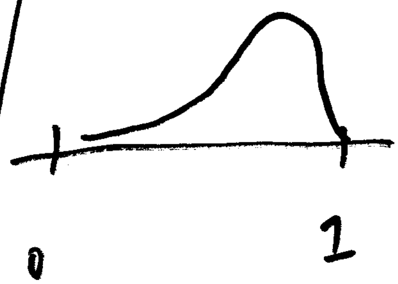
Wed AM discussion section

prior PDF for θ ($U(0,1) = Beta(1,1)$)



likelihood PDF for θ

posterior PDF for θ (identical to likelihood here)

this is a low-information (LI) prior in the conjugate $Beta(\alpha, \beta)$-Bernoulli model: prior sample size $n^* = (\alpha + \beta) = 2$; data sample size $= n$

$(s = 830)$ 921

Haldane prior: Beta$(0, 0)$ (improper)

but posterior proper as long as $s \geq 1$ and $n - s \geq 1$

Some LI priors

→ Beta$(0, 0)$

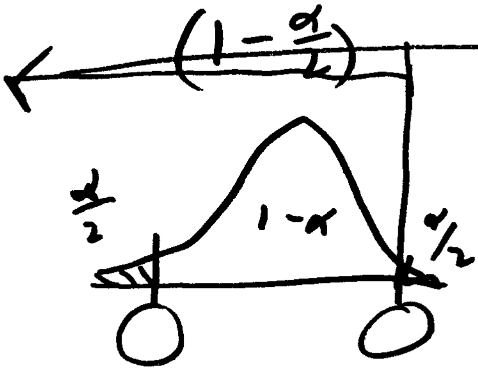
→ Beta$(0, 1)$

→ Beta$(1, 0)$

→ Beta$(1, 1)$

⋮

LI priors are not unique, but with large n they all lead to essentially the same posterior (stability across LI prior specification)



posterior PDF for θ given data & prior info & background

$$P_F(\theta > 0.95) = \text{undefined}$$

frequentist

$$P_B\left(\theta > 0.95 \,\middle|\, \begin{array}{l}\text{data info, } \overset{\text{LI}}{\text{prior info}}, \\ \text{background info}\end{array}\right)$$

perfectly well defined

≥ 0

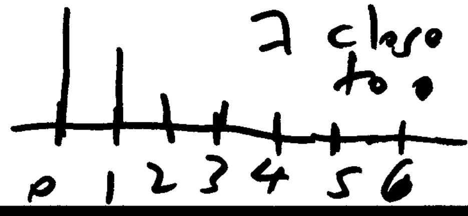
Bayesian sequential learning:

"Tomorrow's prior is today's

<u>posterior</u>"

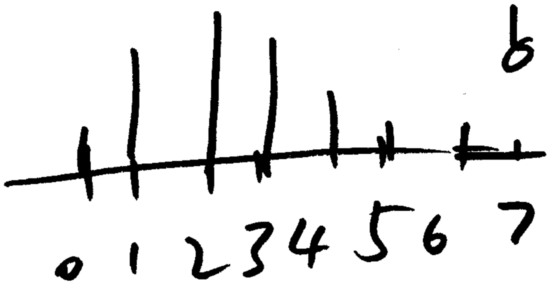
p. 74 of lecture notes part 6

LoS case study:



PMF

Poisson(λ) ($\lambda > 0$)



bigger λ

$\mathbb{P} = (\mathcal{Q}, \mathcal{C})$

$\theta = (\lambda)$ $\quad$ $k=1$

$(\cancel{\theta}, \mathcal{D}, \mathcal{B})$

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

$$\mathcal{M}_{I_{\mathbb{P}}} = \{ \underbrace{p(\lambda \mid \mathcal{B})}_{\text{prior}}, \underbrace{p(y \mid \lambda\, \mathcal{B})}_{\text{sampling dist.}} \}$$

inference and/or prediction

$$\boxed{\ell(\lambda \mid y\, \mathcal{B})}$$

here we have uncertainty about $p(y \mid \lambda\, \mathcal{B})$; it's not uniquely specified by $\mathcal{C}$

This is an example of <u>model uncertainty</u>

one solution: look at the data (*)

to arrive at a plausible sampling (5)
distn specification:

$$p(y_i \mid \lambda\, \mathbb{P}\, \mathcal{B}) = \begin{cases} \dfrac{\lambda^{y_i} e^{-\lambda}}{y_i!}, & \text{for } y_i = 0, 1, \ldots \\ 0 & \text{else} \end{cases}$$

Poisson assumption

but ⊛ involves

cheating: we used the data to

once to specify $p(\lambda \mid y\, \mathcal{B})$ &

again we used data to draw

conclusions about λ

⊛ pays no price for cheating

Leamer (1978)
Draper (1995)
Raftery et. al (1996)

Bayesian model averaging (BMA)

① marginal sampling dist:

$$p(y_i \mid \lambda\, P\, \mathcal{B}) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} I(y_i = 0 \text{ or } 1 \text{ or} \ldots)$$

② joint sampling dist:

$$(Y_i \mid \underline{\lambda}\, \mathbb{P}\, \mathcal{B}) \overset{IID}{\sim} \text{Poisson}(\lambda)$$

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

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

$$= \prod_{i=1}^{n} \frac{\lambda^{y_i} (e^{-\lambda})}{y_i!} I(y_i \in \{0, 1, \ldots\})$$

(**) $$= \frac{\lambda^{y_1 + \cdots + y_n} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!} I(\text{all } y_i \in \{0, 1, \ldots\})$$

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

$$(**) = \frac{\lambda^s e^{-n\lambda}}{\prod y_i!}$$

③ likelihood function

(1)

assume all $y_i \in \{0, 1, \ldots\}$

$$\ell(\lambda \mid \cancel{y}\ \mathbb{P}\ \mathcal{B}) = c\, \frac{\lambda^s e^{-n\lambda}}{\cancel{\prod_{i=1}^{n} y_i!}}$$

3A plot this for interesting (s, n)

$s = \sum_{i=1}^{n} y_i$ is ^minimal sufficient for λ in this sampling model