Assume we want to estimate a parameter θ with data X.

The maximum likelihood approach to estimating θ is to find the value of θ that maximizes $\Pr(X \mid \theta)$.

Before we observe the data, we may have some idea of how plausible are values of θ. This idea is called our prior distribution of θ and we'll denote it $\Pr(\theta)$.

The Bayesian idea is to base our estimate of θ on the posterior distribution $\Pr(\theta \mid X)$.

$$
\begin{aligned}
\Pr(\theta \mid X) &= \frac{\Pr(\theta, X)}{\Pr(X)} \\
&= \frac{\Pr(X \mid \theta)\Pr(\theta)}{\int_\theta \Pr(X, \theta) d\theta} \\
&= \frac{\Pr(X \mid \theta)\Pr(\theta)}{\int_\theta \Pr(X \mid \theta)\Pr(\theta) d\theta} \\
&= \frac{\textbf{likelihood} \times \textbf{prior}}{\textbf{difficult quantity to calculate}}
\end{aligned}
$$

In many situations, determining the exact value of the above integral is difficult.

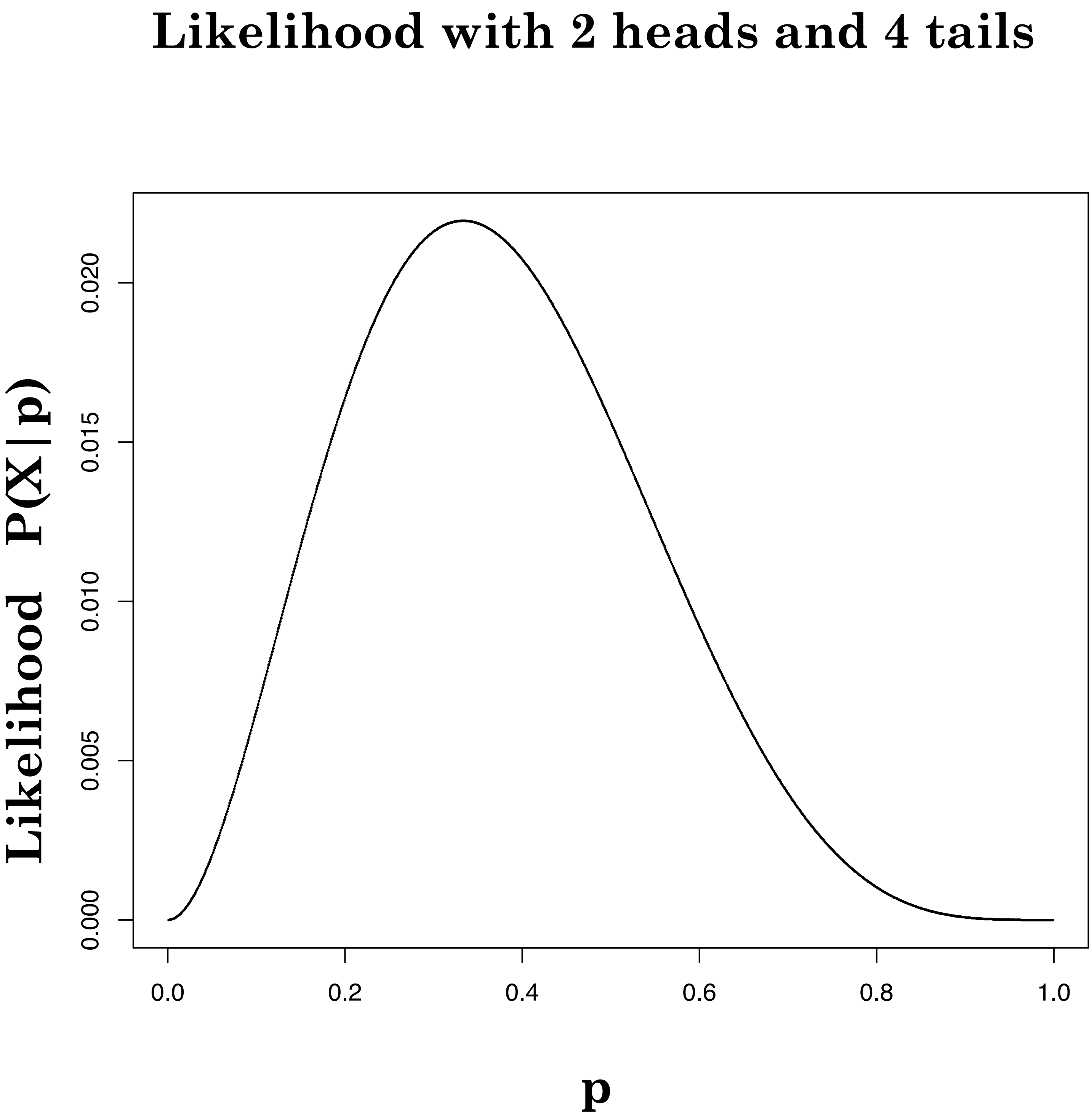
Let p be the probability of heads.

