BIOSTAT 651
Notes #3: Maximum Likelihood

- Lecture Topics:
  - Maximum likelihood estimation (MLE)
  - Hypothesis testing

Data Structure

- The general set-up is described as follows:
  - sample size: n subjects (independent)
  - response: Y_i
  - covariate $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \ldots)$
  - model parameters: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^T$

  - set $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$
  - design matrix:

$$\mathbf{X} \quad = \quad \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}$$

- e.g., linear regression: $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2)^T$

Parameter Estimation

- Consider a model of Y_i based on the parameter $\boldsymbol{\theta}$
  - observed data: $(\mathbf{x}_i^T, Y_i)$ for $i = 1, \ldots, n$
  - fitted values: $\widehat{Y}_i$
- Different choices of $\widehat{\boldsymbol{\theta}}$ will yield different $\widehat{\mathbf{Y}}$

  Q: How to select the "best" $\widehat{\boldsymbol{\theta}}$?
- In linear regression we used LSE, which minimize the following function:

$$S_2 = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$$

- Many other possible criteria exist:

$$\begin{aligned} S_1 &= \sum_{i=1}^{n} |Y_i - \widehat{Y}_i| \\ S_\infty &= \max_{i=1,\ldots,n} |Y_i - \widehat{Y}_i| \end{aligned}$$

- Another well-known method: Maximum Likelihood

Likelihood

- density: $f(Y_i;\boldsymbol{\theta})$
- joint density: $f(\mathbf{Y};\boldsymbol{\theta}) = \prod_{i=1}^{n} f(Y_i;\boldsymbol{\theta})$
  - calculation based on various $\mathbf{Y}$ values, for fixed $\boldsymbol{\theta}$
- likelihood function: $L(\boldsymbol{\theta};\mathbf{X},\mathbf{Y})$
  - often abbreviated to $L(\boldsymbol{\theta})$, or even L
  - viewed as a function of $\boldsymbol{\theta}$, with $(\mathbf{X},\mathbf{Y})$ held constant (at their realized values)
- likelihood is proportional to the joint density,

$$L(\boldsymbol{\theta}) \quad \propto \quad f(\mathbf{Y};\boldsymbol{\theta})$$

  - derived by setting $L(\boldsymbol{\theta}) = f(\mathbf{Y};\boldsymbol{\theta})$, then deleting multiples that are *not* functions of $\boldsymbol{\theta}$

Likelihood Principles (continued)

- Assuming that $Y_1, \ldots, Y_n$ are independent,

$$L(\boldsymbol{\theta}) \quad \propto \quad \prod_{i=1}^{n} f_i(Y_i; \boldsymbol{\theta}),$$

- if, in addition, the Y_i's are identically distributed,

$$L(\boldsymbol{\theta}) \quad \propto \quad \prod_{i=1}^{n} f(Y_i; \boldsymbol{\theta}),$$

- in most cases we consider, the $(\mathbf{x}_i^T, Y_i)$ will be independent and identically distributed

Maximum Likelihood Estimators (MLE)

- A *Maximum Likelihood Estimator* (MLE) is a maximizer of the likelihood function $L(\boldsymbol{\theta}|\mathbf{Y})$, denoted as $\hat{\boldsymbol{\theta}}$, i.e.

$$L(\widehat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}).$$

  where $\boldsymbol{\Theta}$ is the parameter space

  - Note: MLE is also an maximizer of the log-likelihood, $\ell(\boldsymbol{\theta})$.

- For a given parametric model, maximum likelihood identifies the parameter values which make the realized data "most likely"

MLE: Functions

- For convenience, we often maximize the log-likelihood function:

$$\ell(\boldsymbol{\theta}) \quad = \quad \log L(\boldsymbol{\theta})$$

- score function,

$$U(\boldsymbol{\theta}) \quad = \quad \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\theta})$$

- *observed* information,

$$J(\boldsymbol{\theta}) \quad = \quad \frac{-\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \ell(\boldsymbol{\theta})$$

- *expected* information,

$$I(\boldsymbol{\theta}) \quad = \quad E\left[J(\boldsymbol{\theta})\right] = -E\left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \ell(\boldsymbol{\theta})\right]$$

  - $J(\boldsymbol{\theta})$ may be easier to calculate than $I(\boldsymbol{\theta})$
  - In the book, $\mathfrak{I}$ represents the expected information.

Score Function

- When $\ell(\boldsymbol{\theta})$ is differentiable w.r.t. $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$ can typically be obtained as the solution to the score equation, $U(\boldsymbol{\theta}) = \mathbf{0}$, where

$$U(\boldsymbol{\theta}) \quad = \quad \begin{bmatrix} \frac{\partial \ell}{\partial \theta_1} \\ \frac{\partial \ell}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ell}{\partial \theta_q} \end{bmatrix}$$

- This will work if $J(\boldsymbol{\theta})$ is positive-definite, where

$$J(\boldsymbol{\theta}) \quad = \quad - \begin{bmatrix} \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_1} & \cdots & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell}{\partial \theta_q \partial \theta_1} & \cdots & \frac{\partial^2 \ell}{\partial \theta_q \partial \theta_q} \end{bmatrix}$$

Information Matrix

- Expected information is calculated as:

$$I(\boldsymbol{\theta}) = -E\begin{bmatrix} \frac{\partial^2 \ell}{\partial\theta_1\partial\theta_1} & \cdots & \frac{\partial^2 \ell}{\partial\theta_1\partial\theta_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell}{\partial\theta_q\partial\theta_1} & \cdots & \frac{\partial^2 \ell}{\partial\theta_q\partial\theta_q} \end{bmatrix}$$

- We then have

$$\begin{aligned} J(\boldsymbol{\theta}) &= -\frac{\partial U^T}{\partial\boldsymbol{\theta}} \\ I(\boldsymbol{\theta}) &= -E\left[\frac{\partial U^T}{\partial\boldsymbol{\theta}}\right] \end{aligned}$$

MLE: Functions (cont'd)

