this time: multiparameter problems

next time: ↓

Read DS ch. 8

AMS 132

21 Feb 17

$(i = 1, \dots, n)$

$$(Y_i \mid \mu, \sigma, B) \overset{IID}{\sim} N(\mu, \sigma^2) \quad (1)$$

$$\underset{\sim}{Y} = (Y_1, \dots, Y_n) \qquad \underset{\sim}{y} = (y_1, \dots, y_n)$$

$$\underset{\sim}{\theta} = (\mu, \sigma)$$

Last time we worked out that the likelihood function in this model is

$$\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, B) = \ell(\mu, \sigma \mid \underset{\sim}{y}, B)$$

$$= c\, \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right]$$

and the log likelihood function is therefore

$$\ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, B) = \ell\ell(\mu, \sigma \mid \underset{\sim}{y}, B)$$

$$= -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2$$

sufficient statistics?

expand out sum:

$$\sum_{i=1}^{n} (y_i - \mu)^2 = \sum_{i=1}^{n} y_i^2 - 2\mu \sum_{i=1}^{n} y_i + \mu^2$$

so immediately $\left(\sum_{i=1}^{n} y_i, \sum_{i=1}^{n} y_i^2\right)$ is sufficient

But a more familiar set of sufficient statistics can be found with a small trick:

$$\sum_{i=1}^{n}(Y_i-\mu)^2 = \sum_{i=1}^{n}\left[(Y_i-\bar{Y})+(\bar{Y}-\mu)\right]^2$$

$$= \sum_{i=1}^{n}(Y_i-\bar{Y})^2 + 2(\bar{Y}-\mu)\underbrace{\cancel{\sum_{i=1}^{n}(Y_i-\bar{Y})}}_{0}$$

$$+ n(\bar{Y}-\mu)^2$$

$$= \sum_{i=1}^{n}(Y_i-\bar{Y})^2 + n(\bar{Y}-\mu)^2$$

force $\bar{Y}$ into the story

so $\left(\bar{Y}, \sum_{i=1}^{n}(Y_i-\bar{Y})^2\right)$ is also sufficient for $\underset{\sim}{\theta} = (\mu, \sigma)$

Math fact: In repeated sampling from this sampling model, $E_{RS}\left[\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right] = \sigma^2$

Definition If an estimator $\hat{\theta}_n$ of a (1-dimensional) parameter θ is such that $E_{RS}(\hat{\theta}_n) = \theta$, $\hat{\theta}_n$ is said to be an <u>unbiased</u> estimator of θ

Restatement of math fact:

$$s_u^2 \triangleq \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \text{ is}$$

unbiased for σ^2 in the $N(\mu,\sigma^2)$ sampling model

"official" set of sufficient statistics in this model: $(\bar{y}, s_u^2)$

using the trick:

$$\ell(\mu,\sigma \mid \underset{\sim}{y}\, \mathcal{B}) = c\,\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s_u^2 + n(\bar{y}-\mu)^2\right]\right\}$$

$$\ell\ell(\mu,\sigma \mid \underset{\sim}{y}\, \mathcal{B}) = -n\log\sigma - \frac{(n-1)s_u^2}{2\sigma^2} - n\frac{(\bar{y}-\mu)^2}{2\sigma^2}$$

Let's set the MLEs

$$\frac{\partial}{\partial\mu}\ell\ell(\mu,\sigma \mid \underset{\sim}{y}\, \mathcal{B}) = -\frac{2n(\bar{y}-\mu)(-1)}{2\sigma^2}$$

$$= \frac{n(\bar{y}-\mu)}{\sigma^2}$$

(seems ok)

set $= 0$, solve for μ:

$$\mu = \hat{\mu}_{MLE} = \bar{y}$$

$$\frac{\partial}{\partial\sigma}\ell\ell(\mu,\sigma \mid \underset{\sim}{y}\, \mathcal{B}) = -\frac{n}{\sigma} - \frac{(-2)\left[(n-1)s_u^2 + n(\bar{y}-\mu)^2\right]}{2\sigma^3}$$

More carefully

The MLE $\hat{\underset{\sim}{\theta}}_{MLE} = (\hat{\mu}_{MLE}, \hat{\sigma}_{MLE})$ is the solution of $\underline{2}$ equations in 2 unknowns:

$$\left\{ \begin{array}{l} \frac{\partial}{\partial \mu} \ell\ell(\mu, \sigma | \underset{\sim}{y} \mathcal{B}) = 0 \\ \frac{\partial}{\partial \sigma} \ell\ell(\mu, \sigma | \underset{\sim}{y} \mathcal{B}) = 0 \end{array} \right\}$$

