Chapter 8: Prior Distribution

강의 목표

- 사전 분포의 이해
- 사전 분포 선택 방법
- 대표적인 사전 분포의 예
  - 무정보 사전분포 (Non-Informative)
  - 부적합 사전분포 (Non-Proper)
  - 제프리 사전분포
  - 공액 사전분포

Introduction

- 베이지안 추론을 위해서는 우도함수 (Likelihood)와 사전분포 (Prior)가 필요하다.
- Likelihood: 데이터의 유형에 따라 선택.
- Prior: 사전 정보에 따라 선택.
- 우도함수 (Likelihood)와 사전분포(Prior) 모두 베이지안 추론에 큰 영향.

Introduction

- 일반적으로 합리적인, 계산 가능한, 또는 사용 가능한 사전분포가 여러 개 존재
  1. $\theta \propto 1$
  2. $\theta \propto \frac{1}{\theta}$
  3. $\theta \sim Gamma(a, b)$
- 그러므로 적합한 사전분포를 찾는 과정을 알아야 한다. .

Type of Priors

- 정보 사전분포(informative prior):
  - 사전 정보
  - 이론적 지식
- 무정보 사전분포(non-informative prior):
  - 사전정보 (X)
  - 사용하기에 유의하지 못하거나 부적절한 정보
  - Informative prior와의 비교할 경우

Type of Priors

- 정보 사전분포(informative prior):
  - 사전 정보
  - 이론적 지식
- 무정보 사전분포(non-informative prior):
  - 사전정보 (X)
  - 사용하기에 유의하지 못하거나 부적절한 정보
  - Informative prior와의 비교할 경우

Non-informative Prior

- 무정보 사전분포는 사전정보가 거의 없거나 사용하지 않는 경우
- 사후분포에 대한 사전분포의 영향을 최소화. (e.g., $\theta \propto 1$)
- 사전정보를 사용하지 않은 베이지안 분석결과와 고전적 분석결과가 반드시 일치하는 것은 아님.

Example

- $X_1, ..., X_n$이 $Bin(n, \theta)$를 따르는경우를 보자.
- 사전 정보가 없다면, $U(0,1)$ 분포가 θ 의 무정보 사전분포 (non-informative prior)로 적절할 것이다

Another Example

- $X_1, \ldots, X_n$이 $N(\theta, \sigma^2)$를 따르고 σ^2은 알려진 상수라고 하자.
- 사전 정보가 없다면, θ에 대한 무정보 사전분포로 균일분포를 고려할 수 있다.
- 모든 θ값에 동일한 밀도함수값을 가지려면 어떤 상수 C에 대해서 $\pi(\theta) = c, -\infty < \theta < \infty$,이어야 한다.

Another Example

- 사전밀도함수에서 비례상수는 사후밀도함수 유도에 영향을 미치지 않으므로 편의상 $\pi(\theta) = 1, -\infty < \theta < \infty$, 을 θ에 대한 무정보 사전분포(non-informative prior)로 사용할 수 있겠다.
- 이때 $\int \pi(\theta) d\theta = \infty$ 이므로 부적절 사전분포 (improper prior) 이다.
- $\pi(\theta) = 1$ 사전분포를 사용하면 θ의 사후분포가 $N(\bar{x}, \sigma^2/n)$.

Example III

- $X_1, \ldots, X_n$이 $N(\theta, \sigma^2)$를 따르고 σ^2은 알려진 상수이다.
- 모수 θ에 대한 사전정보로 $\theta > 0$을 만족해야 한다고 하자.
- 그렇다면 $\theta > 0$인 구간에서 균일분포인 $\pi(\theta) = I(\theta > 0)$을 사전분포로 가정할 수 있다.

