quiz 2 case study: full employment in Santa Cruz

STAT 206 9 Feb 21

Lecture ①

Algorithm for Bayesian conjugate analysis:

sampling model

$$(Y_i \mid \theta, \mathbb{B}) \overset{IID}{\sim} \text{Bernoulli}(\theta) \quad (0 < \theta < 1)$$

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

$\underset{\sim}{Y} = (Y_1, \dots, Y_n)$ (same as Mr. Fisher)

$\boxed{1}$ specify marginal sampling distribution for a single observation (PMF or PDF)

Here this is

(abbreviation)

$$p_{Y_i}(y_i \mid \theta, \mathbb{B}) = p(y_i \mid \theta, \mathbb{B})$$

$$\begin{cases} \theta & \text{for } y_i = 1 \\ 1 - \theta & y_i = 0 \\ 0 & \text{else} \end{cases}$$

computer science

$$= \theta^{y_i} (1-\theta)^{1-y_i} \, I(y_i = 0 \text{ or } 1)$$

(math)

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

$\boxed{2}$ Construct joint sampling distribution (PMF or PDF)

Same as Mr. Fisher

Here this is (from IID)

(abbreviation)

$$p_{\underset{\sim}{Y}}(\underset{\sim}{y} \mid \theta, \mathbb{B}) = p(\underset{\sim}{y} \mid \theta, \mathbb{B}) = \prod_{i=1}^{n} p(y_i \mid \theta, \mathbb{B})$$

$$p(y \mid \theta\,\mathbb{B}\,\mathcal{B}) = \prod_{i=1}^{n} p(y_i \mid \theta\,\mathbb{B}\,\mathcal{B})$$

$$= \prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i}\, I(y_i = 0 \text{ or } 1)$$

$$= \left[\theta^{y_1} (1-\theta)^{1-y_1}\, I(y_1 = 0 \text{ or } 1)\right] \cdot$$

$$\boxed{\begin{array}{c}\text{T/F prop. } A, B: \\ I(A) \cdot I(B) = \\ I(A \text{ and } B)\end{array}}$$

$$\left[\theta^{y_2} (1-\theta)^{1-y_2}\, I(y_2 = 0 \text{ or } 1)\right] \cdot \ldots$$

$$\cdot \left[\theta^{y_n} (1-\theta)^{1-y_n}\, I(y_n = 0 \text{ or } 1)\right]$$

$$= \theta^{y_1 + \cdots + y_n} (1-\theta)^{n-(y_1 + \cdots + y_n)} \cdot$$

$$I(\text{all } y_i = 0 \text{ or } 1)$$

---

define $s = \sum_{i=1}^{n} y_i$, the data realization of the random variable $S = \sum_{i=1}^{n} Y_i$

[from now on, let's consider only ③]

data sets $\underset{\sim}{y} = (y_1, \ldots, y_n)$ such that

(all $y_i = 0$ or 1); then $I(\text{all } y_i = {}^{0}_{\text{or } 1}) = \underline{1}$

④ we can forget about the $I(\cdot)$ term

---

so finally the joint sampling distribution

is $p(\underset{\sim}{y} \mid \theta \, \mathbb{B} \, \mathcal{B}) = \theta^s (1-\theta)^{n-s}$

(still same as Mr. Fisher)

$\boxed{3}$ define the likelihood function

$$\ell(\theta \mid \underset{\sim}{y} \, \mathbb{B} \, \mathcal{B}) = \overset{(c>0)}{c} \, p(\underset{\sim}{y} \mid \theta \, \mathbb{B} \, \mathcal{B})$$

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

$\boxed{3.1}$ plot this as a function of θ for

interesting values of $(s, n) = (830, 921)$

e.g. Quiz 2

$\boxed{3.2}$ Notice that $s = \sum_{i=1}^{n} y_i$ is a (minimal) sufficient statistic (if, as usual, we adopt the convention that the sample size (n) is also needed to evaluate the likelihood function but we don't include (n) in the vector of (minimal) sufficient statistics), and that $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ is also (minimal sufficient) for θ in this Bernoulli sampling model

still same as Mr. Fisher

This means that we can write

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

$\boxed{4}$ Now we part company with Mr. Fisher:

Bayes's Theorem

~~posterior = $\underset{(c>0)\text{ constant}}{c}$ · (prior) · (likelihood)~~

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

5

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

conditional PDF for θ given both prior & likelihood information ~~& B~~ & B = prior PDF for θ given B · unnormalized likelihood PDF for θ given (y, and B)

to quantify our uncertainty about θ, in the Bayesian story we treat θ as if it were a realization of a random variable living (in this case) continuously on $(0, 1)$ (i.e., $0 < \theta < 1$)

$\boxed{5}$ see if this sampling model yields a conjugate prior: $\boxed{5A}$ Is the likelihood PDF $c\, \theta^{s} (1-\theta)^{n-s}$ a member of a known parametric family?

