# Lecture 9 notes

*Statistical Methods*

---

**Typeset by John Möller**

## Contents

# 1 Page 1

Minimal Variance Unbiased Estimator (MVUE):

**Theorem 1.1** (Lehman-Schally). Let $T(X_1, \dots, X_n)$ be an unbiased estimator of $g(\theta)$.

$$E(T(X_1, \dots, X_n)) = g(\theta)$$

with finite variance. T is MVUE for $\theta$ iff $E(T(X_1, \dots, X_n) = S(X_1, \dots, X_n)) = 0$ for every $S$ s.t. $E(S(X_1, \dots, X_n)) = 0$, $\mathrm{Var}(S) < \infty$, we can use "special estimators" ex BLUE (Best linear unbiased estimator) in lin reg MSE is usually BLUE. CDF:

$$F(x) = P(X \leq x)$$

**Definition 1.1** (Empirical cumulative distribution function).

$$F_n(u, (X_1, \dots, X_n)) = n^{-1} \sum_{i=0}^{n} I_{(-u,u)}(X_i) = \begin{cases} 0, & u \leq X_{(1)} \\ \frac{k}{n} X_{(k)} \leq u \leq X_{k+1} & \\ 1 & u > X_{(n)}. \end{cases}$$

## 1.1 Order statistics

$$X_1, \dots, X_n \Rightarrow X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

[!!image 1]

**Theorem 1.2** (Glivenko-Conteli).

$$F_n \rightarrow_{n \to \infty} F$$

[!add type of convergence]

$$X_i \sim \mathrm{N}(\mu, \sigma^2)$$
$$\overline{X} \sim N(\mu \frac{\sigma^2}{n}$$
$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$
Show:
$$\overline{X} \pm \epsilon$$

$$P(-1.96 \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96) = 95\% / -\frac{\sigma}{\sqrt{n}}$$
$$P(-1.96\frac{\sigma}{\sqrt{n}} \leq \overline{X}-\mu \leq 1.96\frac{\sigma}{\sqrt{n}}) = 95\% / -\overline{X}$$
$$P(-1.96\frac{\sigma}{\sqrt{n}} - \overline{X} \leq -\mu \leq 1.96\frac{\sigma}{\sqrt{n}} - \overline{X}) = 95\% / -(-1)$$
$$P(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 95\%.$$

**Definition 1.2** (Confidence Interval)**.** A RANDOM interval $(L, U)$ $[L, U]$ [?] is a confidence interval for parameter $\theta$ with confidence level $\alpha$ if

$$P(L \leq \theta \leq U) = 1 - \alpha$$

Confidence interval $(L, U)$ covers $\theta$ with $1 - \alpha$ probability.

$$\overline{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

is a $\alpha$-level confidence interval for $\mu$.

Show:

**Problem 1.1.** Show $(-\infty, \overline{X}+Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})$ and $(\overline{X}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \infty)$ are also $(1-\alpha)$-confidence intervals for $\mu$.

### 1.1.1 exe:

**Problem 1.2.** Show sizes of confidence intervals.

**"Statistic:"** $H_\theta = h($ data $, \theta)$, with a known distribution.

$P(L \leq H_\theta \leq U) = 1 - \alpha$.

**Fact:** If $X_i \sim \mathrm{N}(0, 1)$ and $Y = X_1^2 + \cdots + X_n^2$ then $Y \sim \chi^2_k$ where $k$ is the degrees of freedom.

Let $X_i \sim \mathrm{N}(\mu, \sigma^2)$. Since $\frac{X_i - \mu}{\sigma} \sim \mathrm{N}(0, 1)$ that means

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n.$$

"More difficult:"

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}.$$

**Definition 1.3** (Sample Variance)**.**

$$s^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

Thus:

**Theorem 1.3.**

$$\frac{n-1}{\sigma^2}s^2 \sim \chi^2_{n-1}.$$

Quantile of $\chi^2_n$ of $1-\alpha$ level $\chi^2_\alpha(n)$ is a point s.t. $P(Y \geq \chi^2_\alpha(n)) = \alpha$, $Y \sim X^2_n$
[!Insert two graphs]
[!Insert associated calculations]

**Theorem 1.4.** Let $X_1, \dots, X_n$ be on i.i.d sample from $\mathrm{N}(\mu, \sigma^2)$, where $\mu$ and $\delta$ are unknow, then,

$$\left[\frac{(n-1)\cdot s^2}{\chi_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}(n-1)}\right]$$

is an $\alpha$-level confidence interval for $\sigma^2$.

Let $X_i \sim \mathrm{N}(\mu, \sigma^2)$ then

$$\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$

where $s = \sqrt{s^2}$.

**Problem 1.3.** Show $\overline{X} \pm t_{\alpha/2}(n-1)\cdot \frac{s}{\sqrt{n}}$ is a $1-\alpha$ level quantile for $\mu$.

## 1.2 Statistical Hypothesis Testing

$$\begin{aligned} H_0 :& \mu = \mu_0 \\ H_1 :& \mu \neq \mu_0 \end{aligned}$$

Construct a test statistic

$$T = T(X_1, \dots, X_n | H_0) \sim \tilde{f}(\mu_0).$$

$$X_1, \dots, X_n$$

is a random sample

$$T(X_1, \dots, X_n) \in \text{ critical set } \theta_1$$

If $T \in \theta_1$ then we reject the null hypothesis.
If $T \notin \theta_1$ then we fail to reject the null hypothesis.

$$P(T \in \theta_1 | H_0 \text{ is true }) = \alpha.$$