

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

STAT 206
5 Feb 21
Fri AM
disc. sec.

Quiz
case
study

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

$$= c p(\theta | \mathcal{B}) \theta^5 (1-\theta)^{4-5}$$

look what nice thing happens when we use a $\text{Beta}(\alpha, \beta)$ prior for θ :

$$p(\theta | \mathcal{B}) = c \cdot \left[\theta^{\alpha-1} \cdot (1-\theta)^{\beta-1} \right] \left[\theta^5 \cdot (1-\theta)^{4-5} \right]$$

$$= c \cdot \theta^{(\alpha+5)-1} \cdot (1-\theta)^{(\beta+4-5)-1}$$

$$= \text{Beta}(\alpha+5, \beta+4-5) \quad (!)$$

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

② The product of two such functions ② 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 (-)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 is ^{expressing} Beta (\cdot, \cdot) form is slightly awkward-
but correct:
looking λ $l(\theta | \mathcal{Z}, \mathcal{B}) = c \theta^s (1-\theta)^{n-s}$
 $= c \theta^{(s+1)-1} (1-\theta)^{(n-s+1)-1} = \text{Beta}(s+1, n-s+1)$

Whenever the 2^{nice} things at the bottom ③ 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: $\underline{\alpha} = (\alpha_1, \dots, \alpha_n) \mid \underline{y} = (y_1, \dots, y_n)$ observed data

hyperparameters

parameter of the sampling model $(\theta | \underline{\alpha}) \sim \text{Beta}(\underline{\alpha}, \beta)$

the IID $(Y_i | \theta) \sim \text{Bernoulli}(\theta)$ (i=1, ..., n)

$(\theta | \underline{y}, \beta) = \text{Beta}(\underline{\alpha} + \underline{s}, \beta + n - s)$

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

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

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

$$(\alpha, \beta) = (1, \frac{1}{2}) \rightarrow p(\theta) = c \theta^0 (1-\theta)^{-\frac{1}{2}}$$
$$= \frac{c}{\sqrt{1-\theta}} \rightarrow +\infty$$

as $\theta \rightarrow 1$

as either $\alpha \downarrow 0$ or $\beta \downarrow 0$,
the $\text{Beta}(\alpha, \beta)$ PDF becomes
improper \leftrightarrow it integrates to $+\infty$

under some circumstances
improper prior distributions
can work just fine

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

(5)

reverse role of (α, β) :