(6)

(A) (look in Appendix A of Gelman et al.) Yes, $\ell(\theta \mid y\, \mathcal{B}) = c\, \theta^{s} (1-\theta)^{n-s}$

is a member of the Beta(α, β) parametric family of PDFs on $(0, 1)$: $\left(\begin{array}{l}\text{for} \\ \alpha > 0 \\ \beta > 0\end{array}\right)$

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

Thus the likelihood PDF is

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

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

$\boxed{5B}$ Q: Is the product of 2 Beta PDFs another Beta PDF?

A: Yes:

$$p(\theta \mid x\, \mathcal{B}\, \mathcal{B}) = \underbrace{\left[c\, \theta^{\alpha - 1} (1-\theta)^{\beta - 1} \right]}_{\substack{\text{Beta}(\alpha, \beta) \\ \boxed{\text{prior}}}} \cdot \underbrace{\left[c\, \theta^{s} (1-\theta)^{n-s} \right]}_{\substack{\text{Beta}(s+1, \\ n-s+1) \\ \boxed{\text{likelihood}}}}$$

$\boxed{\text{posterior}}$

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

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

This shows that the Beta(α,β) family of prior PDFs is <u>conjugate</u> to the Bernoulli likelihood function

conjugate priors

Advantages: (1) makes the math easy: you don't need to mess with the (normalizing constant)

Disadvantages: ① conjugate priors don't always exist ② even when they *do* exist, we're not compelled to use them; they're mathematically convenient (sufficient) but not necessary

Advantages ② when a conjugate prior exists, there will always be some parameterization [$\eta = g(\theta)$ for some monotonic $g(\cdot)$] such that two really nice things happen (sometimes $g(\theta) = \theta$; ex. quiz 2 case study) on that scale:

Nice thing ① It turns out if $(\theta \mid \alpha, \beta) \sim \text{Beta}(\alpha, \beta)$ that then $E(\theta \mid \alpha, \beta) = \frac{\alpha}{\alpha + \beta}$

prior mean | $(\theta \mid \alpha, \beta) \sim \text{Beta}(\alpha, \beta)$

so $\begin{pmatrix}\text{prior}\\ \text{mean}\end{pmatrix} = \dfrac{\alpha}{\alpha+\beta}$ | data mean | $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{s}{n}$

posterior mean | $(\theta \mid y \; \alpha\beta) \sim \text{Beta}(\alpha+s, \beta+n-s)$

so $\begin{pmatrix}\text{posterior}\\ \text{mean}\end{pmatrix} = \dfrac{\alpha+s}{(\alpha+s)+(\beta+n-s)} = \dfrac{\alpha+s}{\alpha+\beta+n}$

now comes something pretty:

$$\frac{\alpha+s}{\alpha+\beta+n} = \frac{\underbrace{(\alpha+\beta)}_{\text{prior sample size}}\underbrace{\left(\frac{\alpha}{\alpha+\beta}\right)}_{\text{prior mean}} + \underbrace{(n)}_{\text{data sample size}}\underbrace{\left(\frac{s}{n}\right)}_{\text{data mean}}}{(\alpha+\beta)+(n)}$$

$\begin{pmatrix}\text{posterior}\\ \text{mean}\end{pmatrix} =$

$$\begin{pmatrix}\text{posterior}\\ \text{mean}\end{pmatrix} = \text{weighted average of}$$

$$\begin{pmatrix}\text{prior}\\ \text{mean}\end{pmatrix} \text{ and } \begin{pmatrix}\text{data}\\ \text{mean}\end{pmatrix},$$

with weights given by

$$(\alpha + \beta) \quad \text{and} \quad (n)$$

$$\begin{pmatrix}\text{prior}\\ \text{sample}\\ \text{size}\end{pmatrix} \qquad \begin{pmatrix}\text{data}\\ \text{sample}\\ \text{size}\end{pmatrix}$$

---

definition given a (data) vector $\underset{\sim}{x} = (x_1, \dots, x_n)$ of real numbers, the weighted $\left\{\begin{matrix}\text{mean}\\ \text{average}\end{matrix}\right\}$ of $\underset{\sim}{x}$ with weights $\underset{\sim}{w} = (w_1, \dots, w_n)$ satisfying $w_i \geq 0$ for all i and $\sum_{i=1}^{n} w_i > 0$

is $\dfrac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$

"Bayes's Theorem with a conjugate prior works just like voting in a 1-person-1 vote system: in the

$\boxed{\text{Beta}(\alpha, \beta) - \text{Bernoulli likelihood}}$ situation,

it's as if the data set (information) gets n votes and the prior "data set" (information) gets $n^* = (\alpha + \beta)$ votes in computing the posterior mean

$$\frac{\alpha + s}{\alpha + \beta + n} = \frac{\underbrace{(\alpha+\beta)}_{\text{prior sample size}}\left(\frac{\alpha}{\alpha+\beta}\right) + \underbrace{n}_{\text{data sample size}}\left(\frac{s}{n}\right)}{(\alpha+\beta) + (n)}$$

