One Parameter Models

September 22, 2010

Reading: Hoff Chapter 3
Highest Posterior Density Regions

Find

\[ \Theta_{1-\alpha} = \{ \theta : p(\theta \mid Y) \geq h_\alpha \} \text{ such that } P(\theta \in \Theta_{1-\alpha} \mid Y) = 1-\alpha \]

All points in \( \Theta_{1-\alpha} \) have a higher density than any point outside the regions. Often requires iterative solution:

- Find points such that \( p(\theta \mid Y) > h \)
- Find probability of that set
- Adjust \( h \) until reach desired coverage
- may not have symmetric tail areas
- multimodal posterior may not have an interval
solve.HPD.beta

Key ideas:

- use relative posterior by dividing posterior density by the density at the mode (range is 0 to 1)
- for a height $h$ in $(0,1)$ find points where $p(\theta | Y) = h$ on either side of the mode, using the \texttt{uniroot} function
- use \texttt{pbeta} function to calculate area
- move $h$ up or down so that area is closer to coverage $1 - \alpha$
Exponential Example: Rate $\beta$

The time between accidents modeled with an exponential distribution with a rate of $\beta$ accidents per day. $n = 10$ observations:

1.5 15.0 60.3 30.5 2.8 56.4 27.0 6.4 110.7 25.4

$$f(y|\beta) = \beta \exp(-y\beta) \quad y > 0$$

$$L(\beta; y_1, \ldots, y_n) = \prod_i \beta \exp(-y_i\beta) \quad (Likelihood)$$

$$= \beta^n \exp(-\sum_i y_i\beta)$$

Look at plot of $L(\beta)$ for $\sum y_i = 336$. 
Likelihood function

```r
l.exp = function(theta, y) {
  n = length(y)
  sumy = sum(y)
  l = theta^n*exp(-sumy*theta)
  return(l)
}
```

Vectorized: can be used to evaluate $L(\theta)$ at multiple values of $\theta$ rather than using a loop

```r
beta = seq(.00001, .25, length=1000)
l.exp(beta, y)
```
Plot of Likelihood

Exponential Likelihood

$\beta$

$L(\beta)$

$y$-axis: $L(\beta)$

$x$-axis: $\beta$

1.0e-20

1.5e-20

2.0e-20

2.5e-20

5.0e-21

1.0e-20

1.5e-20

2.0e-20

2.5e-20

5.0e-21

0.0e+00

0.00

0.05

0.10

0.15

0.20

0.25

One Parameter Models – p. 6/21
Conjugate Prior Distributions

Recall that a family of distributions is conjugate for a sampling model, if for any prior in the class the posterior is also in the class.

If the uniform distribution is in the class, then that means that the posterior must be proportional to the Likelihood.

Use the form of the likelihood to help identify the conjugate prior:

\[ L(\beta) \propto \beta^n \exp(-\beta \sum y_i) \]
Gamma Distribution

\[ Z \sim \text{Gamma}(a, b) \text{ with mean } \frac{a}{b} \text{ and variance } \frac{a}{b^2} \text{ with density} \]

\[
f(z) = \frac{b^a}{\Gamma(a)} z^{a-1} \exp(-zb) \quad z > 0
\]

In this parameterization \( b \) is a rate parameter.

- \( \text{rgamma}(n, \text{shape}=a, \text{scale}=s) \) \# mean = shape*scale
- \( \text{rgamma}(n, \text{shape}=a, \text{rate}=r) \) \# mean = shape/rate

Mode is at \((a - 1)/b\) if \( a > 1 \) and at 0 if \( a \leq 1 \)
Examples of Gamma Distributions

Density

- $a = 1, r = 1/5$
- $a = 2, r = 1/5$
- $a = 2, r = 1/2$
Distribution of $\beta$

Identify normalizing constant to make this a density for $\beta$

or recognize the form of the distribution

$$p(\beta|y_1, \ldots, y_n) = c\beta^n \exp(-\beta \sum y_i)$$

where

$$c = 1/\int_{0}^{\infty} \beta^n \exp(-\beta \sum y_i) d\beta$$

Looks like a Gamma($a = n + 1, b = \sum y_i$)

- normalized likelihood has same shape has likelihood
- same form that would be obtained using Bayes Theorem with a “uniform” prior
Uniform Prior

- “Uniform” prior $p(\beta) = 1$ in exponential example is not a proper distribution; although the posterior distribution is a proper distribution.
- “Formal Bayes” posterior distribution obtained as a limit of a proper Bayes procedure.
- Be very careful with improper prior distributions, they may not lead to proper posterior distributions!
Gamma Prior/Posterior Distributions

Prior and Posterior distributions are in the same family

\[ \beta \sim G(a, b) \] \hspace{1cm} (1)

\[ p(\beta \mid Y) \propto \frac{b^a}{\Gamma(a)} \beta^{a-1} e^{-\beta b} \beta^n e^{-\beta \sum y_i} \] \hspace{1cm} (2)

\[ \propto \beta^{a+n-1} e^{-\beta(b+\sum y_i)} \] \hspace{1cm} (3)

\[ \beta \mid Y \sim G(a + n, b + \sum y_i) \] \hspace{1cm} (4)