It happens in this problem that the first equation, for $\hat{\mu}_{MLE}$, can be solved without reference to the second equation; this will in general not be true.

$$\frac{\partial}{\partial \sigma} \ell\ell(\mu, \sigma | \underset{\sim}{y} \mathcal{B}) = 0 \iff \frac{-n\sigma^2 + (n-1)s_n^2 + n(\bar{y}-\mu)^2}{\sigma^3} = 0$$

iff $\hat{\sigma}^2_{MLE} = \frac{n-1}{n} s_n^2 + (\bar{y} - \hat{\mu}_{MLE})^2$, i.e.

$$\left(\hat{\mu}_{MLE} = \bar{Y}, \ \hat{\sigma}_{MLE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2} \right).$$

Important property of maximum likelihood

Suppose we parameterize this model in terms of $(\mu, \sigma^2) \triangleq (\mu, \eta)$ instead of (μ, σ)

eta

$\underset{\sim}{\theta}' = (\mu, \sigma^2)$, $\underset{\sim}{\theta} = (\mu, \sigma)$

$$\ell(\underset{\sim}{\theta}' \mid \underset{\sim}{y}, \mathcal{B}) = \ell(\mu, \eta \mid \underset{\sim}{y}, \mathcal{B}) = ?$$

Easy: everywhere you see σ^2 put η; everywhere you see σ put $\sqrt{\eta}$

$$\boxed{\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, \mathcal{B}) = \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]}$$

$$\ell(\underset{\sim}{\theta}' \mid \underset{\sim}{y}, \mathcal{B}) = \ell(\mu, \eta \mid \underset{\sim}{y}, \mathcal{B})$$

μ stays the same

$$= (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]$$

$\sigma^2 = \eta$

$$= \eta^{-\frac{n}{2}} \exp\left[-\frac{1}{2\eta}\sum_{i=1}^{n}(y_i-\mu)^2\right]$$

$$\ell\ell(\underset{\sim}{\theta}' \mid \underset{\sim}{y}, \mathcal{B}) = -\frac{n}{2}\log \eta - \frac{1}{2\eta}\sum_{i=1}^{n}(y_i-\mu)^2$$

$$\frac{\partial}{\partial \mu}\ell\ell(\underset{\sim}{\theta}' \mid \underset{\sim}{y}, \mathcal{B}) = -\frac{1}{2\eta}\sum_{i=1}^{n} 2(y_i-\mu)(-1)$$

$$\overset{(*)}{=} 0 \text{ iff } \sum_{i=1}^{n} y_i = n\mu \text{ so } \hat{\mu}_{MLE} = \bar{Y} \text{ as before}$$

$$\frac{\partial}{\partial \eta}\ell\ell(\underset{\sim}{\theta}' \mid \underset{\sim}{y}, \mathcal{B}) = -\frac{n}{2\eta} + \frac{1}{2\eta^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

~~together with~~

and this is 0 $\overset{(*)}{\text{above}}$ $= \dfrac{-n\eta + \sum_{i=1}^{n}(y_i-\mu)^2}{2\eta^2}$

iff $\eta = \hat{\eta}_{MLE} = \dfrac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu}_{MLE})^2 = \dfrac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$

So: $\hat{\sigma}_{MLE} = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2} = \sqrt{\hat{\eta}_{MLE}}$

$\eta = \sigma^2 \rightarrow \sigma = \sqrt{\eta} = g(\eta) \rightarrow \hat{\sigma}_{MLE} = g(\hat{\eta}_{MLE})$

Math fact: (**) If $g(\cdot)$ is an invertible function and $\lambda = g(\gamma)$ then

$$\hat{\lambda}_{MLE} = g(\hat{\gamma}_{MLE})$$

for any parameters λ and γ

People say that maximum likelihood is invariant with respect to invertible parameter transformation

Ex. 1 In the $N(\mu, \sigma^2)$ sampling model if you were interested for some reason in $\lambda = \mu^3$,

$$\hat{\lambda}_{MLE} = (\hat{\mu}_{MLE})^3 = \bar{Y}^3$$

In this model when $\mu > 0$ people sometimes are interested in $\gamma = \frac{\mu}{\sigma}$ = $\frac{\text{signal}}{\text{noise}}$ ratio;

$$\hat{\gamma}_{MLE} = \frac{\hat{\mu}_{MLE}}{\hat{\sigma}_{MLE}}$$

(this requires a stronger math fact than (**) but is also true).

