

Coursera Notes for Bayesian Statistics: Techniques and Models

Jordan Katz

June 1, 2020

Week 1

Bayesian Modeling

data y , parameter θ

likelihood: $p(y|\theta)$

prior: $p(\theta)$

posterior: $p(\theta|y)$

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta} \propto p(y|\theta)p(\theta)$$

If we do not use conjugate priors, or if the models are more complicated, then the posterior distribution may not have a “standard” or well-known form.

Monte Carlo Estimation

Using simulation to determine some properties of a distribution, e.g. mean, variance, probability of an event, quantiles (which all use integration)

Example: Suppose we have $\theta \sim \text{Ga}(a, b)$ and want to know $E[\theta]$

$$E[\theta] = \int_{-\infty}^{\infty} \theta p(\theta) d\theta = \int_0^{\infty} \theta \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} d\theta = \frac{a}{b}$$

To verify with Monte Carlo, take samples θ_i^* for $i = 1, \dots, m$ from the Gamma distribution. Estimate sample mean as

$$\bar{\theta}^* = \frac{1}{m} \sum_{i=1}^m \theta_i^*$$

Suppose we have some function $h(\theta)$ and we want $E[h(\theta)]$. Can estimate

$$E[h(\theta)] = \int h(\theta)p(\theta)d\theta \approx \frac{1}{m} \sum_{i=1}^m h(\theta_i^*)$$

In particular, if $h(\theta)$ is $I_A(\theta)$, i.e. the indicator function for some event A , then we can approximate probabilities as well: $Pr[\theta \in A]$.

Question: How good is this estimate from sampling? By the Central Limit Theorem we know

$$\bar{\theta}^* \sim N\left(E(\theta), \frac{Var(\theta)}{m}\right)$$

The variance of the estimate is given by

$$\widehat{Var}(\theta) = \frac{1}{m} \sum_{i=1}^m (\theta_i^* - \bar{\theta}^*)^2$$

The standard error (SE) is given by

$$\sqrt{\frac{\widehat{Var}(\theta)}{m}}$$

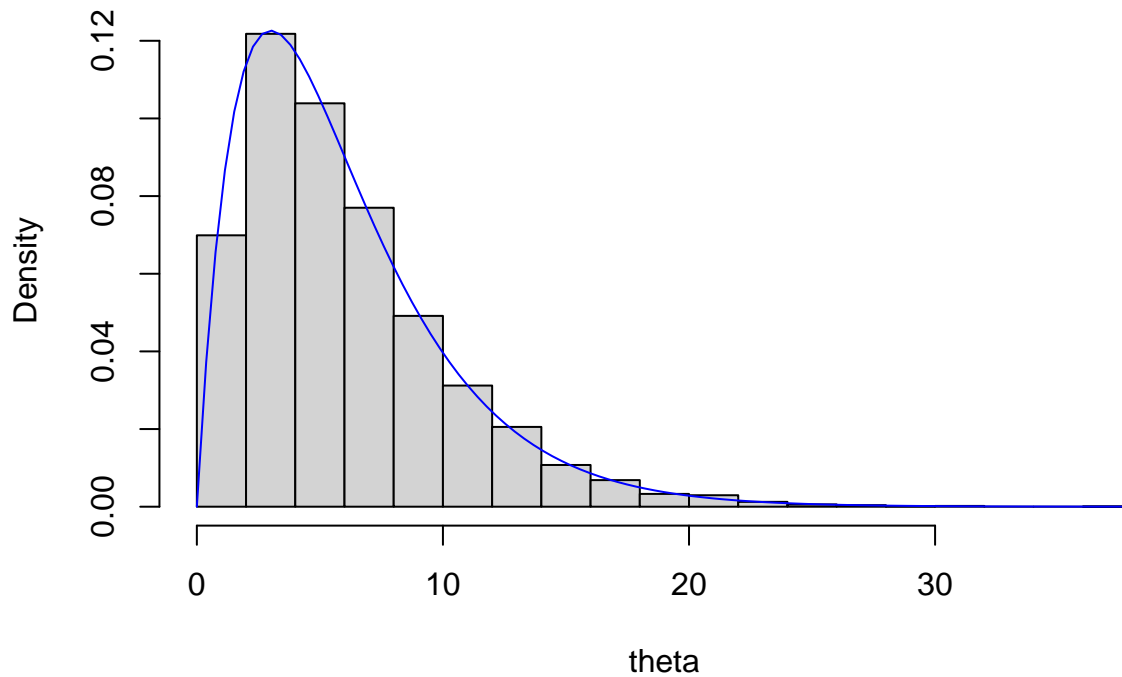
```
set.seed(32)

m=10000
a=2
b=1/3

theta = rgamma(n=m, shape=a, rate=b)

hist(theta, freq=FALSE)
curve(dgamma(x, shape=a, rate=b), col="blue", add=TRUE)
```

