MVUE = Minimum variance unbiased estimators

Information in a sample with respect to \( \theta \)

\[
I(\theta) = -E \left[ \frac{\partial^2 \ell}{\partial \theta^2} \right] = -E \left[ \frac{\partial^2}{\partial \theta^2} \left( \ln L(x; \theta) \right) \right]
\]

Cramér-Rao lower bound

If \( \hat{\theta} = u(x_1, \ldots, x_n) \) is an estimator for \( \theta \) and \( E(\hat{\theta}) = \theta \) then

\[
\text{Var}(\hat{\theta}) \geq \frac{[K'(\theta)]^2}{I(\theta)}
\]

Under regularity conditions

\( \Rightarrow \) Suppose you have an unbiased estimator

\[
\text{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}
\]

Can we find estimators that reach this lower bound? \( \Rightarrow \) MVUE
Generally, that is a hard task. Typically, all you can do is verify that an estimator is the MVUE.

\[ X_1, \ldots, X_n \sim \text{Exp}(\lambda), \quad E(X_i) = \lambda \]

NLE for \( \lambda \)?

\[ \hat{\lambda} = \bar{X} \]

Is the NLE the MVUE?

1) Is it unbiased?

\[ E(\hat{\lambda}) = E(\bar{X}) = E\left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n} n \lambda = \lambda \]

2) \( \text{Var}(\hat{\lambda}) = \text{Var}\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) \]

\[ = \frac{1}{n^2} n \lambda^2 = \frac{\lambda^2}{n} \]
\[ L(\lambda) = \prod_{i=1}^{n} \frac{1}{x_i} \exp\left(-\frac{x_i}{\lambda}\right) \]

Let \( \lambda \) be the parameter.

\[ l(\lambda) = -n \log \lambda - \frac{1}{\lambda} \sum x_i \]

\[ \frac{\partial l}{\partial \lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda^2} \sum x_i \]

\[ \frac{\partial^2 l}{\partial \lambda^2} = \frac{n}{\lambda^2} - \frac{2}{\lambda^3} \sum x_i \]

\[-E \left[ \frac{\partial^2 l}{\partial \lambda^2} \right] = -E \left[ \frac{n}{\lambda^2} - \frac{2}{\lambda^3} \sum x_i \right] = -\frac{n}{\lambda^2} + \frac{2}{\lambda^3} n \lambda \]

Since \( \hat{\lambda} \) is unbiased and \( \text{Var} (\hat{\lambda}) = \frac{1}{\lambda^2} \) then it is the MVUE.
Efficient estimator: one that attains the Cramer-Rao lower bound.

The MVUE is the efficient unbiased estimator.

Assymptotically efficient estimator.

$\hat{\theta}$ is asymptotically efficient.

$$\lim_{n \to \infty} \frac{\text{Var}(\hat{\theta})}{[I(\theta)]^2} = 1$$

Under regularity conditions, the MLE is asymptotically efficient.

More generally, under regularity conditions

$\hat{\theta} - \theta \xrightarrow{D} N(0, I(\theta))$

CLT
\[ \Rightarrow I(\theta) (\hat{\theta} - \theta) \rightarrow N(0,1) \]
\[ I(\theta) (\hat{\theta} - \theta_0) \rightarrow N(0,1) \]

Examples

1) \( X_1, \ldots, X_n \sim \text{Uni}[0, \theta] \)

Is \( \hat{\theta} \) asymptotically normal in this case?

\[ \Rightarrow \text{No, at least from the results that we have discussed because this likelihood does not satisfy regularity conditions.} \]

2) Let \( T_i \) be the lifetime of an electronic component and assume that \( T_i \sim \text{Exp}(\lambda) \) independently for each \( i \).
The data consists of arises from an experiment in which the components are observed for up to D hours. If the component fails, the lifetime is recorded, but if it lives more than D hours, only the maximum lifetime is recorded.

The observation available are

- \( T_1, \ldots, T_m \) = components that failed before D
- \( T_{m+1} \ldots T_n \) = components that did not fail during the study

Get the MLE for \( \lambda \).

\[
L(\lambda) = \prod_{i=1}^{m} \frac{1}{x} \exp(-\frac{1}{x} Ti) \times \prod_{j=m+1}^{n} \int_{0}^{\infty} \frac{1}{x} \exp(-\frac{1}{x} Ti) dTi
\]
\[
\int_0^\infty \frac{1}{x} \exp\left\{- \frac{1}{x} T_i \right\} dT_i = \left[ -\exp\left\{- \frac{1}{x} T_i \right\} \right]_0^\infty = \exp\left\{- \frac{D}{x} \right\}
\]

\[
L(\lambda) = \left( \frac{1}{\lambda} \right)^m \exp\left\{- \frac{1}{\lambda} \sum_{i=1}^{n-m} T_i \right\} \cdot \left[ \exp\left\{- \frac{D}{\lambda} \right\} \right] = (\lambda)^I
\]

\[
l(\lambda) = -m \log \lambda - \frac{1}{\lambda} \left[ \sum_{i=1}^{n-m} T_i + D(n-m) \right]
\]

\[
\frac{d}{d\lambda} = -\frac{m}{\lambda} + \frac{1}{\lambda^2} \left[ \sum_{i=1}^{n-m} T_i + (n-m)D \right] = 0
\]

\[
\lambda = \frac{\sum_{i=1}^{n-m} T_i + (n-m)D}{m}
\]
Can you argue that this MLE is consistent?

Yes, because the regularity conditions are satisfied.