What about $\hat{SE}(\hat{\mu}_{MLE})$ and $\hat{SE}(\hat{\sigma}_{MLE})$?

we need an analogue of Fisher information in more than 1 dimension

Recall:

If your sampling model is

$$(Y_i \mid \mu, \mathcal{B}) \overset{IID}{\sim} N(\mu, \sigma^2) \quad (i = 1, \ldots, n)$$

σ^2 known

so that $\theta = \mu$ (1-dimensional), then the Fisher information is

$$\hat{I}(\hat{\mu}_{MLE}) = \left[-\frac{\partial^2}{\partial \mu^2} \ell\ell(\mu \mid \underset{\sim}{y}, \mathcal{B}) \right]_{\mu = \hat{\mu}_{MLE}}$$

When $\underset{\sim}{\theta}$ has dimension greater than 1, the analogue of this is a <u>matrix</u> of second partial derivatives of $\ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, \mathcal{B})$

Definition $\underset{\sim}{\theta}$ dimension $p \geq 1$: the

<u>Hessian</u> H of $\ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, \mathcal{B})$ is the $(p \times p)$ matrix of second partial derivatives:

$$H_{pp} = \begin{array}{c} \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{array} \begin{array}{c} \begin{array}{cccc} \theta_1 & \theta_2 & \cdots & \theta_p \end{array} \\ \begin{bmatrix} \frac{\partial^2}{\partial\theta_1^2}\ell\ell & \frac{\partial^2}{\partial\theta_1\partial\theta_2}\ell\ell & & \frac{\partial^2}{\partial\theta_1\partial\theta_p}\ell\ell \\ \frac{\partial^2}{\partial\theta_2\partial\theta_1}\ell\ell & \cdot & \cdot & \cdot \\ \vdots & \cdot & \cdot & \cdot \\ & & & \end{bmatrix} \end{array}$$

The (i,j) element of H_{pp} is

$$(H_{pp})_{ij} = \frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, \mathcal{B}).$$

Hessian of log likelihood

Definition $\underset{\sim}{\theta}$ dimension $p \geq 1$: the Fisher information matrix is

$$\hat{I}(\hat{\theta}_{MLE})_{pp} = \left[-H\right]_{\underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}_{MLE}}$$

In the $N(\mu, \sigma^2)$ sampling model, $\underset{\sim}{\theta} = (\mu, \sigma)$

we already worked out that

$$\ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, B) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2$$

$$\frac{\partial}{\partial \mu} \ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, B) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} 2(y_i - \mu)(-1)$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu)$$

$$\frac{\partial^2}{\partial \mu^2} \ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, B) = -\frac{n}{\sigma^2}$$

$$= \frac{\sum_{i=1}^{n} y_i^2}{\sigma^2} - \frac{n\mu}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma} \ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, B) = -\frac{n}{\sigma} - \frac{(-2)}{2\sigma^3} \sum_{i=1}^{n} (y_i - \mu)^2$$

$$= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (y_i - \mu)^2$$

So $\frac{\partial^2}{\partial \sigma^2} \ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, B)$

$$= -\frac{n}{\sigma^2}(-1) - \frac{3}{\sigma^4} \sum_{i=1}^{n} (y_i - \mu)^2 = \frac{n\sigma^2 - 3\sum_{i=1}^{n} (y_i - \mu)^2}{\sigma^4}$$

As usual the mixed partial derivatives

$\frac{\partial^2}{\partial\mu\partial\sigma} \ell\ell$ and $\frac{\partial^2}{\partial\sigma\partial\mu} \ell\ell$ are equal, so

you only need to compute one of them

& you might as well compute the easier

one: $\boxed{\frac{\partial}{\partial\mu} \ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, \mathcal{B}) = \frac{\sum_{i=1}^{n}(y_i-\mu)}{\sigma^2}}$

so

$$\frac{\partial}{\partial\sigma\partial\mu} \ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, \mathcal{B}) = \frac{-2\sum_{i=1}^{n}(y_i-\mu)}{\sigma^3}$$

so $$\underset{2\ 2}{H} = \begin{bmatrix} -\frac{n}{\sigma^2} & \frac{-2\sum_{i=1}^{n}(y_i-\mu)}{\sigma^3} \\ \frac{-2\sum(y_i-\mu)}{\sigma^3} & \frac{n\sigma^2 - 3\sum_{i=1}^{n}(y_i-\mu)^2}{\sigma^4} \end{bmatrix}$$

and $\hat{I}(\hat{\theta}_{MLE}) = \left[- \quad H \right]_{\theta = \hat{\theta}_{MLE}}$

