ST 740: Multiparameter Inference

Alyson Wilson

Department of Statistics
North Carolina State University

September 23, 2013

Multinomial Model

- Generalization of the binomial model, for the case where observations can have more than two possible values.
- Sampling distribution: multinomial with parameters $(\theta_1, ..., \theta_k)$, the probabilities associated to each of the k possible outcomes.
- Example: In a survey, respondents may: Strongly Agree, Agree, Disagree, Strongly Disagree, or have No Opinion when presented with a statement such as "8am is too early to have class".

Multinomial Model

Sampling Distribution

Formally

- $y = (y_1, y_2, \ldots, y_k)$, a k-vector of counts of the number of observations for each outcome
- θ_j: probability of jth outcome
- $\sum_{j=1}^{k} \theta_j = 1$ and $\sum_{j=1}^{k} y_j = n$

The sampling distribution has the form

$$f(\mathbf{y} \mid \theta) \propto \Pi_{j=1}^{k} \theta_j^{y_j}$$

Multinomial Model

Conjugate Prior Distribution

The conjugate prior distribution for the multinomial sampling distribution is the Dirichlet distribution, which is a multivariate generalization of the beta distribution.

$$\pi(\theta \mid \alpha) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1}$$

with $\alpha_j > 0$ and $\alpha_0 = \sum_{j=1}^{k} \alpha_j$.

$$E[\theta_j] = \frac{\alpha_j}{\alpha_0}$$

$$\text{Var}[\theta_j] = \frac{\alpha_j(\alpha_0 - \alpha_j)}{\alpha_0^2(\alpha_0 + 1)}$$

$$\text{Cov}(\theta_i, \theta_j) = -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}$$

Multinomial Model

Other Interesting Properties of the Dirichlet Distribution

- The marginal distribution of any subvector of parameters is also Dirichlet: e.g., $(\theta_i, \theta_j, 1 - \theta_i - \theta_j)$ is Dirichlet$(\alpha_i, \alpha_j, \alpha_0 - \alpha_i - \alpha_j)$
- The marginal distribution of θ_j is Beta$(\alpha_j, \alpha_0 - \alpha_j)$.
- The marginal distribution of $\theta_i + \theta_j$ is Beta$(\alpha_i + \alpha_j, \alpha_0 - \alpha_i - \alpha_j)$
- Common choices for noninformative prior distributions for the Dirichlet are $\alpha_j = 0, 0.5,$ or $1, \forall j$.
- α_0 is commonly thought of as the "prior sample size".

Multinomial Model

Conjugate Prior Distribution

The posterior distribution is also Dirichlet.

$$\pi(\theta \mid \alpha, \mathbf{y}) \propto \prod_{i=1}^{j} \theta_j^{\alpha_j + y_j - 1}$$

Multinomial Model

Example: In a survey, respondents may: Strongly Agree, Agree, Disagree, Strongly Disagree, or have No Opinion when presented with a statement such as "8am is too early to have class".

Our sampling distribution is multinomial, with parameters θ_{SA}, θ_A, θ_D, θ_{SD}, and θ_{NO}, where $\theta_{SA} + \theta_A + \theta_D + \theta_{SD} + \theta_{NO} = 1$.

Our prior distribution on the parameters is a noninformative Dirichlet(0.5,0.5,0.5,0.5,0.5),

$$\pi(\theta) \propto \theta_{SA}^{-0.5}\theta_{A}^{-0.5}\theta_{SD}^{-0.5}\theta_{D}^{-0.5}\theta_{NO}^{-0.5}$$

Multinomial Model

We observe $n = 29$ responses $y = (15, 8, 2, 2, 2)$. Our posterior distribution is

$$\pi(\theta \,|\, \mathbf{y}) \propto \theta_{SA}^{15+0.5-1} \theta_{A}^{8+0.5-1} \theta_{D}^{2+0.5-1} \theta_{SD}^{2+0.5-1} \theta_{NO}^{2+0.5-1}$$

The two probabilities we are interested in are the probability that a student "Strongly Agrees" that 8am is too early for class, and the probability that a student either "Disagrees" or "Strongly Disagrees" that 8am is too early for class.

