Lab 4

PSTAT 115

Objectives

- Posterior credible intervals
- Posterior predictive distribution
- Integral trick

Computing probability intervals with quantile functions

In addition to point summaries, it is nearly always important to report posterior uncentainty. Therefore, as in conventional statistics, an interval summary is desirable. A central interval of posterior probability, which corresponds, in the case of a $100(1-\alpha)\%$ interval, to the range of values above and below which lies exactly $100(\alpha/2)\%$ of the posterior probability.

Example from lab 3:

$$p(\theta|y) \propto p(\theta) * p(y|\theta) = \binom{n}{y} p^y (1-p)^{n-y} \propto p^y (1-p)^{n-y}$$

An early study concerning the sex of newborn Germany babies found that of a total of 98 births, 43 were female. Assume we are using the uniform prior.The posterior is a $Beta(44, 56)$ distribution.

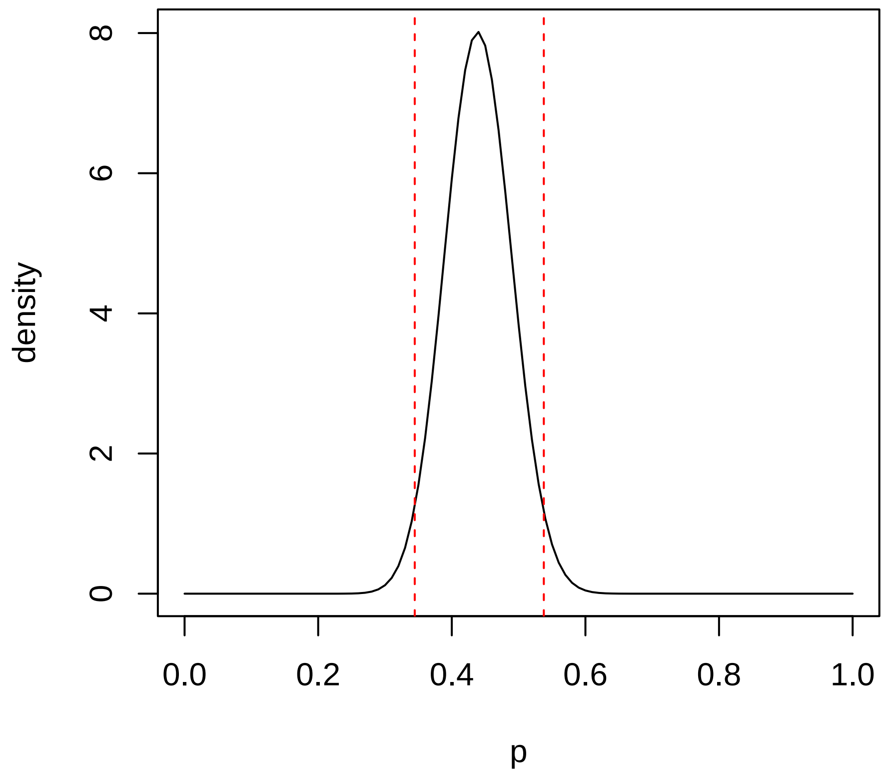
- What is the 95% central interval of the above posterior distribution?

```
a_post = 1 + 43
b_post = 1 + 98 - 43
alpha = 1 - 0.95
low = qbeta(alpha/2, a_post, b_post)
high = qbeta(1 - alpha/2, a_post, b_post)
print(c(low, high))
```

```
## [1] 0.3445430 0.5377312
```

- Visualize the above central interval

```
curve(gamma(a_post + b_post)/gamma(a_post)/gamma(b_post) *
      p^(a_post - 1) * (1-p)^(b_post - 1), from = 0, to = 1, xname = "p",
      xlab = "p", ylab = "density")
abline(v = low, col = "red", lty = 2)
abline(v = high, col = "red", lty = 2)
```



Posterior predictive distribution

- An important feature of Bayesian inference is the existence of a predictive distribution for new observations.
  - Let $\tilde{y}$ be a new (unseen) observation, and $y_1, ...y_n$ the observed data.
  - The Posterior predictive distribution is $p(\tilde{y} \mid y_1, ...y_n)$
- The predictive distribution does not depend on unknown parameters
- The predictive distribution only depends on observed data

The posterior predictive distribution allows us to find the probability distribution for new data given observations of old data.

$$p(\tilde{y} \mid y_1, ...y_n) = \int p(\tilde{y}, \theta \mid y_1, ...y_n) d\theta = \int p(\tilde{y} \mid \theta) p(\theta \mid y_1, ...y_n) d\theta$$

