2: Single-parameter models

Taylor

University of Virginia

Introduction

We discuss prior selection, and demonstrate new ideas with examples of models with only one parameter.

Binomial: from prior to posterior

Let y be the number of n births that are female. Let θ be the population proportion of births that are female

Likelihood:

1. $y \mid \theta \sim \text{Bin}(n, \theta)$
2. $p(y \mid \theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y} \propto \theta^y(1-\theta)^{n-y}$

Chosen prior:

1. $\theta \sim \text{Uniform}(0, 1)$
2. $p(\theta) = \mathbb{1}(0 < \theta < 1)$

Binomial: from prior to posterior

Let y be the number of n births that are female. Let θ be the population proportion of births that are female

Likelihood:

1. $y \mid \theta \sim \text{Bin}(n, \theta)$
2. $p(y \mid \theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y} \propto \theta^y(1-\theta)^{n-y}$

Chosen prior:

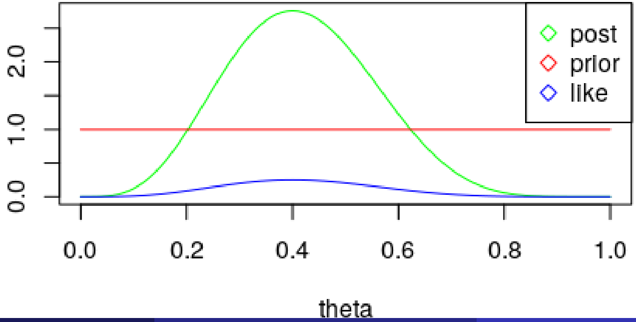
1. $\theta \sim \text{Uniform}(0, 1)$
2. $p(\theta) = \mathbb{1}(0 < \theta < 1)$

Posterior:

1. $p(\theta \mid y) \propto \theta^y(1-\theta)^{n-y}\mathbb{1}(0 < \theta < 1)$
2. $\theta \mid y \sim \text{Beta}(y+1, n-y+1)$

Binomial

```
thetas <- seq(0,1,.01)
priorEvals <- rep(1, length(thetas))
likes <- choose(n,y) * thetas^y * (1-thetas)^(n-y)
posteriors <- dbeta(thetas, shape1 = y+1, shape2 = n-y+1)
```



Binomial

Finding a **posterior/credible interval** is easy. Recall that
$\theta \mid y \sim \text{Beta}(y+1, n-y+1)$

```
# find a posterior interval
left <- qbeta(.025, y+1, n-y+1)
right <- qbeta(.975, y+1, n-y+1)
cat("posterior interval", "(", left, ",", right, ")")
posterior interval ( 0.1674881 , 0.6920953 )
```

Finding **highest posterior density intervals** is a little trickier, but even more useful.

Binomial prediction

We just found $\theta \mid y \sim \text{Beta}(y+1, n-y+1)$.
What's the chance that $\tilde{y} = 1$ if $\tilde{y} \mid \theta \sim \text{Bern}(\theta)$?

Binomial prediction

We just found $\theta \mid y \sim \text{Beta}(y+1, n-y+1)$.
What's the chance that $\tilde{y} = 1$ if $\tilde{y} \mid \theta \sim \text{Bern}(\theta)$?

$$\begin{aligned}
\Pr(\tilde{y} = 1 \mid y) &= \int \Pr(\tilde{y} = 1 \mid \theta) p(\theta \mid y) \mathrm{d}\theta \\
&= \int \theta p(\theta \mid y) \mathrm{d}\theta \\
&= E[\theta \mid y] = (y+1)/(n+2)
\end{aligned}$$

Conjugacy

Conjugate Priors

Let $\mathcal{F}$ be a class of sampling distributions $p(y \mid \theta)$. Let $\mathcal{P}$ be a class of prior distributions $p(\theta)$. Then $\mathcal{P}$ is **conjugate** for $\mathcal{F}$ if

$$p(\theta \mid y) \in \mathcal{P}$$

for all $p(\theta) \in \mathcal{P}$ and $p(y \mid \theta) \in \mathcal{F}$.

