

Multiparameter Models (over two dimensions)

Gwangsu Kim

JBNU

Second semester of 2024

Multinomial Model I

- **Prior:** $\theta = (\theta_1, \dots, \theta_k) \sim \text{Dirichlet}(\alpha = (\alpha_1, \dots, \alpha_k))$.
- **Model:** $y = (y_1, \dots, y_k) \sim \text{Multinomial}(\theta)$.
- **Posterior**

$$\begin{aligned}\pi(\theta|y_1, \dots, y_k) &\propto \left(\prod_{i=1}^k \theta_i^{y_i} \right) \left(\prod_{i=1}^k \theta_i^{\alpha_i - 1} \right) \\ &= \prod_{i=1}^k \theta_i^{\alpha_i + y_i - 1} \left(\text{Dirichlet}(\alpha_1 + y_1, \dots, \alpha_k + y_k) \right).\end{aligned}$$

Multivariate Normal Model: unknown mean and known variance I

- **Prior:** $\mu \sim \mathcal{N}(\mu_0, \Lambda_0)$.
- **Model:** $y_1, \dots, y_n \sim i.i.d. \mathcal{N}(\mu, \Sigma)$ where y_i is a d -dimensional random vector and Σ is given.

Multivariate Normal Model: unknown mean and known variance II

■ Posterior

$$\begin{aligned}
 \pi(\mu|y_1, \dots, y_n) &\propto |2\pi\Lambda_0|^{-1/2} \exp\left(-\frac{1}{2}(\mu - \mu_0)^T \Lambda_0^{-1}(\mu - \mu_0)\right) \\
 &\quad \times \prod_{i=1}^n |2\pi\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(y_i - \mu)^T \Sigma^{-1}(y_i - \mu)\right) \\
 &\propto \exp\left(-\frac{1}{2}\left\{\mu^T \Lambda_0^{-1} \mu - 2\mu_0^T \Lambda_0^{-1} \mu + \mu^T (\Sigma/n)^{-1} \mu - 2n\bar{y}^T \Sigma^{-1} \mu\right\}\right) \\
 &\propto \exp\left(-\frac{1}{2}\left\{\mu^T (\Lambda_0^{-1} + (\Sigma/n)^{-1}) \mu - 2(\mu_0^T \Lambda_0^{-1} + n\bar{y}^T \Sigma^{-1}) \mu\right\}\right) \\
 &\propto \exp\left(-\frac{1}{2}\left\{(\mu - \mu_n)^T (\Lambda_0^{-1} + (\Sigma/n)^{-1}) (\mu - \mu_n)\right\}\right),
 \end{aligned}$$

Multivariate Normal Model: unknown mean and known variance III

where $\mu_n = (\Lambda_0^{-1} + (\Sigma/n)^{-1})^{-1}(\mu_0^T \Lambda_0^{-1} + n\bar{y}^T \Sigma^{-1})$.

■ Note that $\mu|y_1, \dots, y_n \sim N(\mu_n, \Lambda_n)$ where

$$\Lambda_n = (\Lambda_0^{-1} + (\Sigma/n)^{-1})^{-1}.$$

■ If we use precision matrix such as $\Delta_0 = \Lambda_0^{-1}$ and $V = \Sigma^{-1}$, then

$$\mu|y_1, \dots, y_n \sim N(\mu_n, \Delta_n^{-1}),$$

where $\mu_n = \Delta_n^{-1}(\mu_0^T \Delta_0 + n\bar{y}^T V)$ and $\Delta_n = \Delta_0 + nV$.

Predictive Distribution I

- The joint distribution of a new observation \tilde{y} and μ given y is

$$p(\tilde{y}, \mu | y_1, \dots, y_n) = \phi(\tilde{y}; \mu, \Sigma) \phi(\mu; \mu_n, \Lambda_n).$$

- Note that normal p.d.f \times normal p.d.f. gives the normal p.d.f.

