Bayesian analysis of Gaussian sampling models

STAT 206
18 Feb 21

NB10 case study

Lecture three

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

$(y_i \mid \mu\, \sigma\, \mathcal{N}\, \mathcal{B}) \overset{IID}{\sim} N(\mu, \sigma^2)$ $(i=1, \dots, n)$ both unknown

Let's look at a cut-down version of this model, in which we pretend that σ is known (and equal to the sample SD in the NB10 data set [more cheating, temporarily])

Cut-down Gaussian model (k = 1)

$(\mu \mid ?\, \mathcal{B}) \sim p(\mu \mid ?\, \mathcal{B})$

$(y_i \mid \mu\, \mathcal{N}_s\, \mathcal{B}) \overset{IID}{\sim} N(\mu, \sigma_0^2)$ $(i=1, \dots, n)$ known

The likelihood function

$$\ell(\mu \mid y\, \mathcal{N}\, \mathcal{B}) = c\, \cancel{\sigma_0^{-n}} \exp\left[\frac{-1}{2\sigma_0^2} \sum_{i=1}^n (y_i - \mu)^2 \right]$$

$$= c \exp\left[\cancel{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n y_i^2} + \frac{2\mu}{\cancel{2}\sigma_0^2} \sum_{i=1}^n y_i - \frac{n\mu^2}{2\sigma_0^2} \right]$$

Conjugate prior?

After simplification, (ugly algebra) something really nice pops out: ②

lecture notes part 6 p. 126 +

$$\ell(\mu \mid \underline{y}\, \mathbb{N}\, \mathcal{B}) = c \exp\left[-\frac{1}{2\left(\frac{\sigma_0^2}{n}\right)} (\mu - \bar{\underline{y}})^2 \right]$$

$$= N\left(\bar{\underline{y}}, \frac{\sigma_0^2}{n}\right) \text{ as a PDF in } \mu \,(!)$$

$\boxed{q_1:}$ Is this a familiar family of PDFs? $\boxed{A_1}$ yes

$\boxed{q_2:}$ Is the product of 2 Gaussian PDFs another one? $\boxed{A_2}$ Turns out to be (ugly algebra) yes.

So the conjugate prior for μ in the IID $N(\mu, \sigma_0^2)$ ← known sampling model is also Normal:

$$(\mu \mid \mathbb{N}_p\, \mathcal{B}) \sim N(\mu_0, \sigma_\mu^2)$$

$$(y_i \mid \mu\, \mathbb{N}_s\, \mathcal{B}) \overset{IID}{\sim} N(\mu, \sigma_0^2) \leftarrow \text{known}$$

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

$\mathbb{N}_s$ = Normal sampling dist.
$\mathbb{N}_p$ = Normal prior

The conjugate updating rule here is ③
genuinely beautiful: (ugly algebra) $\bar{y}$ is a minimal suff. stat for μ

$$\left\{ \begin{array}{l} (\mu \mid \mathcal{N}_p\, \mathcal{B}) \sim N(\mu_0, \sigma_\mu^2) \\ (Y_i \mid \mu\, \mathcal{N}_s\, \mathcal{B}) \overset{IID}{\sim} N(\mu, \sigma_0^2) \quad \text{(known)} \\ (i = 1, \dots, n) \end{array} \right\}$$

$y = (y_1, \dots, y_n)$ $\longrightarrow$

$$(\mu \mid y\, \mathcal{N}_s\, \mathcal{N}_p\, \mathcal{B}) \sim N(\underline{\underline{\mu_*}}, \sigma_*^2) \text{ with}$$

$$\mu_* = \frac{\boxed{\frac{1}{\sigma_\mu^2}} (\underline{\underline{\mu_0}}) + \boxed{\frac{n}{\sigma_0^2}} (\underline{\underline{\bar{y}}})}{\left(\frac{1}{\sigma_\mu^2}\right) + \left(\frac{n}{\sigma_0^2}\right)} = \underline{\underline{\frac{\left(\frac{\sigma_0^2}{\sigma_\mu^2}\right)(\mu_0) + (\underline{\underline{n}})(\bar{y})}{\left(\frac{\sigma_0^2}{\sigma_\mu^2}\right) + n}}}$$

prior sample size n_0

and $$\sigma_*^2 = \frac{1}{\left(\frac{1}{\sigma_\mu^2}\right) + \left(\frac{n}{\sigma_0^2}\right)} = \frac{\sigma_0^2}{\left(\frac{\sigma_0^2}{\sigma_\mu^2}\right) + \underline{\underline{n}}}$$

data sample size

prior sample size (n_0)

