Bayesian Learning

Lecture 3 - Multi-parameter models

Mattias Villani 🧔

Department of Statistics
Stockholm University

mattiasvillani.com @matvil @matvil mattiasvillani

Lecture overview

- **Multiparameter** models
- **Marginalization**
- **Normal model with unknown variance**
- **Multinomial data**
- **Dirichlet distribution**

Marginalization

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

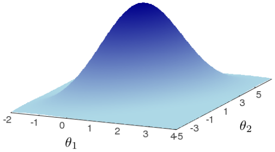
$$p(\theta_1, \theta_2, ..., \theta_p | y) \propto p(y | \theta_1, \theta_2, ..., \theta_p) p(\theta_1, \theta_2, ..., \theta_p).$$

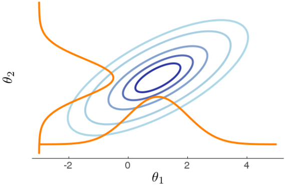
- In vector form

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}).$$

- **Marginalize** out parameters. **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**

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

- **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.

$\text{Inv} - \chi^2(\nu, \tau^2)$

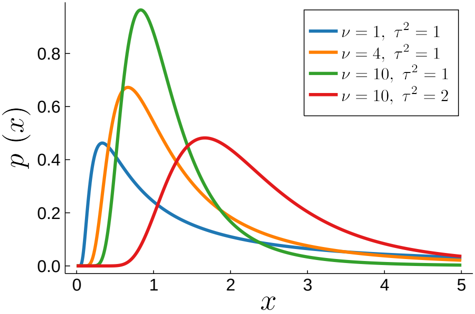
Inv-χ^2 distribution

$X \sim \text{Inv}-\chi^2(\nu, \tau^2)$, $X \in (0, \infty)$

$$p(x) = \frac{(\tau^2\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \frac{\exp\left(\frac{-\nu\tau^2}{2x}\right)}{x^{1+\nu/2}}$$

$$\mathbb{E}(X) = \frac{\nu}{\nu-2}\tau^2$$

$$\mathbb{V}(X) = \frac{2\nu^2\tau^4}{(\nu-2)^2(\nu-4)}$$



Normal model - normal prior

- **Model**

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

- **Conjugate prior**

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

Normal model with normal prior

Posterior

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

where

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

Normal model with normal prior

Posterior

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

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

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

Simulating from posterior - pseudo code

Posterior simulation - iid Gaussian with conjugate prior.

Input: data $\mathbf{x} = (x_1, \ldots, x_n)$
number of posterior draws m.
compute μ_n, σ_n^2, κ_n and ν_n using Figure 50.
for i in 1:m **do**
  $\sigma^2 \leftarrow \textsc{rInvChi2}(\nu_n, \sigma_n^2)$
  $\theta \leftarrow \textsc{rNormal}(\mu_n, \sigma^2 / \kappa_n)$
end
Output: m draws for θ and σ^2 from joint posterior.

Function $\textsc{rInvChi2}(\nu, \tau^2)$
  x = $\textsc{rChi2}(\nu)$
  y = $\nu\tau^2 / x$
  return y

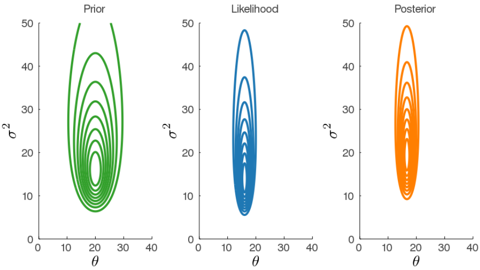
Simulating from posterior - output

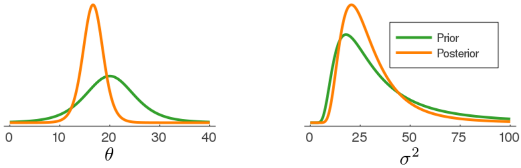
draw	θ	σ^2	σ/θ	$\theta \geq 20$
1	18.165	18.451	0.236	0
2	20.431	29.943	0.267	1
3	15.565	29.094	0.346	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
10,000	16.400	21.668	0.283	0
Mean	16.645	30.813	0.330	0.066