- In the *iid* setting, we can write,

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} L_i(\boldsymbol{\theta}) \\
\ell(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}) \\
U(\boldsymbol{\theta}) &= \sum_{i=1}^{n} U_i(\boldsymbol{\theta}) \\
I(\boldsymbol{\theta}) &= \sum_{i=1}^{n} I_i(\boldsymbol{\theta}) \\
J(\boldsymbol{\theta}) &= \sum_{i=1}^{n} J_i(\boldsymbol{\theta})
\end{aligned}
$$

MLE: Score and Information

- It can be shown that,

$$\begin{aligned} E[U(\boldsymbol{\theta}_0)] &= \mathbf{0} \\ V[U(\boldsymbol{\theta}_0)] &= E[U(\boldsymbol{\theta}_0)^{\otimes 2}] = I(\boldsymbol{\theta}_0), \end{aligned}$$

  where $\boldsymbol{\theta}_0$ is the true underlying value of $\boldsymbol{\theta}$ and $\mathbf{z}^{\otimes 2} = \mathbf{z}\mathbf{z}^T$

- Note:

$$\begin{aligned} V[U(\boldsymbol{\theta}_0)] &= E[U(\boldsymbol{\theta}_0)^{\otimes 2}] \\ &= E\left[\sum_{i=1}^{n} U_i(\boldsymbol{\theta}_0) \sum_{j=1}^{n} U_j(\boldsymbol{\theta}_0)^T\right] \\ &= E\left[\sum_{i=1}^{n} U_i(\boldsymbol{\theta}_0)^{\otimes 2}\right] \\ &= nE\left[U_1(\boldsymbol{\theta}_0)^{\otimes 2}\right] \\ &= nI_1(\boldsymbol{\theta}_0) \end{aligned}$$

Maximum Likelihood Estimation

- Maximum likelihood estimator, $\widehat{\boldsymbol{\theta}}$, computed by solving the score equation,

$$U(\boldsymbol{\theta}) \quad = \quad \mathbf{0}$$

- Note: maximizer may lie on the boundary of $\boldsymbol{\Theta}$, in which case the MLE is ill-behaved.
  - in BIOSTAT 651, we assume that $\ell(\boldsymbol{\theta})$ is *concave*, with information matrix assumed to be positive-definite

MLE Example: Normal

- Example: Suppose that $Y_i \sim N(\mu, \sigma^2)$ with σ^2 known. Determine the MLE of μ.

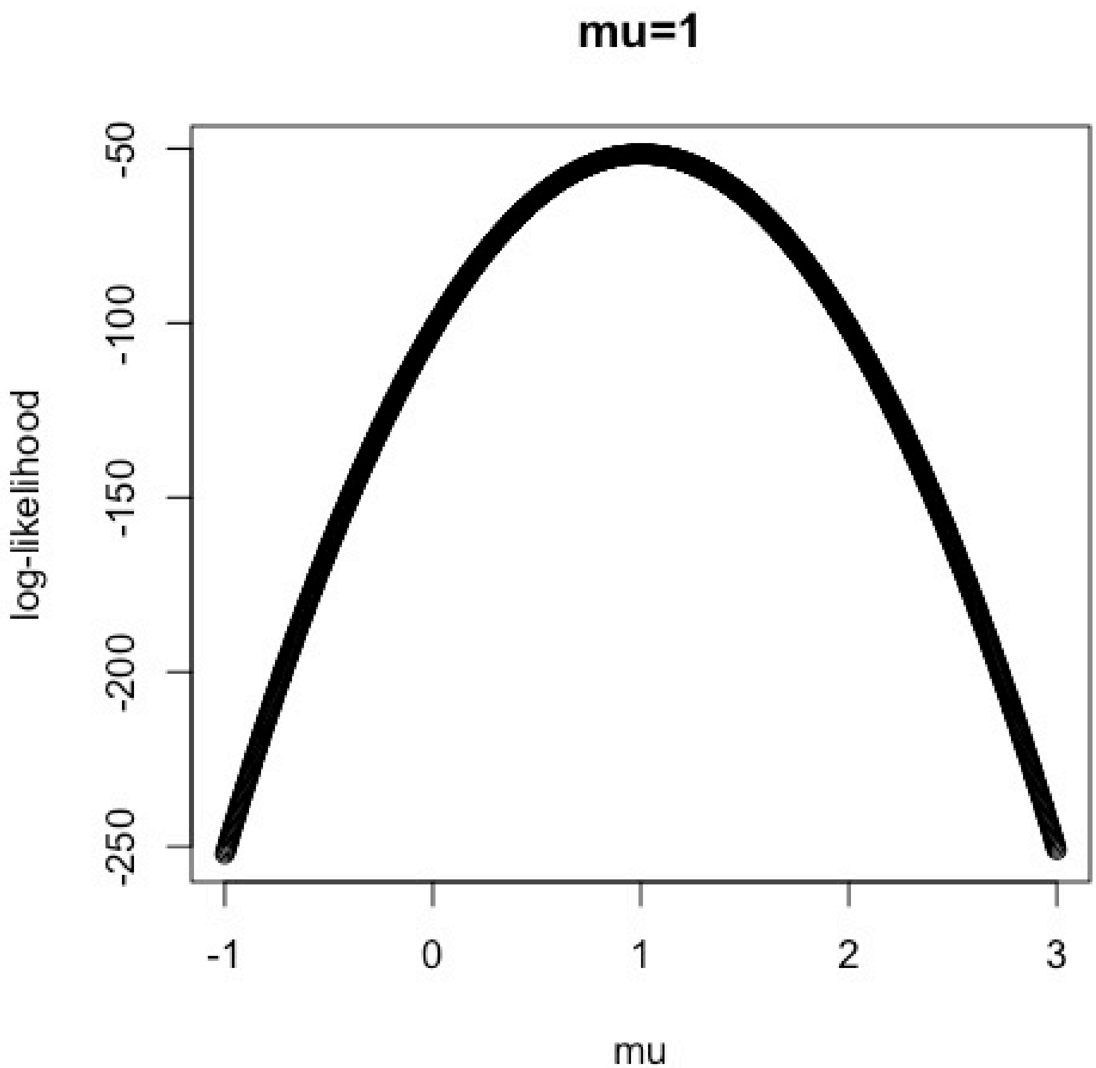
$$
\begin{aligned}
f(Y_i;\mu) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-(Y_i-\mu)^2/(2\sigma^2)} \\
L_i(\mu) &= e^{-(Y_i-\mu)^2/(2\sigma^2)} \\
\ell_i(\mu) &= -(Y_i-\mu)^2/(2\sigma^2) \\
U_i(\mu) &= (Y_i-\mu)/\sigma^2 \\
U(\mu) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\mu) \\
\widehat{\mu} &= \overline{Y}
\end{aligned}
$$