It also satisfies the conditions for the existence of a CLT

\[ I(\lambda) = \mathbb{E} \left( \frac{\partial^2 \ell}{\partial \lambda^2} \right) = \mathbb{E} \left[ \frac{m}{\lambda^2} - \frac{2}{\lambda^3} \left\{ (n-m) \mathbb{D} + \frac{m}{\lambda} \sum_{i=1}^{m} T_i \right\} \right] = \mathbb{E} \left[ \frac{m}{\lambda^2} - \frac{2}{\lambda^3} (n-m) \mathbb{D} + \frac{m}{\lambda} \sum_{i=1}^{m} T_i \right] \]

\( (\hat{\lambda} - \lambda_0) \sim \mathcal{N}(0, \text{I}(\lambda)) \)
What would a 95% confidence interval for \( \lambda \) be?

\[
\frac{\hat{\lambda} - \lambda}{\sqrt{I(\hat{\lambda})}} \sim N(0,1)
\]

\( T = T(\theta) \)

\( t \geq T(\alpha) \)

(a, b) such that

\[ \Pr(a < (\hat{\lambda} - \lambda)I(\hat{\lambda}) < b) = 0.95 \]

-1.96 \( \leq (\hat{\lambda} - \lambda)I(\hat{\lambda}) \leq 1.96 \)

\[ \Rightarrow \lambda - \frac{1.96}{I(\hat{\lambda})} < \lambda < \hat{\lambda} + \frac{1.96}{I(\hat{\lambda})} \]

\[ X \sim N(0,1) \]

\[ X = -X \]

\[ X \not= -X \]

\[ X \sim N(0,1) \]
Likelihood ratio tests

$H_0: \theta = \theta_0 \quad \leftrightarrow \quad H_1: \theta \neq \theta_0$

\[ T = \frac{L(\theta_0)}{L(\hat{\theta})} \quad T \leq 1 \text{ always} \]

$T$ close to 1 favors the null hypothesis

$T$ close to 0 favors the alternative hypothesis

Reject if $T < C$ (0 < C < 1)

where $C$ is chosen to reflect the type I error you want for the test.
If \( c \) is small \( \Rightarrow \) You reject fewer times \( \Rightarrow \) \( c \) moves to the left (closer to zero)

If \( c \) is large \( \Rightarrow \) You don't care about rejecting the null \( \Rightarrow \) \( c \) moves to the right (closer to 1)
Example 6.3.1

H₀: λ = λ₀ \quad H₀: λ ≠ λ₀

X₁, ⋯, Xₙ \sim \text{Exp}(λ)

\[ L(λ) = \left( \frac{1}{λ} \right)^n \exp \left\{ -\frac{1}{λ} \sum_{i=1}^{n} X_i \right\} \]

\[ \hat{λ} = \frac{\bar{X}}{\bar{X}} \]

\[ L(\hat{λ}) = \left( \frac{1}{\bar{X}} \right)^n \exp \left\{ -\frac{1}{\bar{X}} \sum_{i=1}^{n} X_i \right\} = \left( \frac{1}{\bar{X}} \right)^n \exp \left\{ -n \bar{X} \right\} \]

\[ T = \frac{L(λ₀)}{L(\hat{λ})} = \frac{\left( \frac{1}{λ₀} \right)^n \exp \left\{ -\frac{n \bar{X}}{λ₀} \right\}}{\left( \frac{1}{\bar{X}} \right)^n \exp \left\{ -n \bar{X} \right\}} \]

= \left( \frac{\bar{X}}{λ₀} \right)^n \exp \left\{ -\frac{n \bar{X}}{λ₀} \right\} \exp \left\{ n \right\}

Reject \ H₀ \ if \ \ T < C \ where \ C \ is \ to \ be \ chosen.
Now we know that
\[ \sum x_i \sim \text{Gamma}(n, \lambda) \]
\[ \Rightarrow \frac{\sum x_i}{\lambda_0} \sim \text{Gamma}\left(\frac{n}{\lambda_0}, \lambda\right) \]
\[ \Pr(\text{type 1 Error}) = \alpha \]
\[ = \Pr(\text{Rejecting null} \mid \text{Null true}) \]
\[ \lambda = \lambda_0 \]
\[ \Rightarrow \text{If the null is true} \]
\[ \frac{\sum x_i}{\lambda_0} \sim \text{Gamma}(n, 1) \]
\[ 2 \frac{\sum x_i}{\lambda_0} \sim \text{Gamma}(n, 2) = X_{2n}^2 \]
Formally, to get the LRT you need \( a \) and \( b \) to satisfy:

\[
\Pr(\chi^2_{2n} < a) + \Pr(\chi^2_{2n} > b) = \alpha
\]

and

\[
f(2an) = f(2bn)
\]

where

\[
f(t) = t^n \exp\{-nt\}
\]

Once you have \( a \) and \( b \) you have the rejection region.

In practice, the LRT is approximated by taking the probability in both tails to be the same.
The result is a bit more general. 

$$T = -\ln \left( \frac{L(\theta_0)}{\max_{\theta \in \Theta^0} L(\theta)} \right)$$

For the likelihood ratio,

$$H_0: \theta \in \Theta_0 \quad vs \quad H_a: \theta \in \Theta^0.$$ 

Rejection region: 

$A < C$ where $C$ is obtained to match a certain fixed type I error.

What if $-A$ does not have any distribution that I recognize?

Wilk's theorem:

Under regularity conditions:

$$-2 \log A \sim \chi^2_p$$

"$p$ is the difference in the number of tree parameters"
$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$  $\mu, \sigma^2$ are unknown

$H_0: \mu = \mu_0$ vs  $H_a: \mu \neq \mu_0$

Wald-type tests