

# Bayesian Learning

## Lecture 3 - Multi-parameter models

Bertil Wegmann

Department of Computer and Information Science  
Linköping University



# Lecture overview

- **Multiparameter** models
- **Marginalization**
- **Normal model with unknown variance**
- Bayesian analysis of **multinomial data**
- Bayesian analysis of **multivariate normal data**

# Marginalization

- Models with **multiple parameters**  $\theta_1, \theta_2, \dots$

- **Joint posterior distribution**

$$p(\theta_1, \theta_2, \dots, \theta_p | x) \propto p(x | \theta_1, \theta_2, \dots, \theta_p) p(\theta_1, \theta_2, \dots, \theta_p).$$

$$p(\theta | x) \propto p(x | \theta) p(\theta).$$

- **Marginalize** out parameter of no direct interest (**nuisance**).
- Example:  $\theta = (\theta_1, \theta_2)'$ . **Marginal posterior** of  $\theta_1$

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

# Normal model with unknown variance

## ■ Model

$$x_1, \dots, x_n | \theta, \sigma^2 \stackrel{iid}{\sim} N(\theta, \sigma^2)$$

## ■ Prior

$$p(\theta, \sigma^2) \propto (\sigma^2)^{-1}$$

## ■ Posterior

$$\begin{aligned}\theta | \sigma^2, \mathbf{x} &\sim N\left(\bar{x}, \frac{\sigma^2}{n}\right) \\ \sigma^2 | \mathbf{x} &\sim \text{Inv} - \chi^2(n-1, s^2),\end{aligned}$$

where

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

is the usual sample variance.

# Normal model with unknown variance

## ■ Simulating from the posterior :

1. Draw  $X \sim \chi^2(n-1)$
2. Compute  $\sigma^2 = \frac{(n-1)s^2}{X}$  (a draw from  $\text{Inv-}\chi^2(n-1, s^2)$ )
3. Draw a  $\theta$  from  $N\left(\bar{x}, \frac{\sigma^2}{n}\right)$  conditional on the previous draw  $\sigma^2$
4. Repeat step 1-3 many times.

## ■ We may derive the **marginal posterior** analytically as

$$\theta|x \sim t_{n-1}\left(\bar{x}, \frac{s^2}{n}\right).$$

# Normal model - normal prior

## ■ Model

$$y_1, \dots, y_n | \theta, \sigma^2 \stackrel{iid}{\sim} N(\theta, \sigma^2)$$

## ■ Conjugate prior

$$\begin{aligned}\theta | \sigma^2 &\sim N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)\end{aligned}$$

# Normal model with normal prior

## ■ Posterior

$$\theta|y, \sigma^2 \sim N\left(\mu_n, \frac{\sigma^2}{\kappa_n}\right)$$
$$\sigma^2|y \sim \text{Inv-}\chi^2(\nu_n, \sigma_n^2).$$

where

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

## ■ Marginal posterior

$$\theta|y \sim t_{\nu_n}(\mu_n, \sigma_n^2/\kappa_n)$$

# Multinomial model with Dirichlet prior

- **Categorical counts:**  $y = (y_1, \dots, y_K)$ , where  $\sum_{k=1}^K y_k = n$ .
- $y_k$  = number of observations in  $k$ th category. Brand choices.
- **Multinomial model:**

$$p(y|\theta) \propto \prod_{k=1}^K \theta_k^{y_k}, \text{ where } \sum_{k=1}^K \theta_k = 1.$$

- **Dirichlet prior:**  $\text{Dirichlet}(\alpha_1, \dots, \alpha_K)$

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

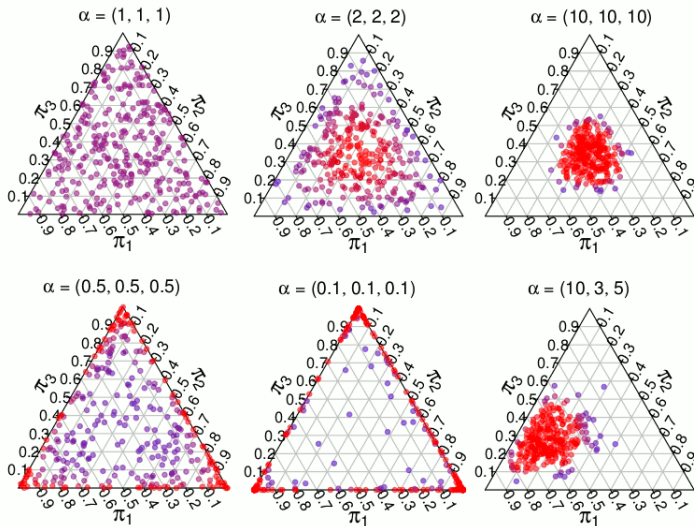
- **Mean and variance** for  $(\theta_1, \dots, \theta_K) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$