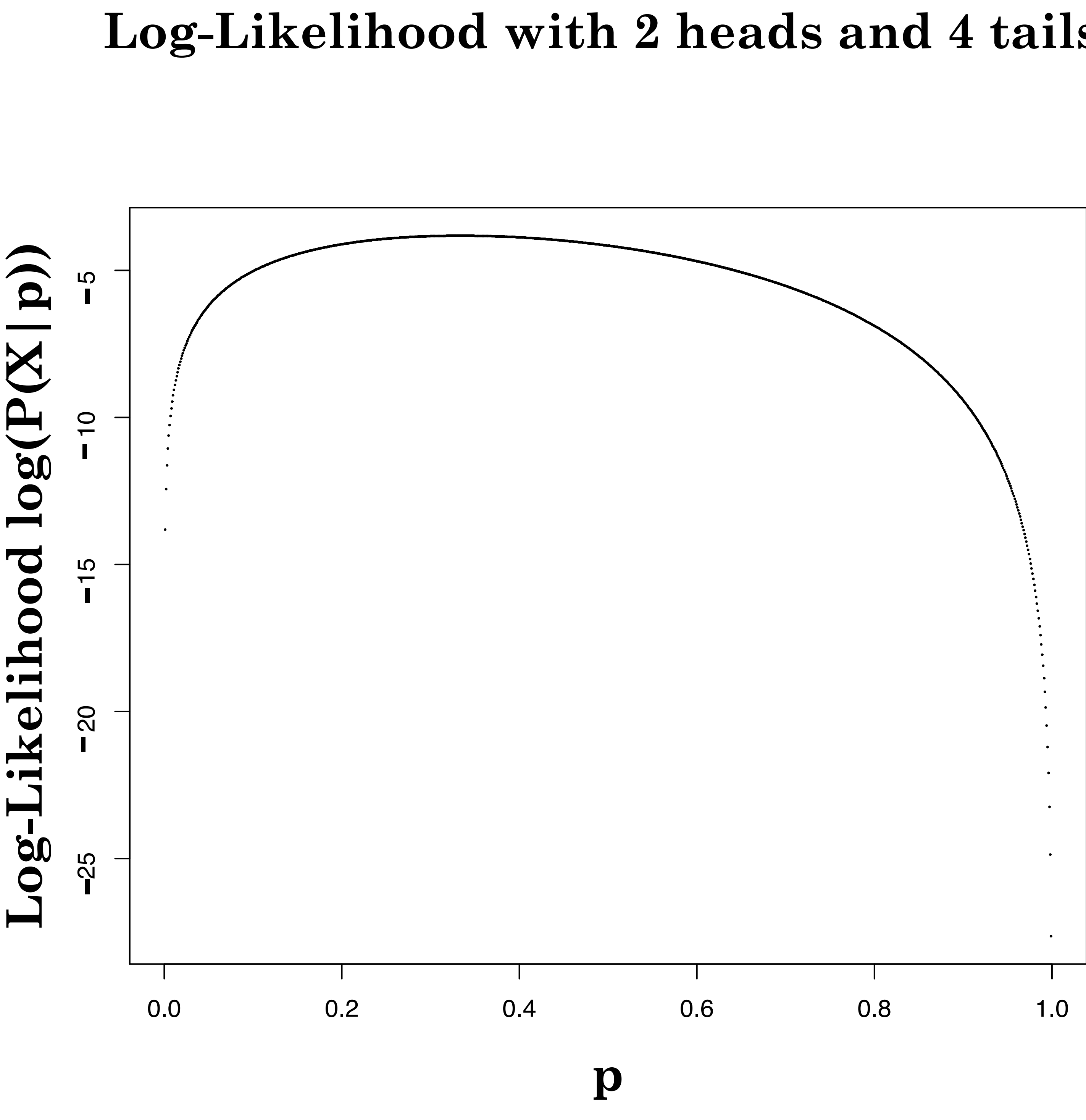
Then 1-p is the probability of tails

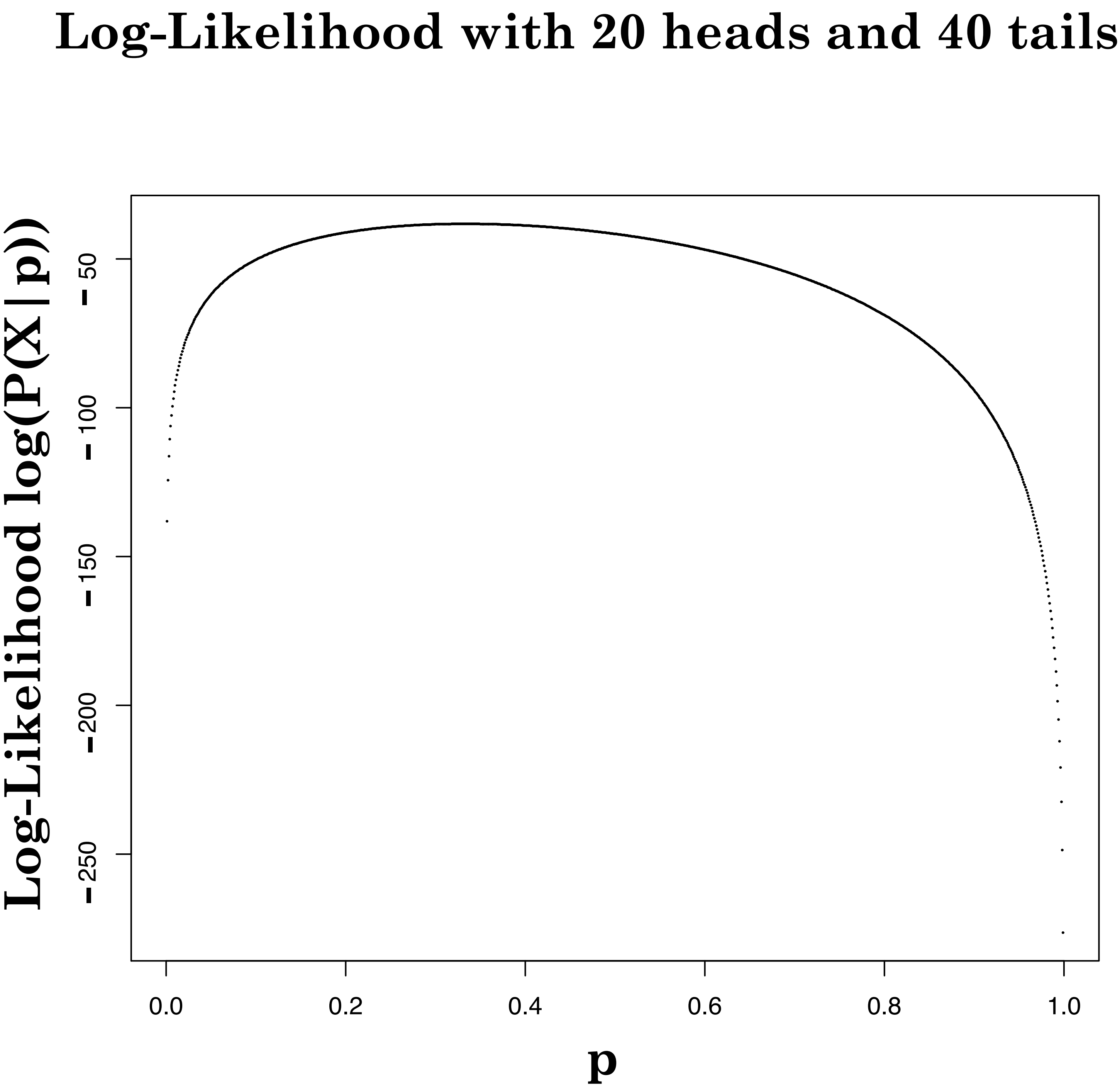
Imagine a data set X with these results from flipping a coin

Toss	1	2	3	4	5	6
Result	H	T	H	T	T	T
Probability	p	1-p	p	1-p	1-p	1-p

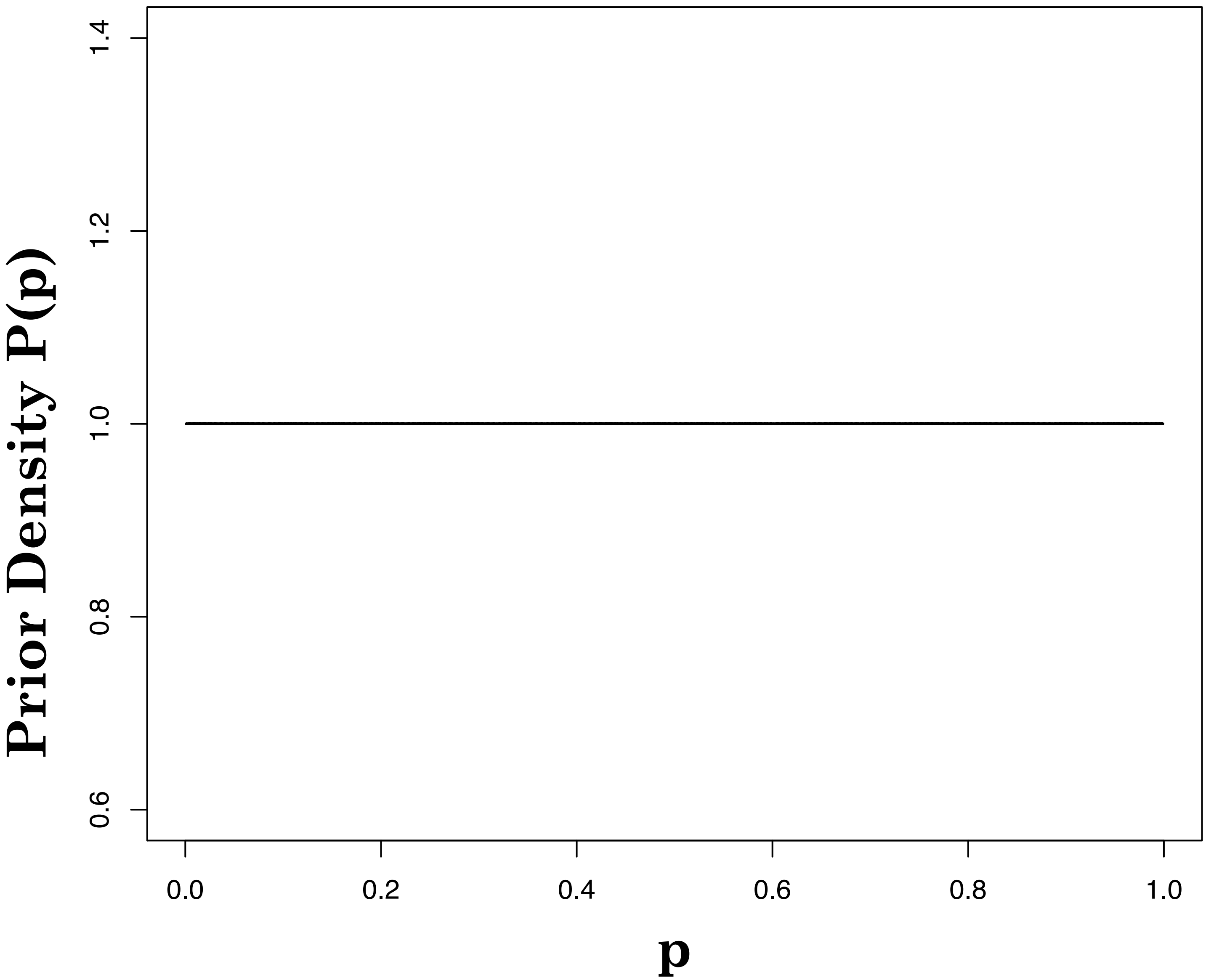
$$P(X|p) = p^2(1-p)^4$$ ← almost binomial distribution form


