Lectures 12: Multi-parameter models

- Recap on Normal model
- The multinomial model
- Prelude to MCMC
Summary of conjugate Normal model

- **Priors on $\mu, \sigma^2$:**

  \[
  \pi(\mu, \sigma^2) = \pi(\mu | \sigma^2)\pi(\sigma^2) \text{ or } \pi(\mu, \tau) = \pi(\mu | \tau)\pi(\tau).
  \]

- **Priors on $\mu | \sigma^2$:**

  \[
  \mu | \sigma^2 \sim \text{N}(\mu_0, \sigma^2/\tau_0)
  \]

  \[
  \pi(\mu | \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2/\tau_0}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma^2/\tau_0}}
  \]

  \[
  \propto \sigma^{-1} e^{-\frac{1}{2\sigma^2} \tau_0 (\mu-\mu_0)^2}
  \]

  \[
  \pi(\mu | \tau) \propto \tau^{\frac{1}{2}} e^{-\frac{\tau}{2} \tau_0 (\mu-\mu_0)^2}.
  \]

- **Priors on $\sigma^2$:**

  \[
  \sigma^2 \sim \text{IG}(\nu_0/2, SS_0/2)
  \]

  \[
  \tau \sim \text{Ga}(\nu_0/2, SS_0/2)
  \]

  \[
  \pi(\tau) = \frac{(SS_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} \tau^{\nu_0/2 - 1} e^{-\frac{SS_0}{2} \tau}
  \]

  \[
  \propto \tau^{\nu_0/2 - 1} e^{-\frac{\tau}{2} SS_0}.
  \]
Derivations

Posterior: \( p(\mu, \sigma^2|x_1, \ldots, x_n) \propto f(x_1, \ldots, x_n|\mu, \sigma^2)\pi(\mu|\sigma^2)\pi(\sigma^2) \)

- Conditional on \( \mu \) given \( \sigma^2 \)
  \[ p(\mu|\sigma^2, x_1, \ldots, x_n) = N(\mu_n, \frac{\sigma^2}{\tau_n}) \]

- Marginal on \( \sigma^2 \)
  \[ p(\sigma^2|x_1, \ldots, x_n) = IG\left(\frac{\nu_n}{2}, \frac{SS^2_n}{2}\right) \]

- Marginal on \( \mu \)
  \[ p(\mu|x_1, \ldots, x_n) = t_{\nu_n}(\mu_n, \frac{\sigma^2_n}{\tau_n}) \]

- Predictive distribution of future outcome \( x_f \)
  \[ p(x_f|x_1, \ldots, x_n) = \int \int \pi(x_f|\mu, \sigma^2)\pi(\mu, \sigma^2|x_1, \ldots, x_n)d\mu d\sigma^2 = t_{\nu_n}(\mu_n, \frac{\sigma^2_n}{\tau_n} + \frac{1}{\tau_n}) \]

\[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad SS = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \]

\[ \mu_n = \frac{\tau_0 \mu_0 + n \bar{x}}{\tau_n}, \quad \tau_n = \tau_0 + n, \quad \nu_n = \nu_0 + n, \quad SS_n = SS_0 + SS + \frac{n \tau_0}{\tau_n} (\bar{x} - \mu_0)^2, \]

\[ \sigma^2_n = SS^2_n/\nu_n \]
Semi-conjugate prior

A semi-conjugate setting is obtained with independent priors
\[ \pi(\mu, \sigma^2) = \pi(\mu)\pi(\sigma^2) \]

\[ \mu \sim N(\mu_0, \sigma_0^2), \quad \sigma^2 \sim IG\left(\frac{\nu_0}{2}, \frac{SS_0^2}{2}\right) \]

then \[ \mu|\sigma^2, x \sim N(\mu_n, \tau_n^2), \quad \mu_n = \frac{\frac{\mu_0}{\sigma_0^2} + \bar{x}\frac{n}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \quad \tau_n^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \]

\[ \sigma^2|x \sim \text{not in closed form} \]

We will solve this with MCMC methods!
The Multinomial Model

- \( x = (x_1, \ldots, x_k) \) vector of counts, with \( x_j \) = number of observations for the \( j \)th category and \( \sum x_j = n \)
- \( p(x|\theta) \propto \prod_{j=1}^{k} \theta_j^{x_j} \) with \( \sum_j \theta_j = 1 \)
- Conjugate prior is Dirichlet, \( \theta \sim D(\alpha) \)

\[
\pi(\theta) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j-1}
\]

\[
p(\theta|x) \propto \prod_{j=1}^{k} \theta_j^{x_j+\alpha_j-1}
\]

that is, \( \theta|x \sim D(\alpha + x) \)

- Uniform prior if \( \alpha_j = 1 \) for every \( j \). Improper prior if \( \alpha_j = 0 \) for every \( j \)
  (uniform on \( \log \theta_j \)) but with proper posterior if there is at least one obs in each of the \( k \) categories (i.e., \( x_j > 0 \) for each \( j \)).
Example

Survey of 1447 US voters to find out their preferences in the upcoming presidential election. Data: $x_1 = 727$ support the republican candidate, $x_2 = 583$ the democratic candidate, $x_3 = 173$ have no preference.

$$p(x|\theta) \propto \theta_1^{727} \theta_2^{583} \theta_3^{173}, \quad \theta \sim D(1, 1, 1), \quad \theta|x \sim D(728, 584, 174)$$

Histogram of $\theta_1 - \theta_2$ indicates more support for republican candidate

```r
thetas <- rdirichlet(10000, c(728, 584, 174))
hist(thetas[,1]-thetas[,2], nclass=50)
```
Prelude to MCMC methods

Often the posterior distribution of a parameter $\theta$ or a function of it, $g(\theta)$, cannot be derived in closed form.

Example: Non-conjugate priors. For the normal model $x \sim N(\mu, \sigma^2)$, the independent prior $\pi(\mu, \sigma^2) = \pi(\mu)\pi(\sigma^2)$ with $\mu \sim N(\mu_0, \sigma_0^2)$ and $\sigma^2 \sim 1/\sigma^2$ (or Inv-Ga) leads to $p(\sigma^2|x)$ which is not of a familiar form.

Need alternative methods for computing posterior distributions and post summaries. Available options:

- Analytical methods based on approximations, e.g., large sample normal approximation of the posterior distribution, Laplace approximation methods, numerical integration (will not be covered).
- Simulation methods bases on direct sampling from the posterior (rejection sampling, importance sampling).
- Simulate from a Markov chain whose stationary distribution is the desired posterior distribution (e.g., via Gibbs sampler and Metropolis-Hastings algorithms) and then calculate MC estimates.