- Note:

$$
J(\mu) = I(\mu) = -\frac{\partial U}{\partial \mu} \quad = \quad \frac{n}{\sigma^2}
$$

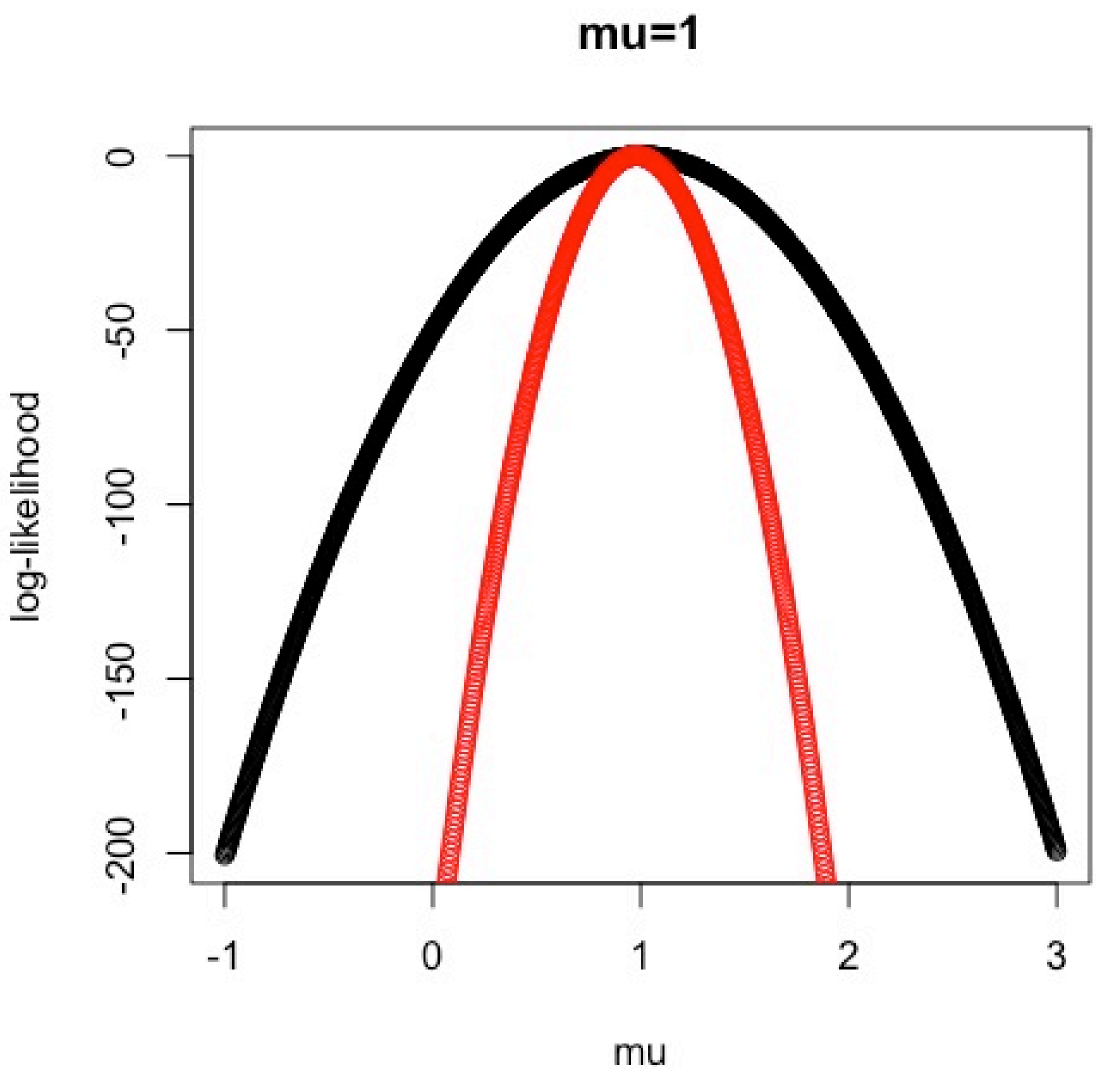
MLE Example: Normal

- Log likelihood function (n=100)



