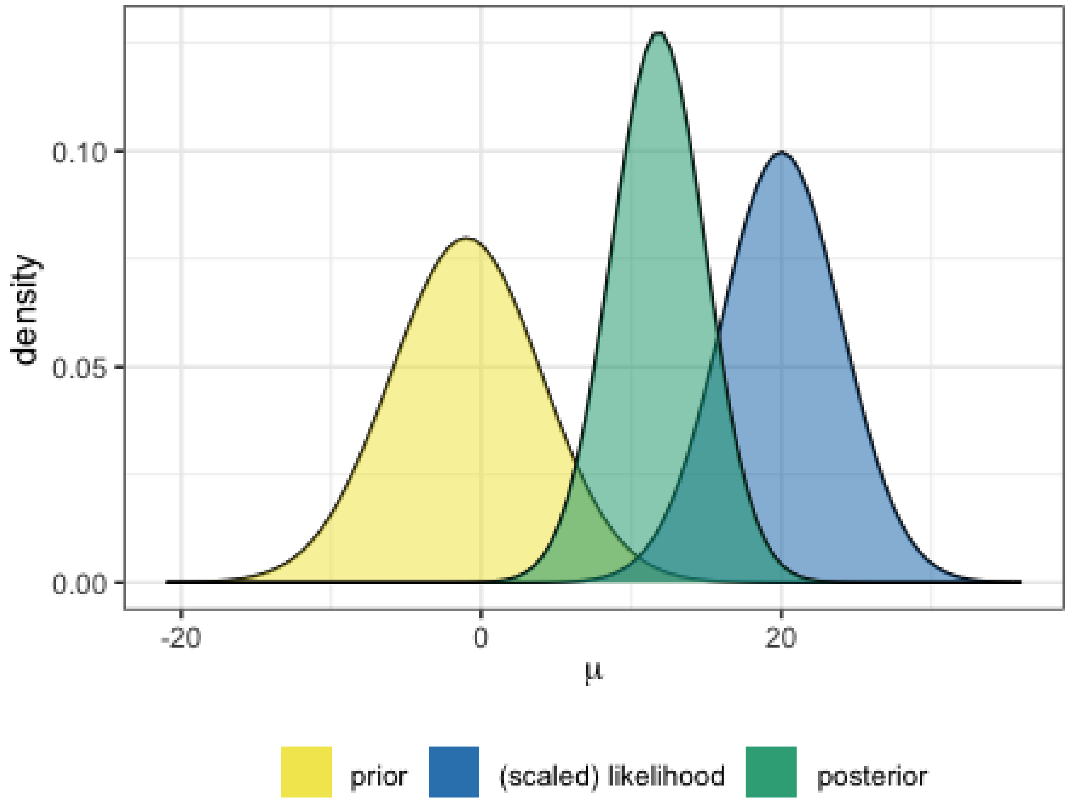
Using a continuous prior distribution

Stat 340: Bayesian Statistics



1. Continuous prior

2. Posterior analysis

3. Prediction

(Problem topics 1-4)

Blindsight design, redux

Data: N N N N **B B** N N N **B** N N N N N N N (14 Ns; **3 Bs**)

Data model (likelihood):

Some true proportion of guesses, p

Toss a coin with probability of heads, p

Belief about p:

Uniform over (0, 1)

Beta distribution

- $f(x|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$
- Parameter space: $a > 0, b > 0$
- Support: $0 < x < 1$

