PS 919: Problem Set 1

Blain Morin

10/6/2020

1. It's only three days into the Major League Blernsball season, but fans are already excited about rookie sensation Casey Lee of the Boston Poindexters. Lee hit a grand slam blern in the first game of the Poindexters' season, none in the second game, and three in the third game. Statistician Jack Johnson sets out to estimate the rate at which Lee hits grand slam blerns. Johnson assumes that the number of gland slam blerns that Johnson hits in a game follows a Poisson distribution with unknown rate parameter λ and that these observations are exchangeable. Since no one had seen Lee play before the start of the season, Johnson uses a flat prior on the rate parameter. Using the results of Lee's first three games, estimate this model in Stan with a flat prior on λ. Please run the model with 2 chains and at least 10,000 iterations per chain (e.g., with the options chains=2, iter=10000).

Here is my .stan model:

```
### model1.stan

data {
  int<lower=0> N;
  int x[N];
}
parameters {
  real<lower=0> lambda;
}
model {
  x ~ poisson(lambda);
}
```

Here is the r code used to run the model:

```
grand.slams = c(1, 0, 3)

standata = list(x = grand.slams, N = length(grand.slams))

model1 = stan(file = 'model1.stan', data = standata, chains = 2, iter = 10000,
              seed = 1234) ### Set seed for replication
```

a. Give an estimate of the rate at which Lee hits grand slam blerns using the Bayes estimator of λ under quadratic loss.

Here is a summary of the posterior distribution of λ:

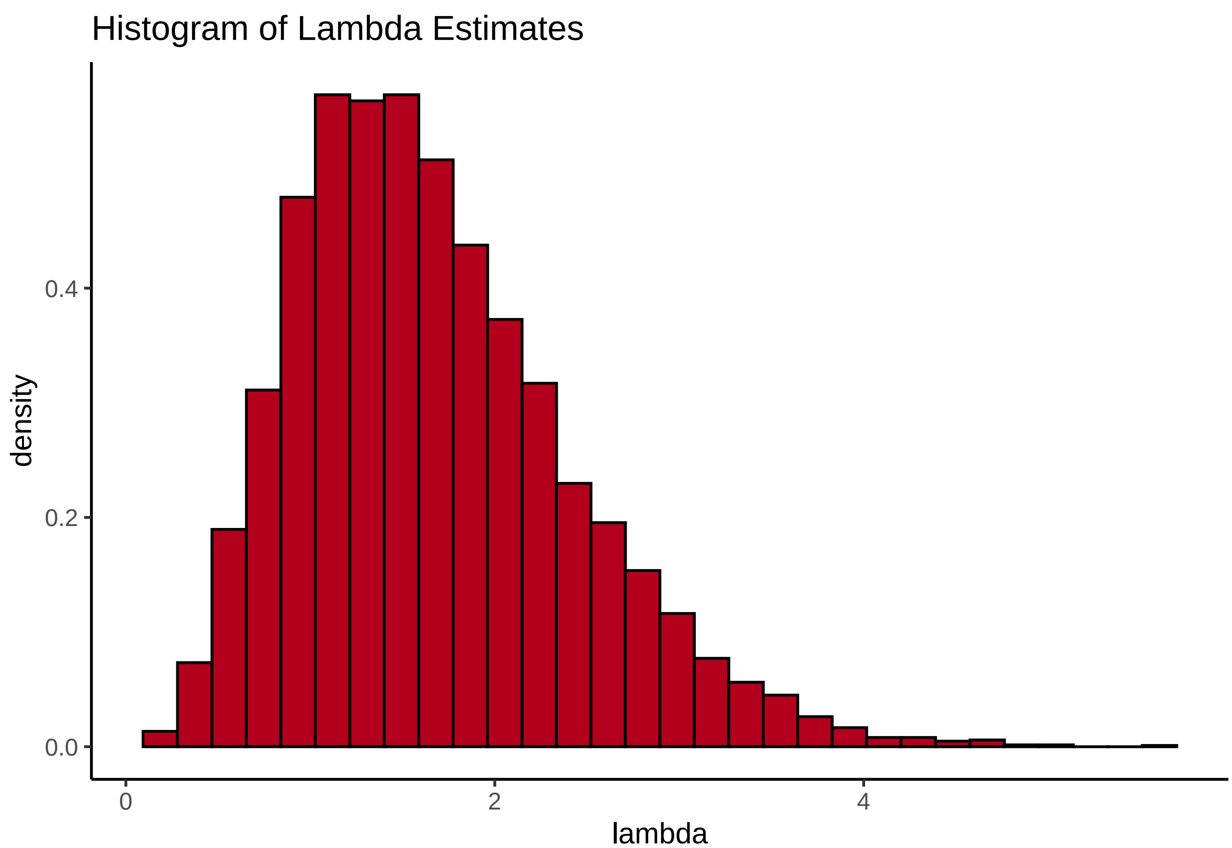
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
lambda	1.669	0.013	0.745	0.524	1.115	1.556	2.106	3.401	3301.101	1
lp___	-2.969	0.013	0.735	-5.049	-3.133	-2.693	-2.499	-2.446	3292.631	1

