Multi-parameter problems

So far $\underset{\sim}{\theta}$ has been a $(k \times 1)$ or $(1 \times k)$ vector with $k = 1$ (i.e., a scalar)

STAT 206
16 Feb 21

Lecture ①

Do the likelihood and Bayesian stories still work with $k > 1$?

Let's see

Ex. NB10 case study (lec. notes 6 p. 23) Ill-fitting but interesting model for practice with $k = 2$

Gaussian sampling model

$$(Y_i \mid \mu, \sigma, NB) \overset{IID}{\sim} N(\mu, \sigma^2) \quad (i = 1, \dots, n) \quad (k = 2) \quad (\mu \in \mathbb{R}, \ \sigma > 0)$$

$\underset{\sim}{Y} = (Y_1, \dots, Y_n)$

The likelihood story: ① marginal PDF

$$p(y_i \mid \mu, \sigma, NB) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y_i - \mu}{\sigma}\right)^2\right]$$

$\underset{\sim}{y} = (y_1, \dots, y_n)$

② joint PDF

$$p(\underset{\sim}{y} \mid \mu, \sigma, NB) = \prod_{i=1}^{n} p(y_i \mid \mu, \sigma, NB)$$

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

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

③ likelihood $l(\mu, \sigma | \underset{\sim}{y}, NB) = c \cdot \sigma^{-n} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \mu}{\sigma}\right)^2\right]$ ②

$(l: \mathbb{R} \times (0,\infty) \to \mathbb{R})$

④ log likelihood $(c = 1)$

$$ll(\mu, \sigma | \underset{\sim}{y}, NB) = -n \log \sigma - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \mu}{\sigma}\right)^2$$

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

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

④.1 Visualize likelihood & log likelihood functions with the NB10 data (see R demo) To do this visualization we have to define a grid for μ and a grid for σ; (chicken & egg problem); let's get the MLEs first, postponing ④.1

⑤ MLEs with $(k=2)$ there are now 2 partial derivatives (with respect to μ and σ), so we now have a system of 2 equations

in 2 unknowns to solve:

$$\frac{\partial}{\partial \mu} ll(\mu, \sigma \mid \underset{\sim}{y}, NB) = 0$$

$$\frac{\partial}{\partial \sigma} ll(\mu, \sigma \mid \underset{\sim}{y}, NB) = 0$$

$$\frac{\partial}{\partial \mu} ll(\mu, \sigma \mid \underset{\sim}{y}, NB) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu) = 0 \quad \text{iff} \quad \mu = \hat{\mu}_{MLE} = \bar{y} \checkmark$$

$$\left(\sum_{i=1}^{n} y_i\right) - n\mu = 0$$

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

$$= 0 \quad \text{iff (simplifying)} \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \mu)^2$$

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

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

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

This is slightly different from the usual estimator of variance in the Gaussian model,

$s^2 \triangleq \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$, which many people prefer because in repeated sampling (frequentist) s^2 is unbiased for σ^2: $E_F(s^2) = \sigma^2$.

This means that MLEs can be biased although the general result (under mild regularity conditions) is that the bias (if any) of an MLE goes to 0 quickly:

$$\text{bias}(\hat{\sigma}^2_{MLE}) = O\left(\frac{1}{n}\right)$$

Here we have $n\hat{\sigma}^2_{MLE} = (n-1)S^2$

so $\hat{\sigma}^2_{MLE} = \left(\frac{n-1}{n}\right)S^2$ and $E_F(\hat{\sigma}^2_{MLE}) = \frac{n-1}{n}\sigma^2$

$$= \sigma^2 - \frac{\sigma^2}{n}$$

and $\text{bias}(\hat{\sigma}^2_{MLE}) = E_F(\hat{\sigma}^2_{MLE}) - \sigma^2$

$$= -\frac{\sigma^2}{n} = O\left(\frac{1}{n}\right),$$ matching the general pattern

Q: How wide should we make the grids?

A1: Trial and error

A2: Standard error calculations

with the ARIA data

$\hat{\mu}_{MLE} = \bar{y} = 404.6$

$\hat{\sigma}_{MLE} = 8.4$

and σ

(6) Estimated standard errors for the (5) MLEs

Q: What's the analogue of observed information when $k>1$?

A: The second partial derivatives of the log likelihood function now are stored in a $(k+1)\times(k+1)$ matrix called the Hessian of $-\ell\ell(\mu,\sigma \mid \underset{\sim}{x}, NB)$:

$$H \triangleq \begin{matrix} \\ \mu \\ \sigma \end{matrix} \overset{\begin{matrix}\mu & \sigma\end{matrix}}{\begin{bmatrix} \frac{\partial^2}{\partial\mu^2}\ell\ell(-) & \frac{\partial^2}{\partial\mu\partial\sigma}\ell\ell(-) \\ \frac{\partial^2}{\partial\sigma\partial\mu}\ell\ell(-) & \frac{\partial^2}{\partial\sigma^2}\ell\ell(-) \end{bmatrix}}$$