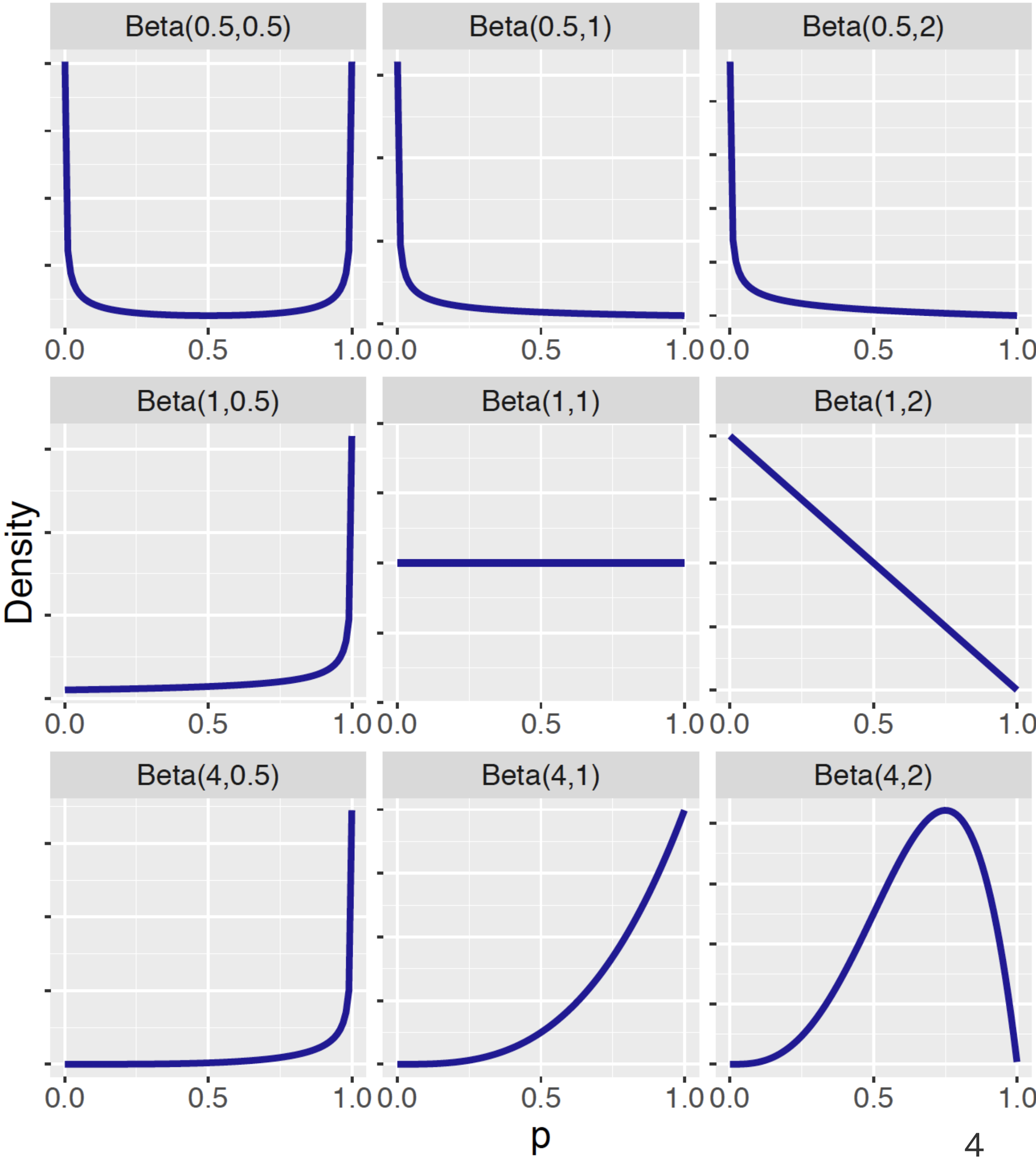
