Bayesian Inference for One Parameter Models

The Exponential Family

Yingbo Li

Clemson University

MATH 9810

Exponential Example: Rate β

Model inter-arrival times, e.g., volcano eruptions, traffic accidents.

The time between accidents modeled with an exponential distribution with a rate of β accidents per day.
Suppose we collect $n = 10$ such times:

```
y = c(1.5, 15, 60.3, 30.5, 2.8, 56.4, 27, 6.4, 110.7, 25.4);
```

1. What is a 95% credible interval for the average waiting time between accidents?
2. What is the probability that the average waiting time between accidents is less than 20 days?

Exploratory Data Analysis

Let's use the exponential distribution to model the accidents waiting times.

$$
\begin{aligned}
f(y|\beta) &= \beta \exp(-y\beta) \qquad y > 0 \\
E(Y|\beta) &= 1/\beta; \;\; Var(Y|\beta) = 1/\beta^2 \\
L(\beta; y_1, \ldots, y_n) &= \prod_i \beta \exp\left(-y_i\beta\right) \quad (Likelihood) \\
&= \beta^n \exp\left(-\sum_i y_i\beta\right)
\end{aligned}
$$

Look at plot of $l(\beta)$ for $\sum y_i = 336$.

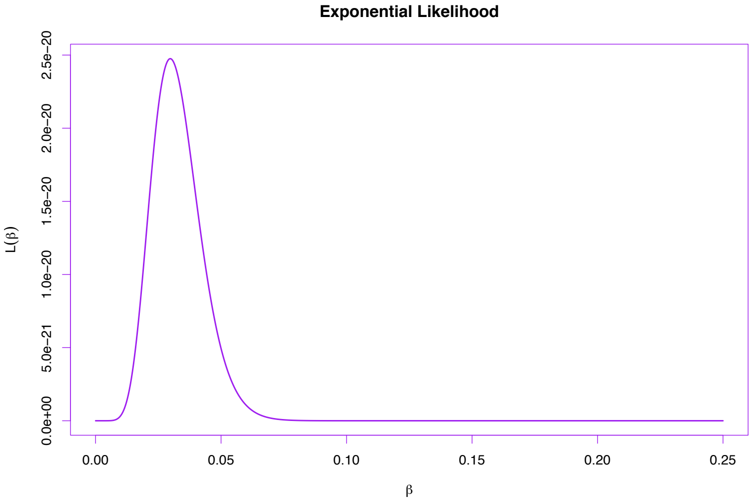
Likelihood function

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

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

```
beta = seq(.00001, .25, length=1000)
likbeta = l.exp(beta, y)
```

Plot of Likelihood



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
- Assume for the moment that $p(\beta) \propto k$, where k is some positive constant. Then, 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

$Z \sim$ Gamma(a, b) with mean a/b and variance a/b^2 with density

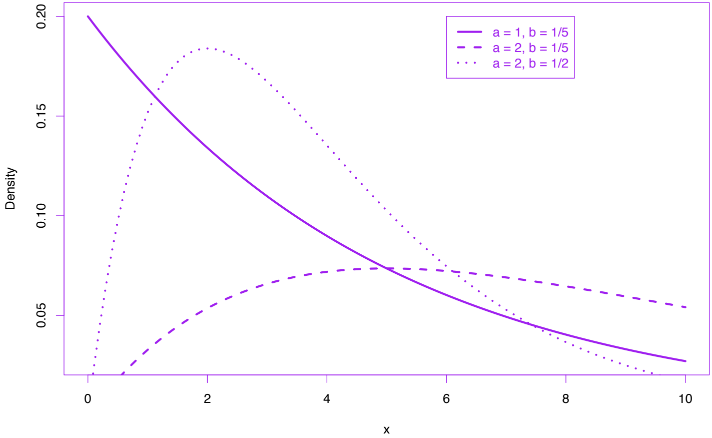
$$f(z) = \frac{b^a}{\Gamma(a)} z^{a-1} \exp(-zb) \qquad z > 0, a > 0, b > 0$$

In this parameterization, b is a rate parameter.

- `rgamma(n, shape=a, scale=s) # mean = shape*scale`
- `rgamma(n, shape=a, 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



Posterior distribution of β

Using $p(\beta) \propto k$, we have

$$p(\beta|y_1,\ldots,y_n) = \frac{p(y_1,\ldots,y_n \mid \beta)p(\beta)}{p(y_1,\ldots,y_n)} \tag{1}$$

$$\propto \beta^n \exp\left(-\sum_{i=1}^{n} y_i\beta\right) \tag{2}$$

which you can recognize as the the kernel of a Gamma$(a = n+1, b = \sum y_i)$, which is the posterior.