Example III

- 사후분포를 유도하면 ϕ와 Φ를 각각 표준정규분포의 확률밀도함수와 누적분포함수라고 할 때,

$$\begin{aligned} \pi(\theta \mid \bar{x}) &= \frac{\frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}} e^{-\frac{n}{2\sigma^2}(\theta-\bar{x})^2} I(\theta > 0)}{\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}} e^{-\frac{n}{2\sigma^2}(\theta-\bar{x})^2} d\theta} \\ &= \phi(\frac{\theta-\bar{x}}{\sigma/\sqrt(n)}) I(\theta > 0)/(1-\Phi(-\frac{\bar{x}}{\sigma/\sqrt{n}})) \\ &= \phi(\frac{\theta-\bar{x}}{\sigma/\sqrt(n)}) I(\theta > 0)/(\Phi(\frac{\bar{x}}{\sigma/\sqrt{n}})) \\ &\sim N(\bar{x}, \sigma^2/n) I(\theta > 0). \end{aligned}$$

- θ의 사후분포는 절단된 정규분포이다. 사전분포의 영향으로 사후분포는 $\theta > 0$인 구간에서만 정의되며 따라서 평균 $\bar{x}$가 어떤 값을 갖든지 θ의 사후 평균은 항상 0보다 클 것이다.

Example III

- θ의 사후구간 (posterior HPD interval)도 0보다 큰 구간만 포함할 것이다.
- 예를 들어 $\sigma^2 = 1, n = 5, \bar{x} = -0.1$로 주어진 경우를 생각해 보자. 이 때 95% HPD 구간은 $(0, c)$이며 c는

$$P(0 < \theta < c \mid \bar{x}) = \frac{\Phi(\frac{c-\bar{x}}{\sigma/\sqrt{n}}) - \Phi(\frac{0-\bar{x}}{\sigma/\sqrt{n}})}{\Phi(\frac{\bar{x}}{\sigma/\sqrt{n}})} = 0.95$$

를 만족하는 상수이다.

Example III

상수 c를 찾기 위해서 다음 R코드가 필요하다.

```
sig=1; n=5; xbar=-0.1
se=sig/sqrt(n)
theta=seq(0, 1.5, length=100)
post= dnorm( theta, xbar, se)/pnorm( xbar/se)
theta.neg=seq(-2,0, length=100)
post.neg= dnorm( theta.neg, xbar,se)/pnorm( xbar/se)
theta.all=c(theta.neg,theta)
post.all=c(post.neg,post)
plot(theta.all, post.all,type="n",xlab="theta",
ylab="posterior")
lines(theta,post,lty=1)
lines(theta.neg,post.neg,lty=2)
##### 95% HPD #####
c= qnorm( pnorm(xbar/se)*0.95 + pnorm(-xbar/se) ) *se + xbar
abline( v=c(0,c), lty=3)
text(0.4, 0.25, "95% HPD")
```

Example III

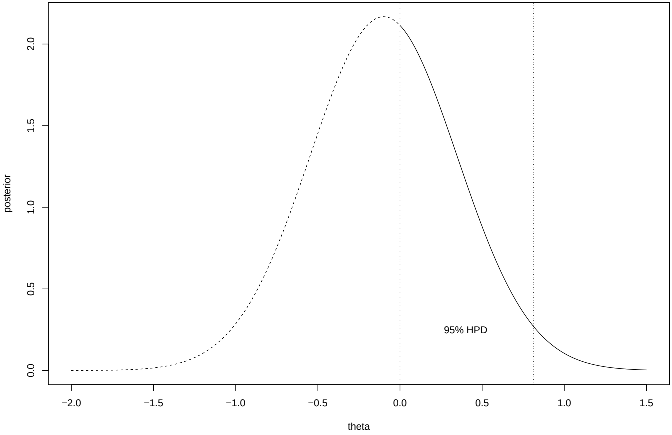


Figure: $\theta > 0$의 제한조건이 있는 경우 θ의 사후분포

Improper Prior

- 모수 공간에서의 적분값이 무한인 사전 밀도 함수:

$$\int \pi(\theta)d\theta = \infty.$$

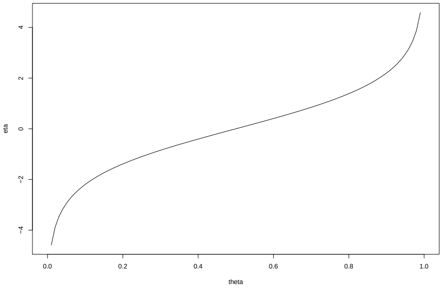
- e.g.,: $X \sim N(\theta, 1)$ and $\pi(\theta) = 1$ on $(-\infty, \infty)$.
- $\pi(\theta) = 1$ 사전분포를 사용하면 θ의 사후분포가 $N(\bar{x}, \sigma^2/n)$ 이므로 사후분포는 적합함(proper).