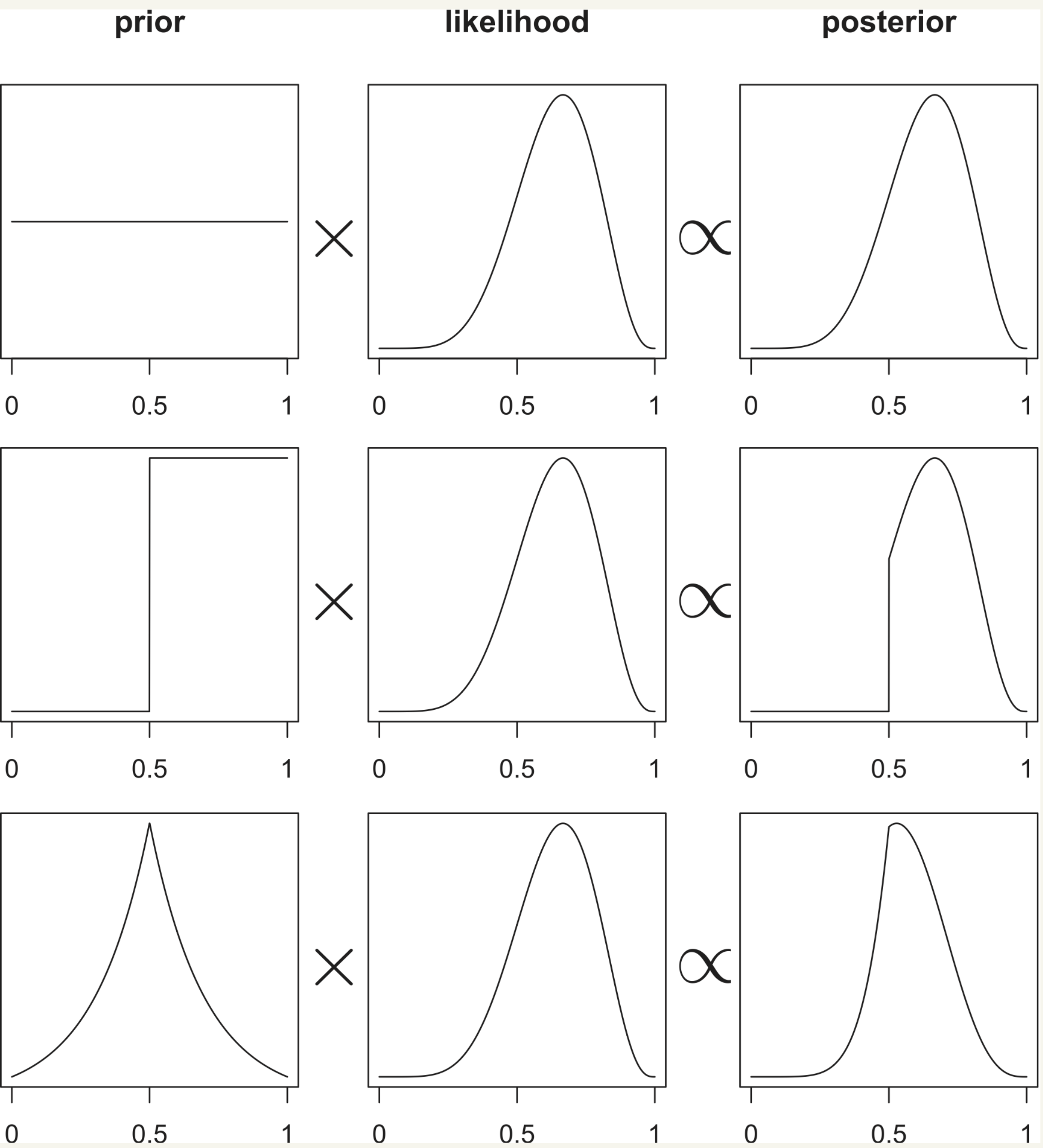