- The prior predictive distribution describes our uncertainty about a new observation before seeing data
- It incorporates uncertainty due to the sampling in a model $p(\tilde{y} \mid \theta)$ *and* our prior uncertainty about the data generating parameter, $p(\theta)$

Example

- $\lambda \sim \text{Gamma}(\alpha, \beta)$
- $\tilde{Y} \sim \text{Pois}(\lambda)$

$$p(\tilde{y}) = \int p(\tilde{y} \mid \lambda)p(\lambda)d\lambda = \int (\frac{\lambda^{\tilde{y}}}{y!}e^{-\lambda})(\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{(\alpha-1)}e^{-\beta\lambda})d\lambda = \frac{\beta^\alpha}{\Gamma(\alpha)y!}\int (\lambda^{(\alpha+y-1)}e^{-(\beta+1)\lambda})d\lambda$$

$\int(\lambda^{(\alpha+y-1)}e^{-(\beta+1)\lambda})d\lambda$ looks like an unormalized $\text{Gamma}(\alpha+y, \beta+1)$

Integral trick (Gamma integral example)

Let $K = \int L(\lambda; y)p(\lambda)d\lambda$ be the integral of the proportional posterior. Then the proper posterior density, i.e. a true density integrates to 1, can be expressed as $p(\lambda \mid y) = \frac{L(\lambda;y)p(\lambda)}{K}$. Compute this posterior density and clearly express the density as a mixture of two gamma distributions.

$$\begin{aligned}
K &= \int e^{-1767\lambda}\lambda^8(\frac{2000^3}{\Gamma(3)}\lambda^2 e^{-2000\lambda} + \frac{1000^7}{\Gamma(7)}\lambda^6 e^{-1000\lambda})d\lambda \\
&= \int \frac{2000^3}{\Gamma(3)}\lambda^{10}e^{-3767\lambda}d\lambda + \int \frac{1000^7}{\Gamma(7)}\lambda^{14}e^{-2767\lambda}d\lambda \\
&= \frac{2000^3}{\Gamma(3)}\frac{\Gamma(11)}{3767^{11}} + \frac{1000^7}{\Gamma(7)}\frac{\Gamma(15)}{2767^{15}}
\end{aligned}$$

$$\begin{aligned}
p(\lambda|y) &= \frac{\frac{2000^3}{\Gamma(3)}\frac{\Gamma(11)}{3767^{11}}}{\frac{2000^3}{\Gamma(3)}\frac{\Gamma(11)}{3767^{11}} + \frac{1000^7}{\Gamma(7)}\frac{\Gamma(15)}{2767^{15}}} * \frac{3767^{11}}{\Gamma(11)}\lambda^{10}e^{-3767\lambda} + \frac{\frac{1000^7}{\Gamma(7)}\frac{\Gamma(15)}{2767^{15}}}{\frac{2000^3}{\Gamma(3)}\frac{\Gamma(11)}{3767^{11}} + \frac{1000^7}{\Gamma(7)}\frac{\Gamma(15)}{2767^{15}}} * \frac{2767^{15}}{\Gamma(15)}\lambda^{14}e^{-2767\lambda} \\
&:= wp_U(\lambda) + (1-w)p_V(\lambda)
\end{aligned}$$

where

$$w = \frac{\frac{2000^3}{\Gamma(3)}\frac{\Gamma(11)}{3767^{11}}}{\frac{2000^3}{\Gamma(3)}\frac{\Gamma(11)}{3767^{11}} + \frac{1000^7}{\Gamma(7)}\frac{\Gamma(15)}{2767^{15}}}, U \sim Gamma(11, \frac{1}{3767}), V \sim Gamma(15, \frac{1}{2767})$$

which means that the posterior density is a mixture of two gamma distributions.

Posterior Predictive Checking

The "hot hand" is the purported phenomenon that a person who experiences a successful outcome has a greater chance of success in further attempts. The concept is originates from basketball whereas a shooter is allegedly more likely to score if their previous attempts were successful. While previous success at a task can indeed change the psychological attitude and subsequent success rate of a player, researchers for many years did not find evidence for a "hot hand" in practice, dismissing it as fallacious. However, later research questioned whether the belief is indeed a fallacy.

