Multiparameter Models

DongYoon Yang

January 8, 2021

Seoul National University

Table of Contents

Table of Contents

- Observation : $\mathbf{x} = (x_1, \cdots, x_n)$
- Prior : $\theta \sim p_\theta$
- Posterior distribution : $p(\theta|\mathbf{x}) \propto \underbrace{p(\theta)}_{\text{prior}} \underbrace{p(\mathbf{x}|\theta)}_{\text{likelihood}}$
- Esitimation
  1. $\hat{\theta}^{\text{MAP}} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, p(\theta|\mathbf{x})$
  2. $\hat{\theta}^{\text{Bayes}} = \underset{\delta \in \Theta}{\operatorname{argmin}}\, \mathbb{E}_{\theta|\mathbf{x}} \ell(\theta, \delta)$
- If $\ell(\delta, \theta) = (\delta - \theta)^2$, $\hat{\theta}^{\text{Bayes}} = \mathbb{E}_{\theta|\mathbf{x}} \theta$ (Mean of posterior distribution)
- In unimodal symmetric posterior distributions, $\hat{\theta}^{\text{Bayes}} = \hat{\theta}^{\text{MAP}}$ under square loss.

Table of Contents

Notation

- $\theta = (\theta_1, \theta_2)$
  - θ_1 : parameter of interest
  - θ_2 : nuisance parameter
- x : observed data
- $p(\theta|x) = p(\theta_1, \theta_2|x)$: Joint posterior density
- $p(x|\theta_1, \theta_2)$: Likelihood
- $p(\theta_1, \theta_2)$: Prior

Averaging over nuisance parameters

$$p(\theta_1, \theta_2|x) \propto p(\theta_1, \theta_2)p(x|\theta_1, \theta_2)$$

$$p(\theta_1|x) = \int p(\theta_1, \theta_2|x)d\theta_2 = \int p(\theta_1|\theta_2, x)p(\theta_2|x)d\theta_2 \qquad (1)$$

- The (1) can be interpreted a mixture of the conditional posterior given the nuisance parameters.
- The weight depends on the posterior of θ_2.
- We rarely evaluate the integral (1) with explicit form.
- In this chapter, we study some conjugate prior-model pairs for which the marginal posterior can be explicitly evaluated.

Table of Contents

The process of finding posterior distribution for interest parameter

- θ_1 : parameter of interest
- θ_2 : nuisance parameter
- $\mathbf{x}$: observed data
- Objective : Find the marginal posterior distribution $p(\theta_1|\mathbf{x})$

1. Set the prior $p(\theta_1, \theta_2)$.
2. Write the likelihood $p(\mathbf{x}|\theta_1, \theta_2)$
3. Write the joint posterior distribution $p(\theta_1, \theta_2|\mathbf{x}) \propto p(\theta_1, \theta_2)p(x|\theta_1, \theta_2)$.
4. Find the marginal posterior distribution $p(\theta_1|\mathbf{x})$
    - With known joint posterior distribution, integrate $p(\theta_1, \theta_2|\mathbf{x})$ w.r.t θ_2.

Table of Contents

Normal data with a noninformative prior distribution

- $(\theta_1, \theta_2) = (\mu, \sigma^2)$

1. Set the prior of $(\mu, \sigma^2) \sim \frac{1}{\sigma^2}$. (uniform on $(\mu, \log \sigma)$)
2. Find the joint posterior distribution $p(\mu, \sigma^2|\mathbf{x}) \propto p(\mu, \sigma^2)p(\mathbf{x}|\mu, \sigma^2)$.
   (1) Find the conditional posterior distribution $p(\mu|\sigma^2, \mathbf{x})$(Section 2.5)
   (2) Find the marginal posterior distribution $p(\sigma^2|\mathbf{x})$.(Integration of $p(\mu, \sigma^2|\mathbf{x})$)
3. Find the marginal posterior distribution $p(\mu|\mathbf{x})$.(Integration)