Predictive Distribution II

■ Thus we have $\tilde{y}|y_1, \dots, y_n \sim N(\mu_n, \Sigma + \Lambda_n)$, where

$$\begin{aligned}\mu_n &= E[\tilde{y}|y_1, \dots, y_n] = E[E[\tilde{y}|\mu, y_1, \dots, y_n] | y_1, \dots, y_n] \\ &= E[\mu | y_1, \dots, y_n],\end{aligned}$$

$$\begin{aligned}\Sigma + \Lambda_n &= E[\text{Var}(\tilde{y}|\mu, y_1, \dots, y_n) | y_1, \dots, y_n] \\ &\quad + \text{Var}(E[\tilde{y}|\mu, y_1, \dots, y_n] | y_1, \dots, y_n) \\ &= E[\Sigma | y_1, \dots, y_n] + \text{Var}(\mu | y_1, \dots, y_n).\end{aligned}$$

Predictive Distribution III

- Non-informative prior: $\pi(\mu) = 1$.

$$\mu|y_1, \dots, y_n \sim N(\bar{y}, \Sigma/n).$$

Wishart Distribution I

- A random $k \times k$ matrix, W follows the Wishart Distribution $\mathcal{Wishart}(\nu, S)$ where

$$p_W(w) = \frac{|w|^{(\nu-k-1)/2} \exp\left(-\frac{1}{2}\text{Tr}(S^{-1}w)\right)}{|S|^{\nu/2} 2^{\nu k/2} \Gamma_d(\nu/2)},$$

and $\Gamma_d(t) = \pi^{d(d-1)/4} \prod_{i=1}^d (t - (i-1)/2)$. Note that S is a positive definite matrix.

* $W \sim \text{Inverse-Wishart}(\nu, S)$ means that $W^{-1} \sim \mathcal{Wishart}(\nu, S)$.

Wishart Distribution II

■ Properties of Wishart Distribution

1 If $Z_i \sim i.i.d. \mathcal{N}(0, \Sigma)$, then $\sum_{i=1}^n Z_i Z_i^T \sim \mathcal{Wishart}(n, \Sigma)$.

2 If $Z_i \sim i.i.d. \mathcal{N}(\mu, \Sigma)$, then

$$\sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})^T \sim \mathcal{Wishart}(n-1, \Sigma), \text{ where } n > k.$$

3 If $A \sim \mathcal{Wishart}(\nu, \Sigma)$, then $E[A] = \nu \Sigma$.

4 If $A_i \sim \mathcal{Wishart}(\nu_i, \Sigma), i = 1, \dots, q$, then

$$\sum_{i=1}^q A_i \sim \mathcal{Wishart}(\sum_{i=1}^q \nu_i, \Sigma).$$

5 Suppose that C is nonsingular $k \times k$ matrix, then

$$CBC^T \sim \mathcal{Wishart}(\nu, \Sigma) \rightarrow B \sim \mathcal{Wishart}(\nu, C^{-1}\Sigma(C^T)^{-1}).$$

Multivariate Normal Model: known mean and unknown variance I

Prior: $V = \Sigma^{-1} \sim \mathcal{Wishart}(v_0, \Delta_0)$.

Model: $y_i \sim i.i.d. \mathcal{N}(\mu, V^{-1})$, where y_i is a d -dimensional random vector and μ is known.

Multivariate Normal Model: known mean and unknown variance II

Posterior:

$$\begin{aligned}
 & \pi(V|y_1, \dots, y_n) \\
 \propto & |V|^{(v_0-d-1)/2} \exp\left(-\frac{1}{2}\text{Tr}(\Delta_0^{-1}V)\right) \prod_{i=1}^n |V|^{1/2} \\
 & \times \exp\left(-\frac{1}{2}(y_i - \mu)^T V (y_i - \mu)\right) \\
 \propto & |V|^{(v_0-d-1)/2} \exp\left(-\frac{1}{2}\text{Tr}(\Delta_0^{-1}V)\right) |V|^{n/2} \exp\left(-\frac{1}{2}\text{Tr}(SV)\right) \\
 \propto & |V|^{(v_0+n-d-1)/2} \exp\left(-\frac{1}{2}\text{Tr}((S + \Delta_0^{-1})V)\right).
 \end{aligned}$$

Multivariate Normal Model: unknown mean and unknown variance I

Prior: $V = \Sigma^{-1} \sim \mathcal{Wichart}(\nu_0, \Delta_0)$, $\mu|V \sim \mathcal{N}(\mu_0, V^{-1}/\kappa_0)$.

Model: $y_i \sim i.i.d. \mathcal{N}(\mu, V^{-1})$, where y_i is a d -dimensional random vector.