Multinomial Model

This means that we are interested in the marginal posterior distribution of θ_{SA}, $\pi(\theta_{SA} \,|\, \mathbf{y})$, and the posterior distribution of $\theta_D + \theta_{SD}$, $\pi(\theta_D + \theta_{SD} \,|\, \mathbf{y})$.

We can find these marginal distributions two ways.

We can find the marginal distributions analytically using properties of the Dirichlet distribution. In particular,

$$\begin{aligned}
\theta_{SA} \,|\, \mathbf{y} &\sim \text{Beta}(\alpha_{SA} + y_{SA}, \sum \alpha_j + n - \alpha_{SA} - y_{SA}) \\
&= \text{Beta}(15.5, 16)
\end{aligned}$$

$$\begin{aligned}
\theta_D + \theta_{SD} \,|\, \mathbf{y} &\sim \text{Beta}(\alpha_D + \alpha_{SD} + y_D + y_{SD}, \\
&\qquad \sum \alpha_j + n - \alpha_D - \alpha_{SD} - y_D - y_{SD}) \\
&= \text{Beta}(5, 26.5)
\end{aligned}$$

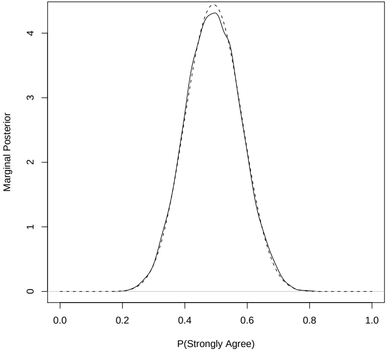
Generating Random Samples from Dirichlet Distributions

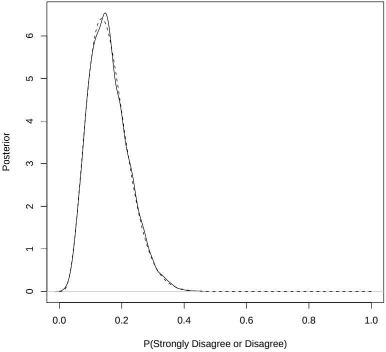
We can simulate directly from the joint posterior Dirichlet distribution (in R, `rdirichlet` from the LearnBayes package).

Gamma method: Two steps for each θ_j:

1. Draw $x_1, x_2, ..., x_k$ from independent Gamma$(\delta, (\alpha_j + y_j))$ for any common δ
2. Set $\theta_j = x_j / \sum_{i=1}^{k} x_i$

Marginal Posterior Distributions





Helpful Hint for Multiparameter Priors

A useful idea is that of the *induced prior*.

- Suppose that I have a sampling distribution $f(y \mid \alpha, \beta)$ that depends on two parameters, α and β.
- However, my prior information is about a function of α and β, say

$$\beta\Gamma(1 + \frac{1}{\alpha})$$

- We want to specify a prior distribution $\pi(\alpha, \beta)$ that reflects our prior information.
- Specifying a prior on $\pi(\alpha, \beta)$ *induces* (specifies) a prior $\beta\Gamma(1 + \frac{1}{\alpha})$.

Weibull Parameters

```
alpha <- rgamma(10000,val1,val2)
beta <- rgamma(10000,val3,val4)
mw <- beta*gamma(1 + 1/alpha)
simy <- rweibull(10000,alpha,beta)
```

Constraints on the Parameters

Start with an unconstrained problem and Bayes' Theorem.

$$\pi(\theta \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{\int f(\mathbf{y} \mid \theta)\pi(\theta)d\theta}$$

Suppose we want to find the posterior distribution when $\theta \in \Theta_C$, where $\int_{\Theta_C} \pi(\theta)d\theta > 0$.

Define the constrained prior density as

$$\pi^C(\theta) = \frac{\pi(\theta)}{\int_{\Theta_C} \pi(\theta)d\theta}, \theta \in \Theta_C$$

Constraints on the Parameters

Start with an unconstrained problem and Bayes' Theorem.

$$\pi(\theta \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{\int f(\mathbf{y} \mid \theta)\pi(\theta)d\theta}$$

Suppose we want to find the posterior distribution when $\theta \in \Theta_C$, where $\int_{\Theta_C} \pi(\theta)d\theta > 0$.