Likelihood with 2 heads and 4 tails
Likelihood P(X|p)
0.000
0.005
0.010
0.015
0.020
0.0
0.2
0.4
0.6
0.8
1.0
p


Log-Likelihood with 2 heads and 4 tails
Log-Likelihood log(P(X|p))
-5
-10
-15
-20
-25
0.0
0.2
0.4
0.6
0.8
1.0
p


Log-Likelihood with 20 heads and 40 tails
Log-Likelihood log(P(X|p))
-50
-100
-150
-200
-250
0.0
0.2
0.4
0.6
0.8
1.0
p

Uniform Prior Distribution (i.e., Beta(1,1) distribution)



For integers a and b, Beta density B(a,b) is

$$P(p) = (a+b-1)!/((a-1)!(b-1)!)\ p^{a-1}(1-p)^{b-1}$$

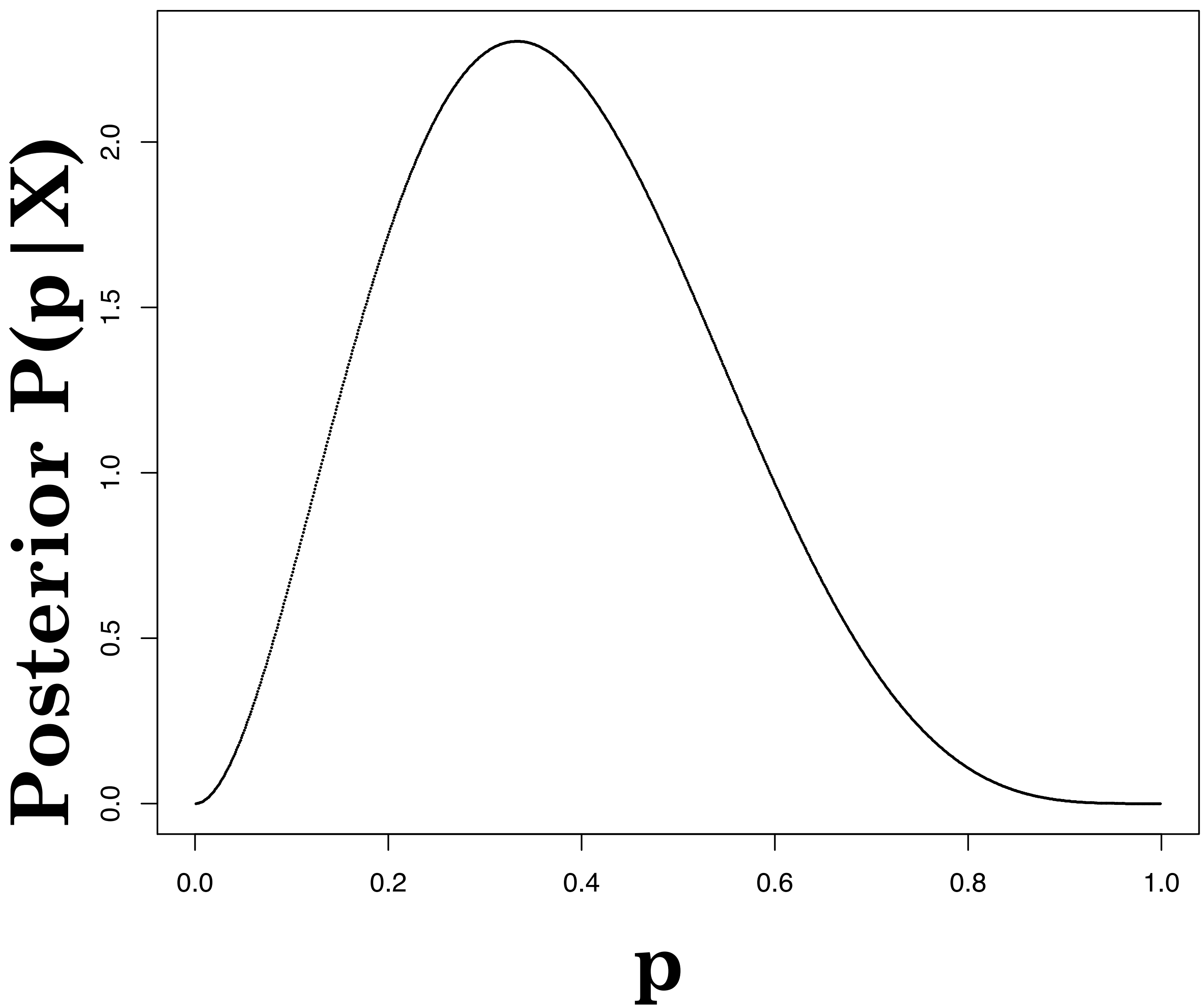
where p is between 0 and 1.

Expected value of p is a/(a+b)

Variance of p is $ab/((a+b+1)(a+b)^2)$

- **Beta distribution is conjugate prior for**
  - **data from binomial distribution**

