

PLSC 503 – Spring 2018

Maximum Likelihood, I

March 29, 2018

$$Y \sim N(\mu, \sigma^2)$$

$$\begin{aligned} E(Y) &= \mu \\ \text{Var}(Y) &= \sigma^2 \end{aligned}$$

Some Data

$Y = 64$

63

59

71

68

Probabilities, Marginal

$$\Pr(Y_i = y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(Y_i - \mu)^2}{2\sigma^2} \right]$$

So

$$\Pr(Y_1 = 64) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(64 - \mu)^2}{2\sigma^2} \right]$$

$$\Pr(Y_2 = 63) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(63 - \mu)^2}{2\sigma^2} \right]$$

...

$$\Pr(A, B \mid A \perp B) = \Pr(A) \times \Pr(B)$$

So:

$$\Pr(Y_1 = 64, Y_2 = 63) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(64 - \mu)^2}{2\sigma^2}\right] \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(63 - \mu)^2}{2\sigma^2}\right]$$

More Generally...

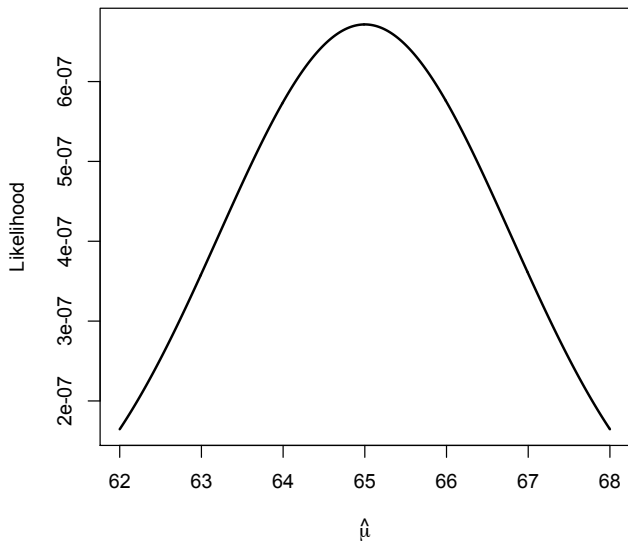
$$\begin{aligned}\Pr(Y_i = y_i \forall i) &\equiv L(Y|\mu, \sigma^2) \\ &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(y_i - \mu)^2}{2\sigma^2} \right]\end{aligned}$$

$$L(\hat{\mu}, \hat{\sigma}^2 | Y) \propto \Pr(Y | \hat{\mu}, \hat{\sigma}^2)$$

For $\hat{\mu} = 68$, $\hat{\sigma} = 4$:

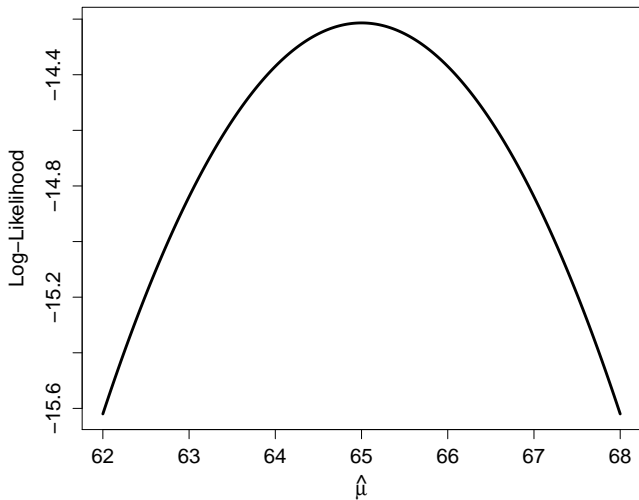
$$\begin{aligned} L &= \frac{1}{\sqrt{2\pi}16} \exp \left[-\frac{(64 - 68)^2}{32} \right] \times \\ &\quad \frac{1}{\sqrt{2\pi}16} \exp \left[-\frac{(63 - 68)^2}{32} \right] \times \\ &\quad \frac{1}{\sqrt{2\pi}16} \exp \left[-\frac{(59 - 68)^2}{32} \right] \times \dots \\ &= \text{some reeeeeally small number...} \end{aligned}$$