$$= \begin{bmatrix} \dfrac{n}{\sigma^2} & \dfrac{2\sum_{i=1}^{n}(y_i-\mu)}{\sigma^3} \\ \dfrac{2\sum_{i=1}^{n}(y_i-\mu)}{\sigma^3} & \dfrac{-n\sigma^2+3\sum(y_i-\mu)^2}{\sigma^4} \end{bmatrix}_{\mu=\bar{y},\ \sigma=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$= \begin{bmatrix} \dfrac{n}{\hat{\sigma}^2_{MLE}} & 0 \\ 0 & \dfrac{2n\,\cancel{\hat{\sigma}_{MLE}}}{\hat{\sigma}^{2}_{MLE}} \end{bmatrix}$$

$\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2$

so

$n\hat{\sigma}^2_{MLE} = \sum_{i=1}^{n}(y_i-\bar{y})^2$

Refined math fact

If $Y_1, \ldots, Y_n$ form an IID random sample from a (population) density such that $E_{PS}(Y_i) = \mu$ exists, $V_{PS}(Y_i) = \sigma^2$ exists, & $V_{PS}(Y_i) < \infty$,

then $$E_{PS}\left[\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] = \sigma^2,$$

where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

Cautionary note:

$$S_u^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$ is unbiased for σ^2 ✓

Q: Is $S_u = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$ unbiased for σ?

A: No: if $\hat{\theta}_u$ is unbiased for θ, $g(\hat{\theta}_u)$ is not unbiased for $g(\theta)$ unless $g(\cdot)$ is linear

$$\begin{matrix} y_1 \\ \vdots \\ y_n \end{matrix}\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \xrightarrow[\bar{y}]{\text{subtract off}} \begin{bmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} = \begin{bmatrix} -3 \\ -2 \\ +5 \end{bmatrix} \xrightarrow{\text{square}} \begin{bmatrix} (-3)^2 \\ (-2)^2 \\ (+5)^2 \end{bmatrix}$$

mean $4 = \bar{y}$

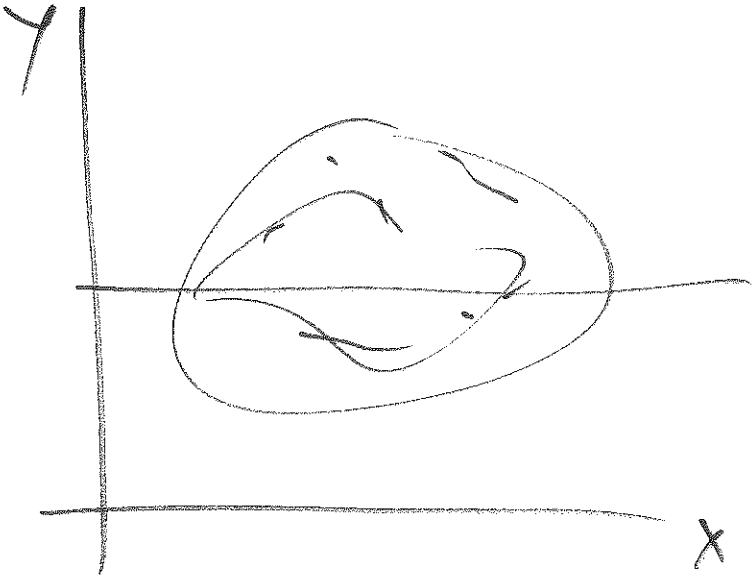
$$\begin{array}{c} \\ n=3 \\ \\ \end{array}\begin{bmatrix} X\checkmark \\ X\checkmark \\ \\ \end{bmatrix} \begin{array}{l} \text{free} \\ \text{free} \\ \text{not free} \end{array}$$

mean $\bar{y} = 4$

there are n elements in this data set, but only $(n-1)$ are free in constructing the deviations $(y_i - \bar{y})$

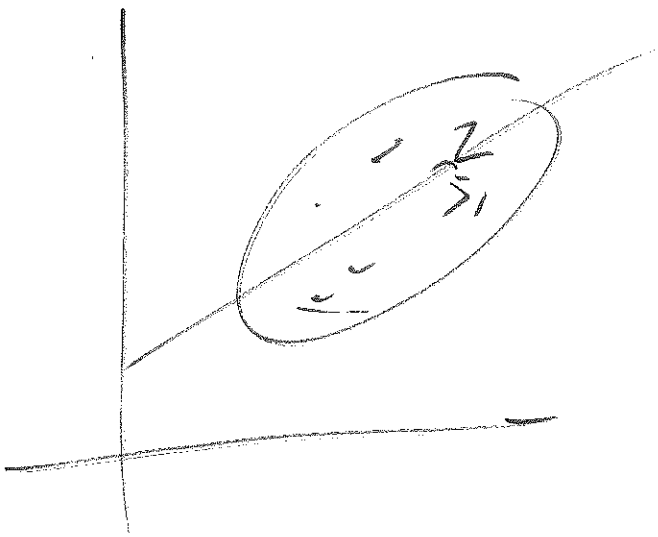
Fisher: degrees of freedom (df)