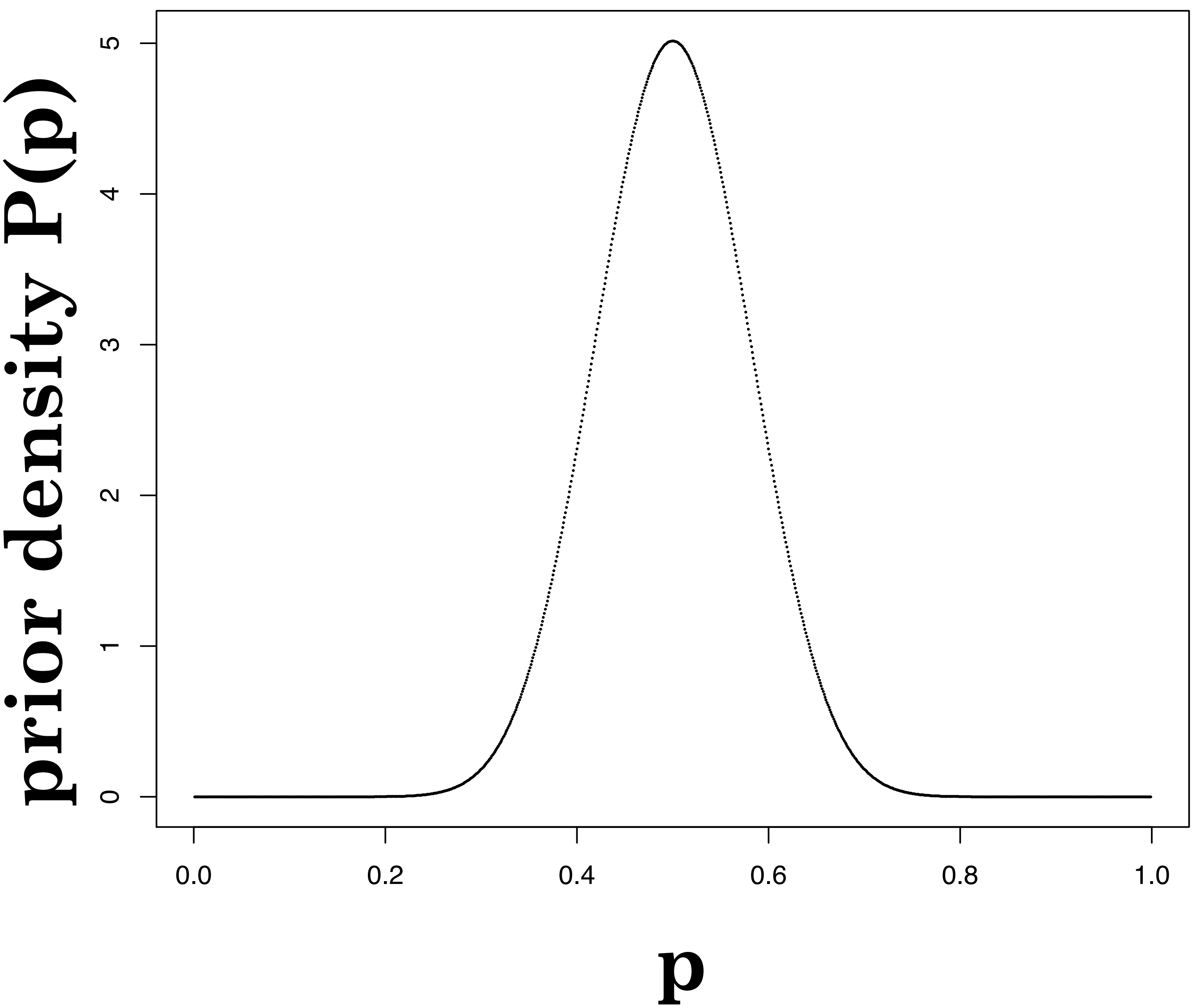
Beta(3,5) posterior from Uniform prior + data (2 heads and 4 tails)



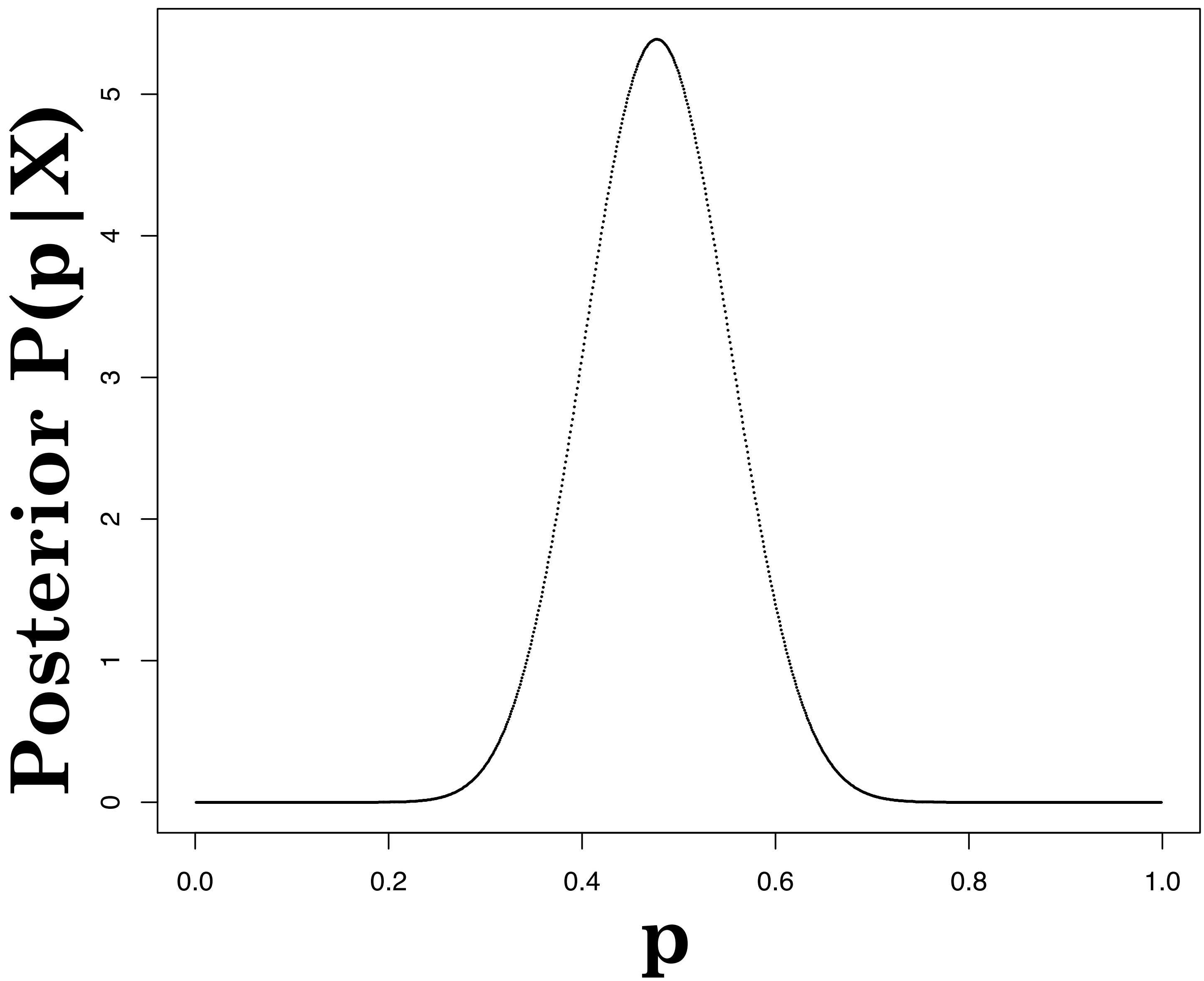
Posterior Mean = 3/(3+5)