Internet speed data - joint and marginal posteriors

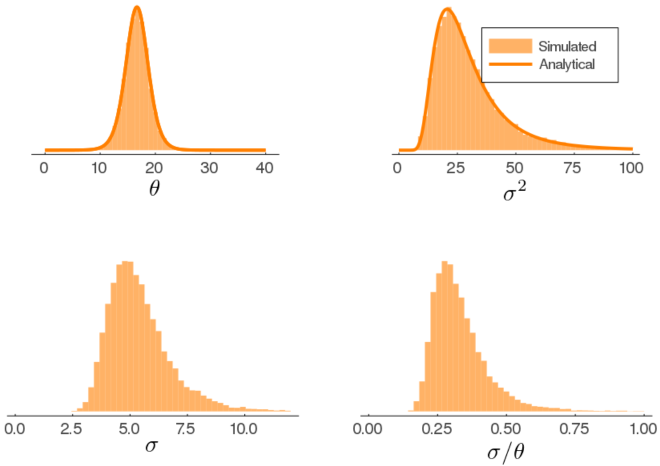
- Prior:

$$\theta|\sigma^2 \sim N\left(20, \frac{\sigma^2}{1}\right) \text{ and } \sigma^2 \sim \text{Inv-}\chi^2\left(\nu_0 = 5, \sigma_0^2 = 5^2\right)$$

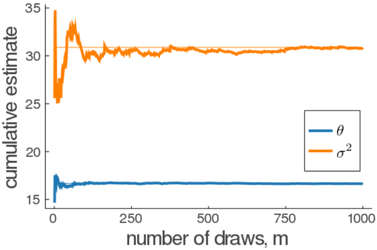


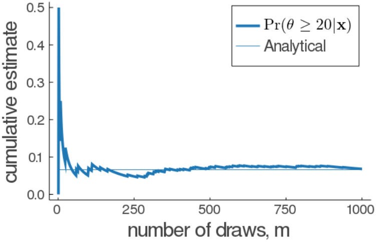


Monte Carlo simulation



Monte Carlo simulation





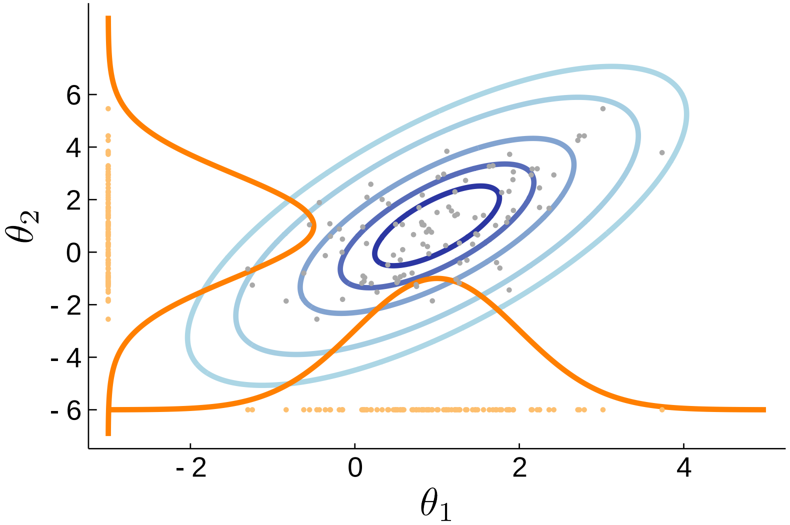
- Law of large numbers for **consistency**:

$$\bar{\theta}_{1:m} \equiv \frac{1}{m}\sum_{i=1}^{m} \theta^{(i)} \overset{\text{a.s.}}{\to} \mathbb{E}(\theta|\mathbf{x}) \text{ as } m \to \infty$$

- Central limit theorem for the **accuracy**:

$$\bar{\theta}_{1:m} \sim N\left(\mathbb{E}(\theta|\mathbf{x}), \frac{\mathbb{V}(\theta|\mathbf{x})}{m}\right)$$

Simulation from marginals by selection



Multinomial model with Dirichlet prior

- **Categorical counts**: $\boldsymbol{y} = (y_1, ... y_C)$, where $\sum_{c=1}^{C} y_c = n$.
- y_c= number of observations in cth category. Brand choices.
- **Multinomial model**:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) \propto \prod_{c=1}^{C} \theta_c^{y_c}, \text{ where } \sum_{c=1}^{C} \theta_c = 1.$$