2×2

recall that

$$\frac{\partial}{\partial\mu}\ell\ell(-) = -\frac{\left(\sum_{i=1}^n x_i\right)}{\sigma^2} + \frac{n\mu}{\sigma^2}$$

so $\frac{\partial^2}{\partial\mu^2}\ell\ell(-) = n\frac{n}{\sigma^2}$

recall that $\frac{\partial}{\partial\sigma}\ell\ell(-) = \frac{n}{\sigma} - \frac{\sum_{i=1}^n (x_i-\mu)^2}{\sigma^3}$

so $\frac{\partial^2}{\partial\sigma^2}\ell\ell(-) = -\frac{n}{\sigma^2} + 3\frac{\sum_{i=1}^n (x_i-\mu)^2}{\sigma^4}$

and the mixed partial becomes $\left(\frac{2\mu n - 2S}{\sigma^3}\right)$ ⊛

To get the Hessian in Wolfram, let's ⑥
rewrite $-ll(\mu,\sigma \mid \neq NB) = -\left[-n\log\sigma - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right]$

$as\ -ll(\mu,\sigma \mid \neq NB) = n\log\sigma + \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2\right)$

$= n\log\sigma + \frac{1}{2\sigma^2}(SS - 2S\mu + n\mu^2)$

of $-ll(\ldots)$ ⑦ Then

the Hessian turns out to be

$$H = \begin{bmatrix} \frac{n}{\sigma^2} & -\frac{2\mu n - 2S}{\sigma^3} \\ -\frac{2\mu n - 2S}{\sigma^3} & \frac{3\mu^2 n + SS - 2\mu S}{\sigma^4} - \frac{n}{\sigma^2} \end{bmatrix}$$

The observed information matrix $\hat{I}_{ik}$ is now obtained by evaluating H at $(\mu,\sigma) = (\hat{\mu}_{MLE}, \hat{\sigma}_{MLE})$; the result, after simplifying

$$\vec{I}_{2\times 2} = \begin{bmatrix} \frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{2n}{\hat{\sigma}^2} \end{bmatrix} = \underline{\underline{O}}(n)$$

and the $(k>1)$ generalization of $\vec{V}(\vec{\theta}_{MLE}) = \vec{I}^{-1}$ (7)

when $k=1$ is the covariance matrix

$$\vec{\Sigma}_{\hat{\theta}_{MLE}} = (\vec{I})^{-1}$$ (the inverse of the information matrix)

Here this is $\vec{\Sigma}_{(\hat{\mu}_{MLE}, \hat{\sigma}_{MLE})} = \begin{bmatrix} \frac{\hat{\sigma}^2}{n} & 0 \\ 0 & \frac{\hat{\sigma}^2}{2n} \end{bmatrix}$

~~(STAT 121)~~

Recall that a covariance matrix looks like this

$$\vec{\Sigma}_{(\hat{\mu}_{MLE}, \hat{\sigma}_{MLE})} = \begin{matrix} \\ \hat{\mu} \\ \hat{\sigma} \end{matrix} \begin{matrix} \begin{matrix} \hat{\mu} & \hat{\sigma} \end{matrix} \\ \begin{bmatrix} \vec{V}(\hat{\mu}) & \vec{C}(\hat{\mu},\hat{\sigma}) \\ \vec{C}(\hat{\mu},\hat{\sigma}) & \vec{V}(\hat{\sigma}) \end{bmatrix} \end{matrix}$$

in which

$\vec{C}(\hat{\mu}, \hat{\sigma})$ = the covariance between $\hat{\mu}_{MLE}$ and $\hat{\sigma}_{MLE}$, here = 0 (!)

NB 10

$\underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ ← 375 ← micrograms of log

$n = 100$

→ 437

Q: How choose good sampling model?

$(Y_i \mid \boxed{?}) \overset{IID}{\sim} \boxed{?}$ $(i = 1, \ldots, n)$

A: A_1 frequently used up through ~1995

histogram, normal qq plot

we cheat (dim. 2)

$S = \sum_{i=1}^n Y_i$ $SS = \sum_{i=1}^n Y_i^2$

(S, SS) form a minimal suff. $(k=2)$ stat.

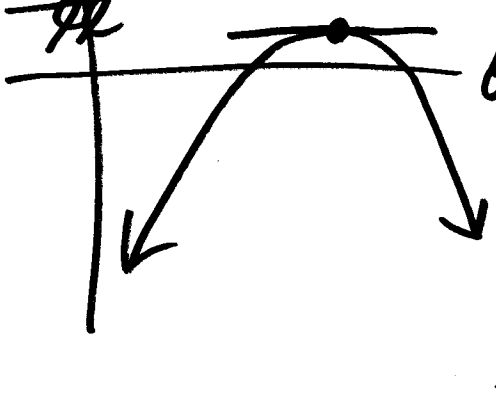
$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ $s^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$