$$\begin{aligned} E(\theta_k) &= \frac{\alpha_k}{\sum_{j=1}^K \alpha_j} \\ V(\theta_k) &= \frac{E(\theta_k) [1 - E(\theta_k)]}{1 + \sum_{j=1}^K \alpha_j} \end{aligned}$$



# Dirichlet distribution

Draws from a 3-dimensional Dirichlet with different  $\alpha$



# Multinomial model with Dirichlet prior

■ 'Non-informative':  $\alpha_1 = \dots = \alpha_K = 1$  (uniform and proper).

■ **Simulating** from the Dirichlet distribution:

- ▶ Generate  $x_1 \sim \text{Gamma}(\alpha_1, 1), \dots, x_K \sim \text{Gamma}(\alpha_K, 1)$ .
- ▶ Compute  $z_k = x_k / (\sum_{j=1}^K x_j)$ .
- ▶ Then  $z = (z_1, \dots, z_K) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$ .

■ **Prior-to-Posterior updating:**

*Model:*  $y = (y_1, \dots, y_K) \sim \text{Multin}(n; \theta_1, \dots, \theta_K)$

*Prior:*  $\theta = (\theta_1, \dots, \theta_K) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$

*Posterior:*  $\theta|y \sim \text{Dirichlet}(\alpha_1 + y_1, \dots, \alpha_K + y_K)$ .

## Example: market shares

- Survey among 513 smartphones owners:
  - ▶ 180 used mainly an iPhone
  - ▶ 230 used mainly an Android phone
  - ▶ 62 used mainly a Windows phone
  - ▶ 41 used mainly some other mobile phone.
- Old survey: iPhone 30%, Android 30%, Windows 20%, Other 20%.
- $\Pr(\text{Android has largest share} \mid \text{Data})$
- Prior:  $\alpha_1 = 15, \alpha_2 = 15, \alpha_3 = 10$  and  $\alpha_4 = 10$  (prior info is equivalent to a survey with only 50 respondents)
- Posterior:  $(\theta_1, \theta_2, \theta_3, \theta_4) \mid y \sim \text{Dirichlet}(195, 245, 72, 51)$ .

# Multivariate normal - known $\Sigma$

## ■ Model

$$y_1, \dots, y_n \stackrel{iid}{\sim} N_p(\mu, \Sigma)$$

where  $\Sigma$  is a known covariance matrix.

## ■ Density

$$p(y|\mu, \Sigma) = |2\pi\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(y - \mu)' \Sigma^{-1}(y - \mu)\right)$$

## ■ Likelihood

$$\begin{aligned} p(y_1, \dots, y_n|\mu, \Sigma) &\propto |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n (y_i - \mu)' \Sigma^{-1}(y_i - \mu)\right) \\ &= |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \text{tr} \Sigma^{-1} S_\mu\right) \end{aligned}$$

where  $S_\mu = \sum_{i=1}^n (y_i - \mu)(y_i - \mu)'$ .

# Multivariate normal - known $\Sigma$

## ■ Prior

$$\mu \sim N_p(\mu_0, \Lambda_0)$$

## ■ Posterior

$$\mu|y \sim N(\mu_n, \Lambda_n)$$

where

$$\begin{aligned}\mu_n &= (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + n\Sigma^{-1}\bar{y}) \\ \Lambda_n^{-1} &= \Lambda_0^{-1} + n\Sigma^{-1}\end{aligned}$$

- Posterior mean is a weighted average of prior and data information.

- **Noninformative prior**: let the precision go to zero:  $\Lambda_0^{-1} \rightarrow 0$ .