Define the constrained prior density as

$$\pi^C(\theta) = \frac{\pi(\theta)}{\int_{\Theta_C} \pi(\theta)d\theta}, \theta \in \Theta_C$$

Constraints on the Parameters

Start with an unconstrained problem and Bayes' Theorem.

$$\pi(\theta \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{\int f(\mathbf{y} \mid \theta)\pi(\theta) d\theta}$$

Suppose we want to find the posterior distribution when $\theta \in \Theta_C$, where $\int_{\Theta_C} \pi(\theta) d\theta > 0$.

Define the constrained prior density as

$$\pi^C(\theta) = \frac{\pi(\theta)}{\int_{\Theta_C} \pi(\theta) d\theta}, \theta \in \Theta_C$$

Constraints on the Parameter

Using Bayes' Theorem

$$
\begin{aligned}
\pi(\theta \,|\, \mathbf{y}, \theta \in \Theta_C) &= \frac{f(\mathbf{y} \,|\, \theta)\pi^C(\theta)}{\int_{\Theta_C} f(\mathbf{y} \,|\, \theta)\pi^C(\theta)d\theta}, \theta \in \Theta_C \\
&= \frac{f(\mathbf{y} \,|\, \theta)\frac{\pi(\theta)}{\int_{\Theta_C} \pi(\theta)d\theta}}{\int_{\Theta_C} f(\mathbf{y} \,|\, \theta)\frac{\pi(\theta)}{\int_{\Theta_C} \pi(\theta)d\theta}d\theta}, \theta \in \Theta_C \\
&= \frac{\frac{f(\mathbf{y} \,|\, \theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{y} \,|\, \theta)\pi(\theta)d\theta}}{\int_{\Theta_C} \frac{f(\mathbf{y} \,|\, \theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{y} \,|\, \theta)\pi(\theta)d\theta}d\theta}, \theta \in \Theta_C \\
&= \frac{\pi(\theta \,|\, \mathbf{y})}{\int_{\Theta_C} \pi(\theta \,|\, \mathbf{y})d\theta}, \theta \in \Theta_C
\end{aligned}
$$

General idea: Do the unconstrained problem, truncate, renormalize.

Constraints on the Parameter

Using Bayes' Theorem

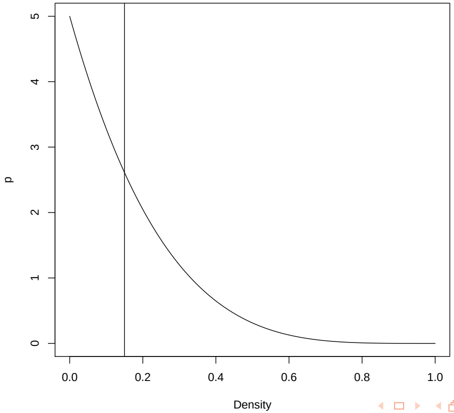
$$
\begin{aligned}
\pi(\theta \mid \mathbf{y}, \theta \in \Theta_C) &= \frac{f(\mathbf{y} \mid \theta)\pi^C(\theta)}{\int_{\Theta_C} f(\mathbf{y} \mid \theta)\pi^C(\theta)d\theta}, \theta \in \Theta_C \\
&= \frac{f(\mathbf{y} \mid \theta)\frac{\pi(\theta)}{\int_{\Theta_C} \pi(\theta)d\theta}}{\int_{\Theta_C} f(\mathbf{y} \mid \theta)\frac{\pi(\theta)}{\int_{\Theta_C} \pi(\theta)d\theta}d\theta}, \theta \in \Theta_C \\
&= \frac{\frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{y} \mid \theta)\pi(\theta)d\theta}}{\int_{\Theta_C} \frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{y} \mid \theta)\pi(\theta)d\theta}d\theta}, \theta \in \Theta_C \\
&= \frac{\pi(\theta \mid \mathbf{y})}{\int_{\Theta_C} \pi(\theta \mid \mathbf{y})d\theta}, \theta \in \Theta_C
\end{aligned}
$$

General idea: Do the unconstrained problem, truncate, renormalize.