Normal data with a noninformative prior distribution

Consider $p(\mu, \sigma^2) \propto 1/\sigma^2$(uniform on $(\mu, \log \sigma)$). Then,

$$
\begin{aligned}
p\left(\mu, \sigma^2 \mid \mathbf{x}\right) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right]\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{x} - \mu)^2\right]\right) \quad (2)
\end{aligned}
$$

where $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})$

Normal data with a noninformative prior distribution

Under uniform prior distribution and normal distribution with known variacne of μ, it easy to show that

$$\mu \mid \sigma^2, \mathbf{x} \sim N(\bar{x}, \sigma^2/n) \tag{3}$$

$$p\left(\sigma^2 \mid \mathbf{x}\right) \propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]\right) d\mu$$

$$\begin{aligned} p\left(\sigma^2 \mid \mathbf{x}\right) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right)\sqrt{2\pi\sigma^2/n} \\ &\propto \left(\sigma^2\right)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \end{aligned} \tag{4}$$

By definition of inverse-χ^2 density,

$$\sigma^2 \mid \mathbf{x} \sim \text{Inv-}\chi^2(n-1, s^2) \tag{5}$$

Normal data with a noninformative prior distribution

$$p(\mu, \sigma^2 \mid \mathbf{x}) = p(\mu \mid \sigma^2, \mathbf{x}) p(\sigma^2 \mid \mathbf{x})$$

In this case, we can sample form the joint posterior distribution as follows:

1. Draw σ^2 from (5)
2. Draw μ from (3)

Normal data with a noninformative prior distribution

Also, we can calculate $p(\mu, \sigma^2 \mid \mathbf{x})$ in closed form.

$$p(\mu \mid \mathbf{x}) = \int_0^\infty p\left(\mu, \sigma^2 \mid \mathbf{x}\right) d\sigma^2 \tag{6}$$

$$\propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz \tag{7}$$

$$\propto \left[(n-1)s^2 + n(\mu - \bar{x})^2\right]^{-n/2} \tag{8}$$

$$\propto \left[1 + \frac{n(\mu - \bar{x})^2}{(n-1)s^2}\right]^{-n/2} \tag{9}$$

So, $\mu \mid \mathbf{x} \sim t_{n-1}(\bar{x}, s^2/n)$

Normal data with a conjugate prior distribution

Definition (Conjugate prior distribution)
$\mathcal{F}$: Class of sampling distribution $p(x|\theta)$

$\mathcal{P}$: Class of prior distributions for θ

$\mathcal{P}$ is conjugate for $\mathcal{F}$ if

$$p(\theta|x) \in \mathcal{P} \text{ for all } p(\cdot|\theta) \in \mathcal{F} \text{ and } p(\cdot) \in \mathcal{P}$$

- In this section, $\mathcal{F}$: normal density function class.
- In short, prior and posterior is in same function class $\mathcal{P}$.

Normal data with a conjugate prior distribution

Assume that

$$\begin{aligned} \mu \mid \sigma^2 &\sim \mathrm{N}\left(\mu_0, \sigma^2/\kappa_0\right) \\ \sigma^2 &\sim \text{Inv-}\chi^2\left(\nu_0, \sigma_0^2\right) \end{aligned}$$

Then

$$p\left(\mu, \sigma^2\right) \propto \sigma^{-1}\left(\sigma^2\right)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2 + \kappa_0\left(\mu_0 - \mu\right)^2\right]\right) \tag{10}$$

In this case, define the distribution of (μ, σ^2) as normal-scaled inverse-χ^2 with parameters $(\mu_0, \kappa, \nu_0, \sigma_0^2)$ and denote N-Inv-$\chi^2(\mu, \sigma^2 \mid \mu_0, \sigma_0^2/\kappa_0, \nu_0, \sigma_0^2)$.

