Bayesian Learning

Lecture 3 - Multi-parameter models

Mattias Villani

Department of Statistics
Stockholm University

Department of Computer and Information Science
Linköping University

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** $\theta_1, \theta_2, ...$
- Examples: $x_i \overset{iid}{\sim} N(\theta, \sigma^2)$; multiple regression ...
- **Joint posterior distribution**

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

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

- **Marginalize** out parameter of no direct interest (**nuisance**).
- Example: $\theta = (\theta_1, \theta_2)'$. **Marginal posterior** of θ_1

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

Normal model with unknown variance

- **Model**

$$x_1, \ldots, x_n \overset{iid}{\sim} N(\theta, \sigma^2)$$

- **Prior**

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

- **Posterior**

$$\theta | \sigma^2, \mathbf{x} \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

$$\sigma^2 | \mathbf{x} \sim \mathrm{Inv} - \chi^2(n-1, s^2),$$

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 $\sigma^2 = \frac{(n-1)s^2}{X}$ (this a draw from $\text{Inv-}\chi^2(n-1, s^2)$)
  3. Draw a θ from $N\left(\bar{x}, \frac{\sigma^2}{n}\right)$ conditional on the previous draw σ^2
  4. Repeat step 1-3 many times.
- The sampling is implemented in the R program `NormalNonInfoPrior.R`
- We may derive the **marginal posterior** analytically as

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

Normal model - normal prior

- **Model**

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

- **Conjugate prior**

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

Normal model with normal prior

- **Posterior**

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

where

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

Normal model with normal prior

Posterior

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

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

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

Multinomial model with Dirichlet prior

- **Categorical counts**: $y = (y_1, ... y_K)$, where $\sum_{k=1}^K y_k = n$.
- y_k= number of observations in kth 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**: $\mathrm{Dirichlet}(\alpha_1, ..., \alpha_K)$

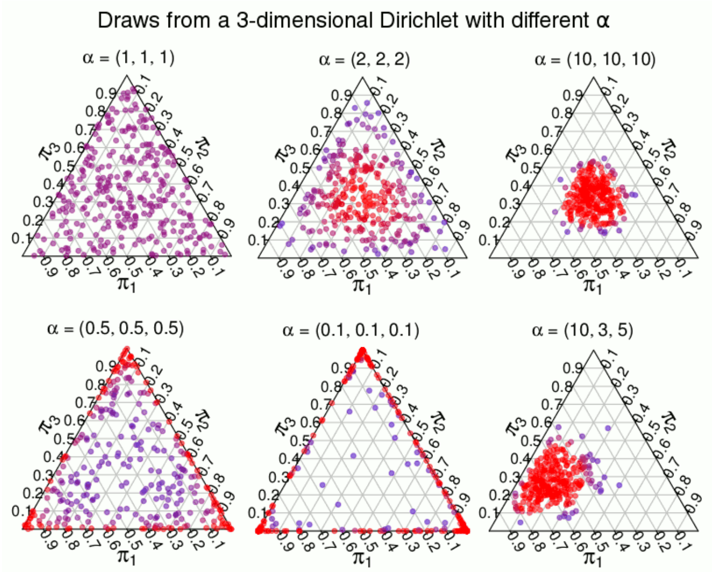
$$p(\theta) \propto \prod_{k=1}^K \theta_k^{\alpha_k - 1}.$$

- **Mean and variance** for $(\theta_1, \ldots, \theta_K) \sim Dirichlet(\alpha_1, ..., \alpha_K)$

$$\mathrm{E}(\theta_k) = \frac{\alpha_k}{\sum_{j=1}^K \alpha_j}$$

$$\mathrm{V}(\theta_k) = \frac{\mathrm{E}(\theta_k)\left[1 - \mathrm{E}(\theta_k)\right]}{1 + \sum_{j=1}^K \alpha_j}$$

Dirichlet distribution



Multinomial model with Dirichlet prior

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

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

- ***Prior-to-Posterior updating***:

$$
\begin{aligned}
\textit{Model}: \quad & y = (y_1, ... y_K) \sim \text{Multin}(n; \theta_1, ..., \theta_K) \\
\textit{Prior}: \quad & \theta = (\theta_1, ..., \theta_K) \sim \text{Dirichlet}(\alpha_1, ..., \alpha_K) \\
\textit{Posterior}: \quad & \theta | y \sim \text{Dirichlet}(\alpha_1 + y_1, ..., \alpha_K + y_K).
\end{aligned}
$$

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(Android has largest share | 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: $(\theta_1, \theta_2, \theta_3, \theta_4)|\mathbf{y} \sim \mathrm{Dirichlet}(195, 245, 72, 51)$.
- `DirichletSurveyData` Rnotebook on web page.

Multivariate normal - known Σ

- **Model**

$$y_1, ..., y_n \overset{iid}{\sim} N_p(\mu, \Sigma)$$

where Σ 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**

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

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

Multivariate normal - known Σ

- **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: $\Lambda_0^{-1} \to 0$.