$= \frac{S}{n}$

sample variance

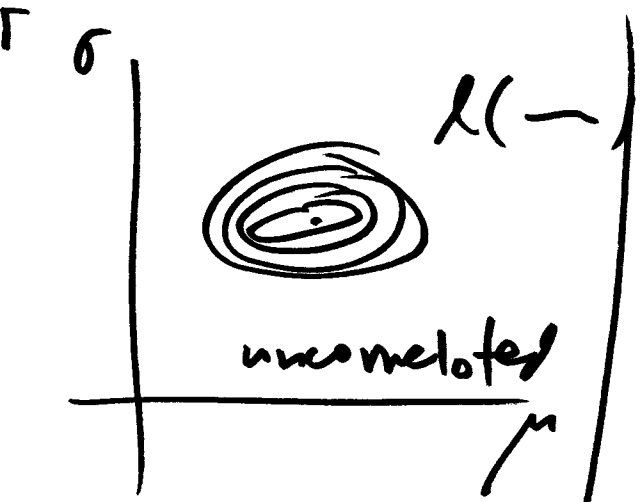
equiv. set of m.ss. is $(\bar{Y}, s^2)$

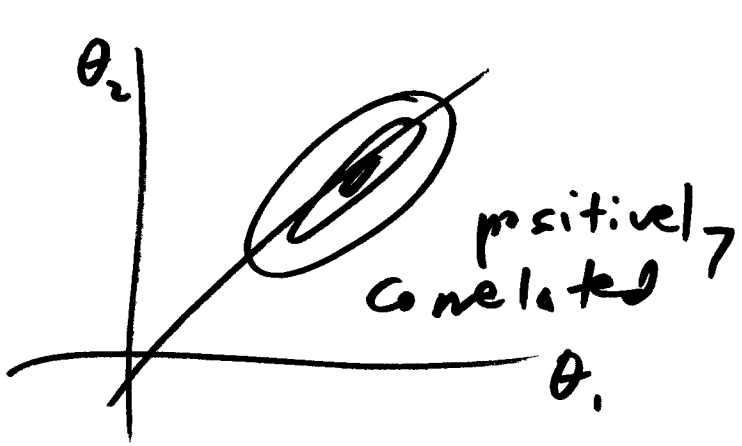
sample mean $(k=1)$

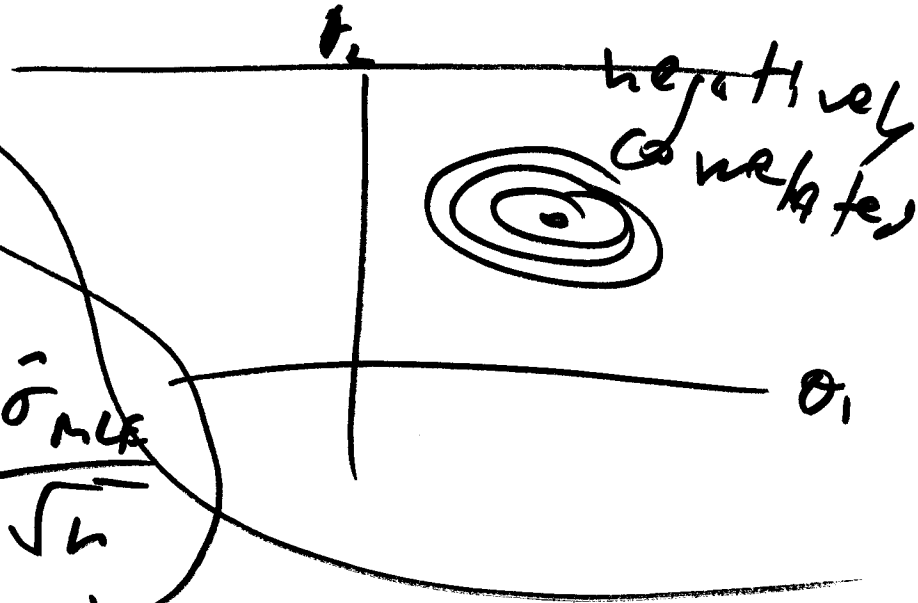


$\left\{ \frac{d}{d\mu} ll(\text{—}) = 0, \right.$ $(k=2)$

$\left. \frac{d}{d\sigma} ll(\text{—}) = 0 \right\}$







$$\widehat{SE}(\hat{\mu}_{MLE}) = \sqrt{\hat{V}(\hat{\mu}_{MLE})} = \frac{\hat{\sigma}_{MLE}}{\sqrt{n}}$$

approx.

$100(1-\alpha)\%$ CI for μ:

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

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

$100(1-\alpha)\%$ CI for σ:

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

for $k \geq 1$ under mild regularity conditions

$$\hat{\underset{\sim}{\theta}}_{MLE} \overset{\cdot}{\sim} N_k\left(\underset{\sim}{\theta}, \hat{\Sigma}_{\hat{\theta}_{MLE}}\right)$$

$\hat{\underset{\sim}{\theta}}_{MLE}$: 1 ... k

$\overset{\cdot}{\sim}$: in repeated sampling