Bayes Goes to Bat

Jo Hardin
Associate Professor of Mathematics
Pomona College
jo.hardin@pomona.edu

July 28, 2009

Acknowledgement

This activity is taken directly from John Spurrier's book *The Practice of Statistics: putting the pieces together*

Appropriate Courses

- stat theory
- introduction to statistics
- modeling course
- statistical computing

Student Background

- estimation (Bayesian & frequentist)
- bias, variance, mean squared error
- priors
- beta distribution

Learning Goals

1. For the students to get an idea of how priors are set.
2. To demonstrate how Bayesian and frequentist estimators can be compared.
3. To reinforce the ideas of bias and variability **of an estimator**.

Activity Background

- Goal: to estimate the true batting average (BA) in the major leagues for a high school player (that is, *parameter estimation*)
- Spurrier gives some basic info about BAs (best BA ever is Ty Cobb with 0.366, good BA is 0.3, barely adequate BA is 0.2)
- The baseball player gets 10 at bats. We can either use the frequentist estimate or we can incorporate our prior knowledge.

Frequentist Estimation

$$X \sim Bin(n = \text{number of at bats}, \theta = P(\text{getting a base hit}))$$

$$\theta_f = \frac{X}{n}$$

$$E[\theta_f] = \theta \qquad Var[\theta_f] = \frac{\theta(1-\theta)}{n}$$

Bayesian Estimation

$$X \sim Bin(n = \text{number of at bats}, \theta = P(\text{getting a base hit}))$$
$$\theta \sim beta(\alpha, \beta)$$

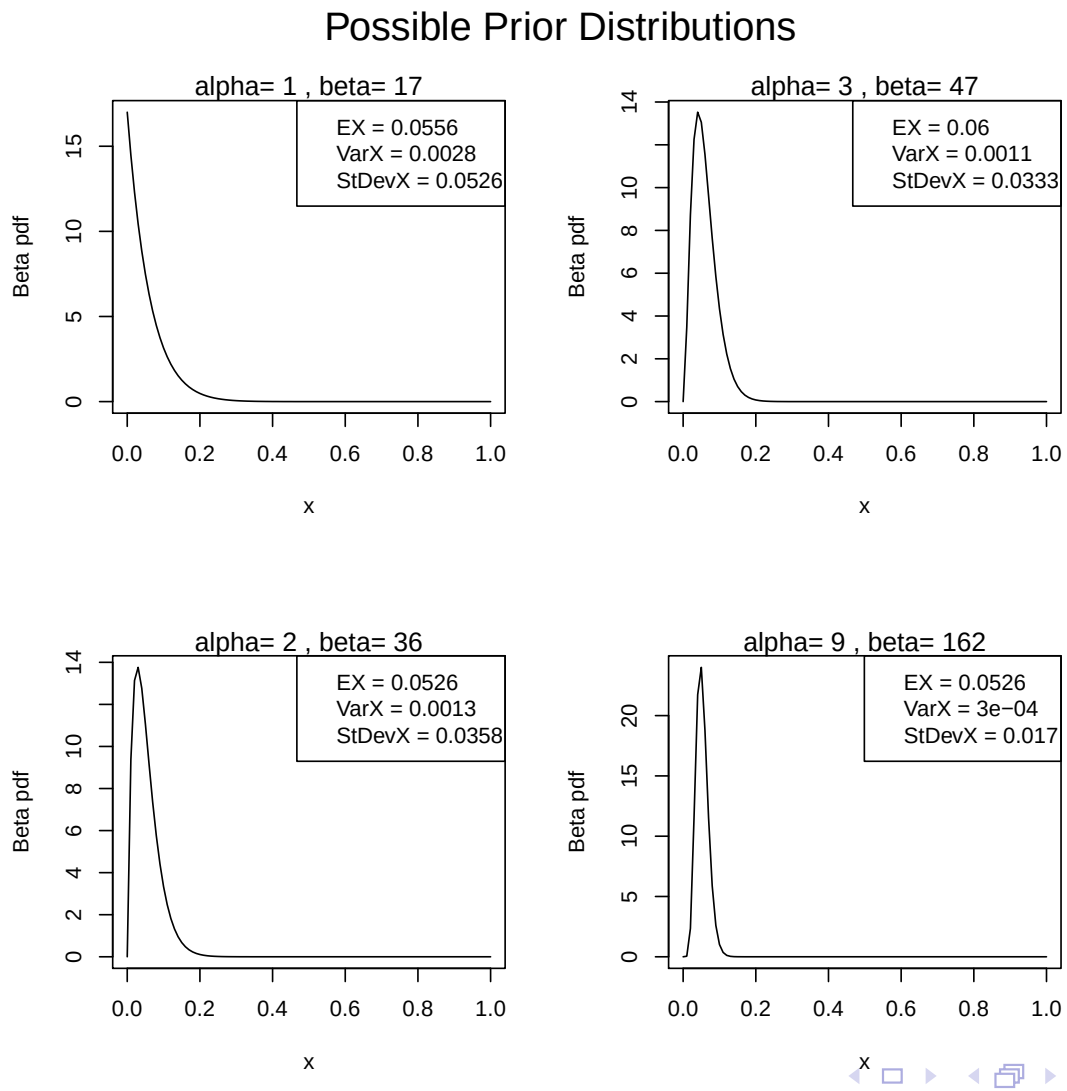
$$\theta_B = \frac{X+\alpha}{n+\alpha+\beta}$$

