(continued from p. 5 of 4 Feb 21 document camera lecture notes)

STAT 206
5 Feb 21

Fri AM disc. sec.

Quiz 2 case study

$$p(\theta \mid y\, \mathcal{B}) = c \cdot p(\theta \mid \mathcal{B}) \cdot \ell(\theta \mid y\, \mathcal{B}) \quad (1)$$

$$= c\, p(\theta \mid \mathcal{B})\, \theta^{5} (1-\theta)^{n-5}$$

Look what nice thing happens when we use a Beta(α, β) prior for θ:

$$p(\theta \mid y\, \mathcal{B}) = c \cdot \left[\theta^{\alpha-1} \cdot (1-\theta)^{\beta-1}\right]\left[\theta^{5} \cdot (1-\theta)^{n-5}\right]$$

$$= c \cdot \theta^{(\alpha+5)-1} \cdot (1-\theta)^{(\beta+n-5)-1}$$

$$= \text{Beta}(\alpha+5, \beta+n-5) \quad (!)$$

What just happened? 2 excellent things: (1) the prior has the same mathematical form as the likelihood, and

The product of two such functions is also of the same form; this will ensure that all of {prior, likelihood, posterior} are of the ~~same~~ same parametric PDF form

(* Continued * on p. (3) below)

Take care with the (-1)s in the exponents of θ and $(1-\theta)$ in the Beta(α, β) family: $c\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$

This means that the likelihood in Beta$(\cdot,\cdot)$ form, expressing is slightly awkward-looking but correct:

$$\ell(\theta \mid y\, \mathcal{B}) = c\,\theta^{s}(1-\theta)^{n-s}$$

$$= c\,\theta^{(s+1)-1}(1-\theta)^{(n-s+1)-1} = \text{Beta}(s+1,\, n-s+1)$$

[(*) (*) Whenever the 2 nice things at the bottom of p.(1) and the top of p.(2) happen, people say that you've found a **conjugate prior**: ~~we've shown that~~ the Beta(α,β) family of prior PDFs is conjugate to the IID Bernoulli(θ) sampling distribution/likelihood.

we have our first **conjugate updating rule**:

observed data

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

the parameter of the sampling model

hyperparameters

$$\left.\begin{array}{l} (\theta \mid \mathcal{B}) \sim \text{Beta}(\alpha, \beta) \\ (Y_i \mid \theta\, \mathcal{B}) \overset{IID}{\sim} \text{Bernoulli}(\theta) \\ \quad (i = 1, \ldots, n) \end{array}\right\} \begin{array}{l} (\theta \mid \underset{\sim}{y}\, \mathcal{B}) = \\ (\theta \mid s\, \mathcal{B}) = \end{array}$$

$$\text{Beta}(\alpha + s, \beta + n - s)$$

in which $s = \sum_{i=1}^{n} y_i$ is minimal sufficient

$\theta \sim \text{Beta}(\alpha, \beta) \to p(\theta) = c\, \theta^{\alpha-1}(1-\theta)^{\beta-1}$

$(\alpha, \beta) = (1, 1) \to \theta \sim \text{Uniform}(0, 1)$

$(\alpha, \beta) = (1, \frac{1}{2}) \to p(\theta) = c\, \theta^{0} (1-\theta)^{-\frac{1}{2}}$

$= \dfrac{c}{\sqrt{1-\theta}} \to +\infty$ as $\theta \to 1$

as either $\alpha \downarrow 0$ or $\beta \downarrow 0$,

the $\text{Beta}(\alpha, \beta)$ PDF becomes

<u>improper</u> $\leftrightarrow$ it integrates to $+\infty$

under some circumstances

improper prior distributions

can work just fine

$$p(\theta) = c\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

reverse role of (α, β):

reflect PDF around $\theta = \frac{1}{2}$

$(\alpha = \beta) \to$ symmetric around $\theta = \frac{1}{2}$

$(\alpha, \beta) = (2, 2) \to c\,\theta^{2-1}(1-\theta)^{2-1}$

$= c\,\theta(1-\theta)$

(concave parabola) $= c\,[\theta - \theta^2]$

$(\alpha < \beta) \to$ long right-hand tail

positive skew

$(\alpha > \beta) \to$ long left-hand tail ($\ominus$ skew)

6

'prior dataset'

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{matrix} \uparrow \\ \beta \\ \downarrow \\ \uparrow \\ \alpha \\ \downarrow \end{matrix}$$

$\underline{\underline{n}}^* = (\alpha+\beta)$ 'prior sample size'

$\theta \sim \text{Beta}(\alpha, \beta)$

$E(\theta) = \dfrac{\alpha}{\alpha+\beta}$

mean $\dfrac{\alpha}{\alpha+\beta}$ ✓ (prior mean)

posterior: $(\alpha+s)$ votes for 1

$(\beta+n-s)$ — 0

sample dataset

$$\begin{array}{c} y_1 \\ | \\ y_n \end{array} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{matrix} \uparrow \\ n-s \\ \downarrow \\ \uparrow \\ s \\ \downarrow \end{matrix}$$

921 s votes for 1

$(n-s)$ —— 0

$\underline{\underline{n}}$ ← data sample size

sum $s = \sum_{i=1}^{n} y_i$

data mean $\bar{y} = \dfrac{s}{n}$

$= \dfrac{s}{s+(n-s)}$ ⟶

$(\theta \mid \mathcal{B}) \sim \text{Beta}(\alpha, \beta)$

$(y_i \mid \theta\, \mathcal{B}) \overset{IID}{\sim} \text{Bernoulli}(\theta)$

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

$(\theta \mid \mathbf{y}\, \mathcal{B}) \sim \text{Beta}(\alpha+s, \beta+n-s)$

2 different analysis plans:

$\boxed{1}$ do Beta prior to Beta posterior updating using Bayes's Theorem

or

$\boxed{2}$ merge 'prior', sample data sets into one big dataset

$$\begin{array}{cl} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline 1 \\ \vdots \\ 1 \end{bmatrix} & \begin{array}{l} \beta + n - s \\ \\ \\ \alpha + s \end{array} \end{array}$$

& do a likelihood analysis on the merged dataset

~~when prior is~~ conjugate

→ you get the same results with 2 different analysis plans

posterior sample size $= (\alpha + \beta) + n = n^* + n$