ESC-21WINTER

Bayesian Statistics

Week 2. Conjugacy

Bayesian Statistics
- 0. Recap
- 1. Single Parameter Conjugacy
  - 1.1 Binomial Model
  - 1.2 Poisson Model
  - 1.3 Normal Model
    - 1.3.1 unknown μ
    - 1.3.2 unknown σ^2
- 2. Two Parameter Conjugacy
  - 2.1 Normal Model : unknown $\mu,\ \sigma^2$
  - 2.2 Normal Model : semi-conjugacy

0. Recap

What is Bayesian Inference?

Bayes Rule : $p(\theta|\mathcal{D}) = \dfrac{p(\mathcal{D},\theta)}{p(\mathcal{D})} = \dfrac{p(\mathcal{D}|\theta)p(\theta)}{\int p(\mathcal{D}|\theta)p(\theta)d\theta}$

- $p(\mathcal{D}|\theta)$: Likelihood (data)

- $p(\theta)$: Prior

- $p(\theta|\mathcal{D})$: Posterior

closed form 구하기 어렵... :/

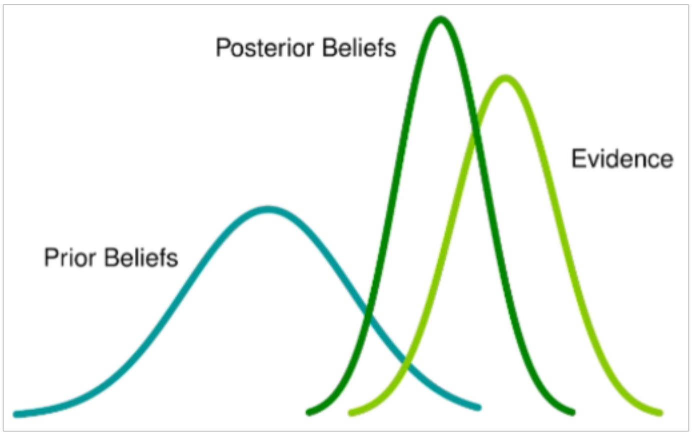
normalizing constant ($\int p(\mathcal{D}|\theta)p(\theta)d\theta$)의 적분 어려움 :(

=> Solution? Given that $p(\theta|\mathcal{D}) \propto p(\mathcal{D},\theta)$...

Sol 1) Conjugacy **

Sol 2) Sampling from unnormalized posterior (ex. MCMC)

1. Single Parameter Conjugacy



Conjugacy? 활용하다, 켤레

(ex. conjugate complex number)

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int p(\mathcal{D}|\theta)p(\theta)d\theta} \propto p(\mathcal{D}|\theta)p(\theta)$$

요 개형을 $p(\theta)$의 개형과 같게 되도록 틀을 맞춰주기

$$p(\theta)$$

conjugacy

* 이런 식으로 맞춰줄 수 있는 이유??

- $p(\theta|\mathcal{D}), p(\theta)$는 모두 pdf 이다. (sum up to 1)

- 특수한 꼴 : parameter의 지수연산 가능 (ex. Binomial, Exponential families) or exponential of a quadratic form of θ (ex. Normal families)

$$p(\theta) = e^{a\theta^2+b\theta+c}$$

*** Posterior as compromise between data(likelihood) and prior information**

우리가 이미 알고 있던 정보 (prior)에 새로운 정보, 단서가 주어졌을 때 (likelihood) 우리의 믿음이 "어떻게" 수정되는지 (posterior)

Binomial, Normal dist 등은 비교적 간단한 분포이고 근사, 응용 등이 쉬움 :)

(conjugate fmily is mathematically convenient in that the posterior distribution follows a known parametric form)

→ closed parametric form of dist. 쉽게 계산·해석가능

1.1 Binomial Model

"Beta prior distribution is a conjugate family for the binomial data(likelihood)"

Likelihood : $Y|\theta \sim Binom(n, \theta)$

Prior : $p(\theta) \sim Beta(a, b)$

$$E(\theta) = \frac{a}{a+b}$$

* uniform prior = Beta(1, 1)

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

e.g. (성공횟수의 합)

Posterior : $p(\theta|y) \sim Beta(a+y, b+n-y)$ where $y = Y = \sum y_i$

$$= \frac{p(\theta|y)p(y|\theta)}{p(y)} = \frac{1}{p(y)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} \times \binom{n}{y} \theta^{y}(1-\theta)^{n-y}$$

sufficient statistics (충분통계량)

C가 뭔진모르지만 암튼 p(θ|y)는 pdf임!

$$= C \times \theta^{a+y-1}(1-\theta)^{b+n-y+1}$$

$$E(\theta|y) = \frac{a+y}{a+b+n}$$

" weighted average "

$$= \frac{a+b}{a+b+n}\frac{a}{a+b} + \frac{n}{a+b+n}\frac{y}{n}$$

$$= \frac{a+b}{a+b+n} * prior\ expectation + \frac{n}{a+b+n} * data\ average$$

interpratability with respect to additional data :)

-> posterior 알았으면 그걸 바탕으로 inference, prediction 등 할 수 있음 (ex. $\tilde{y}$, C.I, etc.)