Beta(20,20) prior distribution

Prior Mean = 0.5

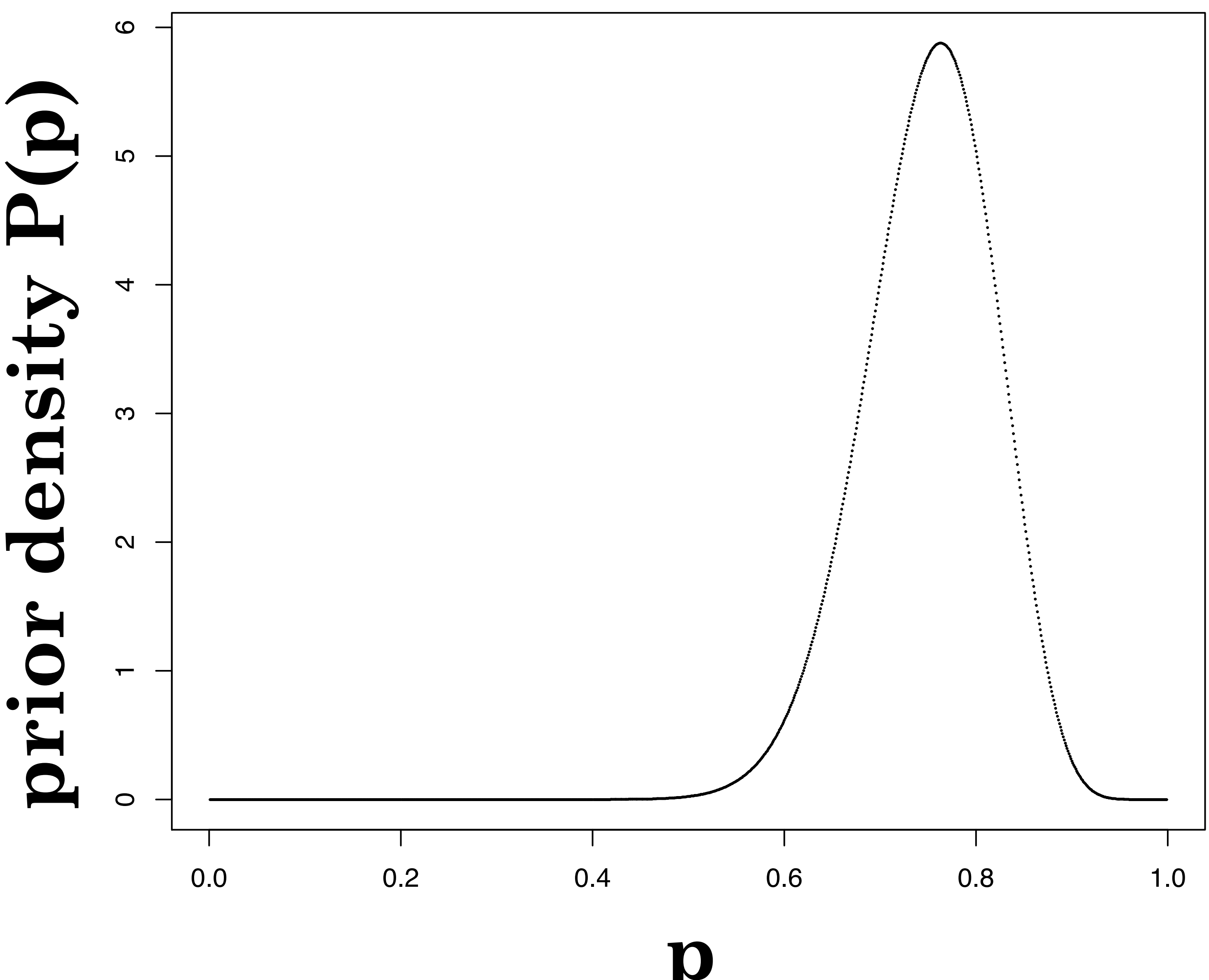


Beta(22,24) posterior from Beta(20,20) prior + data (2 heads and 4 tails)