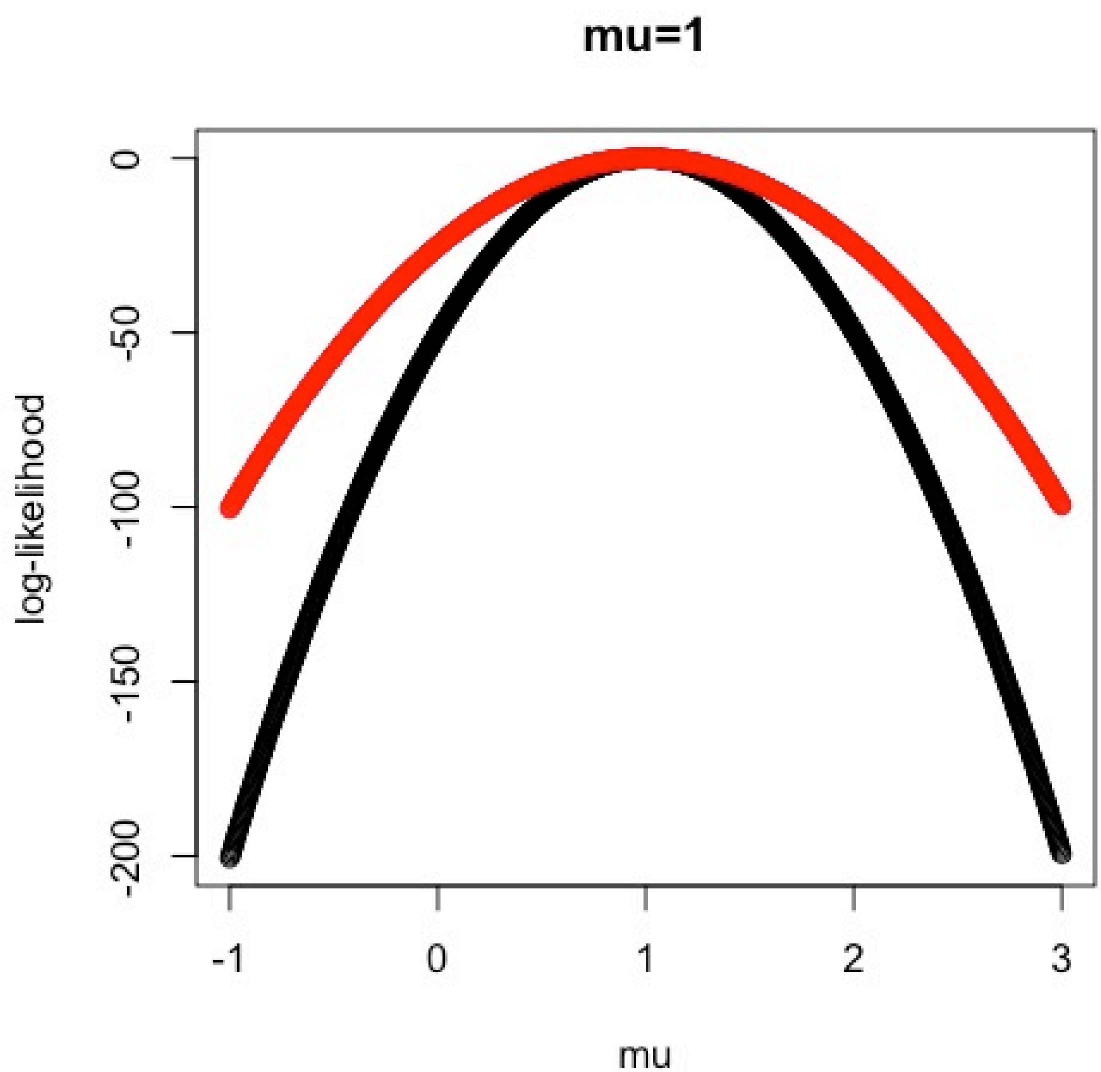
MLE Example: Normal - Fisher Information

- $J(\mu) = I(\mu) = -\frac{\partial U}{\partial \mu} = \frac{n}{\sigma^2}$
- $n = 100$(black) vs. $n = 500$(red)



MLE Example: Normal - Fisher Information

- $J(\mu) = I(\mu) = -\frac{\partial U}{\partial \mu} = \frac{n}{\sigma^2}$
- $\sigma^2 = 1$(black) vs. $\sigma^2 = 2$(red)



MLE Example: Binomial

- <u>Example</u>: Suppose that $Y_\bullet = Y_1 + \ldots + Y_n$ follows a Binomial distribution with parameter π. Compute the MLE of π.

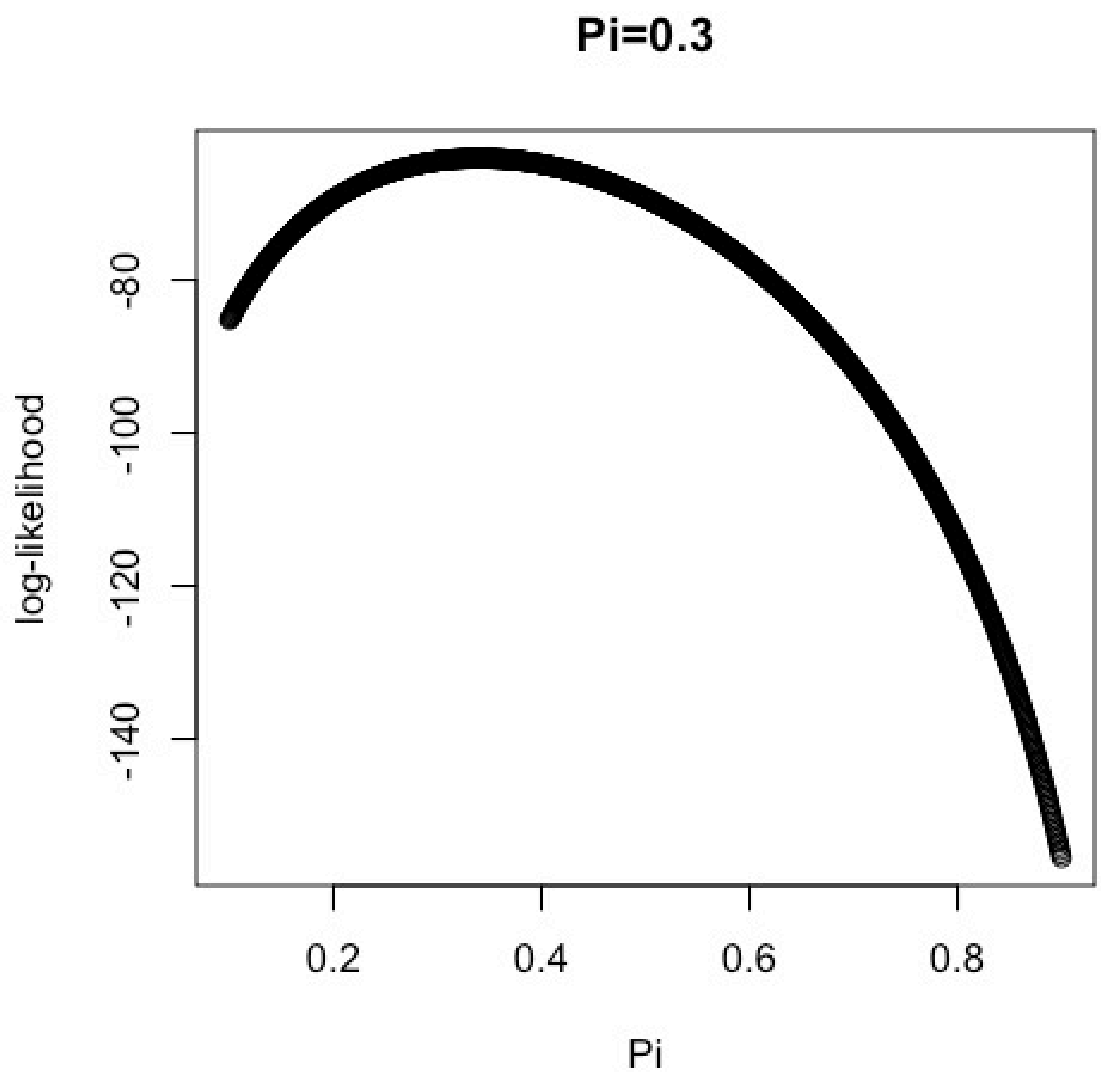
$$
\begin{aligned}
p(Y;\pi) &= \binom{n}{Y} \pi^Y (1-\pi)^{n-Y} \\
L(\pi) &= \pi^Y (1-\pi)^{n-Y} \\
\ell(\pi) &= Y \log(\pi) + (n-Y)\log(1-\pi) \\
U(\pi) &= \frac{Y}{\pi} - \frac{n-Y}{1-\pi} \\
\widehat{\pi} &= \overline{Y}
\end{aligned}
$$

