Single Parameter Model II:

Review:

$\theta \in (0,1)$

prior: $\theta \sim \text{uniform}(0,1) = \text{Beta}(1,1)$

model: $y_i \overset{iid}{\sim} \text{Bernoulli}(\theta) \Rightarrow$ likelihood $p(y_1, y_2, \cdots, y_n \mid \theta)$

$$= \theta^{\sum_{i=1}^n y_i} (1-\theta)^{n-\sum_{i=1}^n y_i}$$

posterior: $\text{Beta}(1+\underbrace{\sum_{i=1}^n y_i}_{\text{head}},\ \underbrace{1+n-\sum_{i=1}^n y_i}_{\text{tails}})$

Generalization: Conjugate priors

Extension: Binomial model with beta prior

$$\begin{cases} \text{prior:} & \theta \sim \text{Beta}(a,b) \Longleftrightarrow p(\theta) \boxed{\propto} \theta^{a-1}(1-\theta)^{b-1} \\ \text{likelihood:} & p(y_1, \cdots, y_n \mid \theta) = \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i} \\ \text{posterior:} & p(\theta \mid \underbrace{y_1, y_2, \cdots, y_n}_{\text{observation/data}}) = \dfrac{p(\theta)\, p(y_1, y_2, \cdots y_n \mid \theta)}{p(y_1, \cdots, y_n)} \end{cases}$$

parameter of interest | unknown quantity

$$\propto p(\theta)\, p(y_1, \cdots, y_n \mid \theta)$$

$$\propto \theta^{a-1}(1-\theta)^{b-1}\, \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i}$$

$$= \theta^{a+\sum_{i=1}^n y_i - 1}(1-\theta)^{b+n-\sum_{i=1}^n y_i - 1}$$

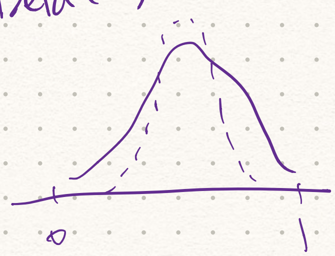
$$\propto \text{Beta}(\theta;\ \underbrace{a+\sum_{i=1}^n y_i}_{\text{head}},\ \underbrace{b+n-\sum_{i=1}^n y_i}_{\text{tail}})$$

(in 抽样中)

a: pseudo-count for head

b: pseudo-count for tail.

$\text{Beta}(\theta; a, b)$



Beta(1,1)

a.b越大，形状越陡.

When can we use "$\propto$"? (But make sure the constant is positive!).

when we want to eliminate a "constant"

e.g. $p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, if $p(x) \propto \exp(-x^2/2)$

$\to x \sim N(0,1)$

if the term contains $\theta \Rightarrow$ cannot reduce it!

generally ignore $p(y_1, \dots, y_n)$

we just use posterior $\propto$ prior * likelihood.

Quiz 2.

$$\underbrace{\underset{}{\overset{\text{Beta}(a,b)}{\text{prior}}} \xrightarrow{\text{data 1}} \overset{\text{Beta}(a+\Sigma,\ b+n-\Sigma)}{\text{posterior}}}_{\text{Beta - Binomial}} \xrightarrow{\text{data 2}} \underbrace{\overset{\text{Beta}(a'+\Sigma,\ b'+n_2-\Sigma)}{\text{posterior}}}_{\text{Beta - Binomial}}$$

$a+\Sigma \to a'$, $b+n-\Sigma \to b'$

since it's another Beta dis'

Gaussian Conjugate Model:

conjugacy: if $\mathcal{F}$ is a class of sampling distributions $p(y|\theta)$ → Binomial (Bernoulli)

and $\mathcal{P}$ is a class of prior distribution for θ. → Beta

then class $\mathcal{P}$ is conjugate for $\mathcal{F}$ if $\underbrace{p(\theta|y) \in \mathcal{P}}_{\text{posterior}}$ Beta for all $p(y|\theta) \in \mathcal{J}$ & $\underbrace{p(\theta)}_{\text{prior}} \in \mathcal{P}$

prior	likelihood	posterior	
Beta (uniform)	Binomial	Beta	✓
Gaussian	Gaussian	Gaussian	(in class)
Gamma	Poisson	Gamma	(quiz)
⋮	⋮	⋮	

for exponential families $\longrightarrow$ natural conjugate prior

△ ① General form $p(y|\theta)$ for exponential family

& its natural conjugate prior:

② Use the Beta-Binomial & match to the Beta distribution.

Conjugate Gaussian - Gaussian :

Example 2. normal with variance known

data $\{y_1, \dots, y_n\} \in \mathbb{R}$

parameter of interest : $\theta \in \mathbb{R}$

infere $\underset{(\text{unknown})}{\theta}$ given $y_1, y_2, \dots, y_n$ (consider <u>mean</u>)

Likelihood : $y_i \overset{iid}{\sim} N(\underset{\substack{\downarrow \\ \text{parameter of interest}}}{\theta}, \overset{\text{known constant (so far)}}{\sigma^2})$

△ prior distribution: $\theta \sim N(\underbrace{\mu_0, \tau_0^2}_{\text{known}})$

posterior distribution: (p.d.f) :

$$p(\theta \mid y_1, \dots, y_n) \propto p(\theta)\, p(y_1, y_2, \dots, y_n \mid \theta)$$

$$\propto \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right) \prod_{i=1}^{n} \exp\left[-\frac{(y_i-\theta)^2}{2\sigma^2}\right]$$

$$\propto \exp\left[-\frac{(\theta-\mu_0)^2}{2\tau_0^2} - \frac{\sum_{i=1}^{n}(y_i-\theta)^2}{2\sigma^2}\right]$$

$$= \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\theta^2 + \left(\frac{\mu_0}{\tau_0^2}+\frac{\sum_{i=1}^{n} y_i}{\sigma^2}\right)\theta \right.$$

$$\left. \underbrace{-\left(\frac{\mu_0^2}{\tau_0^2}+\frac{\sum_{i=1}^{n} y_i^2}{2\sigma^2}\right)}_{\text{throw! since not related with } \theta} \right]$$

$$\left(a\theta^2 + b\theta = a\left(\theta+\frac{b}{2a}\right)^2 - \frac{b^2}{4a} \right)$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta - \frac{\mu_0/\tau_0^2 + \sum y_i/\sigma^2}{1/\tau_0^2 + n/\sigma^2}\right)^2 \right]$$

$$\propto N(\mu_n, \tau_n^2)$$

$$\Rightarrow \mu_n = \frac{\mu_0/\tau_0^2 + \sum y_i/\sigma^2}{1/\tau_0^2 + n/\sigma^2} = (1-w_n)\mu_0 + w_n \bar{y}_n$$

posterior mean

$$\tau_n^2 = \frac{1}{1/\tau_0^2 + n/\sigma^2}$$

posterior variance

$(\tau_n^2)^{-1}$: posterior precision $= \frac{1}{\text{variance}}$

$$= \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$$

$\frac{1}{\tau_0^2}$ → prior precision

$\frac{n}{\sigma^2}$ → likelihood precision (since need to times n)

for μ_n, $w_n = \dfrac{n/\sigma^2}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}$ $\propto \frac{n}{\sigma^2}$ (从 likelihood)

→ $1 - w_n = \dfrac{\frac{1}{\tau_0^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}} \propto \frac{1}{\tau_0^2}$ (从 prior)