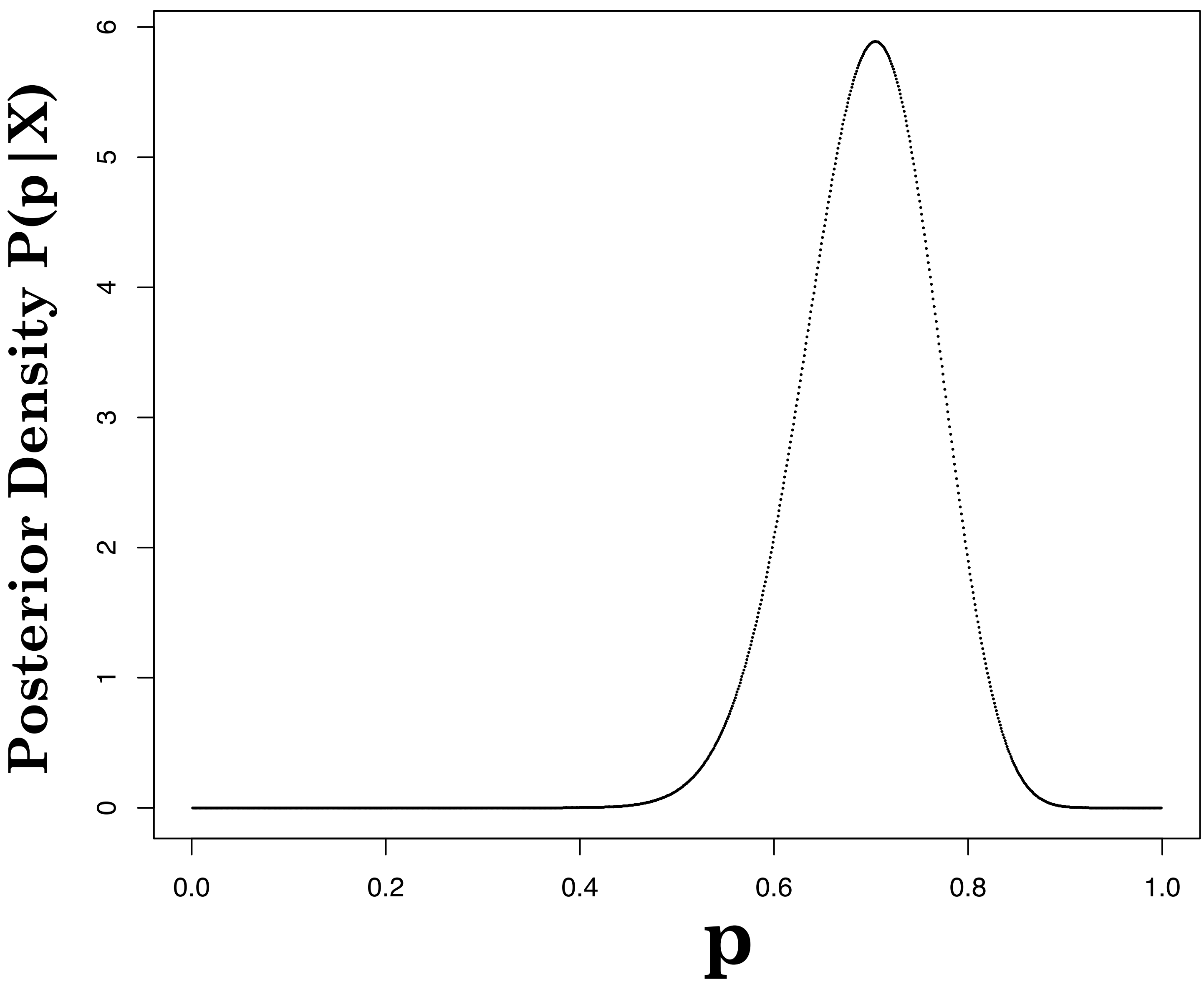


Beta(30,10) prior distribution

Prior Mean = 0.75

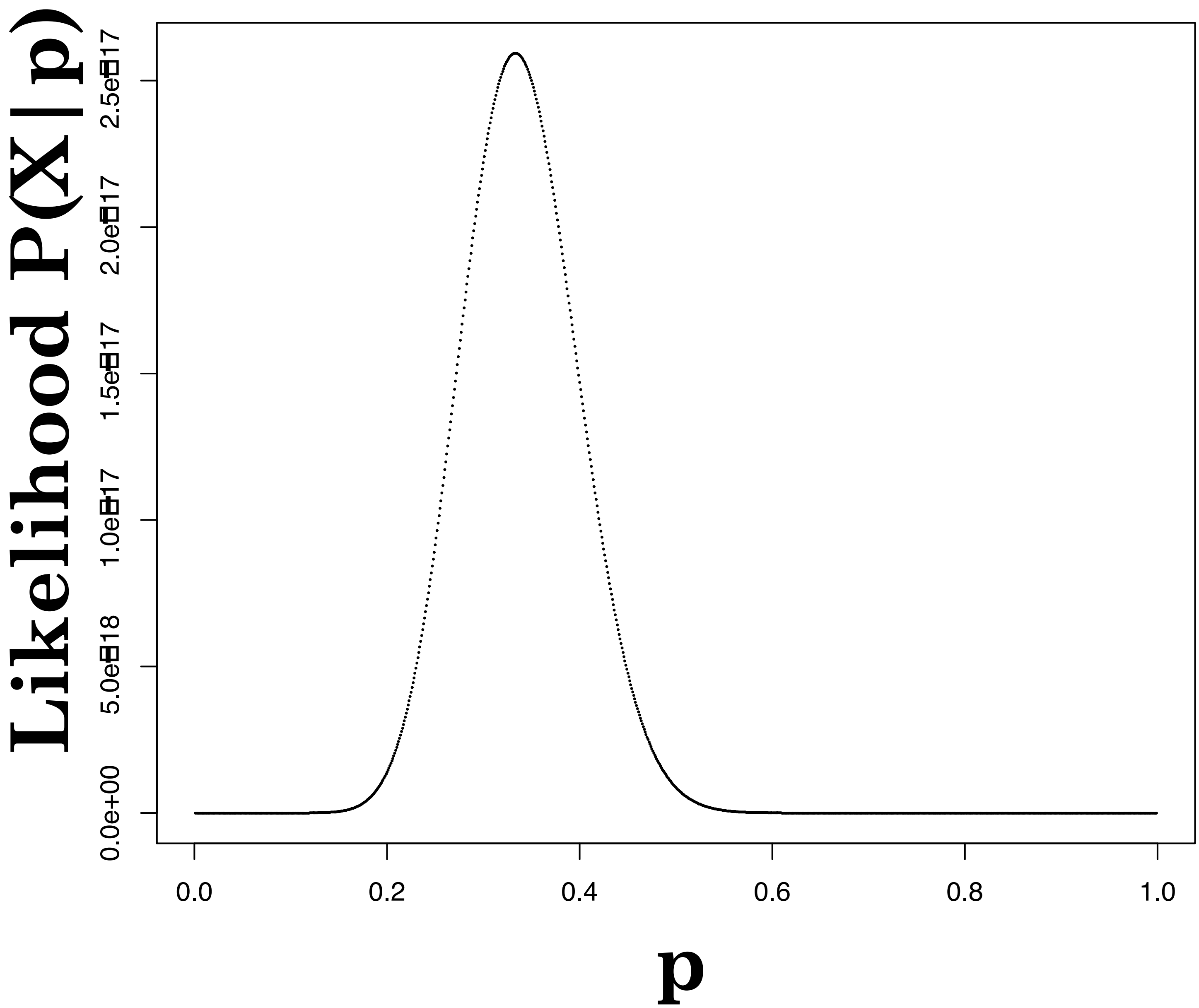


Beta(32,14) posterior from Beta(30,10) prior + data (2 heads and 4 tails)

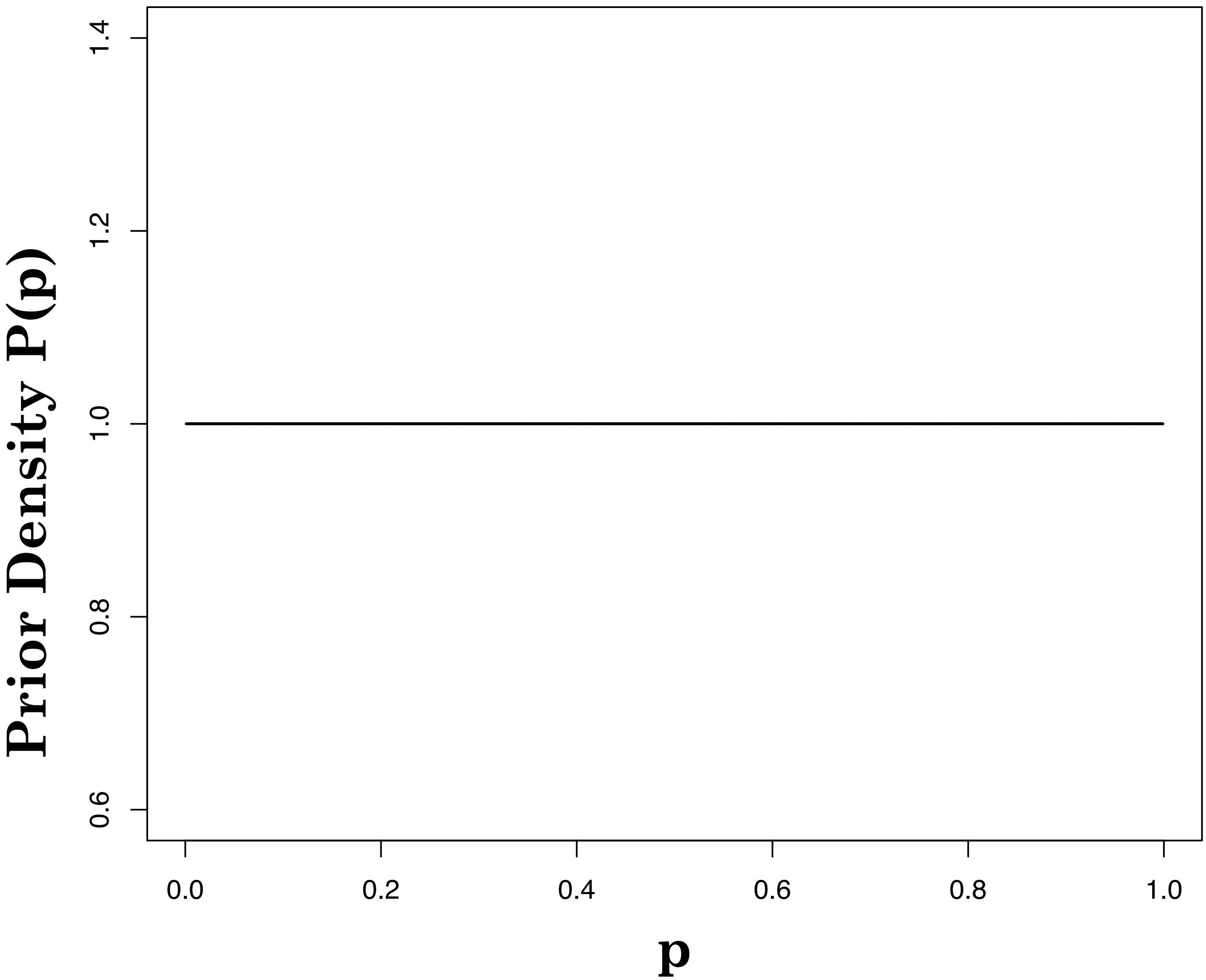


Posterior Mean = 32/(32+14)