Histogram of theta



```
mean(theta) # Estimated mean
```

```
## [1] 6.022368
```

```
a/b # True mean
```

```
## [1] 6
```

```

var(theta) # Estimated variance

## [1] 18.01033
a/b^2 # True variance

## [1] 18
ind = theta < 5
mean(ind) # Estimated Prob[theta < 5]

## [1] 0.4974
pgamma(q=5, shape=a, rate=b) # True Prob[theta < 5]

## [1] 0.4963317
quantile(theta, probs=0.9) # Estimated quantile

##      90%
## 11.74426
qgamma(p=0.9, shape=a, rate=b) # True quantile

## [1] 11.66916
se = sd(theta) / sqrt(m) # Standard error of mean
mean(theta) - 2*se # Lower bound CI

## [1] 5.937491
mean(theta) + 2*se # Upper bound CI

## [1] 6.107245

```

As we can see, Monte Carlo does a pretty good job.

Example: Suppose we have

$$y|\phi \sim \text{Bin}(10, \phi)$$

$$\phi \sim \text{Beta}(2, 2)$$

and we want to simulate from marginal distribution of y (which can be difficult to do in general). Can do the following procedure:

1. Draw $\phi_i^* \sim \text{Beta}(2, 2)$
2. Given ϕ_i^* , draw $y_i^* \sim \text{Bin}(10, \phi_i^*)$

Results in a list of independent pairs (y_i^*, ϕ_i^*) drawn from the joint distribution. Discarding the ϕ_i^* s effectively results in a sample from the marginal distribution of y .

```

m = 1e5

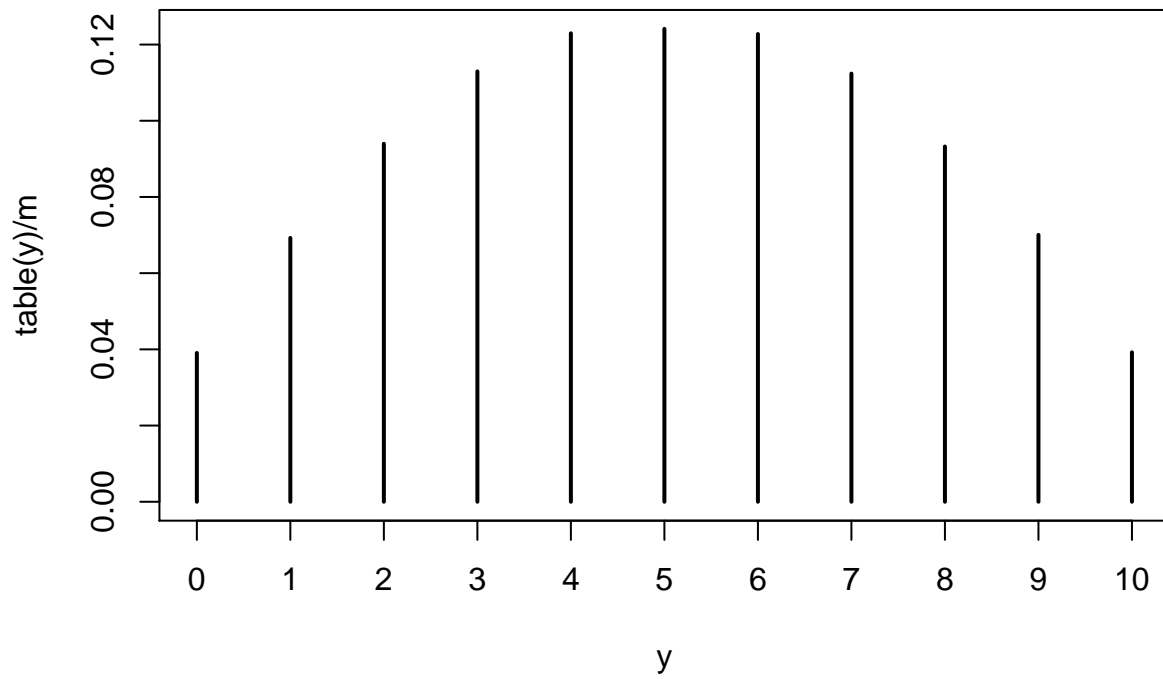
phi = rbeta(m, shape1=2, shape2=2)
y = rbinom(m, size=10, prob=phi)

table(y) / m

## y
##      0      1      2      3      4      5      6      7      8      9
## 0.03906 0.06925 0.09398 0.11296 0.12296 0.12412 0.12277 0.11238 0.09325 0.07005
##      10
## 0.03922

```

```
plot(table(y) / m) # Estimated marginal distribution of y
```



```
mean(y) # Estimate mean of y
```

```
## [1] 5.00046
```