Image credit: Probability and Bayesian Modeling



"The prior is proportional to the prior times the likelihood"

Your turn 1

- Work with your neighbors
- Work through the R code to simulate kernels of the beta distribution
- You can copy/paste the code from the course webpage
- Develop your understanding of the kernel of a distribution

05:00

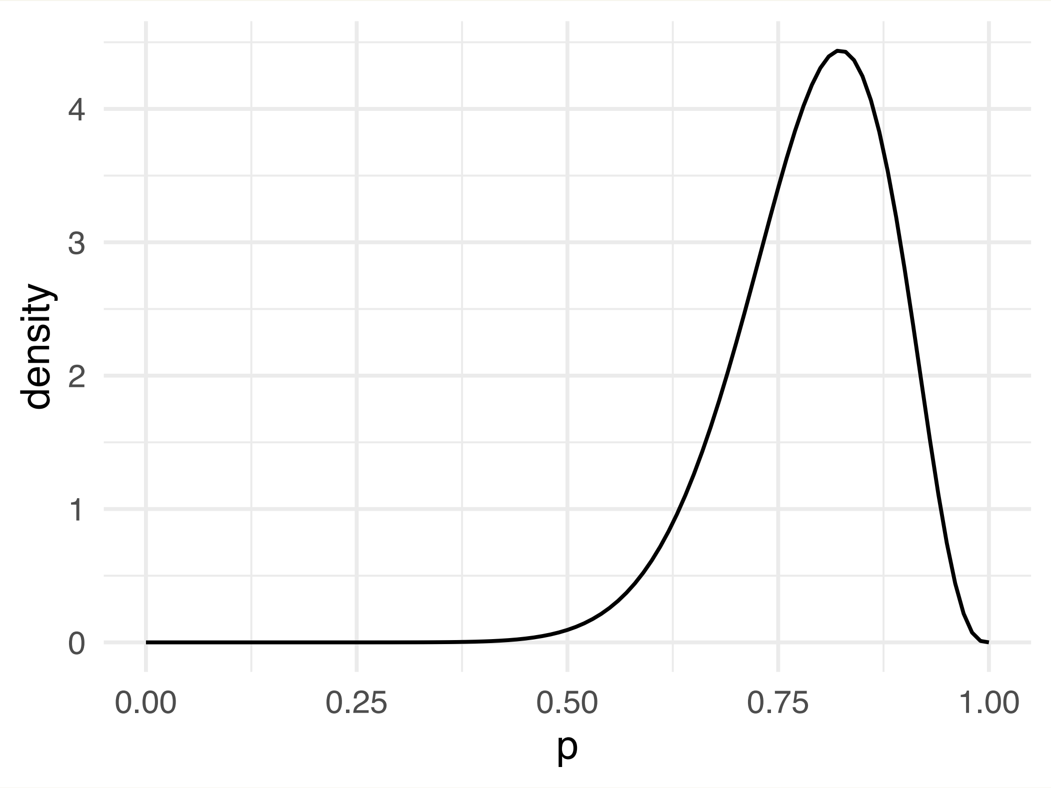
Your turn 2

- Work with your neighbors
- Derive the posterior
- Are you working with a conjugate family?

05:00

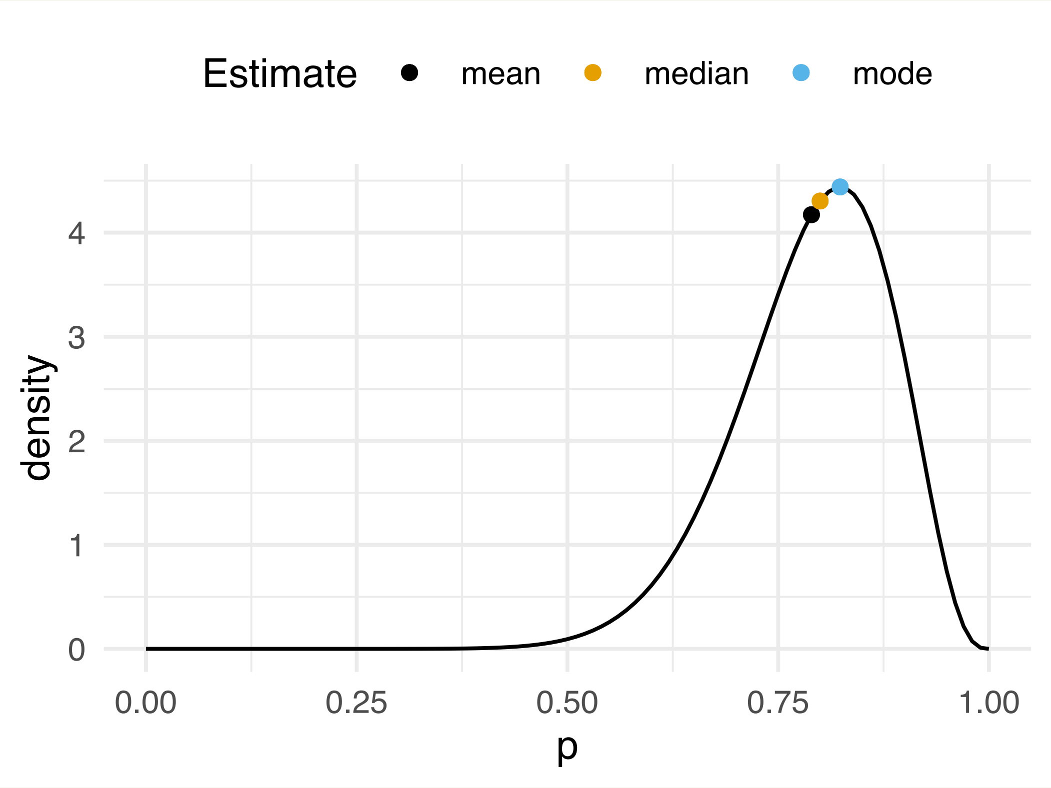
Posterior analysis

> *To a Bayesian, the best information one can ever have about θ is to know the posterior density.*
>
> *— Christensen, et al; Bayesian Ideas and Data Analysis, p. 31*

