An Introduction to Bayesian Methods for Survival Analysis

Sarah Lotspeich, Elizabeth Sigworth

Vanderbilt University

5 December 2017

Outline

"Be Bayesian or go home."
- Dr. Frank Harrell, 2017

The Bayesian Paradigm

A Refresher on Posterior Distributions

Begin with a probability model for your data given a set of unknown parameters $\boldsymbol{\theta}$, with corresponding likelihood function $L(\boldsymbol{\theta}|\text{data})$. Next, assume that $\boldsymbol{\theta}$ is random and has a prior distribution $\pi(\boldsymbol{\theta})$. The posterior distribution combining this prior distribution and observed data is obtained from Bayes' Theorem and has the form

$$\pi(\boldsymbol{\theta}|\text{data}) = \frac{L(\boldsymbol{\theta}|\text{data})\pi(\boldsymbol{\theta})}{\int_{\Theta} L(\boldsymbol{\theta}|\text{data})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

where Θ denotes the parameter space of $\boldsymbol{\theta}$. [1]

- Note: the denominator here is the marginal distribution of the data, which often does not have a closed form.
- Without this, it is difficult to analytically solve for $\pi(\boldsymbol{\theta}|\text{data})$.

[1] Ibrahim, et al. (2001)

Connecting the Prior and Posterior

We can see that $\pi(\boldsymbol{\theta}|\text{data})$ is proportional to the likelihood $L(\boldsymbol{\theta}|\text{data})$ multiplied by the prior $\pi(\boldsymbol{\theta})$

$$\pi(\boldsymbol{\theta}|\text{data}) \propto L(\boldsymbol{\theta}|\text{data})\pi(\boldsymbol{\theta})$$

The posterior distribution receives contributions from the data and the prior through the likelihood and prior distributions, respectively.

MCMC Posterior Sampling

The Problem with Marginal Distributions

Marginal distributions and posteriors often do not have closed forms. Methods to overcome this and sample from the posterior distribution without knowing the marginal include Gibbs sampler and other Markov chain Monte Carlo sampling algorithms.

- The idea in Gibbs sampling is to generate posterior samples by sweeping through each variable (or block of variables) to sample from its conditional distribution with the remaining variables fixed to their current values.[2]

[2] Yildirim (2012)

Gibbs Sampler

We have $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ be a p-dimensional vector of parameters, and $\pi(\boldsymbol{\theta}|\text{data})$ is the posterior distribution of $\boldsymbol{\theta}$ given our data.

Step 0. Choose an arbitrary starting point $\boldsymbol{\theta}_0 = (\theta_{1,0}, \theta_{2,0}, \ldots, \theta_{p,0})'$, and set $i = 0$.

Step 1. Generate $\boldsymbol{\theta}_{i+1} = (\theta_{1,i+1}, \theta_{2,i+1}, \ldots, \theta_{p,i+1})'$ as follows:

- Generate $\boldsymbol{\theta}_{1,i+1} \sim \pi(\theta_1|\theta_{2,i}, \ldots, \theta_{p,i}, \text{data})$
- Generate $\boldsymbol{\theta}_{2,i+1} \sim \pi(\theta_2|\theta_{1,i+1}, \theta_{3,i}, \ldots, \theta_{p,i}, \text{data})$
- $\vdots$
- Generate $\boldsymbol{\theta}_{p,i+1} \sim \pi(\theta_p|\theta_{1,i+1}, \theta_{2,i+1}, \ldots, \theta_{p-1,i+1}, \text{data})$

Step 2. Set $i = i + 1$, and go to Step 1. Repeat until distribution becomes apparent ($n > 1000$).[1]

[1] Ibrahim, et al. (2001)

Gibbs Sampler Example: Setup

Consider 20 subjects in an observational study. Denote the number dead due to natural causes as random variable $D \sim Bin(20, \theta)$. Suppose a previous study found that, on average, 40% of people die due to natural causes. Based on this, we impose a beta prior on θ, i.e. $\theta \sim Beta(a, b)$. We choose parameters a and b such that

$$E(\theta) = \frac{a}{a+b} = 0.40$$

This gives us $a = 4$ and $b = 6$. We want to be able to sample from the posterior distribution for $\theta|\text{data}$.