What a Likelihood Looks Like



$$\begin{aligned}\ln L(\hat{\mu}, \hat{\sigma}^2 | Y) &= \ln \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(Y_i - \mu)^2}{2\sigma^2} \right] \\&= \sum_{i=1}^N \ln \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(Y_i - \mu)^2}{2\sigma^2} \right] \right\} \\&= -\frac{N}{2} \ln(2\pi) - \left[\sum_{i=1}^N \frac{1}{2} \ln \sigma^2 + \frac{1}{2\sigma^2} (Y_i - \mu)^2 \right]\end{aligned}$$

What a Log-Likelihood Looks Like



The “Maximum” Part

For $L = f(Y, \theta)$,

- Calculate $\frac{\partial \ln L}{\partial \theta}$,
- Set $\frac{\partial \ln L}{\partial \theta} = 0$, solve for $\hat{\theta}$,
- Calculate $\frac{\partial^2 \ln L}{\partial \theta^2}$,
- Verify $\frac{\partial^2 \ln L}{\partial \theta^2} < 0$.

Example: Normal Y

$$\ln L(\hat{\mu}, \hat{\sigma}^2 | Y) = -\frac{N}{2} \ln(2\pi) - \left[\sum_{i=1}^N \frac{1}{2} \ln \sigma^2 + \frac{1}{2\sigma^2} (Y_i - \mu)^2 \right]$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^N (Y_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = \frac{-N}{2\sigma^2} + \frac{1}{2} \sigma^4 \sum_{i=1}^N (Y_i - \mu)^2$$

Example: Normal Y (continued)

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N Y_i$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2$$

Example: Linear Regression

$$\begin{aligned}E(Y) \equiv \mu &= \beta_0 + \beta_1 X_i \\ \text{Var}(Y) &= \sigma^2\end{aligned}$$

$$L(\beta_0, \beta_1, \sigma^2 | Y) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2} \right]$$

Linear Regression (continued)

$$\begin{aligned}\ln L(\beta_0, \beta_1, \sigma^2 | Y) &= \ln \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2} \right] \\ &= -\frac{N}{2} \ln(2\pi) - \sum_{i=1}^N \left[\frac{1}{2} \ln \sigma^2 + \frac{1}{2\sigma^2} (Y_i - \beta_0 - \beta_1 X_i)^2 \right]\end{aligned}$$

Kernel:

$$-\sum_{i=1}^N \left[\frac{1}{2} \ln \sigma^2 + \frac{1}{2\sigma^2} (Y_i - \beta_0 - \beta_1 X_i)^2 \right]$$

$$\Pr(Y) = f(\mathbf{X}, \theta)$$

$$L = \prod_{i=1}^N f(Y_i | \mathbf{X}_i, \theta)$$

$$\ln L = \sum_{i=1}^N \ln f(Y_i | \mathbf{X}_i, \theta)$$

$$\ln L(\hat{\theta} | Y, \mathbf{X}) = \max_{\theta} \{ \ln L(\theta | Y, \mathbf{X}) \}$$

The Gradient

$$\mathbf{g}(\hat{\theta}) = \frac{\partial \ln L(\hat{\theta})}{\partial \hat{\theta}}$$

Taylor series:

$$\frac{\partial \ln L}{\partial \hat{\theta}} \approx \frac{\partial \ln L}{\partial \theta} + \frac{\partial^2 \ln L}{\partial \theta^2}(\hat{\theta} - \theta)$$

$$\begin{aligned}\hat{\theta} - \theta &= \left(-\frac{\partial^2 \ln L}{\partial \theta^2} \right)^{-1} \frac{\partial \ln L}{\partial \theta} \\ &= -\mathbf{H}(\theta)^{-1} \mathbf{g}(\theta)\end{aligned}$$