$\begin{pmatrix}\text{prior} \\ \text{variance} \\ \text{for } \mu\end{pmatrix} = \underline{\underline{\sigma_\mu^2}}$ $\quad\begin{pmatrix}\text{likelihood} \\ \text{variance} \\ \text{for } \mu\end{pmatrix} = \underline{\underline{\frac{\sigma_0^2}{n}}}$

Notice that these quantities enter into μ_* as <u>reciprocals</u>: $\underline{\underline{\left(\frac{1}{\sigma_\mu^2}\right)}}$ and $\underline{\underline{\left(\frac{n}{\sigma_0^2}\right)}}$ are the weights in the weighted average representation of the posterior mean:

Bayesian definition: $\forall$ any random variable with <u>variance</u> $\sigma^2 > 0 \rightarrow$ the <u>precision</u> of Σ is just the reciprocal $\left(\frac{1}{\sigma^2} = \text{precision} = \tau\right)$ of the variance

"tau"

$$\begin{pmatrix}\text{posterior}\\ \text{mean}\\ \text{for } \mu\end{pmatrix} = \mu_* = \frac{\begin{pmatrix}\text{prior}\\ \text{precision}\end{pmatrix}\begin{pmatrix}\text{prior}\\ \text{mean}\end{pmatrix} + \begin{pmatrix}\text{likelihood}\\ \text{precision}\end{pmatrix}\cdot\begin{pmatrix}\text{data}\\ \text{mean}\end{pmatrix}}{\begin{pmatrix}\text{prior}\\ \text{precision}\end{pmatrix} + \begin{pmatrix}\text{likelihood}\\ \text{precision}\end{pmatrix}}$$

$$= \frac{\left(\frac{1}{\sigma_\mu^2}\right)\mu_0 + \left(\frac{n}{\sigma_0^2}\right)\bar{y}}{\left(\frac{1}{\sigma_\mu^2}\right) + \frac{n}{\sigma_0^2}}$$

$$\begin{pmatrix}\text{posterior}\\ \text{precision}\\ \text{for } \mu\end{pmatrix} = \frac{1}{\sigma_*^2} = \left(\frac{1}{\sigma_\mu^2}\right) + \left(\frac{n}{\sigma_0^2}\right)$$

$$= \begin{pmatrix}\text{prior}\\ \text{precision}\end{pmatrix} + \begin{pmatrix}\text{likelihood}\\ \text{precision}\end{pmatrix}$$

(precision)

on this scale, information is additive

$$\left\{\begin{array}{l}(\mu\,\sigma \mid ?\,\mathcal{B}) \sim p(\mu\,\sigma \mid ?\,\mathcal{B}) \\ (Y_i \mid \mu\,\sigma\,\mathcal{N}_S\,\mathcal{B}) \stackrel{IID}{\sim} N(\mu, \sigma^2) \\ (i = 1, \ldots, n)\end{array}\right.$$

$(k = 2)$

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

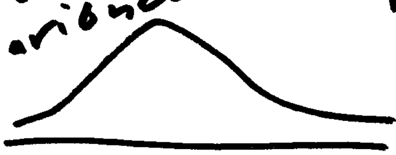
or $\underset{1 \times k}{\underset{\sim}{\theta}} = (\mu, \sigma^2)$

$(\mathbb{R}^2 \to \mathbb{R})$ ✓ $(\mathbb{R}^2 \to \mathbb{R})$ · $\mathbb{R}^2 \to \mathbb{R}$

$$p(\underset{\sim}{\theta} \mid y\,\mathcal{N}\,?\,\mathcal{B}) = c\,\boxed{p(\underset{\sim}{\theta} \mid ?\,\mathcal{B})}\,\ell(\underset{\sim}{\theta} \mid \overset{y}{\mathcal{N}}\,\mathcal{B})$$

this will work for any finite $k \geq 1$

SD σ
variance σ^2



PDF for Y

$V(Y) = \sigma^2 > 0$

$SD(Y) = \sigma > 0$

$\left\{\begin{array}{l}\text{SD} \\ \text{variance} \\ \text{of } Y\end{array}\right\} \leftarrow \left(\begin{array}{l}\text{what we} \\ \underline{\text{don't}} \text{ know} \\ \text{about } Y\end{array}\right) \leftrightarrow$ uncertainty