- Second derivative,

$$
\frac{\partial U}{\partial \pi} = \frac{-Y}{\pi^2} - \frac{n-Y}{(1-\pi)^2}
$$

MLE Example: Binomial

- Log likelihood function (n=100)



MLE Example: Poisson Case

- Example: Suppose that Y_i is distributed as Poisson(θ) for $i = 1, \ldots, n$. Determine the maximum likelihood estimator of θ.

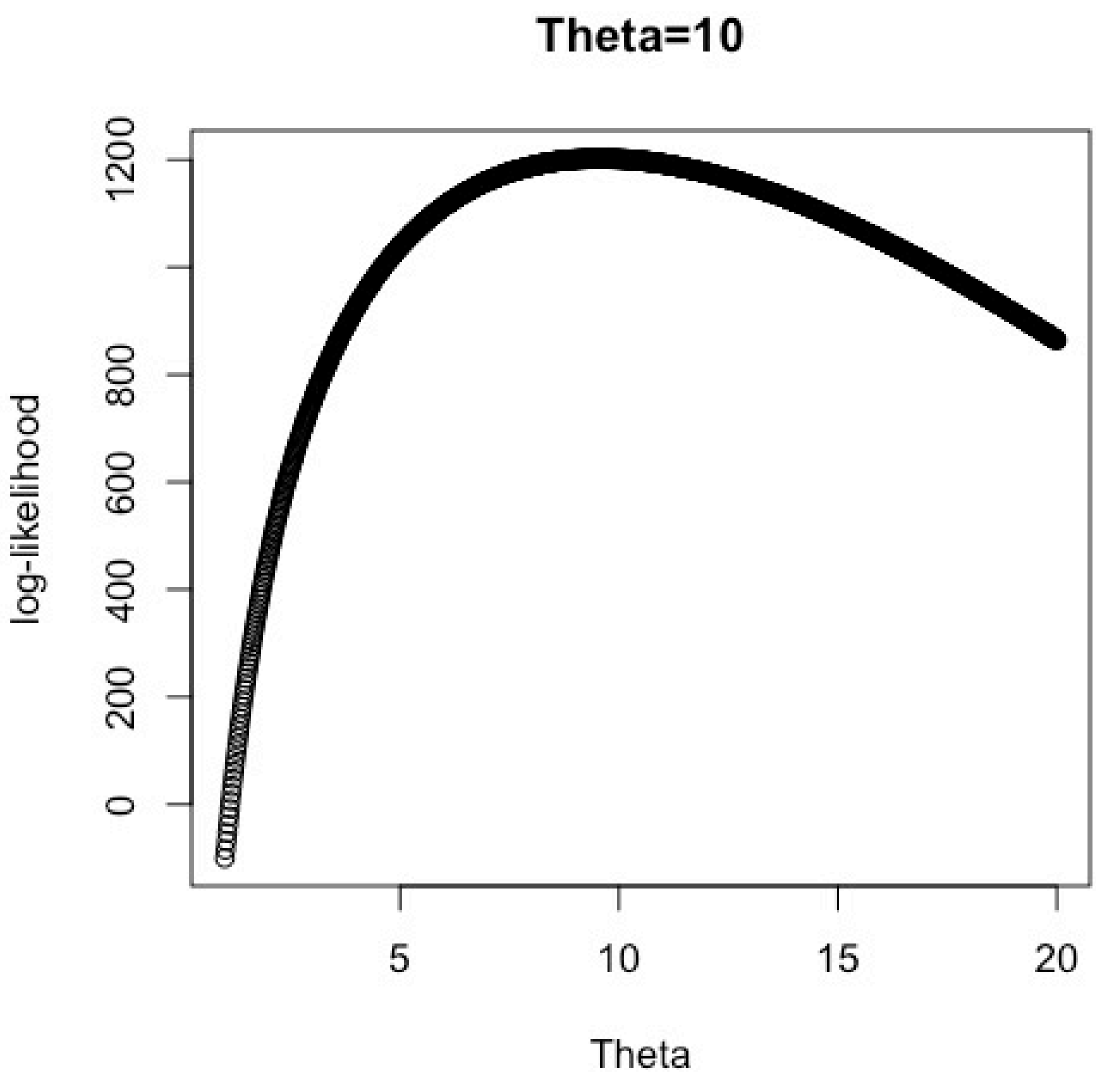
$$
\begin{aligned}
f(Y_i; \theta) &= \frac{e^{-\theta}\theta^{Y_i}}{Y_i!} \\
L_i(\theta) &= e^{-\theta}\theta^{Y_i} \\
\ell_i(\theta) &= -\theta + Y_i \log \theta \\
U_i(\theta) &= -1 + \frac{Y_i}{\theta} \\
J_i(\theta) &= \frac{Y_i}{\theta^2} \\
U(\theta) &= \ldots \\
\widehat{\theta} &= \ldots
\end{aligned}
$$

- Expected and observed information:

$$
\begin{aligned}
J(\theta) &= \\
I(\theta) &=
\end{aligned}
$$

MLE Example: Poisson

- Log likelihood function (n=100)



Maximum Likelihood Estimation

- Usually, a closed-form solution for $\widehat{\boldsymbol{\theta}}$ is not available
  - ex) Logistic regression
- Need to solve $U(\boldsymbol{\theta}) = \mathbf{0}$ through iterative methods

  e.g., Newton-Raphson ...