The Bayes estimator for λ under quadratic loss is the mean, which we see here = 1.669.

b. Give a 95% credible interval for λ.

Using the summary output above, the 95% interval for λ is $(.524, 3.401)$.

c. Plot the posterior distribution of λ (as approximated by a histogram or kernel density plot of the samples).



d. The Poindexters are hoping that Lee is that rare blernsball player that averages a rate of one grand slam blerns per game or more. Give the posterior probability that this is the case (i.e., $\lambda \geq 1$).

```
### Extract all lambda estimates
lambdas = extract(model1)$lambda

### Probability that lambda is > 1 is
### the count of lambda estimates > 1 divided by the number of estimates
sum(lambdas > 1) / length(lambdas)
```

```
## [1] 0.8177
```

The probability that λ is greater than or equal to 1 is .818.

2. Johnson's bitter rival, John Jackson, criticizes Johnson's choice of prior. While Jackson agrees with Johnson's model, Jackson argues that the appropriate prior on λ should be based on the distribution of the grand-slam-blerns rate exhibited by other blernsball players. This distribution is similar to a Weibull distribution with a shape parameter of 1.6 and a scale parameter of 0.3. Estimate this same model in Stan using Jackson's Weibull in place of the Johnson's uniform prior.

Here is the .stan code for the new model:

```
### model2.stan

data {
  int<lower=0> N;
  int x[N];
}
parameters {
  real<lower=0> lambda;
}
model {
  x ~ poisson(lambda);
  lambda ~ weibull(1.6, .3);
}
```

Here is the r code used to run the new model:

```
grand.slams = c(1, 0, 3)

standata = list(x = grand.slams, N = length(grand.slams))

model2 = stan(file = 'model2.stan', data = standata, chains = 2, iter = 10000,
              seed = 1234) ### Set seed for replication
```

a. Give an estimate of the rate at which Lee hits grand slam blerns using the Bayes estimator of λ under quadratic loss.

Here is a summary of the posterior distribution of λ for the new model:

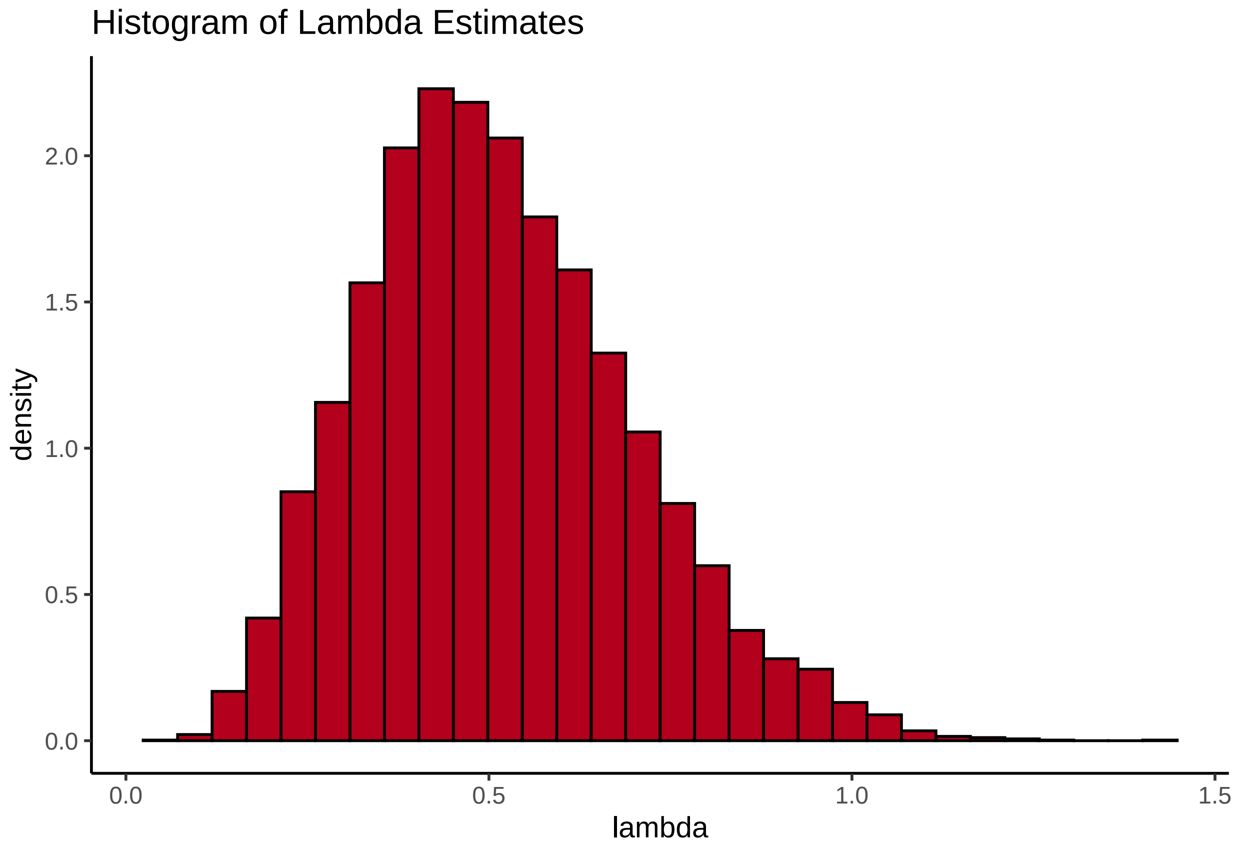
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
lambda	0.515	0.003	0.184	0.207	0.382	0.496	0.631	0.926	3487.950	1.001
lp__	-8.163	0.012	0.733	-10.259	-8.329	-7.874	-7.688	-7.632	3663.102	1.000

The Bayes estimator for λ under quadratic loss is the mean, which we see here = .515.

b. Give a 95% credible interval for λ.

Using the summary output above, the 95% interval for λ is (.207, .926).

c. Plot the posterior distribution of λ (as approximated by a histogram or kernel density plot of the samples).



d. The Poindexters are hoping that Lee is that rare blernsball player that averages a rate of one grand slam blerns per game or more. Give the posterior probability that this is the case (i.e., $\lambda \geq 1$).

```
### Extract all lambda estimates
lambdas = extract(model2)$lambda

### Probability that lambda is > 1 is
### the count of lambda estimates > 1 divided by the number of estimates
sum(lambdas > 1) / length(lambdas)
```

```
## [1] 0.0102
```

The probability that λ is greater than or equal to 1 is .01.