Nice thing (2) | As I said in a recent course meeting, we can make this story tangible:

$$\boxed{\text{I}} \quad \begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array} \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hline 1 \\ \vdots \\ 1 \end{array}\right] \begin{array}{c} \uparrow \\ n-s \\ \downarrow \\ \uparrow \\ s \\ \downarrow \end{array} \Bigg\} \; n$$

sample data set

data sample size

mean $\bar{y} = \frac{s}{n} = \frac{s}{s+(n-s)}$

$$\boxed{\text{II}} \quad \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hline 1 \\ \vdots \\ 1 \end{array}\right] \begin{array}{c} \uparrow \\ \beta \\ \downarrow \\ \uparrow \\ \alpha \\ \downarrow \end{array} \Bigg\} \; n^* = (\alpha+\beta)$$

prior data set

prior sample size

mean $\frac{\alpha}{\alpha+\beta}$

$$\left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hline 1 \\ \vdots \\ 1 \end{array}\right] \begin{array}{c} \\ \beta+n-s \\ \\ \\ \alpha+s \\ \\ \end{array} \Bigg\} \; (n+n^*)$$

mean $\frac{\alpha+s}{\alpha+\beta+n}$

merged data set

Combine $\boxed{\text{I}}$ and $\boxed{\text{II}}$ into a single column vector

Here are 2 different-looking analysis plans in the Quiz 2 context + ~~Analysis~~ plan

(1) use (AP) conjugate prior-to-posterior updating via Bayes's Theorem & summarize your total information about θ with the resulting posterior distribution:

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

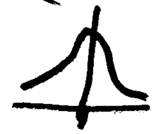
where $s = \sum_{i=1}^{n} y_i$ is minimal sufficient for θ

this is our first conjugate updating rule (among quite a few)

Analysis plan 2 (AP) do a full Mr.-Fisher-style likelihood analysis on the merged data set on p. 12 of today's notes

Superb fact: Analysis plans 1 & 2 will essentially (large n) coincide in their inferential conclusions

Example: posterior mean $= \dfrac{\alpha + s}{\alpha + \beta + n}$ (AP 1)



$$\hat{\theta}_{MLE} = \left(\begin{array}{c} \text{sample mean of} \\ \text{merged data} \\ \text{set} \end{array} \right) = \frac{\alpha + s}{\alpha + \beta + n}$$

Fact about Beta(α, β) PDF

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

$$V(\theta | \alpha, \beta) = \frac{\alpha \cdot \beta}{(\alpha+\beta)(\alpha+\beta)(\alpha+\beta+1)}$$

$$= \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\beta}{\alpha+\beta}\right)\frac{1}{\alpha+\beta+1}$$

$\left(\text{like } \hat{\theta}_{MLE}\right)\left(\text{like } 1-\hat{\theta}_{MLE}\right)\left(\frac{1}{n^*+1}\right)$ like $\frac{1}{n}$

i.e. like $\frac{\hat{\theta}_{MLE}(1-\hat{\theta}_{MLE})}{n}$

so posterior SD is

$$SD(\theta | y\, \alpha\, \beta\, \mathcal{B}\mathcal{B}) = \sqrt{\frac{(\alpha+s)(\beta+n-s)\cdot 1}{(\alpha+s+\beta+n-s)^2(\alpha+s+\beta+n-s+1)}}$$

and the information-based standard error is

$$\sqrt{\frac{\hat{\theta}_{MLE}(1-\hat{\theta}_{MLE})}{\alpha+\beta+n}} =$$

identical to Bayesian answer except for the extra 1 in denom.

A $\underbrace{\text{proper prior}}$ for $\theta \in (0,1)$:

$\boxed{A}$ $p(\theta) \geq 0$ $\boxed{B}$ $\displaystyle\int_0^1 p(\theta)\,d\theta = 1$

An $\underline{\text{improper prior}}$ for $\theta \in (0,1)$:

$\boxed{A}$ $p(\theta) \geq 0$ $\boxed{B}$ $\displaystyle\int_0^1 p(\theta)\,d\theta = +\infty$

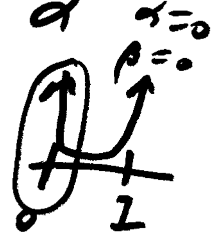
Beta (α, β) conjugate prior for Bernoulli likelihood

---

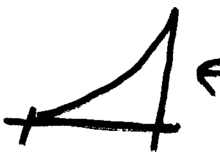
data set gets n votes	prior gets $(\alpha + \beta)$ votes

Suppose that context $\mathcal{C}$ in $\mathbb{P} = (\mathcal{P}, \mathcal{C})$ implies that the prior should have low information (LI) content (i.e., we don't know much about the full-employment rate in SC last month

we can express this (LI) by choosing α and β so that $(\alpha+\beta)$ is small



(positive but close to 0)



to get a proper posterior (which we must have for Bayes to work well), we need

$$\left\{ \begin{array}{c} \alpha+s>0 \\ \beta+n-s>0 \end{array} \right\}$$

e.g. even for ~~[illegible]~~ $s=1$ and $n=2$, $(\alpha=\beta=0)$ (improper prior) yields a proper posterior