}}$$

- Note: The index of $Y$ in finding 100$u$ percentile point: $\boxed{i = \lfloor (n+1)\ u \rfloor}$

# Confidence Interval for the Percentile Point

- Recall the notation $P[X \le x_{[u]}] \triangleq u$. Then

$$P[Y_k \le x_{[u]}] = P[\text{at least } k \text{ of } \{X_i\} \text{ are } \le x_{[u]}]$$
$$= \sum_{i=k}^{n} \binom{n}{i} u^i (1-u)^{n-i}$$
$$P[Y_{k+r} > x_{[u]}] = P[\text{no more than } (k+r-1) \text{ of } \{X_i\} \text{ are } \le x_{[u]}]$$
$$= \sum_{i=0}^{k+r-1} \binom{n}{i} u^i (1-u)^{n-i}$$

- Since $\{Y_k \le x_{[u]}\} \cap \{Y_{k+r} > x_{[u]}\} = \{Y_k \le x_{[u]} < Y_{k+r}\}$, we have

$$\boxed{P[Y_k \le x_{[u]} < Y_{k+r}] = \sum_{i=k}^{k+r-1} \binom{n}{i} u^i (1-u)^{n-i}}$$

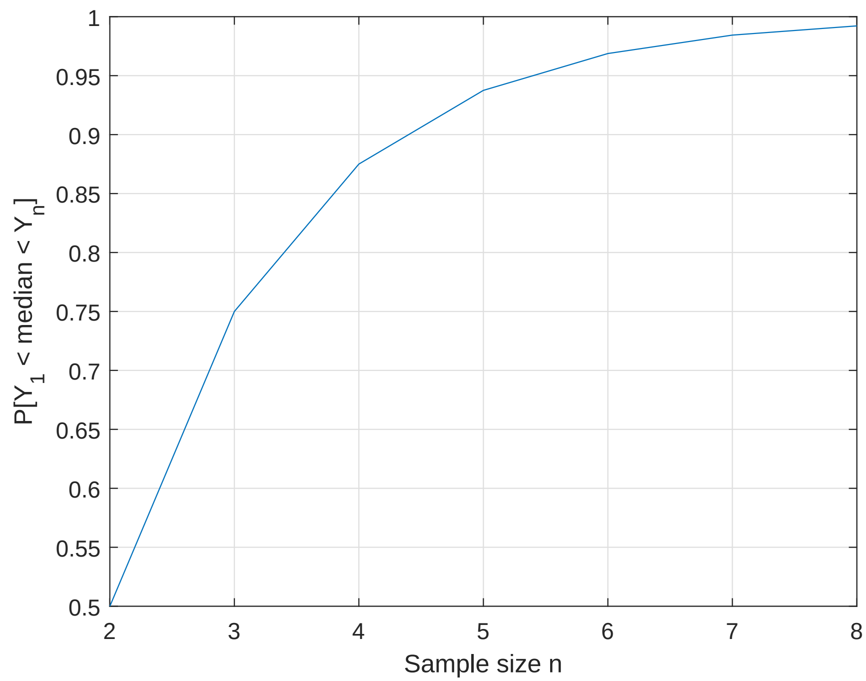
which can be computed without knowledge of $f_X(x)$.

- Example 6.8-5 We seek the end points $Y_1, Y_n$ of a random interval $[Y_1, Y_n]$ so that the event $\{Y_1 < x_{[0.5]} < Y_n\}$ occurs with probability $\sim 0.95$. Here $Y_1 \triangleq \min(X_1, X_2, \ldots, X_n)$, $Y_n \triangleq \max(X_1, X_2, \ldots, X_n)$. How large should $n$ be?

**Solution**: The answer is furnished by computing

$$P[Y_1 \leq x_{[0.5]} < Y_n] = \sum_{i=1}^{n-1} \binom{n}{i} (1/2)^n \approx 0.95$$

and find that for $n = 5$, $P[Y_1 \leq x_{[0.5]} < Y_n] \approx 0.94$.