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

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

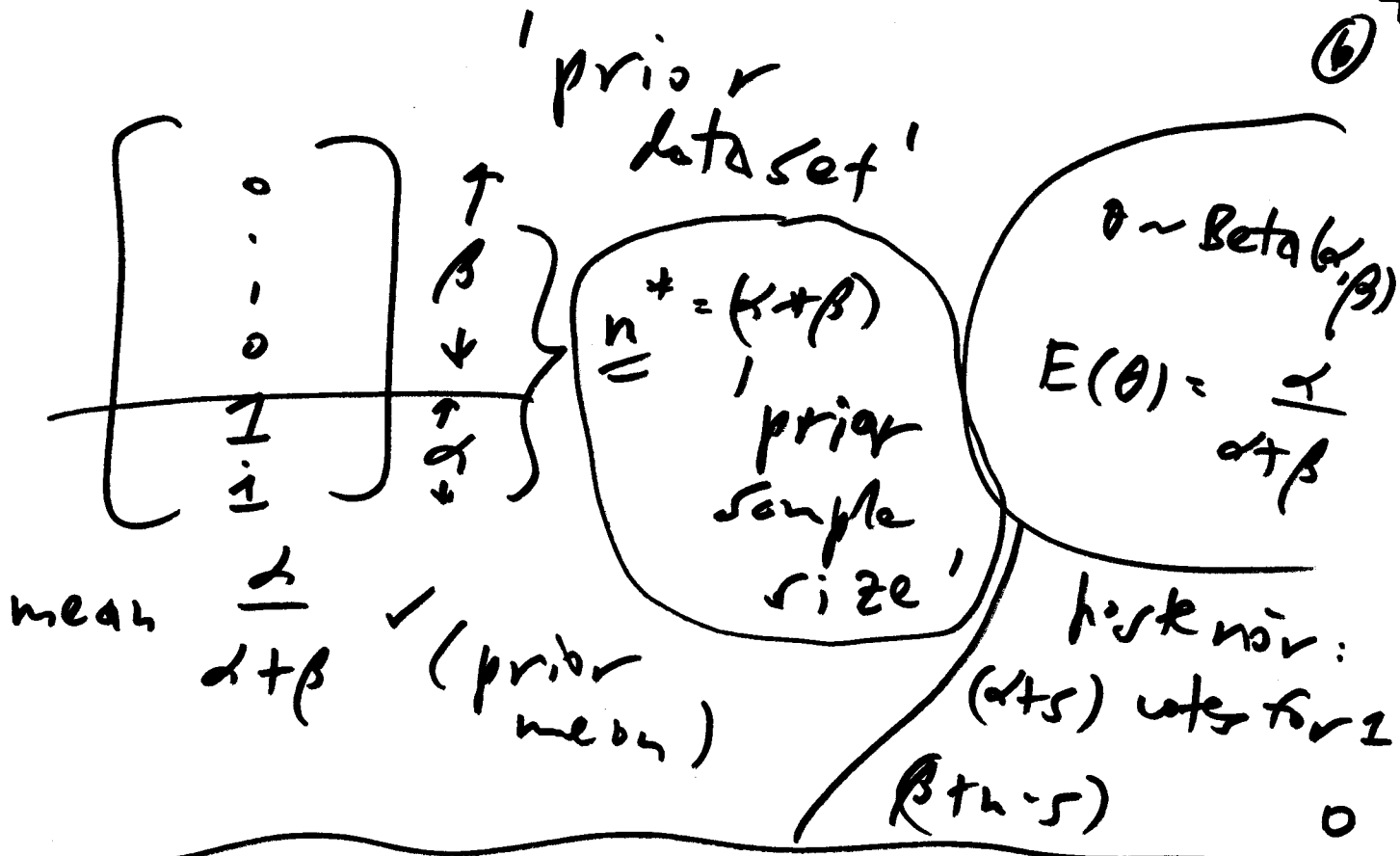
$$(\alpha, \beta) = (2, 2) \rightarrow c \theta^{2-1} (1-\theta)^{2-1}$$

$$\begin{aligned} &= c \theta (1-\theta) \\ \text{(concave parabola)} &= c (\theta - \theta^2) \end{aligned}$$