Multivariate Normal Model: unknown mean and unknown variance II

Posterior:

$$\begin{aligned}
 & \pi(\mu, V | y_1, \dots, y_n) \\
 \propto & |V|^{(v-d-1)/2} \exp\left(-\frac{1}{2} \text{Tr}(\Delta_0^{-1} V)\right) |\kappa_0 V|^{1/2} \exp\left(-\frac{1}{2} (\mu - \mu_0)^T (\kappa_0 V) (\mu - \mu_0)\right) \\
 & \times \prod_{i=1}^n |V|^{1/2} \exp\left(-\frac{1}{2} (y_i - \mu)^T V (y_i - \mu)\right) \\
 \propto & |V|^{(v+n-d)/2} \exp\left(-\frac{1}{2} \text{Tr}(\Delta_0^{-1} V)\right) \\
 & \times \exp\left(-\frac{1}{2} \left\{ \kappa_0 \mu^T V \mu - 2 \kappa_0 \mu_0^T V \mu + \kappa_0 \mu_0^T V \mu_0 + n \mu^T V \mu - 2 n \bar{y}^T V \mu + \sum_{i=1}^n y_i^T V y_i \right\}\right).
 \end{aligned}$$

Multivariate Normal Model: unknown mean and unknown variance III

$$\begin{aligned}
 &\propto |V|^{(v+n-d)/2} \exp\left(-\frac{1}{2} \text{Tr}(\Delta_0^{-1} V)\right) \\
 &\quad \times \exp\left(-\frac{1}{2} \left\{ \mu^T (\kappa_0 + n) V \mu - 2(\kappa_0 V \mu_0 + n V \bar{y}) \mu + \kappa_0 \mu_0^T V \mu_0 + \sum_{i=1}^n y_i^T V y_i \right\}\right) \\
 &\propto |V|^{(v+n-d)/2} \exp\left(-\frac{1}{2} \text{Tr}(\Delta_0^{-1} V)\right) \\
 &\quad \times \exp\left(-\frac{1}{2} \left\{ (\mu - \mu_n)^T (\kappa_0 + n) V (\mu - \mu_n) - (\kappa_0 + n) \mu_n^T V \mu_n + \kappa_0 \mu_0^T V \mu_0 + \sum_{i=1}^n y_i^T V y_i \right\}\right),
 \end{aligned}$$

where $\mu_n = (\kappa_0 + n)^{-1} V^{-1} (\kappa_0 V \mu_0 + n V \bar{y}) = \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y}.$

Multivariate Normal Model: unknown mean and unknown variance IV

■ Thus, we have

$$\mu|V, y_1, \dots, y_n \sim N\left(\mu_n, [(\kappa_0 + n)V]^{-1}\right).$$

Multivariate Normal Model: unknown mean and unknown variance V

■ Integrating out μ , we have

$$\begin{aligned} & \pi(V|y_1, \dots, y_n) \\ \propto & |V|^{(\nu_0+n-d)/2} \exp\left(-\frac{1}{2}\text{Tr}(\Sigma_0^{-1}V)\right) |2\pi(\kappa_0+n)^{-1}V^{-1}|^{1/2} \\ & \times \exp\left(-\frac{1}{2}\left\{-\mu_n^T[(\kappa_0+n)V]\mu_n + \kappa_0\mu_0^T V\mu_0 + \sum_{i=1}^n y_i^T V y_i\right\}\right) \end{aligned}$$

Multivariate Normal Model: unknown mean and unknown variance VI

$$\begin{aligned} &\propto |V|^{(v_0+n-d-1)/2} \exp\left(-\frac{1}{2}\text{Tr}(\Sigma_n^{-1}V)\right) \\ &\equiv \mathcal{Wishart}(v_0 + n, \Sigma_n), \end{aligned}$$

where

$$\begin{aligned} \Sigma_n^{-1} &= \Delta_0^{-1} - (\kappa_0 + n)\mu_n\mu_n^T + \kappa_0\mu_0\mu_0^T + \sum_{i=1}^n y_i y_i^T \\ &= \Delta_0^{-1} + S + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)(\bar{y} - \mu_0)^T, \end{aligned}$$

Multivariate Normal Model: unknown mean and unknown variance VII

$$\text{and } S = \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})^T.$$

Example: Logistic Regression I

■ Modeling the dose-response relationship

$$y_i | \theta_i \sim B(n_i, \theta_i).$$

■ Link

$$\log \frac{\theta_i}{1 - \theta_i} = \alpha + \beta x_i.$$

Example: Logistic Regression II

■ Likelihood

$$L(\alpha, \beta) \propto \prod_{i=1}^n p(y_i | n_i, x_i, \alpha, \beta) = \left[\frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right]^{y_i} \left[\frac{1}{1 + \exp(\alpha + \beta x_i)} \right]^{n_i - y_i}.$$

Example: Logistic Regression III

■ Posterior

$$\pi(\alpha, \beta | y_1, \dots, y_n,) \propto \pi(\alpha, \beta) \prod_{i=1}^n p(y_i | n_i, x_i, \alpha, \beta).$$

■ Obtaining the posterior by the numerical methods.