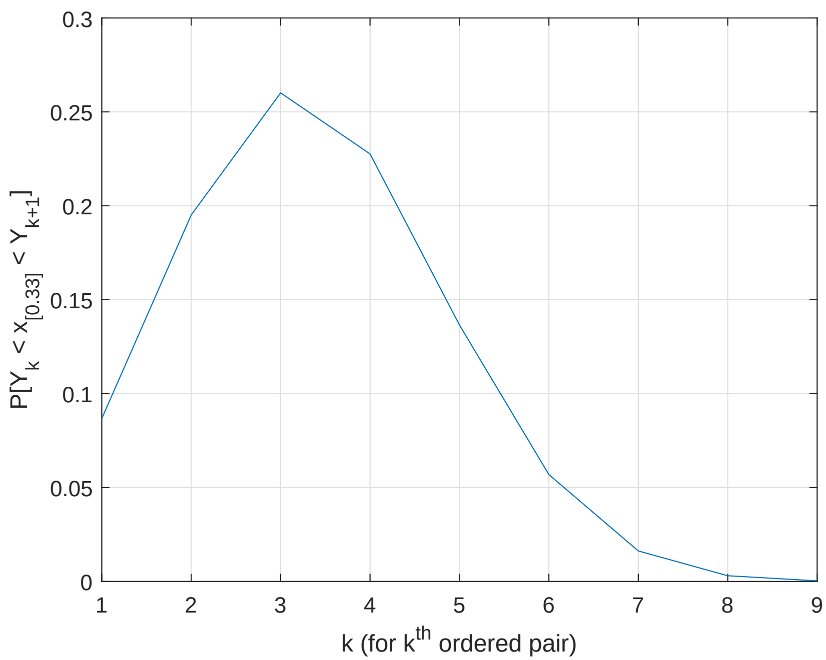


- Example 6.8-6. We have a set of ordered samples $\{Y_1, Y_2, \ldots, Y_n\}$ and wish to find the pair $\{Y_i, Y_{i+1}, i = 1, \ldots, n-1\}$ that maximizes the probability of covering the 33.33rd percentile point.

**Solution**: The 33.33rd percentile point is defined by $u = 1/3 = F_X(x_{[0.33]})$. For specificity we assume $n = 10$. Then

$$P[Y_k \leq x_{[u]} < Y_{k+r}] = \sum_{i=k}^{k+r-1} \binom{n}{i} u^i (1-u)^{n-i},$$

$$P[Y_k \leq x_{[0.33]} < Y_{k+1}] = \frac{10!}{k!(10-k)!}(1/3)^k(2/3)^{10-k}, k = 1, \ldots, 9$$




Clearly the interval $[Y_3, \ Y_4)$ is most likely to cover $x_{[0.33]}$. The probability of the event $\{Y_3 \leq x_{[0.33]} < Y_4\}$ is 0.26.

# Confidence Interval for the Median When $n$ is Large

- If $n$ is large, we can use Normal approximation to Binomial as

$$\boxed{P[\alpha \le S_n \le \beta] \approx F_{\mathrm{SN}}(\beta_n) - F_{\mathrm{SN}}(\alpha_n)}$$

$$\text{where } P[\alpha \le S_n \le \beta] = \sum_{i=\alpha}^{\beta} \binom{n}{i} p^i (1-p)^{n-i}$$

$$\boxed{\alpha_n \triangleq \frac{\alpha - np - 0.5}{\sqrt{np(1-p)}} \quad \text{and} \quad \beta_n \triangleq \frac{\beta - np + 0.5}{\sqrt{np(1-p)}}}$$