* compromise 확인하기 (다음 page Figure)

1.2 Poisson Model

"Gamma prior distribution is a conjugate family for the Poisson sampling model"

Likelihood : $Y|\theta \sim Pois(\theta) = \frac{1}{y!}\theta^{y}e^{-\theta}$

$Y_1, \ldots, Y_n|\theta \sim Pois(\theta) = \prod_{i=1}^{n} \frac{1}{y_i!}\theta^{y_i}e^{-\theta} \quad \propto \theta^{\sum y_i} e^{-n\theta}$

Prior : $p(\theta) \sim Gamma(a, b) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}$

$$E(\theta) = \frac{a}{b}$$

Posterior : $p(\theta|Y_1, \ldots Y_n) = Gamma(a + \sum Y_i, b+n)$

$$E(\theta|y) = \frac{a+\sum y_i}{b+n}$$

sum of counts from b prior obs.

$$= \frac{b}{b+n}\frac{a}{b} + \frac{n}{b+n}\frac{\sum y_i}{n}$$

\# of prior obs.

*how to compute...

$$p(\theta|y) \propto p(\theta)p(y|\theta) = \{\theta^{a-1}e^{-b\theta}\} \times \{\theta^{\sum y_i}e^{-n\theta}\} \times C$$

$$= \theta^{a+\sum y_i - 1} e^{-(b+n)\theta} \times C$$

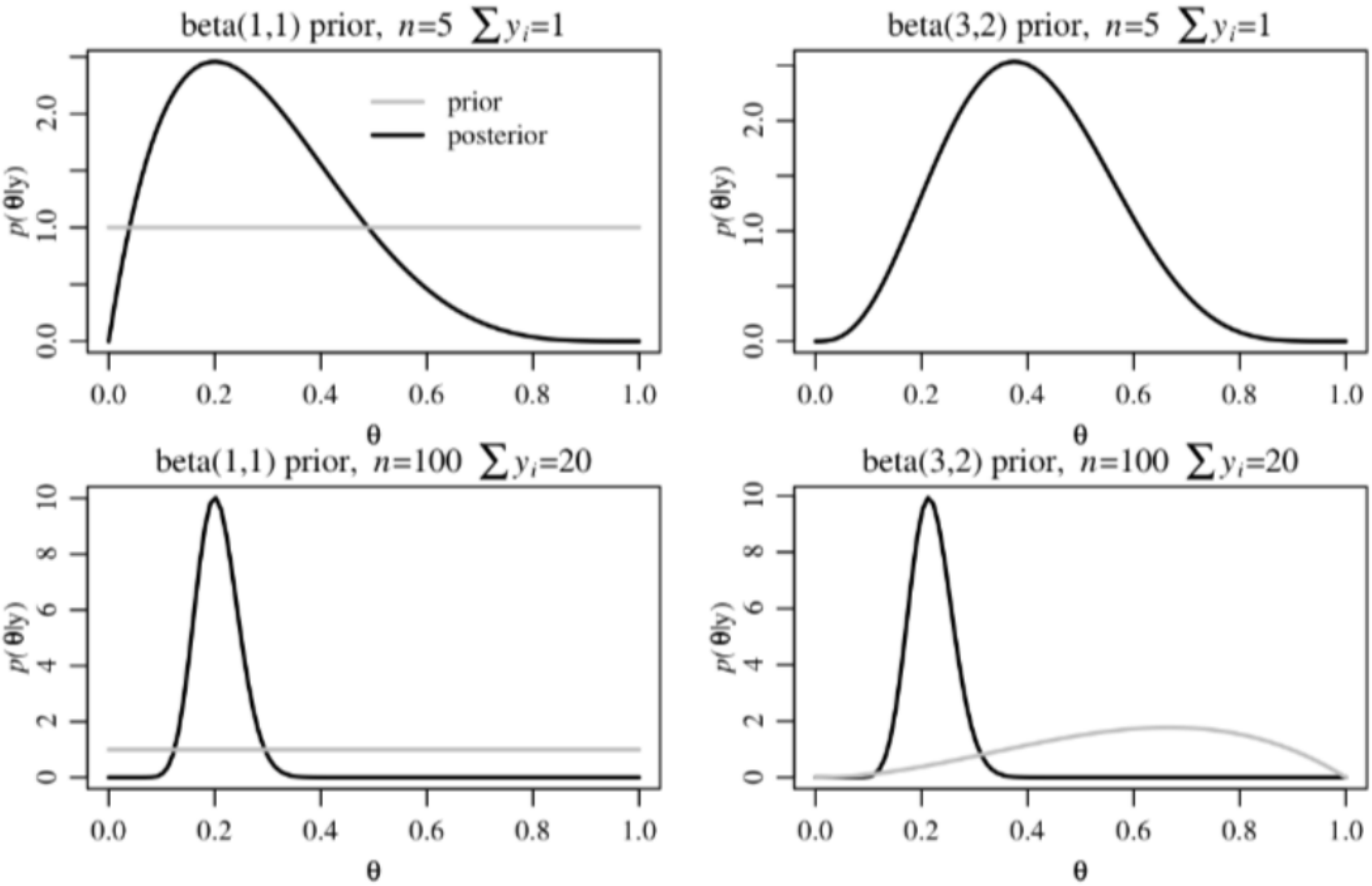


Fig. 3.4. Beta posterior distributions under two different sample sizes and two different prior distributions. Look across a row to see the effect of the prior distribution, and down a column to see the effect of the sample size.

1.3 Normal Model

1.3.1 unknown μ

$$\text{Likelihood}: \quad p(y|\mu\ (,\sigma)) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\sum(y-\mu)^2} \quad where \ \ y = y_1, \dots y_n$$

$$y \mid \mu, \sigma^2 \sim N(\mu, \underbrace{\sigma^2}_{\text{known}})$$

$$\text{Prior}: \quad p(\theta) \propto exp[-\frac{1}{2\tau_0^2}(\mu-\mu_0)^2]$$

$$\mu \sim \underbrace{N(\mu_0, \tau_0^2)}_{\text{known}}$$

$$\text{Posterior}: \quad p(\mu|y) \propto p(y|\mu)p(\mu) \propto exp[-\frac{1}{2}\underbrace{(\frac{\sum(y_i-\mu)^2}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\tau_0^2})}_{}]$$