$\left(\begin{array}{l}\text{reciprocal} \\ \text{of variance}\end{array}\right) \leftrightarrow \left(\begin{array}{l}\text{what} \\ \text{we } \underline{\text{do}} \\ \text{know} \\ \text{about } Y\end{array}\right) \leftrightarrow$ information

$\begin{pmatrix}\text{general} \\ k \geq 2\end{pmatrix}$ $\underset{1 \times k}{\underset{\sim}{\theta}} = (\theta_1, \ldots, \theta_k)$

$p(\underset{\sim}{\theta} \mid D(\cdots)\mathcal{B}) = c \, p(\underset{\sim}{\theta} \mid (\cdots)\mathcal{B}) \cdot \ell(\underset{\sim}{\theta} \mid \cancel{\mathcal{B}}(\cdots)\mathcal{B})$

normalizing constant

$$c^{-1} = \underbrace{\int \cdots \int}_{\leftarrow k \rightarrow} p(\underset{\sim}{\theta} \mid (\cdots)\mathcal{B}) \cdot \ell(\underset{\sim}{\theta} \mid \cancel{\mathcal{B}}D(\cdots)\mathcal{B}) \, d\theta_1 \, d\theta_2 \cdots d\theta_k$$

marginals k of them, each requiring a $(k-1)$-dimensional integral:

e.g. $p(\theta_2 \mid D(\cdots)\mathcal{B}) = \underset{\theta_1}{\int} \underset{\theta_3}{\int} \cdots \underset{\theta_k}{\int}$ posterior

predictive dist.

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

$p(y_{n+1} \mid \underset{\sim}{y}(\cdots)\mathcal{B}) =$ another k-dimensional integral

new (future) data

$(\mu, \sigma^2 | ? \, \mathcal{B}) \sim p(\mu, \sigma^2 | ? \, \mathcal{B})$ conj. prior exists

$(I_i | \mu \, \sigma^2 \, \mathcal{N}_S \, \mathcal{B}) \stackrel{IID}{\sim} N(\mu, \sigma^2)$

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

("N-$\underline{\underline{IG}}$ prior")

$p(\mu, \sigma^2 | ? \, \mathcal{B}) = \cancel{p(\mu | ? \, \mathcal{B}) \cdot p(\sigma^2 | \mu \, ? \, \mathcal{B})}$ correct but not useful

$= \boxed{\underline{p(\sigma^2 | ? \, \mathcal{B})} \cdot p(\mu | \underline{\underline{\sigma^2}} \, ? \, \mathcal{B})}$

intuitive

hierarchical prior (2 $\underline{\text{levels}}$ or layers to the hierarchy)

$\left.\begin{array}{l} ① \; (\sigma^2 | ? \, \mathcal{B}) \\ ② \; (\mu | \sigma^2 \, ? \, \mathcal{B}) \end{array}\right\} = p(\underline{\underline{\mu}}, \sigma^2 | ? \, \mathcal{B})$

$(SI-\chi^2)$

$(\sigma^2 | CP \, \mathcal{B}) \sim \chi^{-2}(\nu_0, \sigma_0^2)$

(prior sample size for σ^2) (prior estimate of σ^2)

χ^2 dist: see Gelman et al. App. A

$(\mu, \sigma^2 \mid CP \; \mathcal{B}) \sim CP(\mu, \sigma^2 \mid \mathcal{B})$

$(y_i \mid \mu \, \sigma^2 \, \mathbb{N}_s \, \mathcal{B}) \stackrel{IID}{\sim} N(\mu, \sigma^2)$

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

LI conj. prior:

$\nu_0 \doteq 0$ ← prior sample size for σ^2

$> \to 0$

$k_0 \doteq 0$

$> \to 0$

prior sample size for μ

$(\mu \mid y \, \mathbb{N}_s \, (CP) \, \mathcal{B}) \sim$

$t_n\left(\mu_n, \frac{\sigma_n^2}{k_n}\right)$

$\nu_n = \nu_0 + n \to n$

$\mu_n \doteq \bar{y}$

$k_n = k_0 + n \to n$

$\sigma_n^2 \doteq \frac{n-1}{n} s^2$

$t_{100}\left(\bar{y}, \frac{s^2}{n}\right)$

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

$\downarrow \leftrightarrow (\mu \mid \sim) \stackrel{\cdot}{\sim} N\left(\bar{y}, \frac{s^2}{n}\right)$

$\hat{SE}(\hat{\mu}_{MLE}) = \frac{s}{\sqrt{n}}$ ✓