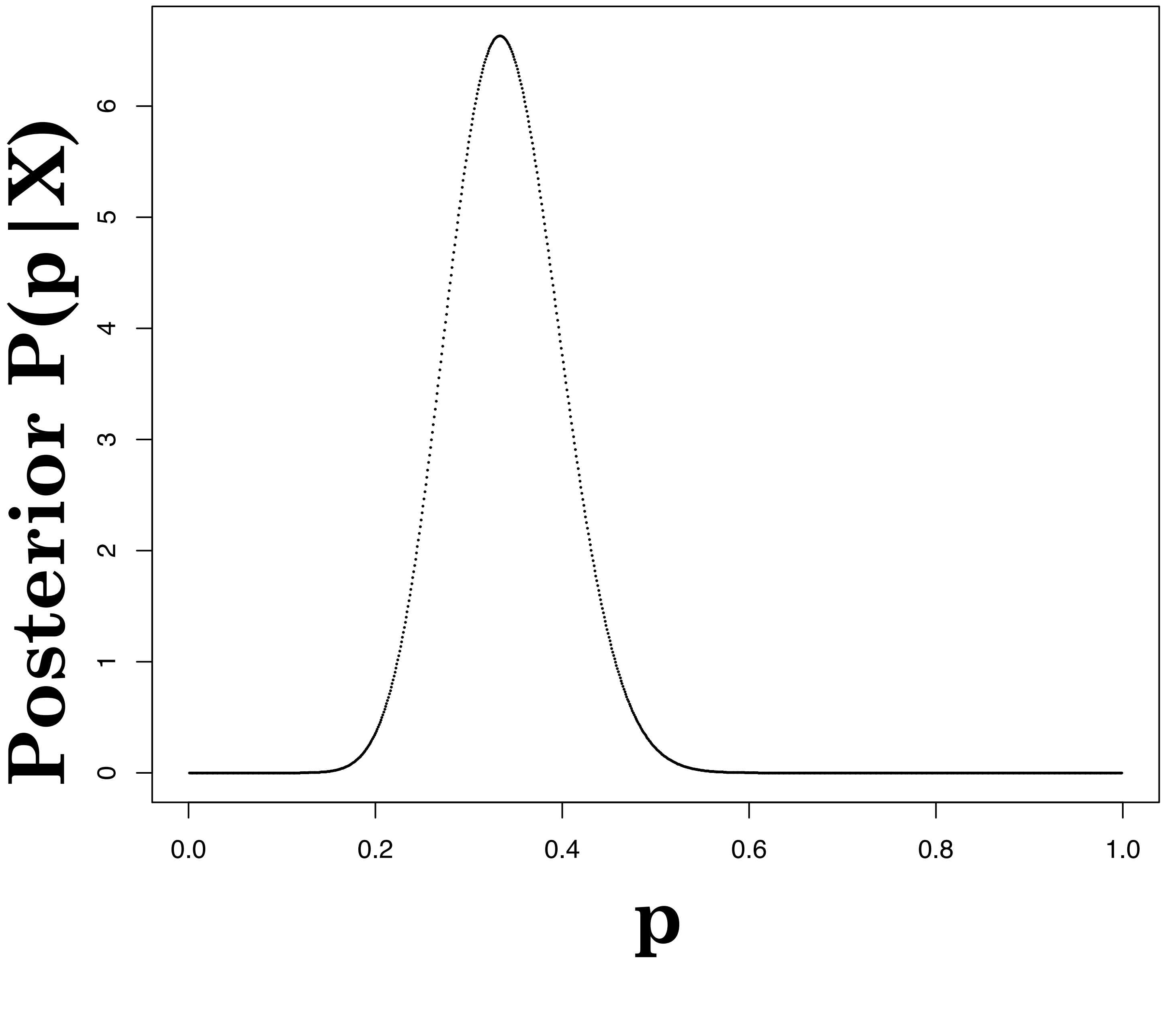
Likelihood with 20 Heads and 40 Tails



Uniform Prior Distribution (i.e., Beta(1,1) distribution)

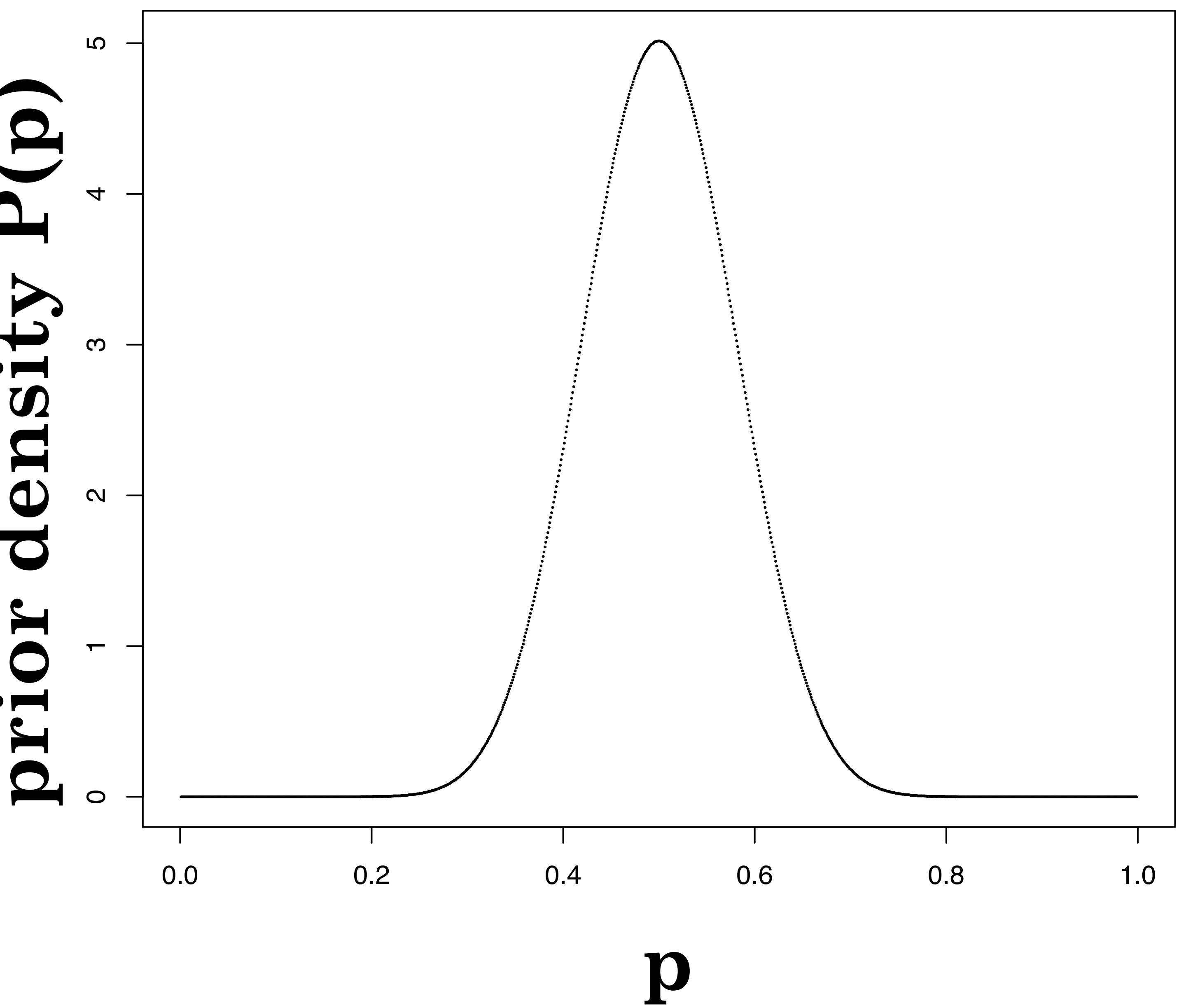


Beta(21,41) posterior from Uniform prior + data (20 heads and 40 tails)