Newton-Raphson Procedure

- Pre-specify tolerance, ξ; start with an initial "estimate", $\widehat{\boldsymbol{\theta}}_{(0)}$, and
  - e.g., $\widehat{\boldsymbol{\theta}}_{(0)} = \mathbf{0}$, with $\xi = 10^{-4}$
- Update the estimate,

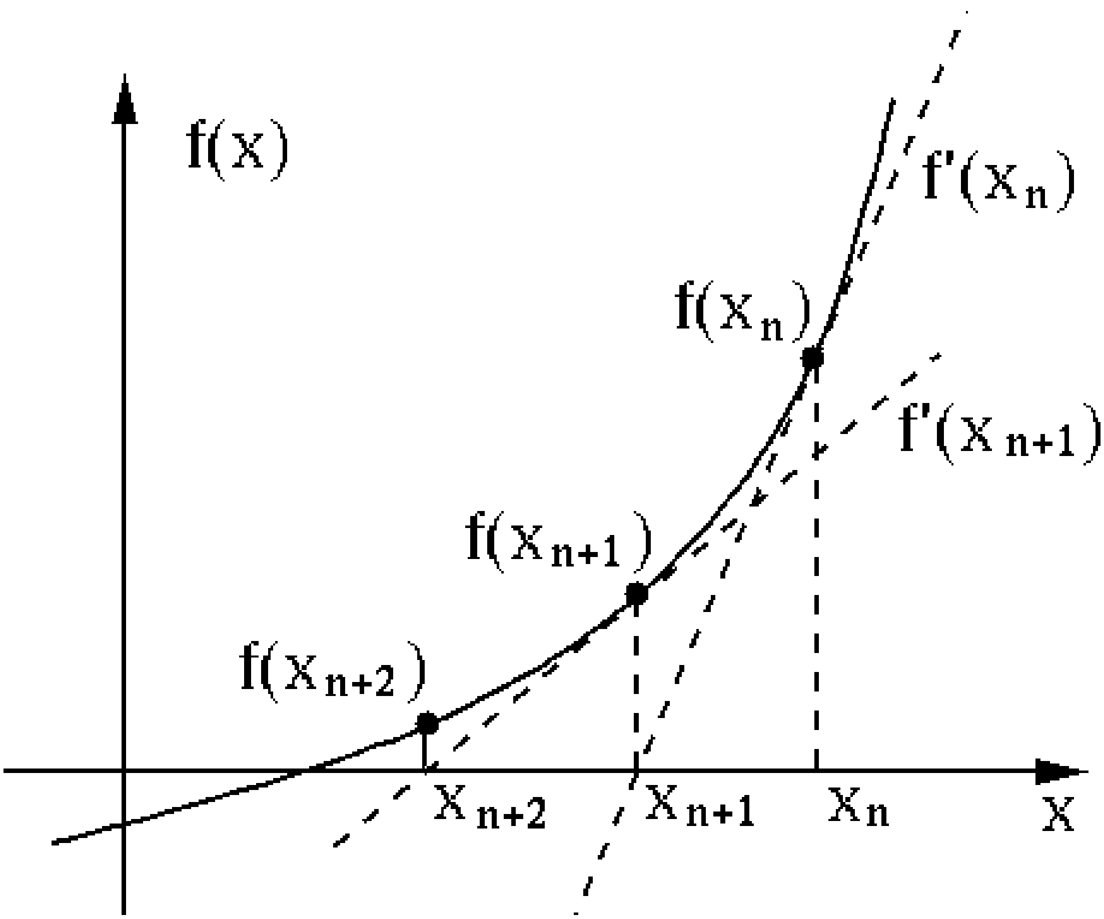
$$\widehat{\boldsymbol{\theta}}_{(j+1)} = \widehat{\boldsymbol{\theta}}_{(j)} + J^{-1}(\widehat{\boldsymbol{\theta}}_{(j)})U(\widehat{\boldsymbol{\theta}}_{(j)})$$

- Continue until until convergence is attained; e.g.,

$$\begin{aligned} ||\widehat{\boldsymbol{\theta}}_{(j+1)} - \widehat{\boldsymbol{\theta}}_{(j)}|| &< \xi \\ ||U(\widehat{\boldsymbol{\theta}}_{(j)})|| &< \xi \end{aligned}$$

where $||\mathbf{z}|| = (\mathbf{z}^T\mathbf{z})^{1/2}$

Newton-Raphson Procedure



From
fourier.eng.hmc.edu/e176/lectures/NM/node20.html

Properties of MLEs

Properties of MLEs

- Under certain regularity conditions:
  - $\widehat{\boldsymbol{\theta}}$ is the unique maximizer of $\ell(\boldsymbol{\theta})$
  - $\widehat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$

$$\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$$

  - $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges to a mean zero Normal with a covariance $I_1(\boldsymbol{\theta}_0)^{-1}$

$$n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \Rightarrow N(0, I_1(\boldsymbol{\theta}_0)^{-1})$$

  - $n^{-1}J(\widehat{\boldsymbol{\theta}}) \xrightarrow{p} I_1(\boldsymbol{\theta}_0)$