A dataset $\underset{\sim}{y} = (y_1, \ldots, y_n)$ only has $(n-1)$ df for measuring spread (scale), if center (mean) must also be estimated

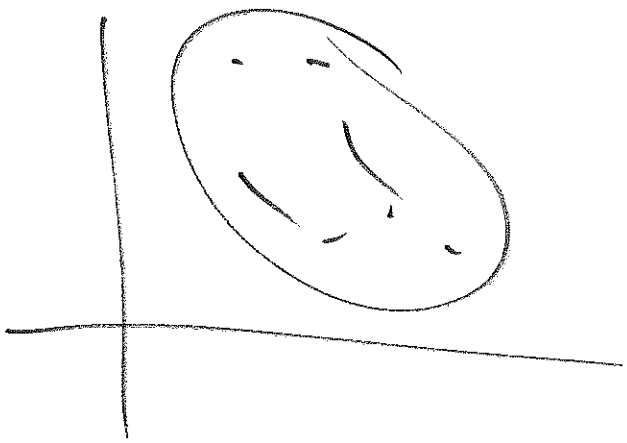


r = correlation

$\doteq 0$



$r > 0$



$r < 0$

Facts about MLE

① $\hat{\theta}_{MLE}$ is not necessarily unbiased, but

$$\text{Bias}(\hat{\theta}) = O\left(\frac{1}{n}\right)$$

$\downarrow 0$ as $n \to \infty$

$\boxed{\text{Def.}}$ $\text{Bias}(\hat{\theta}) = E_{R_s}(\hat{\theta}) - \theta$

Math fact: to get $\hat{SE}((\hat{\theta}_i)_{MLE})$, invert the Fisher information matrix to get $\hat{I}^{-1}(\hat{\underset{\sim}{\theta}}_{MLE})$ and take the square root of diagonal element i in $\hat{I}^{-1}$

ex. $N(\mu, \sigma^2)$ model

$$\hat{I}^{-1}(\hat{\underset{\sim}{\theta}}_{MLE}) = \begin{bmatrix} \frac{\hat{\sigma}^2_{MLE}}{n} & 0 \\ 0 & \frac{\hat{\sigma}^2_{MLE}}{2n} \end{bmatrix}$$

so $\hat{SE}(\hat{\mu}_{MLE}) = \sqrt{\frac{\hat{\sigma}^2_{MLE}}{n}}$

$\hat{SE}(\bar{Y})$

$\hat{V}_{RS}(\bar{Y}) = \frac{\hat{\sigma}^2}{n}$

$\hat{SE}(\bar{Y}) = \hat{\sigma}/\sqrt{n}$

Sidebar: Irregular inference problem:

ex. $(Y_i \mid \theta) \overset{IID}{\sim} \text{uniform}(0, \theta)$

& $\hat{SE}(\hat{\sigma}_{MLE}) = \sqrt{\dfrac{\hat{\sigma}^2_{MLE}}{2n}} = \dfrac{\hat{\sigma}_{MLE}}{\sqrt{2n}}$.

Finally when n is large, the components of $\hat{\theta}_{MLE}$ follow normal distributions:

$$(\hat{\theta}_i)_{MLE} \underset{RS}{\sim} N\left(\theta, I^{-1}(\hat{\theta}_{MLE})_{ii}\right)$$

Thus: to get an approx. (large-sample) $100(1-\alpha)\%$ CI for μ,

$$\hat{\mu}_{MLE} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \hat{SE}(\hat{\mu}_{MLE})$$

ex. $\alpha = .05$ $\left(\bar{Y} \pm 1.96 \dfrac{\hat{\sigma}}{\sqrt{n}}\right)$ familiar CI

ex. 95% CI for σ:

large-sample approx.

$$\hat{\sigma}_{MLE} \pm 1.96 \frac{\hat{\sigma}_{MLE}}{\sqrt{2n}}$$

Not so good when n is small: correct interval for σ should be skewed.