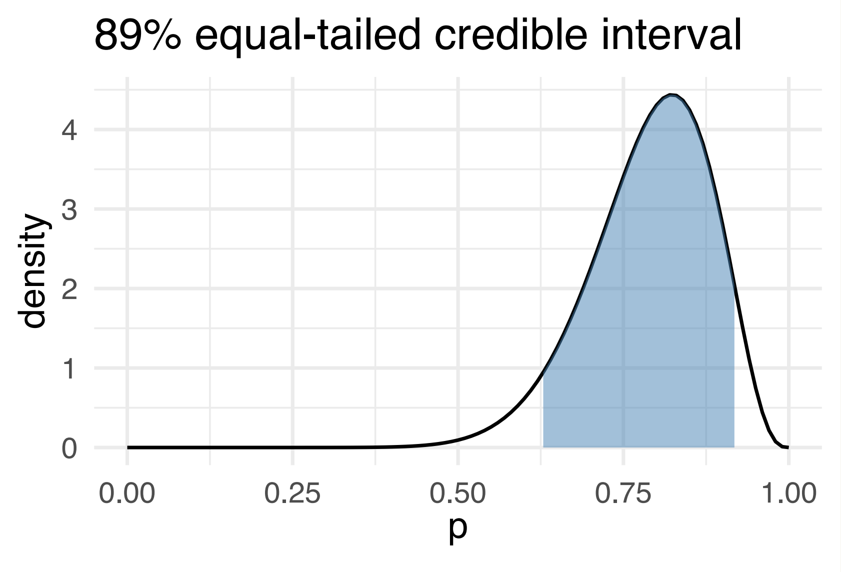


Point estimates

- **Posterior mean**
- **Posterior median**
- **Posterior mode**
  i.e. *maximum a posteriori* (MAP) estimate



Credible intervals

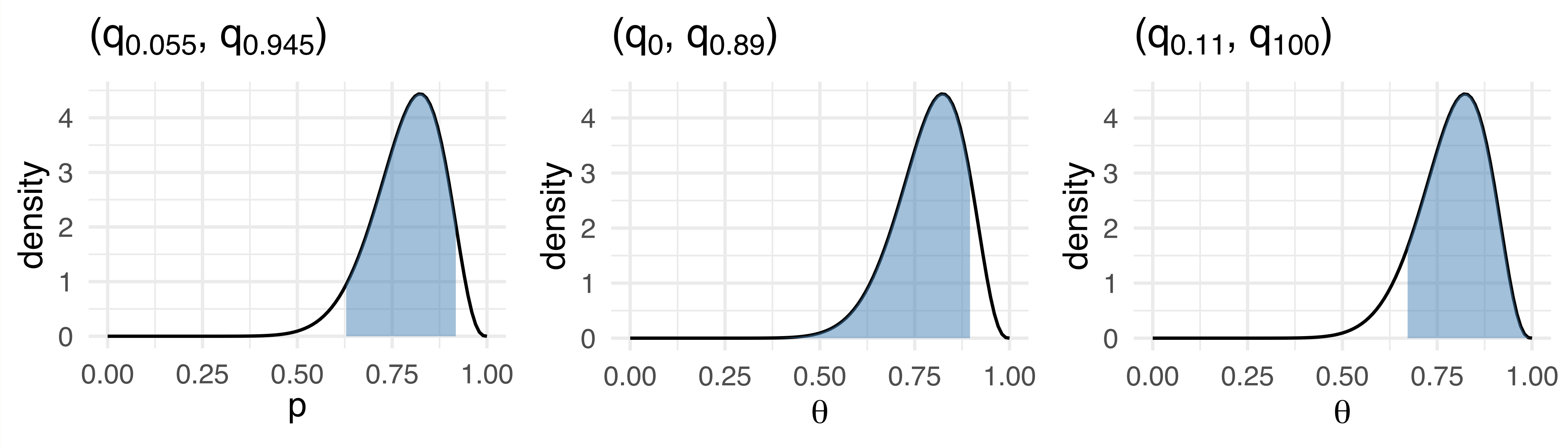



```r
# q*() functions calculate quantiles from the specified distribution
c(lower = qbeta(0.055, 15, 4), upper = qbeta(1 - 0.055, 15, 4))
```

```
##     lower     upper
## 0.6288166 0.9177940
```