- “Uniform” is a limiting case with \( a = 1 \) and \( b \to 0 \)
- “default” prior is \( p(\beta) \propto 1/\beta \) – a “Gamma(0,0)”
Posterior Quantities

- posterior distribution $G(10, 336)$
- posterior mode $\hat{\beta} = 0.0268$
- posterior mean $E(\beta|Y) = 10/336 = 0.0298$

Interval Estimates – Probability Intervals or Credible Regions
- Highest Posterior Density region $[0.012, 0.049]$
- Equal Tail area $[0.014, 0.051]$

After observing the data, we believe there is a 95% chance of $[0.01, 0.05]$ accidents per day (or 1 to 5 accidents per every 100 days)
Highest Posterior Density Region

95% Highest Posterior Density
[0.012, 0.05]

\[ P(0.012 < \beta < 0.049 \mid Y) = 0.959 \]
Equal Tail Area Intervals

Easier alternative is to find points such that

- \( P(\theta < \theta_l \mid Y) = \alpha/2 \)
- \( P(\theta > \theta_u \mid Y) = \alpha/2 \)
- \((\theta_l, \theta_u)\) is a \(1 - \alpha\) 100\% Credible Interval (or Posterior Probability Interval) for \(\theta\).

\[
\begin{align*}
> & \ \texttt{qgamma(.025, 10, 336)} \\
& \quad [1] 0.01427199 \\
> & \ \texttt{qgamma(.975, 10, 336)} \\
& \quad [1] 0.05084763
\end{align*}
\]
# Examples of Conjugate Families

<table>
<thead>
<tr>
<th>Data Distributions</th>
<th>Prior/Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>Gamma</td>
</tr>
<tr>
<td>Poisson</td>
<td>Gamma</td>
</tr>
<tr>
<td>Binomial</td>
<td>Beta</td>
</tr>
<tr>
<td>Normal (known variance)</td>
<td>Normal</td>
</tr>
<tr>
<td>Normal (unknown mean/variance)</td>
<td>Normal-Gamma</td>
</tr>
</tbody>
</table>

Always available in exponential families
Exponential Family

One parameter exponential family is expressed as

\[ p(y \mid \theta) = h(y)c(\theta) \exp(\theta t(y)) \]

where \( \theta \) is the natural parameter of the exponential family and \( t(y) \) is the sufficient statistic

Likelihood:
\[ L(\theta) \propto c(\theta)^n \exp(\theta n \sum_i t(y_i) / n) \]

prior:
\[ p(\theta) \propto c(\theta)^{n_0} \exp(\theta n_0 t_0) \]

where \( n_0 \) may be interpreted as the number of prior observations and \( t_0 \) is the prior expected value of \( t(Y) \)
Conjugate Prior & Posterior

- Likelihood: \( L(\theta) \propto c(\theta)^n \exp(\theta n \sum_i t(y_i) / n) \)
- Prior: \( p(\theta) \propto c(\theta)^{n_0} \exp(\theta n_0 t_0) \)
- Posterior: 
  \( p(\theta \mid Y) \propto c(\theta)^{n_0+n} \exp(\theta(n_0 t_0 + n \sum_i t(y_i) / n)) \) subject to finite normalizing constant
- Updating of hyperparameters:
  \[
  \begin{align*}
  n_0 & \rightarrow n_0 + n \\
  t_0 & \rightarrow \frac{n_0 t_0 + n t(y)}{n_0 + n} \\
  t(y) & \equiv \sum_{i=1}^n t(y_i)
  \end{align*}
  \]
Exponential

\[ p(y \mid \beta) = \beta \exp(-\beta y) = \theta \exp(\theta(-y)) \]

- natural parameter \( \theta = \beta \)
- sufficient statistic \( -\bar{y} \)
- conjugate prior \( p(\theta) \propto \theta^{n_0} \exp(\theta n_0 t_0) \) or \( \text{Gamma}(n_0, -n_0 t_0) \) or \( \text{Gamma}(n_0, n_0 \bar{y}_0) \)

Note: Exponential distribution is NOT the same as an exponential family
Binomial

Rearrange Bernoulli (a Binomial with n = 1) to get in to exponential family form:

\[ p(y \mid \pi) = \pi^y (1 - \pi)^{1-y} \]

\[ = \frac{\pi^y}{1 - \pi} (1 - \pi) \]

\[ = \left[ \exp \left\{ \log \left( \frac{\pi}{1 - \pi} \right) y \right\} \right] (1 - \pi) \]

\[ = c(\theta) \exp(\theta y) \]

where \( \theta = \log(\pi/(1 - \pi)) \)

\[ c(\theta) = (1 + \exp(\theta))^{-1} \]

Natural parameter \( \theta \) is the log-odds.
Conjugate Prior

\[ p(\theta) \propto \left( \frac{1}{1 + \exp(\theta)} \right)^{n_0} \exp(\theta n_0 t_0) \]

where \( t_0 \) is prior expected value for the probability that \( Y \) is 1

- Normalizing constant?
- Implied Prior distribution for \( \pi \)
- Review change of variables