- $\delta$-confidence interval for the median is obtained from

$$P[Y_r \le x_{[0.5]} < Y_{n-r+1}] = \sum_{i=r}^{n-r} \binom{n}{i} (1/2)^n \approx F_{\mathrm{SN}}(\beta_n) - F_{\mathrm{SN}}(\alpha_n) = \delta$$

$$(\alpha_n = -\beta_n \text{ due to symmetry}), \qquad \boxed{2F_{\mathrm{SN}}(\beta_n) - 1 = \delta \quad \Rightarrow \quad \beta_n = z_{[(1+\delta)/2]}}$$

and with $\alpha = r$, $\beta = n - r$, $p = 0.5$, we can find $r$ as:

$$\beta_n = z_{[(1+\delta)/2]} = \frac{n - r - 0.5n + 0.5}{\sqrt{n}/2} \quad \Rightarrow \quad \boxed{r = \left\lfloor (n - \sqrt{n}\, z_{[(1+\delta)/2]} + 1)/2 \right\rfloor}$$

and the confidence interval for the median is $\boxed{[Y_r,\ Y_{n-r+1})}$.

- Example 6.8-7
  Find the 95 percent confidence interval for the median for $n = 20$.

  **Solution**:
  We make 20 observations on an RV $X$ and label these $\{X_i, i = 1, \ldots, 20\}$.

  We order them by signed magnitude so that $Y_1 < Y_2 < \ldots < Y_n$. Then, we have

$$P[Y_r \leq x_{[0.5]} < Y_{n-r+1}] = \sum_{i=r}^{n-r} \binom{n}{i} (1/2)^n$$

$$\approx F_{\mathrm{SN}}(\beta_n) - F_{\mathrm{SN}}(\alpha_n) = \delta = 0.95$$

$$(\alpha_n = -\beta_n \text{ due to symmetry}),\ 2F_{\mathrm{SN}}(\beta_n) - 1 = \delta = 0.95$$

$$\Rightarrow \quad \beta_n = z_{[(1+\delta)/2]} = 1.96$$

  We use $r = \left\lfloor (n - \sqrt{n}\, z_{[(1+\delta)/2]} + 1)/2 \right\rfloor$ to obtain $r = 6$.

  Then, $\quad P[Y_6 \leq x_{[0.5]} < Y_{15}] \geq 0.95$,

  i.e., the 95% confidence interval is $[Y_6,\ Y_{15})$.

# Estimation of Vector Means and Covariance Matrices

- Consider a real-valued random vector $\boldsymbol{X} \triangleq [X_1, \cdots, X_p]^T$ with the mean vector $\boldsymbol{\mu_X} \triangleq E[\boldsymbol{X}]$ and the covariance matrix $\boldsymbol{K_{XX}} \triangleq E[(\boldsymbol{X} - \boldsymbol{\mu_X})(\boldsymbol{X} - \boldsymbol{\mu_X})^T]$. Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ denote the $n$ i.i.d. observations of $\boldsymbol{X}$.
- The unbiased mean vector estimator is

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{X}} \triangleq \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i$$

- The unbiased covariance matrix estimator with known $\boldsymbol{\mu_X}$ is

$$\hat{\boldsymbol{K}}_{\boldsymbol{XX}} \triangleq \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{X}_i - \boldsymbol{\mu_X})(\boldsymbol{X}_i - \boldsymbol{\mu_X})^T$$

- The unbiased covariance matrix estimator with unknown $\boldsymbol{\mu_X}$ is

$$\hat{\boldsymbol{K}}_{\boldsymbol{XX}} \triangleq \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}})(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}})^T$$

# Least-Squares (LS) Estimator

- Consider the following (real-valued) linear signal model

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{N}$$

where $\boldsymbol{Y}$ is the $n \times 1$ observation vector, $\boldsymbol{\theta}$ is the $k \times 1$ unknown parameter vector ($k \leq n$), $\boldsymbol{H}$ is a known $n \times k$ matrix, and $\boldsymbol{N}$ is the random noise (measurement error) vector with $E[\boldsymbol{N}] = \mathbf{0}$.

- The LS estimator for $\boldsymbol{\theta}$ minimizes $\left\| \boldsymbol{Y} - \boldsymbol{H}\hat{\boldsymbol{\theta}} \right\|^2$ (i.e., least-squares fit to the data) and is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{Y}$$

- If $\boldsymbol{K}_{\boldsymbol{NN}} = \sigma^2\boldsymbol{I}$, then $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ is a minimum variance unbiased estimator.