Credible intervals are not unique

Here are three 89% credible intervals

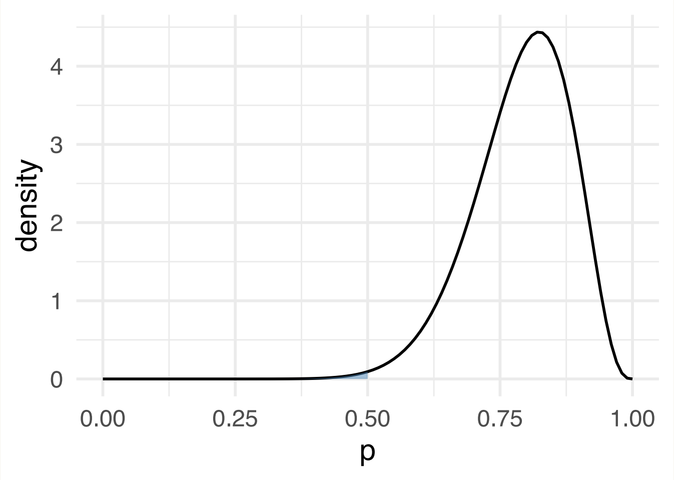


Testing a hypothesis

Suppose the researchers were interested in testing

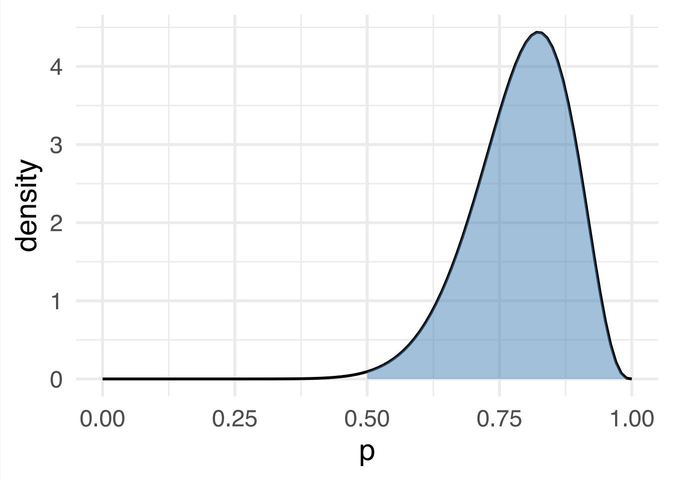
$$H_0 : p \leq 0.5$$

$$\mathrm{P}(p \leq 0.5 | Y = 14) = 0.004$$



$$H_1 : p > 0.5$$

$$\mathrm{P}(p > 0.5 | Y = 14) = 0.996$$



Predicting a new observation

To make predictions, we need to work with the **posterior predictive** distribution:

$$
\begin{aligned}
f(\widetilde{Y} = \tilde{y} | Y = y) &= \int_0^1 f(\widetilde{Y} = \tilde{y}, p | Y = y) dp \\
&= \int_0^1 f(\widetilde{Y} = \tilde{y} | p, Y = y) \pi(p | Y = y) dp \\
&= \int_0^1 f(\widetilde{Y} = \tilde{y} | p) \pi(p | Y = y) dp
\end{aligned}
$$