Constraints on the Parameter

Example

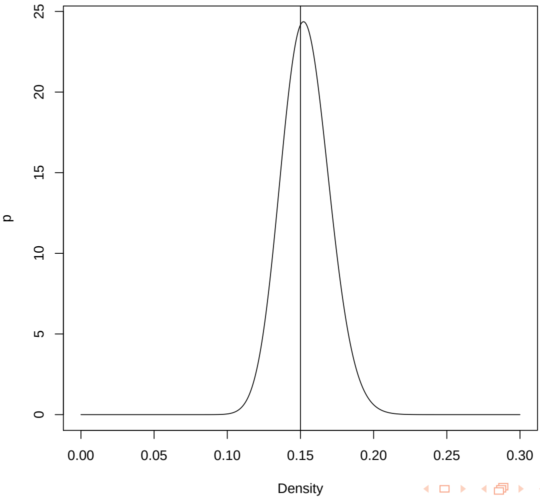
We are trying to estimate the probability that Barry Bonds hits a home run in 2001. We observe 73 home runs in 476 at bats. Our prior distribution for p is Beta(1,5) with $p \in (0, 0.15)$.



Constraints on the Parameter

Example

The unconstrained posterior distribution is Beta$(1 + 73, 403 + 5)$.



Less Nice Example

Suppose that our sampling distribution is Gamma(α, β) and we choose independent marginal prior distributions $\pi(\alpha) \sim$ Gamma$(2, 1)$ and $\pi(\beta) \sim$ Gamma$(5, 1)$.

$$
\begin{aligned}
\pi(\alpha, \beta \,|\, \mathbf{y}) &\propto \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} \exp(-\beta \sum y_i)(\prod y_i)^{\alpha-1} \exp(-(\alpha+\beta))\alpha\beta^4 \\
&\propto \frac{\beta^{n\alpha+4}}{\Gamma(\alpha)^n} \exp(-\beta(\sum y_i + 1))\alpha \exp(-\alpha)(\prod y_i)^{\alpha-1}
\end{aligned}
$$

We want to make posterior inferences about α, β, and the predictive distribution for the next observation.

Predictive Distribution

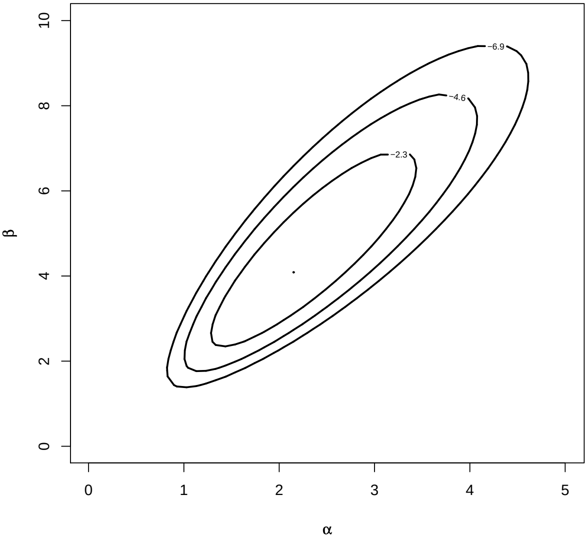
The predictive distribution is the distribution of the next observation conditional on the data observed so far. The uncertainty about the parameters is integrated out.

$$f(y_{n+1} \mid \mathbf{y}) = \int f(y_{n+1} \mid \alpha, \beta)\pi(\alpha, \beta \mid \mathbf{y}) d\alpha d\beta$$

where

$$f(y_{n+1} \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \exp(-\beta y_{n+1})(y_{n+1})^{\alpha-1}$$

Contour Plot of Posterior Distribution



Joint, Marginal, and Conditional Distributions

Posterior?

$$\pi(\alpha, \beta \mid \mathbf{y}) \propto \frac{\beta^{n\alpha+4}}{\Gamma(\alpha)^n} \exp(-\beta(\sum y_i + 1))\alpha \exp(-\alpha)(\prod y_i)^{\alpha-1}$$

Conditionals?

$\pi(\alpha \mid \beta, \mathbf{y})$ hard.

$\beta \mid \alpha, \mathbf{y} \sim \text{Gamma}(n\alpha + 5, \sum y_i + 1)$

Marginals?

$\pi(\beta \mid \mathbf{y})$ hard.

$\pi(\alpha \mid \mathbf{y}) \propto \frac{\alpha \exp(-\alpha)}{\Gamma(\alpha)^n} \frac{\Gamma(n\alpha+5)}{(\sum y_i+1)^{n\alpha+5}} (\prod y_i)^{\alpha-1}$