- **Dirichlet prior**: $\boldsymbol{\theta} \sim \mathrm{Dirichlet}(\alpha_1, ..., \alpha_C)$

$$p(\boldsymbol{\theta}) \propto \prod_{c=1}^{C} \theta_c^{\alpha_c - 1}.$$

- **Marginal distributions**

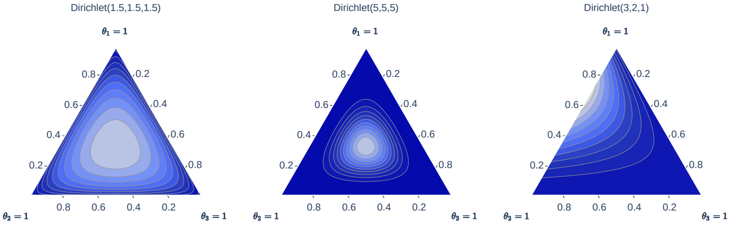
$$\theta_c \sim \mathrm{Beta}\left(\alpha_c, \alpha_+ - \alpha_c\right), \text{ where } \alpha_+ = \sum_{c=1}^{C} \alpha_c$$

Dirichlet prior

$$(\theta_1, \ldots, \theta_C) \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_C)$$

$$\mathbb{E}(\theta_c) = \frac{\alpha_c}{\sum_{j=1}^{C} \alpha_j}$$

$$\mathbb{V}(\theta_c) = \frac{\mathbb{E}(\theta_c)(1-\mathbb{E}(\theta_c))}{1+\sum_{j=1}^{C} \alpha_j}$$



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

Multinomial model with Dirichlet prior

- **Simulation** from a $\text{Dirichlet}(\boldsymbol{\alpha})$ with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_C)$:

```
Function rDirichlet(α)
    for c in 1:C do
        y[c] ← rGamma(α[c], 1)
    end
    return y/Sum(y)
```

- **Prior-to-Posterior:**

Multinomial data with Dirichlet prior

Model: $\mathbf{n}|\boldsymbol{\theta} \sim \text{Multinomial}(\boldsymbol{\theta})$, where
$\mathbf{n} = (n_1, \ldots, n_C)$ are counts in C categories
$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_C)$ are category probabilities.

Prior: $\boldsymbol{\theta} \sim \text{Dirichlet}(\boldsymbol{\alpha})$, for $\alpha = (\alpha_1, \ldots, \alpha_C)$

Posterior: $\boldsymbol{\theta} \sim \text{Dirichlet}(\boldsymbol{\alpha} + \mathbf{n})$

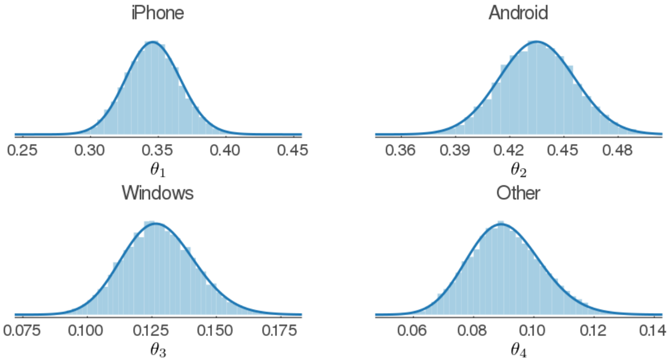
Example: smartphone market shares

- 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: $(\theta_1, \theta_2, \theta_3, \theta_4)|\mathbf{y} \sim \mathrm{Dirichlet}(195, 245, 72, 51)$.
- **R Notebook**: Multinomial.Rmd
- Julia **Pluto Notebook**: multinom.jl

Posterior simulation output

draw	θ_1	θ_2	θ_3	θ_4	I
1	0.33	0.47	0.10	0.09	1
2	0.34	0.44	0.11	0.09	1
3	0.36	0.41	0.13	0.08	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
10,000	0.35	0.43	0.14	0.08	1
Mean	0.34	0.43	0.13	0.09	0.99

Example: smartphone market shares



- **Pr(Android has largest share | Data) = 0.991**