Normal data with a conjugate prior distribution

$$
\begin{aligned}
p\left(\mu, \sigma^{2} \mid \mathbf{x}\right) \propto\ & \underbrace{\left(\sigma^{2}\right)^{-(\nu_0+3)/2} \exp\left(-\frac{1}{2\sigma^{2}}\left[\nu_0\sigma_0^{2} + \kappa_0\left(\mu-\mu_0\right)^{2}\right]\right)\}}_{\text{joint distribution of } (\mu, \sigma^2)} \\
& \times \underbrace{\left(\sigma^{2}\right)^{-n/2} \exp\left(-\frac{1}{2\sigma^{2}}\left[(n-1)s^{2} + n(\bar{x}-\mu)^{2}\right]\right)}_{\textit{likelihood}} \qquad (11) \\
=& N\text{-Inv-}\chi^{2}\left(\mu, \sigma^{2} \mid \mu_n, \sigma_n^{2}/\kappa_n; \nu_n, \sigma_n^{2}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0+n}\mu_0 + \frac{n}{\kappa_0+n}\bar{x} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\nu_n\sigma_n^{2} &= \nu_0\sigma_0^{2} + (n-1)s^{2} + \frac{\kappa_0 n}{\kappa_0+n}\left(\bar{x}-\mu_0\right)^{2}
\end{aligned}
$$

Normal data with a conjugate prior distribution

The conditional posterior distribution

$$p(\mu \mid \sigma^2, \mathbf{x}) \sim N(\mu_n, \sigma^2/\kappa_n) \tag{12}$$

The marginal posterior distribution

$$p(\sigma^2 \mid \mathbf{x}) \sim \text{Inv-}\chi^2(\nu_n, \sigma_n^2) \tag{13}$$

Marginal posterior distribution of μ

$$p(\mu|\mathbf{x}) \propto \left(1 + \frac{\kappa_n(\mu - \mu_n)^2}{\nu_n \sigma_n^2}\right)^{-(\nu_n+1)/2} = t_{\nu_n}(\mu_n, \sigma_n^2/\kappa_n) \tag{14}$$

Table of Contents

Multinomial model for categorical data

- Parameter : $\theta \in \mathbb{R}^K$ with $\sum_{k=1}^{K} \theta_k = 1$
- Observations $\mathbf{x} = (x_1, \cdots, x_K)$
- Likelihood : $p(\mathbf{x}|\theta) = \prod_{k=1}^{K} \theta_k^{x_k}$
- Prior : $\theta \sim \text{Dirichlet}(\alpha_1, \cdots, \alpha_K)$

$$p(\theta) \propto \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \tag{15}$$

- Posterior : $\theta|\mathbf{x} \sim \text{Dirichlet}(\alpha_1 + x_1, \cdots, \alpha_K + X_K)$

Table of Contents

Multivariate normal model with known variance

- Parameter : $\boldsymbol{\mu} \in \mathbb{R}^d$
- $\mathsf{x}|\boldsymbol{\mu}, \Sigma \sim N(\boldsymbol{\mu}, \Sigma)$
- Observations $\mathsf{x} = (\mathsf{x}_1, \cdots, \mathsf{x}_n)$ with $\mathsf{x}_i \in \mathbb{R}^K$
- Likelihood :
  $p(\mathsf{x}|\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-n/2} \exp(-\frac{1}{2} \sum_{i=1}^{n} (\mathsf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathsf{x}_i - \boldsymbol{\mu}))$
- Prior : $\boldsymbol{\mu} \sim \mathrm{N}(\boldsymbol{\mu}_{\mathbf{0}}, \Lambda_0)$
- Posterior : $\boldsymbol{\mu}|\Sigma, \mathsf{x} \sim N(\boldsymbol{\mu}_n, \Lambda_n)$

$$p(\boldsymbol{\mu}|\Sigma, \mathsf{x}) \propto \exp(-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^\top \Lambda_n^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_n)) \tag{16}$$

$$\text{where } \boldsymbol{\mu}_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\boldsymbol{\mu} + n\Sigma^{-1}\bar{\mathsf{x}})$$
$$\Lambda_n^{-1} = \Lambda_0^{-1} + n\Sigma^{-1}$$