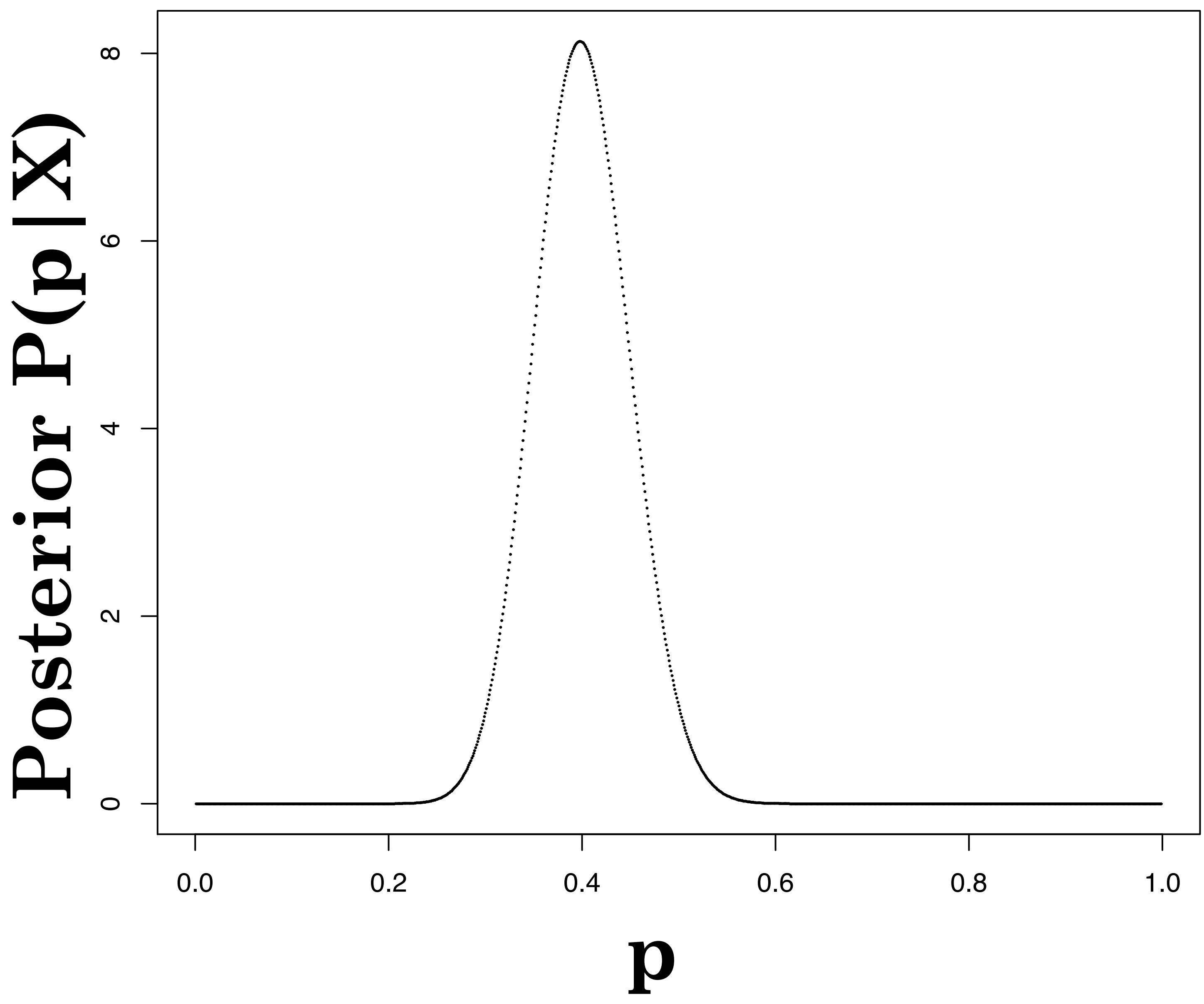


Beta(20,20) prior distribution

Prior Mean = 0.5

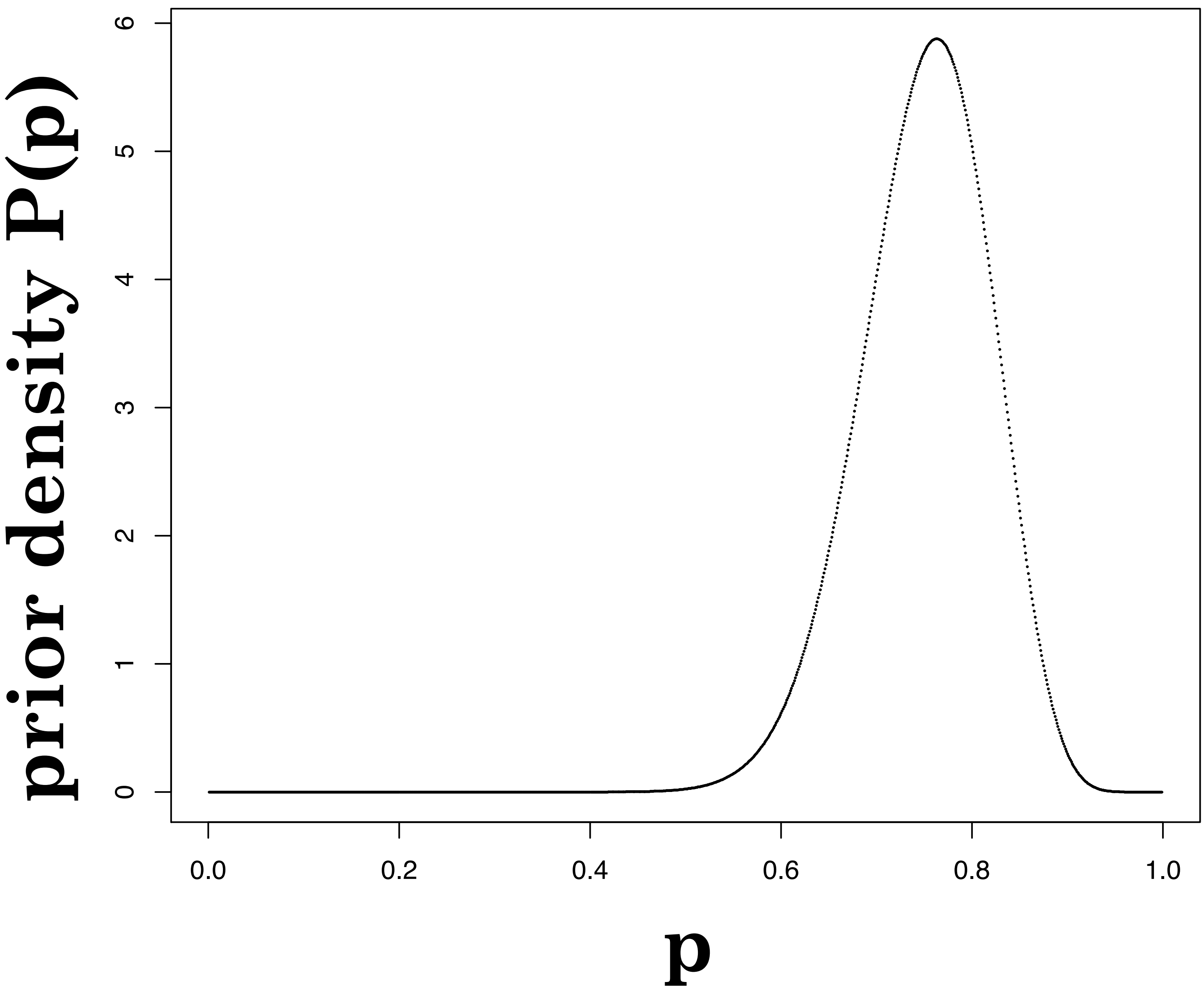


Beta(40,60) posterior from Beta(20,20) prior + data (20 heads and 40 tails)



Beta(30,10) prior distribution

Prior Mean = 0.75



Beta(50,50) posterior from Beta(30,10) prior + data (20 heads and 40 tails)