Posterior distribution of β

Let's derive the posterior formally,

$$p(\beta|y_1,\ldots,y_n) = \frac{\beta^n \exp\left(-\beta \sum y_i\right)(k)}{\int_0^\infty \beta^n \exp(-\beta \sum y_i) k d\beta} \tag{3}$$

The k cancels. If we multiply numerator and denominator by $\frac{(\sum_i y_i)^{n+1}}{\Gamma(n+1)}$, the denominator integrates to one. So, we are left with

$$p(\beta|y_1,\ldots,y_n) = \frac{(\sum_i y_i)^{n+1}}{\Gamma(n+1)} \beta^n \exp(-\beta \sum y_i)$$

which is a Gamma($a = n+1, b = \sum y_i$)

Uniform Prior

- "Uniform" prior $p(\beta) = 1$ in exponential example is not a proper distribution (does not integrate to one), 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 (abbreviate Gamma with G)

$$\beta \sim G(a, b) \tag{4}$$

$$p(\beta \mid Y) \propto \frac{b^a}{\Gamma(a)} \beta^{a-1} e^{-\beta b} \beta^n e^{-\beta \sum y_i} \tag{5}$$

$$\propto \beta^{a+n-1} e^{-\beta(b+\sum y_i)} \tag{6}$$

$$\beta \mid Y \sim G(a+n, b+\sum y_i) \tag{7}$$

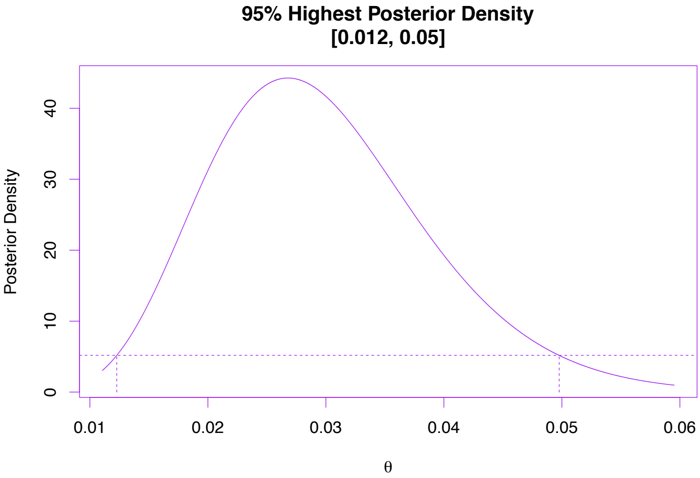
- "Uniform" is a limiting case with $a = 1$ and $b \to 0$
- "Default" prior is $p(\beta) \propto 1/\beta$ $--$ "Gamma(0,0)"
  Actually, this is Jeffreys' prior (see exercise 3.12).

Posterior Quantities (Default Prior)

- posterior distribution for β: $G(10, 336)$
- posterior mode $\hat{\beta} = 0.0268$
- posterior mean $E(\beta|Y) = 10/336 = 0.0298$
- Interval Estimates – Probability Intervals or Credible Regions
  - 95% highest posterior density region for β: $[0.012, 0.049]$
  - 95% central interval (equal tail area) for β: $[0.014, 0.051]$

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)

HPD Interval



Equal Tail Area Intervals

Easier alternative is to find points such that

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

```
> qgamma(.025, 10, 336)
[1] 0.01427199
> qgamma(.975, 10, 336)
[1] 0.05084763
```

Inference about $1/\beta$

Given 95% central credible interval for β, the 95% central credible region for $1/\beta$ is simply the limits inverted.

$$\begin{aligned} .95 &= p(0.014 < \beta < 0.051) \\ &= p(1/0.014 < 1/\beta < 1/0.051) \\ &= p(19.6 < 1/\beta < 71.4) \end{aligned}$$

So, there is a 95% probability that the average number of months to wait is between 19.6 and 71.4 months. And, there is a 0.029 probability that the average wait time is less than two years:

$$p(1/\beta < 20) = p(\beta > 0.05) = 0.029$$

Examples of Conjugate Families

Data Distributions	Prior/Posterior
Binomial	Beta
Exponential	Gamma
Poisson	Gamma
Normal (known variance)	Normal
Normal (unknown mean/variance)	Normal-Gamma

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 θ is the natural parameter of the exponential family and $t(y)$ is the sufficient statistic

$$\text{Likelihood:} \quad L(\theta) \propto c(\theta)^n \exp^{\theta\, n \sum_i t(y_i)/n}$$

$$\text{prior:} \quad p(\theta) \propto c(\theta)^{n_0} \exp(\theta\, n_0 t_0)$$

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:

$$\begin{aligned} n_0 &\quad\rightarrow\quad n_0 + n \\ t_0 &\quad\rightarrow\quad \frac{n_0 t_0 + n\overline{t(y)}}{n_0 + n} \\ \text{where } \overline{t(y)} &\quad= \frac{\sum_{i=1}^n t(y_i)}{n} \end{aligned}$$