Invariance Property

- If $\widehat{\boldsymbol{\theta}}$ is the MLE for $\boldsymbol{\theta}_0$, then $g(\widehat{\boldsymbol{\theta}})$ will be the MLE for $g(\boldsymbol{\theta}_0)$
  - i.e, assuming $g(\cdot)$ is a well-behaved function
  - e.g., continuous (differentiable)
- Application: depending on the specifics of the likelihood, it may be more convenient to maximize $L(g(\boldsymbol{\theta}))$ than $L(\boldsymbol{\theta})$
  - obtain $\widehat{g(\boldsymbol{\theta})}$,

    then obtain $\widehat{\boldsymbol{\theta}} = g^{-1}\{\widehat{g(\boldsymbol{\theta})}\}$
- e.g., set $g(\boldsymbol{\theta}) = \log \boldsymbol{\theta}$

MLE: Interval Estimation

- Given the afore-listed large-sample properties of MLEs, interval estimators can be computed using the Normal approximation ...
  - e.g., 95% confidence interval for θ_0:

  - e.g., 95% confidence interval for $g(\theta_0)$:

- Note: could use *Delta Method* to compute interval estimate of a function of θ_0
  - e.g., 95% confidence interval for $g(\theta_0)$:

Example: CI, Normal

- Example: We return to the case where $Y_i \sim N(\mu, \sigma^2)$ with σ^2 known. Compute a 95% confidence interval for μ.

  Recall that

  $$\begin{aligned} \widehat{\mu} &= \overline{Y} \\ \frac{\partial U}{\partial \mu} &= \frac{-n}{\sigma^2} \end{aligned}$$

  We then have

  $$\begin{aligned} J(\mu) &= \\ V(\widehat{\mu}) &= \end{aligned}$$

  such that a 95% CI is then given by:

CI Example: Binomial

- Example: We revisit the setting in which $Y_\bullet = Y_1 + \ldots + Y_n$ follows a Binomial distribution with parameter π. Compute an interval estimate of π.

  Recall that:

$$\begin{aligned} \widehat{\pi} &= \overline{Y} \\ \frac{\partial U}{\partial \pi} &= \frac{-Y}{\pi^2} - \frac{n-Y}{(1-\pi)^2} \end{aligned}$$

  such that

$$\begin{aligned} J(\pi) &= \frac{Y}{\pi^2} + \frac{n-Y}{(1-\pi)^2} \\ I(\pi) &= \frac{n}{\pi} + \frac{n}{(1-\pi)} = \frac{n}{\pi(1-\pi)} \end{aligned}$$

- Therefore, the CI is given by:

CI Example: Binomial (continued)

- Q1: What is one limitation of the CI just derived?

- Q2: How to remedy?

CI Example: Poisson Case

- Example: Determine an interval estimator for the case where Y_i is distributed as Poisson(θ) for $i = 1, \ldots, n$.

- Based on previous calculations,
$$\begin{aligned} \widehat{\theta} &= \overline{Y} \\ I(\theta) &= \frac{n}{\theta} \end{aligned}$$
such that we obtain the 95% CI as

- Q1: Problem with this estimator?

- Q2: Solution?

Hypothesis Testing

- For the next few slides, we consider the following setting:
  - let $\boldsymbol{\theta}$ be a $q \times 1$ parameter, partitioned as, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$, where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are of dimension $q_1 \times 1$ and $q_2 \times 1$, respectively
  - we wish to test $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1H}$ versus $H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{1H}$
  - estimation is based on ML
  - let $\widehat{\boldsymbol{\theta}}_H$ be the MLE, constrained by H_0; i.e., $\widehat{\boldsymbol{\theta}}_H = (\boldsymbol{\theta}_{1H}^T, \widehat{\boldsymbol{\theta}}_{2H}^T)^T$,
- Three most commonly used tests: Score, Wald and Likelihood ratio

Score Test

- Score test makes use of asymptotic result that $U(\boldsymbol{\theta}_0) \sim N(\mathbf{0}, I(\boldsymbol{\theta}_0))$
- Score test statistic:

$$U(\widehat{\boldsymbol{\theta}}_H)^T J(\widehat{\boldsymbol{\theta}}_H)^{-1} U(\widehat{\boldsymbol{\theta}}_H) \quad \sim \quad \chi^2_{q_1}$$

- Properties:
  - only the restricted (H_0) MLE is computed, not the unrestricted (H_1)
  - computationally very fast

Wald Test

- The Wald test exploits the result that $\widehat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}_0, I(\boldsymbol{\theta}_0)^{-1})$
- Wald statistic

$$(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1H})^T V(\widehat{\boldsymbol{\theta}_1})^{-1}(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1H}) \quad \sim \quad \chi^2_{q_1}$$

- Properties:
  - most intuitive
  - only the unrestricted (or "full model") MLE is computed
- Most frequently used test, especially when $q_1 = 1$