Ex 9.4

Jeffrey's Prior Motivation

- 무정보 사전밀도 함수가 모수에 따라 바뀔 수 있다.
- $\theta \in (0,1)$에서 non-informative prior $= \pi(\theta) \sim U(0,1)$.
- 만약 관심 모수가 odds ratio ($\eta = \log \frac{\theta}{1-\theta}$)라면,

  non-informative prior $= \pi(\eta) = 1$ on $(-\infty, \infty)$.



Jeffrey's Prior Motivation

- Consider the following hypothesis test:

$$H_0 : \theta \leq \frac{1}{2} \quad v.s \quad H_1 : \theta > \frac{1}{2}.$$

- Equivalently,

$$H_0 : \eta \leq 0 \quad v.s \quad H_1 : \eta > 0.$$

- It is natural to expect that

$$P(H_0 \mid x) = P\left(\theta \leq \frac{1}{2} \mid x\right) = P\left(\eta \leq \frac{1}{2} \mid x\right).$$

- It is called an invariance property.

Jeffrey's Prior Motivation

- 사전분포 (prior distribution)이 모수 (prameter)의 변환 (transformation)에 불변성 확보.
- 이를 위해 non-informative prior에 피셔의 정보 상수 (Fisher's information)를 이용.

$$I_1(\theta) = \mathbb{E}\left[\left(\frac{\partial^2 \log f(X \mid \theta)}{\partial \theta}\right)^2\right] = -\mathbb{E}\left[\frac{\partial^2 \log f(X \mid \theta)}{\partial \theta^2}\right].$$

- 만약 자료변수 $X_1, \ldots, X_n$이 서로 독립이고 같은 확률밀도함수 $f(X \mid \theta)$를 가지면 정보상수는 $I_n(\theta) = n \times I_1(\theta)$이다.

Jeffrey's Prior: Fisher's Information

- Let $h(\theta) = \eta$ for some differentiable $h(.)$. Then,

$$\left(\frac{\partial \log f(X \mid \eta)}{\partial \eta}\right) = \left(\frac{\partial \log f(X \mid \theta)}{\partial \theta}\right)\left(\frac{d\theta}{d\eta}\right).$$

- Therefore, $J(\eta) = I(\theta)\left(\frac{d\theta}{d\eta}\right)^2$.
- Let the prior $\pi_1(\theta) = \sqrt{I(\theta)}$. Then the prior $\pi_2(\eta)$ is:

$$\pi_2(\eta) = \pi_1(h^{-1}(\theta))\left|\frac{d\theta}{d\eta}\right| = \sqrt{I(h^{-1}(\eta))}\left|\frac{d\theta}{d\eta}\right| = \sqrt{J(\eta)}.$$

- 모수와 관계 없이 불변성을 가진다.

Jeffrey's Prior

- Jeffrey's Prior: $\pi(\theta) = \sqrt{I(\theta)}$.
- It is popular for a non-informative prior due to the invariance of parameters.
- It might be improper prior.

Jeffrey's Prior: Multi-parameters

- $\theta = (\theta_1, ..., \theta_p)$일 경우 vector θ에 대해서도 제프리 사전분포는 적용 가능하다.
- $(i, j)^{th}$ entry of Fisher's information:

$$I_{ij}(\theta) = -\mathbb{E}\left[\frac{\partial^2 \log f(X \mid \theta)}{\partial\theta_i \partial\theta_j}\right].$$

- Jeffrey's Prior: $\pi(\theta) = \sqrt{det(I(\theta))}$.