Need

$$\text{plim}(\hat{\theta} - \theta) = 0$$

So:

- Assume $\mathbf{H}(\theta) \xrightarrow{a} \mathbf{A} < \infty$
- Show $E[\mathbf{g}(\theta)] \rightarrow \mathbf{0}$ as $N \rightarrow \infty$

Consistency (continued)

$$\begin{aligned} \mathbb{E}[\mathbf{g}(\theta)] &= \frac{1}{N} \mathbb{E} \left(\frac{\partial \ln L_1}{\partial \theta} + \frac{\partial \ln L_2}{\partial \theta} + \dots + \frac{\partial \ln L_N}{\partial \theta} \right) \\ &= \frac{1}{N} \left[\mathbb{E} \left(\frac{\partial \ln L_1}{\partial \theta} \right) + \mathbb{E} \left(\frac{\partial \ln L_2}{\partial \theta} \right) + \dots \right] \\ &\stackrel{a}{=} \mathbf{0} \end{aligned}$$

Cramer-Rao say:

$$\text{Var}(\hat{\theta}) \geq \left[-\text{E} \left(\frac{\partial^2 \ln L(\theta)}{\partial \theta^2} \right) \right]^{-1}$$

Efficiency, continued

$$\begin{aligned}\text{Var}(\hat{\theta}) &= \text{E}[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] \\ &= \text{E} \left[\left(-\frac{\partial^2 \ln L}{\partial \theta^2} \right)^{-1} \frac{\partial \ln L}{\partial \theta} \frac{\partial \ln L'}{\partial \theta} \left(-\frac{\partial^2 \ln L}{\partial \theta^2} \right)^{-1} \right]\end{aligned}$$

For MLE:

$$\text{E} \left[\frac{\partial \ln L}{\partial \theta} \frac{\partial \ln L'}{\partial \theta} \right] = \text{E} \left[\frac{\partial^2 \ln L}{\partial \theta^2} \right]$$

So,

$$\begin{aligned}\text{Var}(\hat{\theta}) &= \left[-\text{E} \left(\frac{\partial^2 \ln L}{\partial \theta^2} \right) \right]^{-1} \\ &= [\mathbf{I}(\theta)]^{-1}\end{aligned}$$

By LLN:

$$\frac{\hat{\theta} - \theta}{\sqrt{\mathbf{I}(\theta)^{-1}}} \sim N(\mathbf{0}, \mathbf{1})$$

Or:

$$\hat{\theta} \sim N(\theta, \mathbf{I}(\theta)^{-1})$$

Invariance: Parameters

For

$$\gamma = h(\theta)$$

$$\hat{\gamma}_{ML} = h(\hat{\theta}_{ML})$$

Suppose

$$\phi^2 = 1/\sigma^2$$

so that

$$Y \sim N(\mu, \phi^2).$$

Invariance: Example

Then:

$$\ln L(\hat{\mu}, \hat{\phi}^2) = - \left[\sum_{i=1}^N \frac{1}{2} \ln \phi^2 - \frac{1}{2\phi^2} (Y_i - \mu)^2 \right]$$

and:

$$\frac{\partial \ln L}{\partial \phi^2} = \frac{-N}{2\phi^2} + \frac{1}{2} \phi^4 \sum_{i=1}^N (Y_i - \mu)^2$$

and:

$$\begin{aligned} \hat{\phi}^2 &= \frac{N}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \\ &= \frac{1}{\hat{\sigma}^2} \end{aligned}$$

MLEs:

- Maximize $L(\theta|Y, \mathbf{X})$
- Are consistent in N
- Are asymptotically efficient
- Are asymptotically Normal
- Are invariant to (injective) transformations and varying sampling methods