Conjugacy example 1

beta - binomial:

1. $p(\theta) = [\mathrm{B}(\alpha, \beta)]^{-1} \theta^{\alpha-1}(1-\theta)^{\beta-1}$
2. $p(y \mid \theta) = \prod_i \theta^{y_i}(1-\theta)^{1-y_i}$
3. $p(\theta \mid y) \propto \theta^{\sum_i y_i}(1-\theta)^{n-\sum_i y_i} \theta^{\alpha-1}(1-\theta)^{\beta-1}$

$$p(\theta \mid y) = \frac{1}{\mathrm{B}(\alpha + \sum_i y_i, \beta + n - \sum_i y_i)} \theta^{\alpha+\sum_i y_i - 1}(1-\theta)^{\beta+n-\sum_i y_i}$$

Conjugacy example 2

inverse gamma-normal (known mean):

1. $p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} \exp\left(-\beta/\sigma^2\right)$
2. $p(y \mid \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_i (y_i - \theta)^2\right]$
3. $p(\sigma^2 \mid y) = \text{Inverse-Gamma}\left(\frac{n+\nu_0}{2}, \frac{\nu_0\sigma^2+nv}{2}\right)$

where θ is some known mean, and $v = \sum_i (y_i - \theta)^2/n$.

Conjugacy example 2

$$
\begin{aligned}
& p(\sigma^2 \mid y) \\
& \propto (\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{\sigma_0^2 \nu_0}{2\sigma^2}\right) (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_i (y_i - \theta)^2\right] \\
& \vdots \qquad \text{(homework)} \\
& \propto (\sigma^2)^{-(\frac{n+\nu_0}{2}+1)} \exp\left[-\frac{1}{2\sigma^2} \left(\nu_0 \sigma^2 + nv\right)\right]
\end{aligned}
$$

Conjugacy example 3

gamma-Poisson:

1. $p(\theta) \propto \theta^{\alpha-1} \exp(-\beta\theta)$
2. $p(y \mid \theta) \propto e^{-n\theta}\theta^{\sum_i y_i}$
3. $p(\theta \mid y) = \text{Gamma}(\alpha + n\bar{y}, \beta + n)$

note the rate parameterization

Conjugacy example 3

$$
\begin{aligned}
&p(\theta \mid y) \\
&\propto \theta^{\alpha-1} \exp\left(-\beta\theta\right) \exp(-n\theta)\theta^{\sum_i y_i} \\
&\propto \theta^{\sum_i y_i + \alpha - 1} \exp\left(-(\beta + n)\theta\right)
\end{aligned}
$$

Conjugacy example 4

gamma-exponential:

1. $p(\theta) \propto \theta^{\alpha-1} \exp(-\beta\theta)$
2. $p(y \mid \theta) \propto \theta^{n} \exp(-\theta \sum_i y_i)$
3. $p(\theta \mid y) = \text{Gamma}(\alpha + n, \beta + \sum_i y_i)$

Conjugacy example 5

normal-normal (known variance):

1. $p(\theta) \propto \exp\left[-\frac{1}{2\tau_0}(\theta - \mu_0)^2\right]$
2. $p(y \mid \theta) \propto \prod_i \exp\left[-\frac{1}{2\sigma^2}\left(y_i - \theta\right)^2\right]$
3. $p(\theta \mid y) = \text{Normal}(\mu_n, \tau_n^2)$

Conjugacy example 5

$$\begin{aligned}
&p(\theta \mid y) \\
&\propto \exp\left[-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right]\prod_i \exp\left[-\frac{1}{2\sigma^2}(y_i-\theta)^2\right] \\
&= \exp\left[-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right]\exp\left[-\frac{1}{2\sigma^2}\sum_i (y_i-\theta)^2\right] \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(homework)} \\
&\propto \exp\left[-\frac{1}{2\tau_n^2}(\theta-\mu_n)^2\right]
\end{aligned}$$

where $\frac{1}{\tau_n^2} = \frac{1}{\sigma^2} + \frac{1}{\tau_0^2}$ and
$\mu_n = \frac{\mu_0\sigma^2 + n\tau_0^2\bar{y}}{n\tau_0^2+\sigma^2} = \mu_0\left(\frac{1/\tau_0^2}{n/\sigma^2+1/\tau_0^2}\right) + \bar{y}\left(\frac{n/\sigma^2}{n/\sigma^2+1/\tau_0^2}\right)$.