$$E[\theta_B] = \frac{n\theta+\alpha}{n+\alpha+\beta} \qquad Var[\theta_B] = \frac{n\theta(1-\theta)}{(n+\alpha+\beta)^2}$$

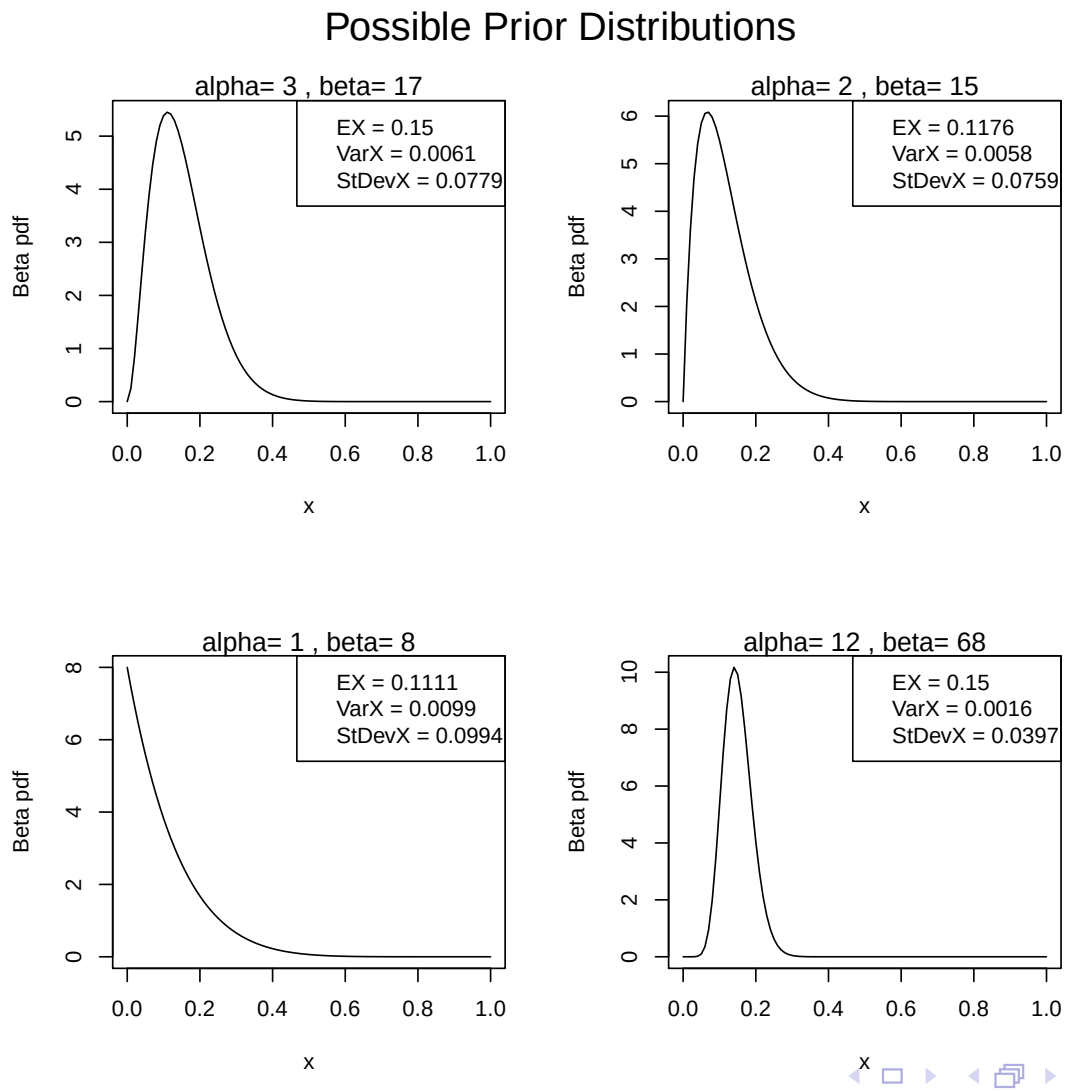
Personal Correspondence with John Spurrier:

- The Bayes estimator has an interesting interpretation.
- It is the usual batting average (sample proportion) with
  - $\alpha + \beta$ imaginary extra at bats
  - yielding α imaginary hits added to the data.
- $$\theta_B = \frac{X + \alpha}{n + \alpha + \beta}$$
- I sense that many major league managers are Bayesians without knowing what that means.

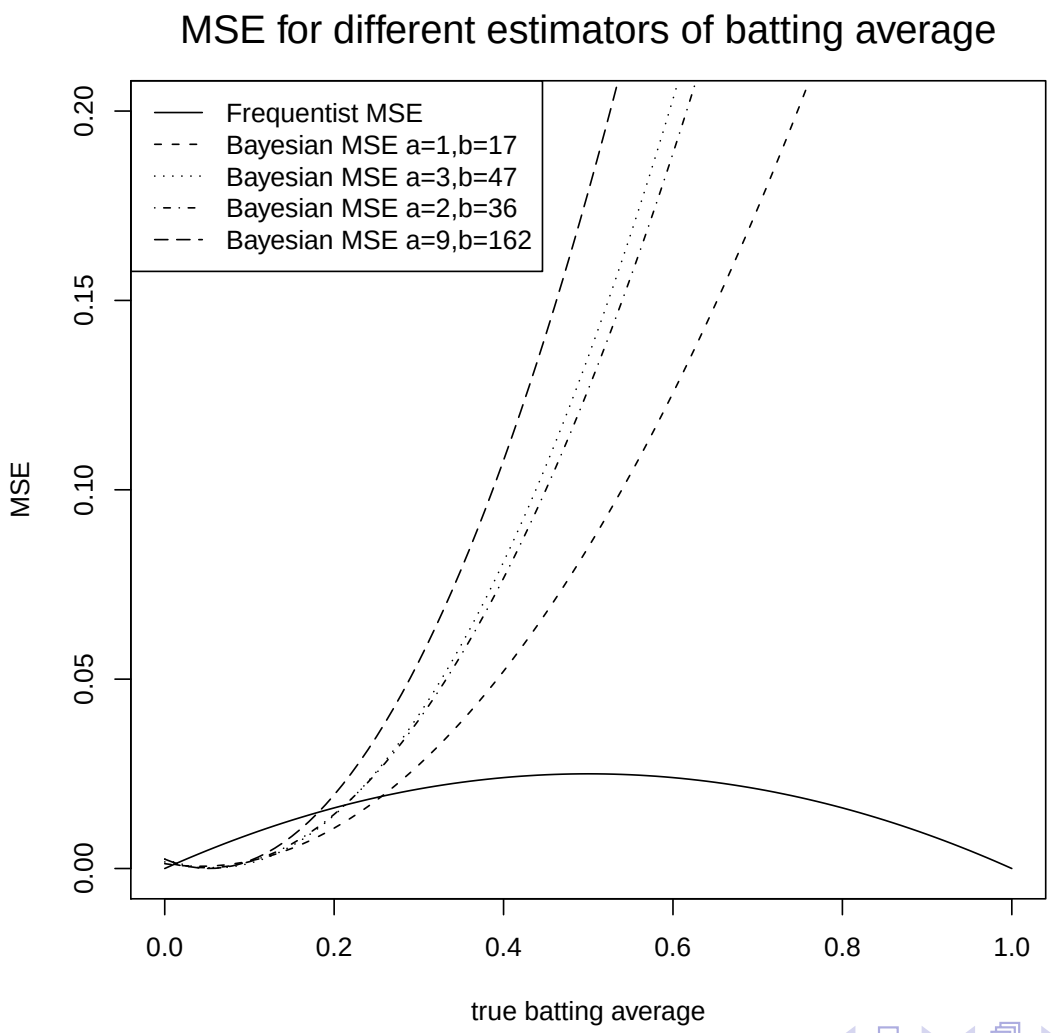
Priors


Possible Prior Distributions
alpha= 1 , beta= 17
EX = 0.0556
VarX = 0.0028
StDevX = 0.0526
alpha= 3 , beta= 47
EX = 0.06
VarX = 0.0011
StDevX = 0.0333
alpha= 2 , beta= 36
EX = 0.0526
VarX = 0.0013
StDevX = 0.0358
alpha= 9 , beta= 162
EX = 0.0526
VarX = 3e−04
StDevX = 0.017
Beta pdf
x