See Appendix B for algebraic work specific to the Binomial distribution

Monte Carlo simulation for prediction

Suppose we wish to make predictions for a new set of 20 "guesses" made by PS

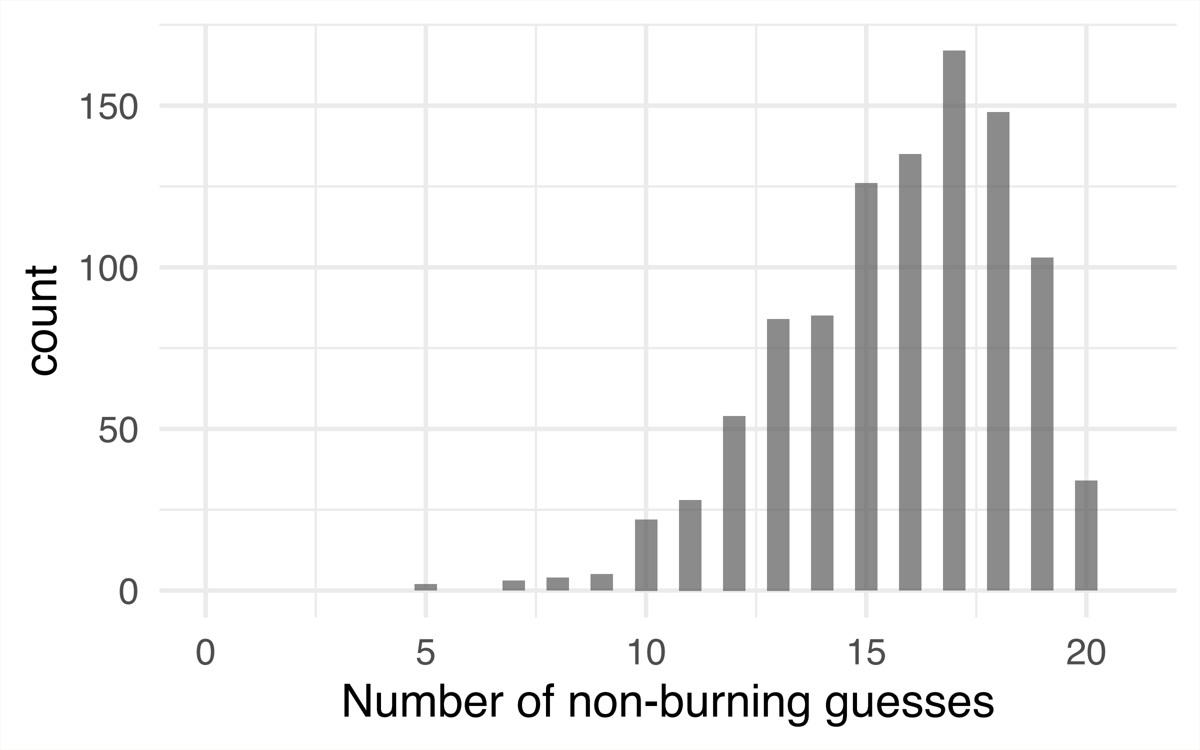
Posterior predictive

$$f(\widetilde{Y} = \tilde{y} | Y = 14) = \int_0^1 f(\widetilde{Y} = \tilde{y} | p) \pi(p | Y = 14) dp$$

Integration via simulation:

```
n <- 20   # No. of new binomial trials
S <- 1000 # No. simulations
sim_p <- rbeta(S, 15, 4)
sim_y <- rbinom(S, size = n, prob = sim_p)
```

Posterior predictive distribution




```
janitor::tabyl(sim_y)
```

```
##  sim_y   n percent
##      5   2   0.002
##      7   3   0.003
##      8   4   0.004
##      9   5   0.005
##     10  22   0.022
##     11  28   0.028
##     12  54   0.054
##     13  84   0.084
##     14  85   0.085
##     15 126   0.126
##     16 135   0.135
##     17 167   0.167
##     18 148   0.148
##     19 103   0.103
##     20  34   0.034
```

Prediction intervals

How can we construct an 89% prediction interval?

Put in the most likely values until the probability is **at least** 0.89

```
post_pred_dsn <- janitor::tabyl(sim_y)[, -2]
LearnBayes::discint(post_pred_dsn, prob = 0.89)
```

```
## $prob
## [1] 0.902
##
## $set
## [1] 12 13 14 15 16 17 18 19
```

Your turn 3

Let p denote the proportion of U.S. adults that do not believe in climate change. Of 1000 survey respondents, 150 responded that it was "not real at all".

1. Using a Beta(1, 2) prior distribution, what is the posterior distribution of p?
2. Simulate 1000 draws from the posterior distribution.
3. Use your simulated draws to calculate a 93% credible interval equal-tailed for p. Interpret this interval in context.
4. Suppose you were to survey 100 more adults. Approximate the probability that at least 20 of the 100 people don't believe in climate change.

https://thepulseofthenation.com/, data from September 2017.