Example I

- Suppose that $X \sim B(n, \theta)$. Then,

$$\log f(x \mid \theta) = \log \binom{n}{x} + x \log(\theta) + (n - x) \log(1 - \theta)$$

이고

$$\frac{d \log f(x \mid \theta)}{d\theta} = \frac{x}{\theta} - \frac{n - x}{1 - \theta}, \quad \frac{d^2 \log f(x \mid \theta)}{d\theta^2} = -\frac{x}{\theta^2} - \frac{n - x}{(1 - \theta)^2}.$$

- Fisher's Information:

$$I(\theta) = \frac{n}{\theta(1 - \theta)}.$$

Example I

- θ에 대한 제프리 사전밀도함수는

$$\pi_1(\theta) = \sqrt{\frac{n}{\theta(1-\theta)}}.$$

- 사전밀도함수에서 비례상수는 무시해도 좋으므로

$$\pi_1(\theta) = \frac{1}{\sqrt{\theta(1-\theta)}}.$$

- 성공확률 θ가 아닌 로그 오즈 $\eta = \log(\frac{\theta}{1-\theta})$가 관심의 대상이라고 하자.
- 그렇다면

$$\theta = \frac{e^\eta}{1+e^\eta}.$$

Example I

- η 에 대한 정보상수 $J(\eta)$로부터 제프리 사전밀도함수를 구하면

$$\pi_2(\eta) = \frac{e^{\eta/2}}{1+e^{\eta}}.$$

- 그런데 제프리 사전밀도함수는 변환에 대하여 불변이므로 η에 대한 정보상수를 구하지 않고 θ와 η의 관계식으로부터

$$\pi_2(\eta) = \pi_1(h^{-1}(\eta)) \left| \frac{d\theta}{d\eta} \right| = \pi_1(e^{\eta}/(1+e^{\eta}))e^{\eta}/(1+e^{\eta})^2 = \frac{e^{\eta/2}}{1+e^{\eta}}$$

Example I

- Posterior distribution for θ:

$$
\begin{aligned}
P(\theta \mid x) &\propto \frac{1}{\sqrt{\theta(1-\theta)}} \theta^{x}(1-\theta)^{n-x} \\
&= \theta^{-1/2+x}(1-\theta)^{-1/2+n-x}.
\end{aligned}
$$

- Posterior distribution for η:

$$
\begin{aligned}
P(\eta \mid x) &\propto \frac{e^{\eta/2}}{1+e^{\eta}} \left(\frac{e^{\eta}}{1+e^{\eta}}\right)^{x} \left(\frac{1}{1+e^{\eta}}\right)^{n-x} \\
&= \left(\frac{e^{\eta}}{1+e^{\eta}}\right)^{1/2+x} \left(\frac{1}{1+e^{\eta}}\right)^{1/2+n-x}.
\end{aligned}
$$

Another Example

- $\mathbf{X} = (X_1, \ldots, X_n)$이 $N(\mu, \sigma^2)$을 따른다.
- $\Theta = (\mu, \sigma^2)$의 제프리 사전밀도함수를 구해보자.
- Log-likelihood:

$$\log f(\mathbf{x} \mid \Theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{\sum(\mu - x_i)^2}{2\sigma^2}$$

- For μ

$$\frac{\partial \log f(\mathbf{x} \mid \Theta)}{\partial \mu} = -\frac{\sum(\mu - x_i)}{\sigma^2}$$

$$\frac{\partial^2 \log f(\mathbf{x} \mid \Theta)}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

Another Example

- For σ^2,

$$\frac{\partial \log f(\mathbf{x} \mid \Theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum(\mu - x_i)^2}{\sigma^4}$$

$$\frac{\partial^2 \log f(\mathbf{x} \mid \Theta)}{\partial \sigma^4} = \frac{n}{2\sigma^4} - \frac{\sum(\mu - x_i)^2}{\sigma^6}$$

- For μ and σ^2,

$$\frac{\partial^2 \log f(\mathbf{x} \mid \Theta)}{\partial \mu \partial \sigma^2} = \frac{\sum(\mu - x_i)}{\sigma^4}$$