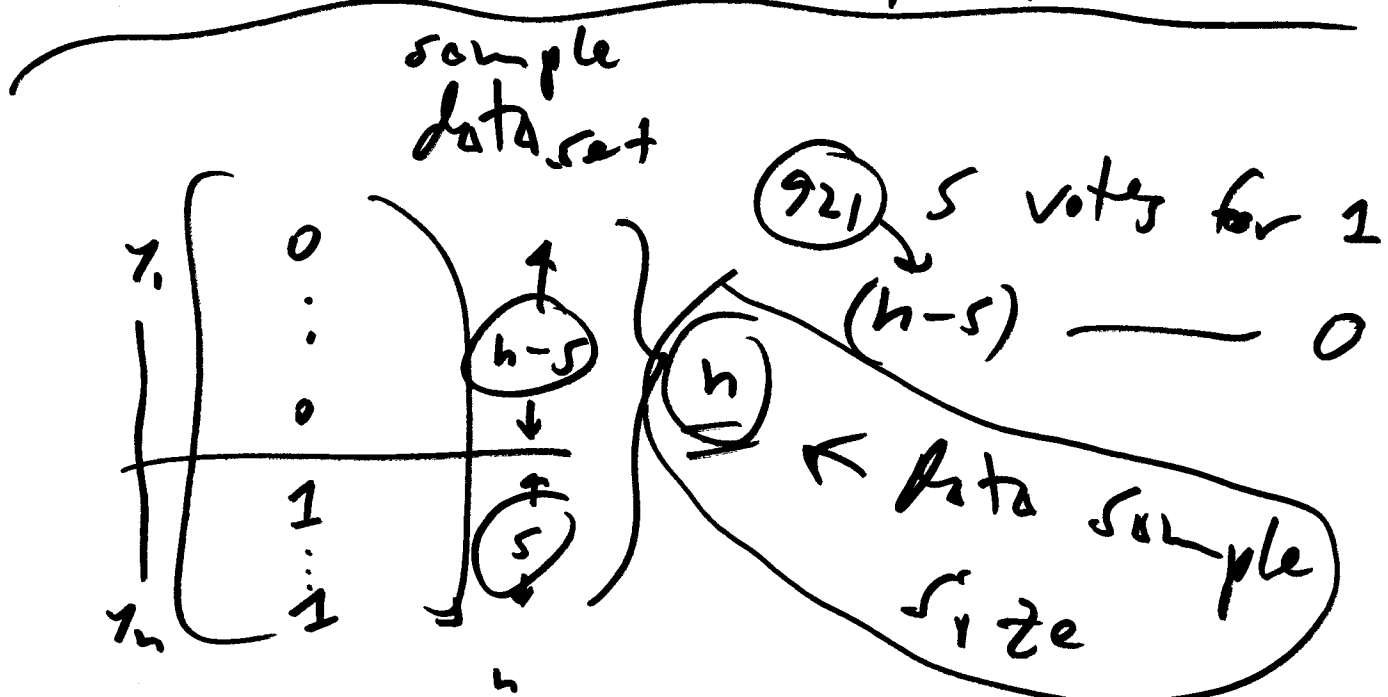
$(\alpha < \beta) \rightarrow$ long right-hand tail
positive skew

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

⑥



mean $\frac{\alpha}{\alpha + \beta}$ (prior mean)



Sum $S = \sum_{i=1}^n y_i$
 data mean $\bar{y} = \frac{S}{n}$
 $= \frac{S}{s + (n-s)}$

$(\theta | \mathcal{D}) \sim \text{Beta}(\alpha, \beta)$
 $(z_i | \theta) \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(\theta)$
 $(i=1, \dots, n)$

 $(\theta | z, \mathcal{B}) \sim \text{Beta}(\alpha + s, \beta + n - s)$

2 different analysis plans:

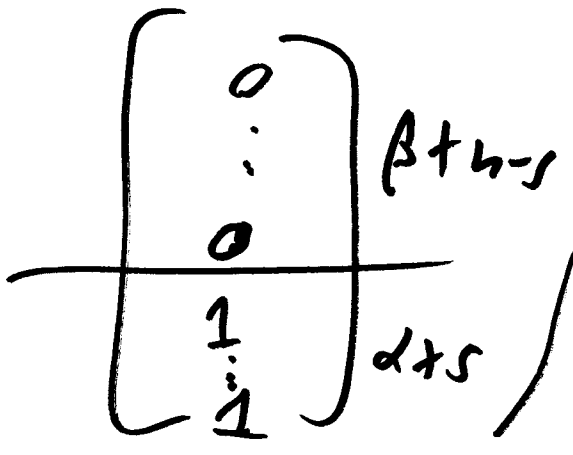
1 do Beta prior to Beta posterior updating using Bayes's Theorem

or

merge 'prior', sample data

2

sets into one big dataset



do a likelihood analysis on the merged dataset

when prior conjugate → you get the same results with 2 different analysis plans

posterior sample size = (d + β) + h = h* + h