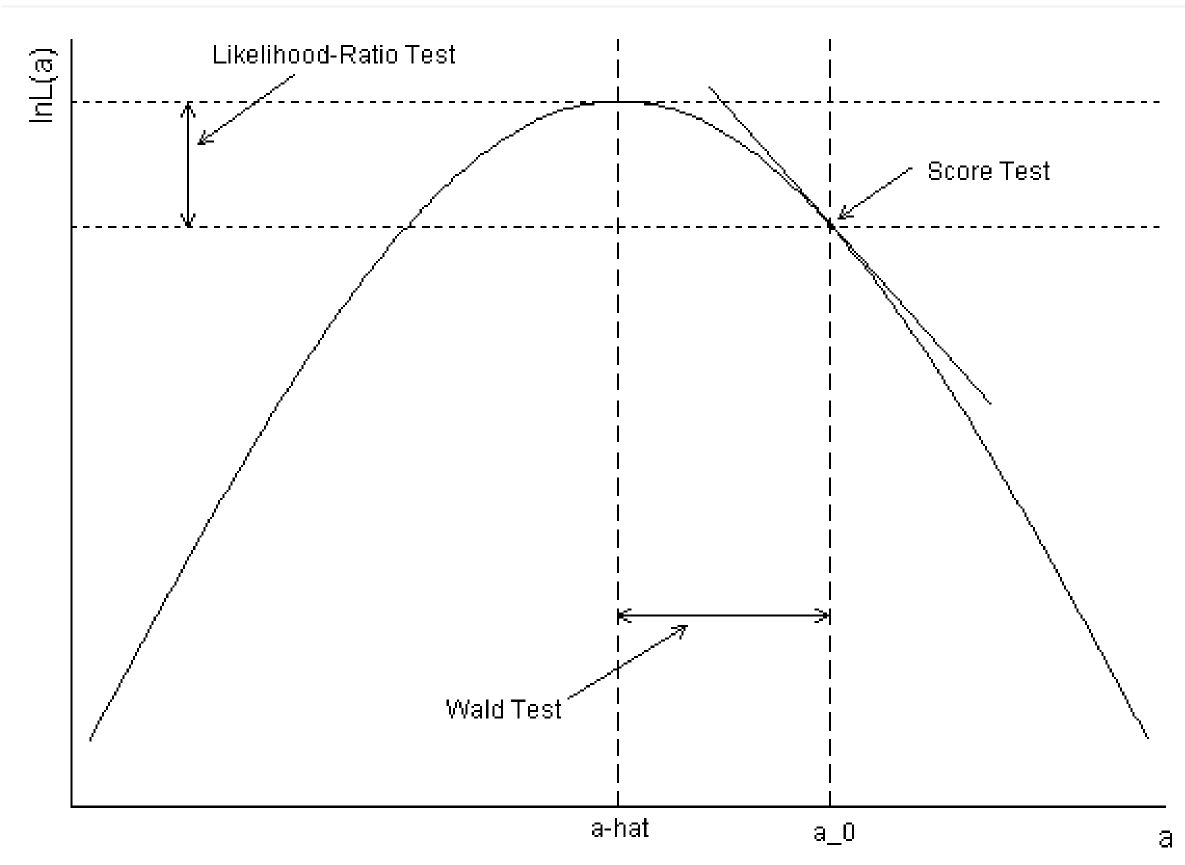
Likelihood Ratio Test

- LR Statistic:

$$\log\left\{\left[\frac{L(\widehat{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}}_H)}\right]^2\right\} \quad \sim \quad \chi^2_{q_1}$$

- Properties of LRT:
  - requires computation of both full and restricted MLEs
  - of the three tests, the LRT is considered the best
- Often written as $-2 \times \{\ell(\widehat{\boldsymbol{\theta}}_H) - \ell(\widehat{\boldsymbol{\theta}})\}$

Three tests



www.ats.ucla.edu

MLE Example: Exponential Model

1. Example: The following $n = 10$ failure times are observed and assumed to arise from an Exponential(θ) distribution.

i	1	2	3	4	5	6	7	8	9	10
Y_i	10	12	8	7	2	4	15	6	5	19

- Summary statistic: $S \equiv \sum_{i=1}^{n} Y_i = 88$

Example: (a) Computing MLE

(a) Estimate θ using maximum likelihood.

$$
\begin{aligned}
f(Y_i;\theta) &= \theta e^{-\theta Y_i} \\
L(\theta) &= \theta^n \exp\left\{-\theta \sum_{i=1}^{n} Y_i\right\} \\
\ell(\theta) &= n \log\theta - S\theta \\
U(\theta) &= \frac{n}{\theta} - S \\
\widehat{\theta} &= \frac{n}{S} = \frac{10}{88} = 0.114
\end{aligned}
$$

Example: (b) Asymptotic CI

(b) Derive a 95% confidence interval for θ by referring to the asymptotic properties of MLE's.

- Recall: $U(\theta) = n/\theta - S$

$$\begin{aligned}
\frac{\partial U}{\partial \theta} &= \frac{-n}{\theta^2} \\
I(\theta) &= -E\left[\frac{-n}{\theta^2}\right] = \frac{n}{\theta^2} \\
I(\widehat{\theta}) &= \frac{10}{(0.114)^2} = 769.47 \\
\widehat{SE}(\widehat{\theta}) &= (769.47)^{-1/2} = 0.036 \\
CI(\theta) &= 0.114 \pm (1.96)(0.036) = (0.043,\ 0.185)
\end{aligned}$$

Example: Hypothesis Testing

- A previous study, conducted under similar conditions but in a different university, estimated $\widehat{\theta} = 0.15$.

  (c) Conduct a Wald test of whether or not the results of the current investigation are consistent with those of the previous study.

  $H_0 : \theta = 0.15$ vs. $H_1 : \theta \neq 0.15$

  from (b), $\widehat{SE}(\widehat{\theta}) = 0.036$

$$
\begin{aligned}
X_W^2 &= \left\{ \frac{\widehat{\theta} - \theta_H}{\widehat{SE}(\widehat{\theta})} \right\}^2 \\
&= \left\{ \frac{0.114 - 0.15}{0.036} \right\}^2 \\
&= 1.00 \\
&< \chi^2_{0.95} = 3.84
\end{aligned}
$$

  fail to reject $H_0 : \theta = 0.15$.

Example: (d) Score Test

(d) Test $H_0 : \theta = 0.15$ vs. $H_1 : \theta \neq 0.15$ using the score test.

$$
\begin{aligned}
U(\theta_H) &= U(0.15) \\
&= \frac{10}{0.15} - S = -21.33 \\
I(0.15) &= \frac{10}{0.15^2} = 444.44 \\
X_S^2 &= (-21.33)(444.44)^{-1}(-21.33) = 1.02 \\
&< 3.84
\end{aligned}
$$

- fail to reject $H_0 : \theta = 0.15$

Example: (e) Likelihood Ratio Test

(e) Test the same hypothesis using the likelihood ratio test.

- computing the maximized and restricted log likelihoods,

$$
\begin{aligned}
\ell(\widehat{\theta}) &= 10\log(0.114) - (0.114)(88) \\
&= -31.75 \\
\ell(\theta_H) &= 10\log(0.15) - (0.15)(88) \\
&= -32.17 \\
2\{\ell(\widehat{\theta}) - \ell(\theta_H)\} &= 2(32.17 - 31.75) = 0.84
\end{aligned}
$$

- fail to reject H_0

Likelihood: Additional Comments

- Exact inference only available for select (and really simple) cases
  - asymptotic results usually employed
  - if applicability of large-sample results is in question (e.g., low n), re-sampling algorithm could be used
    - bootstrap
    - jackknife
- LR, score and Wald tests are asymptotically equivalent and usually yield similar results for even moderate size n