Another Example

- Fisher's Information:

$$I(\Theta) = -E\begin{bmatrix} \frac{\partial^2 \log f}{\partial \mu^2} & \frac{\partial^2 \log f}{\partial \sigma^2 \partial \mu} \\ \frac{\partial^2 \log f}{\partial \mu \partial \sigma^2} & \frac{\partial^2 \log f}{\partial \sigma^4} \end{bmatrix} = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

- Therefore

$$det(I(\Theta)) = n^2/(2\sigma^6).$$

- 그러므로 $\Theta = (\mu, \sigma^2)$에 대한 제프리 사전밀도함수는 $\pi(\mu, \sigma^2) = n/\sqrt{2}\sigma^{-3}$이고 비례상수를 무시하면

$$\pi(\mu, \sigma^2) = \sigma^{-3}.$$

Another Example

- 모수를 (μ, σ^2)대신 (μ, σ)를 고려하자 .
- (μ, σ^2)에서 (μ, σ)으로의 변환에 대한 Jacobian이 2σ이므로 (μ, σ)에 대한 제프리 사전밀도함수는

$$\pi(\mu, \sigma) = \pi(\mu, \sigma^2) 2\sigma = \sigma^{-3} 2\sigma = 2\sigma^{-2}$$

- Therefore,

$$\pi(\mu, \sigma) = \sigma^{-2}.$$

Conjugate Prior

모수 θ에 대한 우도함수를 $f(x \mid \theta)$ 라고 할 때 어떤 사전분포의 집합 Π가 임의의 밀도함수 $\pi(\theta) \in \Pi$와 임의의 x에 대하여 $\pi(\theta \mid x) \in \Pi$를 만족시키면 Π는 우도함수 $f(x \mid \theta)$에 대한 공액 사전분포집합이라고 정의한다.

Conjugate Prior

$$\text{Prior} \times \text{Likelihood} = \text{Posterior}$$

- Beta + Binomial $\Longrightarrow$ Beta
- Gamma + Poisson $\Longrightarrow$ Gamma
- Nomral + Nomral $\Longrightarrow$ Normal

Example I

- Likelihood:

$$X_1, \ldots, X_n \mid \sigma^2 \sim N(0, \sigma^2).$$

- Prior for σ^2: $IG(\alpha, \beta)$.
- Posterior for σ^2:

$$\begin{aligned} \pi(\sigma^2 \mid x_1, \ldots, x_n) &\propto (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum x_i^2} (\sigma^2)^{-\alpha-1} e^{-\beta/\sigma^2} \\ &= (\sigma^2)^{-\frac{n}{2}-\alpha-1} e^{-\frac{1}{\sigma^2}[\sum x_i^2/2+\beta]} \end{aligned}$$

$$\pi(\sigma^2 \mid x_1, \ldots, x_n) \sim IG(\frac{n}{2} + \alpha, \frac{1}{2}\sum x_i^2 + \beta)$$

- 모수값만 다를 뿐 사전과 사후분포가 모두 역감마 분포이므로 역감마 분포는 σ^2에 대한 공액(Conjugate) 사전분포이다.

Example II

- Likelihood:

$$X \mid \theta \sim Poi(\theta)$$

- Prior for θ:

$$Gamma(\alpha, \beta)$$

- Posterior for θ:

$$Gamma(x + \alpha, \beta + 1)$$

- 따라서 포아송 평균에 대한 공액(Conjugate) 사전분포는 감마 분포이다.

Example III

- Likelihood: $X_1, \ldots, X_n \mid \theta \sim N(\theta, \sigma^2)$
- Suppose that σ^2 is known.
- Prior for θ:

$$N(\mu_0, \sigma_0^2)$$

- Posterior for θ:

$$N\left(\frac{n\bar{x}/\sigma^2 + \mu_0/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2}, \frac{1}{n/\sigma^2 + 1/\sigma_0^2}\right)$$

- 따라서 정규평균에 대한 공액(Conjugate) 사전분포는 정규분포이다.