Gibbs Sampler Example: Posterior Distribution

For a beta-binomial model, we have the following conjugate posterior

$$
\begin{aligned}
\pi(\theta|D) &= \frac{f(d;\theta)\pi(\theta)}{f(d)} \\
&= \frac{\binom{20}{d}\theta^{d}(1-\theta)^{20-d}\frac{\Gamma(4+6)}{\Gamma(4)\Gamma(6)}\theta^{4-1}(1-\theta)^{6-1}}{\binom{20}{d}\frac{\Gamma(4+6)}{\Gamma(4)\Gamma(6)}\frac{\Gamma(4+d)\Gamma(6+20-d)}{\Gamma(4+6+20)}} \\
&= \frac{\Gamma(4+6+20)}{\Gamma(4+d)\Gamma(6+20-d)}\theta^{4+d-1}(1-\theta)^{26-d-1}
\end{aligned}
$$

Which we recognize as posterior distribution $\theta|D \sim Beta(4+d, 26-d)$.

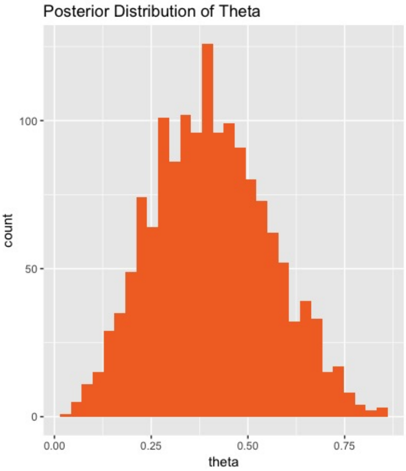
Gibbs Sampler Example: R Code[3]

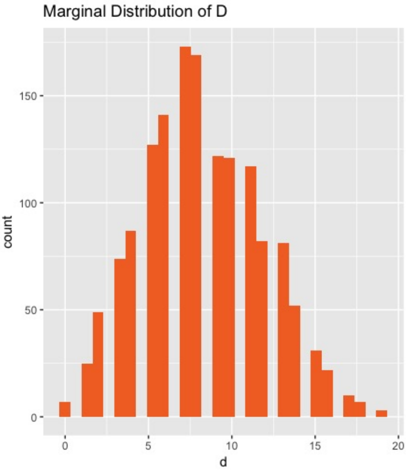
```
n <- 20; a <- 4; b <- 6 #setup parameters

it <- 1500 #number of iterations
d <- rep(NA,it); theta <- rep(NA,it)
d[1] <- 1; theta[1] <- 0.5 # set arbitrary initial values

# Perform Gibbs iterations
for (i in 2:it)
{
  d[i] <- rbinom(1,size=n,prob=th[i-1])
  theta[i] <- rbeta(1,a+d[i],b+n-d[i])
}
```

[3] Huerta (2012)


Posterior Distribution of Theta
count
100
50
0
0.00
0.25
0.50
0.75
theta


Marginal Distribution of D
count
150
100
50
0
0
5
10
15
20
d

Power Priors

We've Got the Power [Prior]

The power prior is defined to be the likelihood function based on historical data raised to a power a_0 where $0 \leq a_0 \leq 1$.

- Scalar parameter a_0 controls the influence of the historical data on the current data.

The power prior distribution of $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{\theta}|\text{data}_0, a_0) \propto L(\boldsymbol{\theta}|\text{data}_0)^{a_0}\pi_0(\boldsymbol{\theta}|c_0)$$

- data_0: historical data from a similar previous study
- a_0 : relative precision parameter for historical data, data_0
- c_0: a specified hyperparameter for the initial prior that controls the impact of $\pi_0(\boldsymbol{\theta}|c_0)$ on the entire prior $\pi(\boldsymbol{\theta}|\text{data})$

Controlling the Influence of Past Data

- When $a_0 = 1$, the power prior distribution corresponds to the posterior distribution of $\boldsymbol{\theta}$ from the historical study (no change for current study).
- When $a_0 = 0$, the power prior distribution does not depend on the historical data at all (i.e. no incorporation of historical data for current study).
- Thus, a_0 allows the investigator to select the influence of historical data on the current study.
  - Want a_0 closer to 1 when the studies are very similar.
  - Want a_0 closer to 0 when sample sizes are different or there is heterogeneity between studies.

Hierarchical Power Prior Specification

Incorporate a prior for a_0, $\pi(a_0|\boldsymbol{\gamma}_0)$, into the prior for $\boldsymbol{\theta}|\text{data}_0, a_0$ to get

$$\pi(\boldsymbol{\theta}, a_0|\text{data}_0) \propto L(\boldsymbol{\theta}|\text{data}_0)^{a_0}\pi_0(\boldsymbol{\theta}|c_0)\pi(a_0|\boldsymbol{\gamma}_0)$$

where $\boldsymbol{\gamma}_0$ is a specified vector of hyperparameters. A natural choice for $\pi(a_0|\boldsymbol{\gamma}_0)$ is a beta prior.