Let "1" denotes a valid shot and "0" denotes a invalid. Suppose we observe the following results of a player:

```r
# observations #
set.seed(123)
y <- c(rep(1, 18), rep(0, 3), rep(1, 6), rep(0, 2),
       rbinom(67, 1, prob = 0.25), rep(1, 4))
y
```

```
##   [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 0 1 0 1 1 0 0 1
##  [38] 0 0 1 0 0 0 0 1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 1 1 0 1 0 0 1 0 0 0 0 0 0 0 0
##  [75] 0 0 0 0 1 0 0 1 0 0 0 0 1 1 0 0 0 0 0 1 0 1 1 1 1 1
```

Suppose $Y_i \sim Bernoulli(p)$ and $p \sim Beta(3, 7)$

Find the posterior using conjugacy:

```r
# prior #
a <- 3
b <- 7
#posterior #
a_post <- a + sum(y)
b_post <- b + (length(y) - sum(y))
a_post; b_post
```

```
## [1] 50
```

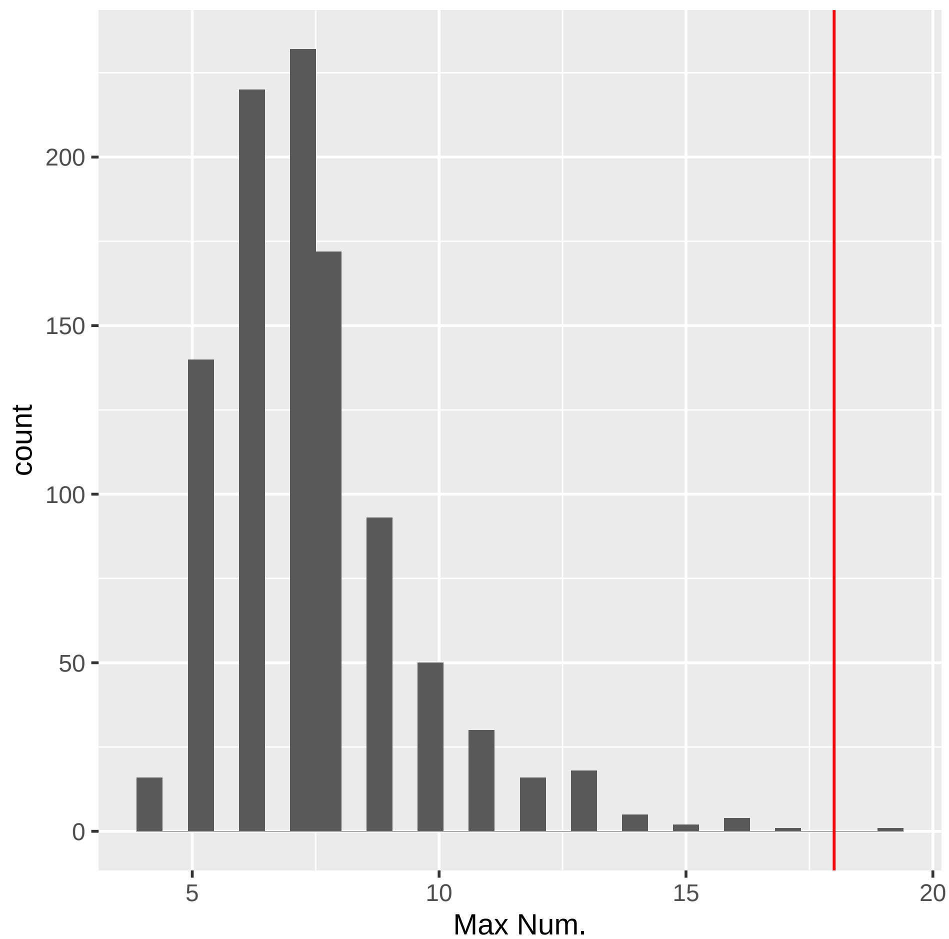
```
## [1] 60
```

Let the test stat. be the maximum number of the same consecutive results.

```r
# observed test stat. #
test_stat_obs <- max(rle(y)$lengths)

# test stat. based on simulation #
nsim <- 1000
test_stat_rep <- rep(NA, nsim)
for (i in 1:1000) {
  p_post <- rbeta(1, a_post, b_post)
  y_rep <- rbinom(100, size = 1, prob = p_post)
  test_stat <- max(rle(y_rep)$lengths)
  test_stat_rep[i] <- test_stat
}

ggplot(tibble(test_stat_rep), aes(test_stat_rep)) +
  geom_histogram() + xlab("Max Num.") +
  geom_vline(xintercept = test_stat_obs, colour = "red")
```



200
150
100
50
0
count
5
10
15
20
Max Num.