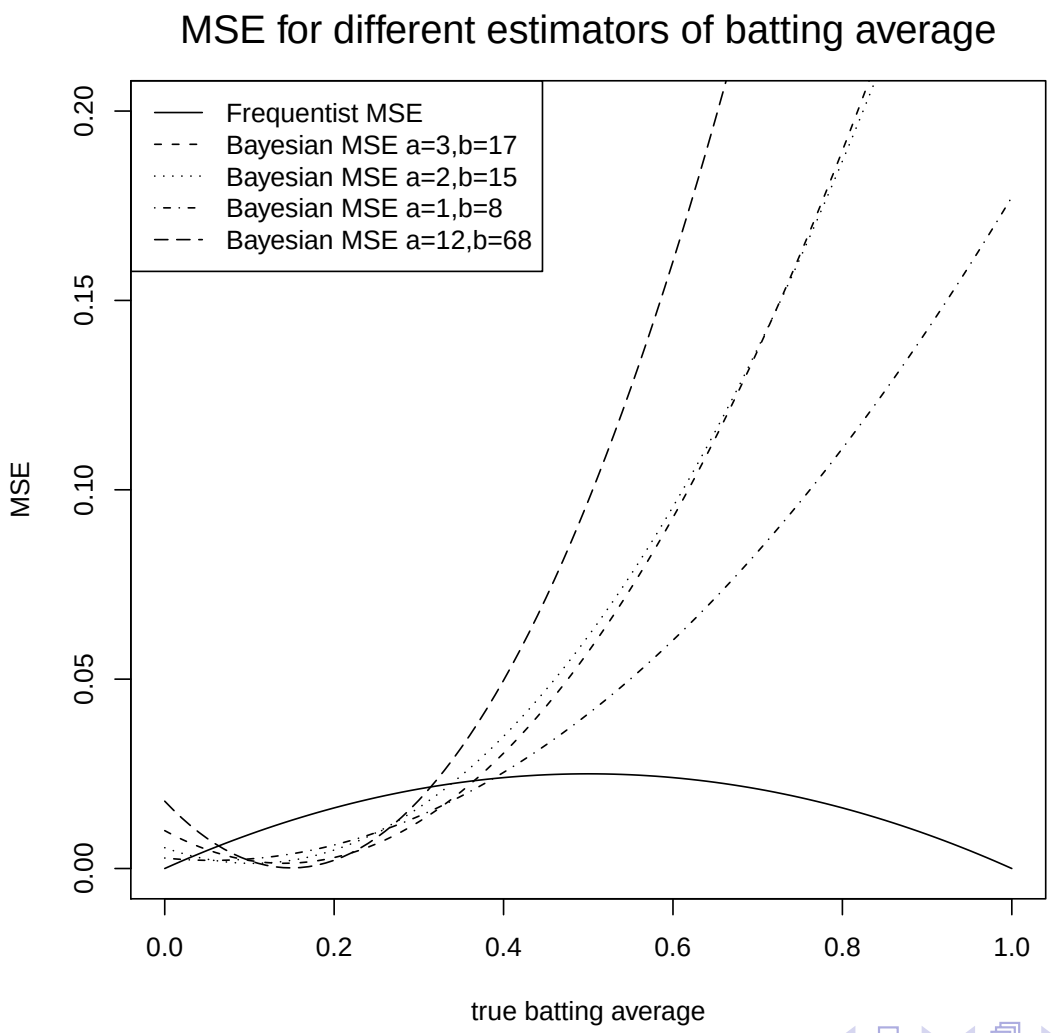
Priors


Possible Prior Distributions
alpha= 3 , beta= 17
EX = 0.15
VarX = 0.0061
StDevX = 0.0779
alpha= 2 , beta= 15
EX = 0.1176
VarX = 0.0058
StDevX = 0.0759
alpha= 1 , beta= 8
EX = 0.1111
VarX = 0.0099
StDevX = 0.0994
alpha= 12 , beta= 68
EX = 0.15
VarX = 0.0016
StDevX = 0.0397
Beta pdf
x

Evaluating


MSE for different estimators of batting average
Frequentist MSE
Bayesian MSE a=1,b=17
Bayesian MSE a=3,b=47
Bayesian MSE a=2,b=36
Bayesian MSE a=9,b=162
MSE
0.00
0.05
0.10
0.15
0.20
0.0
0.2
0.4
0.6
0.8
1.0
true batting average

Evaluating


MSE for different estimators of batting average
Frequentist MSE
Bayesian MSE a=3,b=17
Bayesian MSE a=2,b=15
Bayesian MSE a=1,b=8
Bayesian MSE a=12,b=68
MSE
0.00
0.05
0.10
0.15
0.20
0.0
0.2
0.4
0.6
0.8
1.0
true batting average

Theoretical Extensions

- Deriving expected value, variance, MSE
- MLE of θ, MOM of θ
- Minimum Variance Unbiased Estimator (θ_f)
- θ_B is a weighted average of the prior expected value and θ_f

Computational Aspects

- Figuring out the prior on their own
- Computing MSE as a function of different values (true θ, α, β, $n =$ sample size)

Things that go well

- The students see the process of creating a prior (no more black box)
- You can continue to refer to the example when discussing
  - different distributions / priors
  - Bayesian hypothesis testing
  - Bayesian posterior intervals
- The students love this activity!

Difficulties

- MSE is sort of apples and oranges
- Hard to find time to teach both theory and computation
- Not all students are familiar with baseball

Thanks

- John Spurrier – great book!
- CAUSE for this webinar