Noninformative prior for μ

- Prior : $\boldsymbol{\mu} \sim$ const(Improper prior)
- Posterior : $\boldsymbol{\mu}|\Sigma, \mathsf{x} \sim N(\bar{\mathsf{x}}, \Sigma/n)$(Proper posterior distribution)
- In this case, $\hat{\boldsymbol{\mu}}^{bayes} = \hat{\boldsymbol{\mu}}^{MAP} = \hat{\boldsymbol{\mu}}^{MLE} = \bar{\mathsf{x}}$

Table of Contents

Conjugate inverse-Wishart family of prior distributions

- Parameter : $\boldsymbol{\mu} \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d}$
- $\mathsf{x}|\boldsymbol{\mu}, \Sigma \sim N(\boldsymbol{\mu}, \Sigma)$
- Observations $\mathsf{x} = (\mathsf{x}_1, \cdots, \mathsf{x}_n)$ with $\mathsf{x}_i \in \mathbb{R}^K$
- Likelihood :
  $p(\mathsf{x}|\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-n/2} \exp(-\frac{1}{2} \sum_{i=1}^{n} (\mathsf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathsf{x}_i - \boldsymbol{\mu}))$
- Prior :

$$\Sigma \sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1})$$
$$\boldsymbol{\mu}|\Sigma \sim N(\boldsymbol{\mu}_0, \Sigma/\kappa_0)$$

Define $(\boldsymbol{\mu}, \Sigma) \sim \text{N-inv-Wishart}(\boldsymbol{\mu}_0, \kappa_0, \nu_0, \Lambda_0)$

Conjugate inverse-Wishart family of prior distributions

- Joint posterior $\boldsymbol{\mu}, \Sigma|\mathbf{x}$: N-inv-Wishart$(\boldsymbol{\mu}_n, \kappa_n, \nu_n, \Lambda_n)$
- Marginal posterior :

$$\boldsymbol{\mu}|\mathbf{x} \sim t_{v_n-d+1}(\boldsymbol{\mu}_n, \Lambda_n/(\kappa_n(\nu_n - d + 1))) \tag{17}$$

$$\begin{aligned}
\text{where } \boldsymbol{\mu}_n &= \frac{\kappa_0}{\kappa_0 + n}\boldsymbol{\mu}_0 + \frac{n}{\kappa + n}\bar{\mathbf{x}} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\Lambda_n &= \Lambda_0 + S + \frac{\kappa n}{\kappa + n}(\bar{\mathbf{x}} - \mu_0)(\bar{\mathbf{x}} - \mu_0)^T \\
S &= \sum_{i=1}^{n}(x_i - \bar{\mathbf{x}})(x_i - \bar{\mathbf{x}})^\top
\end{aligned}$$

Summary

- Multiparameter model에서 posterior distribution의 계산이 어렵더라도 크게 문제가 안됨
  1. Posterior distribution는 simulation을 통해서 계산할 수 있음.
  2. 복잡한 모형은 hierachy로 나타 낼 수 있고(Chapter 5), 이를 어렵지 않게 계산할 수 있는 툴이 있음. (Chpater. 10∼13)
  3. Normal approximation을 통해서 posterior distribution을 근사할 수 있음.(Chapter 4)

The process of bayesian inference

1. Set the prior $p(\theta)$.
2. Write the likelihood $p(\mathbf{x}|\theta)$
3. Write the posterior density, $p(\theta|\mathbf{x}) \propto p(\theta)p(x|\theta,)$.
4. Create a crude estimate of θ. For example, $\hat{\theta}^{MLE}$.
5. Draw $\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(S)}$ from posterior distribution.
6. If we are interested in predictive quantities, $\tilde{y}$, draw the $\tilde{y}^s$ for each $p(\tilde{y}|\theta^{(s)})(s = 1, \cdots, S)$.

- For nonconjugate models, step (5) is difficult.(Chapter. 10 $\sim$ 13)

References i

A. G. et.al., *Bayesian Data Analysis 3rd*. CRC Press, 2013, ISBN: 9781439840955.