Conjugacy example 6

general exponential families:

1. $p(\theta) \propto g(\theta)^{\eta} \exp\left[\phi(\theta)'\nu\right]$
2. $p(y \mid \theta) \propto g(\theta)^{n} \exp\left[\phi(\theta)'t(y)\right]$
3. $p(\theta \mid y) \propto g(\theta)^{\eta+n} \exp\left[\phi(\theta)'\{t(y)+\nu\}\right]$

Improper/Proper Priors

Proper Prior

A prior $p(\theta)$ is **improper** if it integrates to ∞.

Improper priors may still be used, as long as the *posterior* integrates to 1. If one is using an improper prior, one must check that

$$\int p(y \mid \theta) p(\theta) \mathrm{d}\theta < \infty$$

Interpretation should also be justified.

Improper/Proper Priors example 1

$p(\sigma^2) \propto (\sigma^2)^{-1}$ is a common example of an improper (and noninformative) prior:

$$\begin{aligned} &p(\sigma^2)p(y \mid \sigma^2) \\ &\propto (\sigma^2)^{-1}(\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_i (y_i - \theta)^2\right] \\ &= (\sigma^2)^{-[(n/2)+1]} \exp\left[-\frac{nv}{2\sigma^2}\right] \end{aligned}$$

with $v = \sum_i (y_i - \theta)^2/n$. So

$$\sigma^2 \mid y \sim \text{Inv-Gamma}(n/2, nv/2)$$

Jeffreys' Prior

Setting: we want $p(\theta)$ to be as noninformative as possible. Perhaps we don't have any pre-existing scientific knowledge about θ.

If you don't have any "information" about θ, then you also don't have any "information" about $\phi = h(\theta)$ (1-to-1), right?

Jeffreys' Prior

Setting: we want $p(\theta)$ to be as noninformative as possible. Perhaps we don't have any pre-existing scientific knowledge about θ.

If you don't have any "information" about θ, then you also don't have any "information" about $\phi = h(\theta)$ (1-to-1), right?

Uh oh:

$$p(\phi) = p(\theta)\left|\frac{d\theta}{d\phi}\right|$$

$|h'(\theta)|^{-1}$ isn't always flat!

Jeffreys' Prior

Should we make $p(\theta)$ flat, or should we make $p(\phi)$ flat? Wouldn't it be nice if we had a rule to tell us?

Jeffreys Prior

The Jeffreys' prior for θ is

$$p(\theta) \propto [J(\theta)]^{1/2}$$

where $J(\theta)$ is the Fisher Information of one data point.

Jeffreys' Prior

You will show in your homework that

$$J(\phi) = J(\theta)\left|\frac{\mathrm{d}\theta}{\mathrm{d}\phi}\right|^2$$

so

$$p(\phi) = \sqrt{J(\phi)} = \sqrt{J(\theta)}\left|\frac{\mathrm{d}\theta}{\mathrm{d}\phi}\right| = p(\theta)\left|\frac{\mathrm{d}\theta}{\mathrm{d}\phi}\right|.$$

No more ambiguity! If we find $p(\theta)$ first, then transform to ϕ to find $p(\phi)$, then that's the same as if we just found $p(\phi)$ straight away following Jeffreys' strategy. This prior is "invariant to parameterization."

Note that this does not mean that one parameterization is guaranteed to be flat, nor does it mean that the densities are the same functions.

Jeffreys' Prior Example 1

Example $p(y \mid \theta) = (2\pi)^{-1/2}\theta^{-1/2}\exp\left[-\frac{1}{2\theta}y^2\right]$. $J(\theta) = \frac{1}{\theta^2}$.

So Jeffreys' prior is

$$\sqrt{J(\theta)} \propto 1/\theta$$

If you define $\phi = \log\theta$ then

$$p(\phi) = p_\theta(\theta[\phi])\left|\frac{d\theta[\phi]}{d\phi}\right| \propto e^{-\phi+\phi} = 1$$

Noninformative Priors

Example $p(y \mid \theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$. We fix a parameterization.

1. Bayes'-Laplace: $p(\theta) \propto 1$
2. Jeffreys': $\sqrt{J(\theta)} \propto \theta^{1/2-1}(1-\theta)^{1/2-1}$
3. $p(\theta) \propto \theta^{0-1}(1-\theta)^{0-1}$ (if $p(\text{logit}\theta) \propto 1$)

Hopefully our likelihood will stay away from those side peaks.