Bayesian Parametric Models

Exponential Model

Consider event times $X_i \overset{iid}{\sim} Expo(\lambda)$ for $i = 1, \ldots, n$.
Define:

- X_i: time on study for observation i
- δ_i: event indicator for observation i ($= 0$ if right censored, $= 1$ if event observed)
- T_i : time to event for observation i (if $\delta_i = 1$, $X_i = T_i$)
- C_i : time to right censoring for observation i (if $\delta_i = 0$, $X_i = C_i$)

Exponential Model: Likelihood

$$
\begin{aligned}
L(\lambda|\text{data}) &= \prod_{i=1}^{n} f(x_i|\lambda)^{\delta_i} S(x_i|\lambda)^{1-\delta_i} \\
&= \prod_{i=1}^{n} (\lambda \exp\{-\lambda x_i\})^{\delta_i} (\exp\{-\lambda x_i\})^{1-\delta_i} \\
&= (\lambda \exp\{-\lambda x_i\})^{\sum_{i=1}^{n} \delta_i} (\exp\{-\lambda x_i\})^{n-\sum_{i=1}^{n} \delta_i} \\
&= (\lambda^{\sum_{i=1}^{n} \delta_i} \exp\{-\sum_{i=1}^{n} \delta_i \lambda x_i\})(\exp\{-(n-\sum_{i=1}^{n} \delta_i)\lambda x_i\}) \\
&= \lambda^{d} \exp\{-\lambda \sum_{i=1}^{n} x_i\} \text{ where we let } d = \sum_{i=1}^{n} \delta_i
\end{aligned}
$$

Exponential Model: Prior Distribution

For exponential event times, the conjugate prior is the gamma distribution: $\lambda \sim Gamma(\alpha_0, \lambda_0)$ with density

$$\pi(\lambda|\alpha_0, \lambda_0) \propto \lambda^{\alpha_0 - 1} exp\{-\lambda_0 \cdot \lambda\}.$$

Exponential Model: Posterior Distribution

Based on this prior, we have posterior distribution for λ

$$
\begin{aligned}
\pi(\lambda|\text{data}) &\propto L(\lambda|\text{data})\pi(\lambda|\alpha_0, \lambda_0) \\
&\propto \left(\lambda^{\sum_{i=1}^{n}\delta_i} \exp\left\{-\lambda \sum_{i=1}^{n} x_i\right\}\right)\left(\lambda^{\alpha_0 - 1}\exp(-\lambda_0\lambda)\right) \\
&\propto \lambda^{\alpha_0 + d - 1}\exp\left\{-\lambda(\lambda_0 + \sum_{i=1}^{n} x_i)\right\}
\end{aligned}
$$

We recognize this as a $Gamma(\alpha_0 + d, \lambda_0 + \sum_{i=1}^{n} x_i)$ distribution.

Exponential Model: Posterior Mean and Variance

Based on the posterior $\lambda|\text{data} \sim Gamma(\alpha_0 + d, \lambda_0 + \sum_{i=1}^{n} x_i)$ we have the following estimates for the mean and variance of the parameter λ in $X_i \sim Expo(\lambda)$.

- $\text{E}(\lambda|\text{data}) = \frac{\alpha_0+d}{\lambda_0+\sum_{i=1}^{n} x_i}$
- $\text{Var}(\lambda|\text{data}) = \frac{\alpha_0+d}{(\lambda_0+\sum_{i=1}^{n} x_i)^2}$

Conclusion

So, Why Bayes?

- Historical data can be formally incorporated into the current analysis through a power prior.
- Model comparisons can be made between non-nested models, and without asymptotic assumptions, via Bayes Factors,

$$BF_{\text{model A,model B}} = \frac{P(\text{data}|\text{model A})}{P(\text{data}|\text{model B})}$$

- MCMC sampling techniques allow us to make exact inference for any sample size without relying on asymptotic approximations.

Questions?

References

1 Ibrahim, Joseph George, et al. Bayesian Survival Analysis. Springer, 2010.

2 Yildirim, Ilker. Bayesian Inference: Gibbs Sampling. 2012, Bayesian Inference: Gibbs Sampling.

3 Huerta, Gabriel. Some Examples on Gibbs Sampling and Metropolis-Hastings methods. 2012, Some Examples on Gibbs Sampling and Metropolis-Hastings methods.