θ, 즉 μ에 대한 이차식

$a\mu^2 + b\mu + c$

$$\mu|y, \sigma^2 \sim N(\mu_n, \tau_n^2)$$

... 계산파티 ...

$$\mu_n = \frac{\frac{1}{\tau_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}\mu_0 + \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}\bar{y}$$

$\bar{y}$ sufficient statistic

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

$$\downarrow$$

$$\underbrace{\frac{1}{\tau_n^2}}_{\text{posterior precision}} = \underbrace{\frac{1}{\tau_0^2}}_{\text{prior precision}} + \underbrace{\frac{1}{\sigma^2/n}}_{\text{data precision}}$$

$$\tilde{y} \mid y \sim N(\mu_n, \underbrace{\sigma^2}_{\text{y가 갖는 변동성}} + \underbrace{\tau_n^2}_{\mu \mid y}) \leftarrow \underbrace{p(\tilde{y}|y)}_{\text{posterior predictive}} = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$

* 분산의 역수 = precision

분산 ↑ 변동성 ↑ precision ↓

precision ↑

1.3.2 unknown σ^2

1.3.1과 달리 normalizing constant 살아있음

Likelihood $\quad y \mid \mu, \sigma^2 \sim N(\mu, \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum(y_i-\mu)^2\right)$

Prior $\quad \sigma^2 \sim \chi^{-2}(\nu_0, \sigma_0^2)$

Posterior $\quad \sigma^2 \mid y, \mu \sim \chi^{-2}(\nu_n, \sigma_n^2)$

$$\nu_n = \nu_0 + n$$

$$\sigma_n^2 = \frac{\nu_0\sigma_0^2 + ns(y)}{\nu_0 + n}$$

$$s(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2$$

2. Two Parameter Conjugacy

2.1 Normal Model : unknown $\mu,\ \sigma^2$

- Conditional prior $\quad p(\mu, \sigma^2) = p(\sigma^2)p(\mu|\sigma^2)$

  1) Uninformative prior

  2) Conjugate prior

- Posterior

  1) Joint posterior

  2) Conditional posterior

  3) Marginal posterior

  4) Posterior Predictive

2.2 Normal Model : semi-conjugacy

- Independent prior
- Gibbs sampler for posterior draws