Now, let's suppose that we've chosen to use a beta prior having specific parameters a and b. Then the posterior density is such that

$$\pi(\theta|y) \propto \theta^y (1-\theta)^{n-y} \theta^{a-1} (1-\theta)^{b-1},$$

i.e.,

$$\pi(\theta|y) \propto \theta^{y+a-1} (1-\theta)^{n-y+b-1}.$$

There is one and only one density proportional to the very last function, and this is the Beta(y $+a, n-y+b$) density.

So, when one uses a beta prior in a binomial experiment, the resulting posterior is also beta, but with different parameters. This type of prior is called a *conjugate prior*.

<u>*Definition*</u> Let $\mathcal{F}$ be the class of densities of which it is assumed that $f(\cdot|\boldsymbol{\theta})$ is a member. A class $\mathcal{P}$ of prior distributions is said to be a *conjugate family for* $\mathcal{F}$ if $\pi(\cdot|\boldsymbol{y}) \in \mathcal{P}$ for all $\pi \in \mathcal{P}$ and for all $\boldsymbol{y} \in \mathcal{Y}$.

Now that we know what the posterior is, how do we use it to make inferences about θ? In classical, or frequentist, statistics the following are the "big three" of inference:

- Point estimation
- Confidence intervals
- Hypothesis testing

Each of these has its analog in the Bayesian world. For the moment, we'll just describe, in the context of the binomial experiment, a couple of ways that a Bayesian would obtain point estimates.

One popular Bayes point estimate is the mode of the posterior distribution. This is analogous to a maximum likelihood estimate.

Note that, in general, the mode of the posterior may be found without determining $m(\boldsymbol{y})$.

In the binomial experiment with Beta(a, b) prior, we found that the posterior is Beta$(y + a, n - y + b)$.

It's easy to check that a Beta(c, d) density has mode $(c - 1)/(c + d - 2)$ so long as $c > 1$ and $d > 1$. So, if $a > 1$ and $b > 1$, the prior and posterior have respective modes $(a - 1)/(a + b - 2)$ and $(y + a - 1)/(n + a + b - 2)$. (The latter expression is actually valid whenever $n > 1$ and $0 < y < n$, regardless of the values of a and b.)

A classical frequentist estimator of θ in the binomial experiment is the sample proportion

$$\widehat{\theta} = \frac{Y}{n}.$$

This estimator is both the MLE and the uniformly minimum variance unbiased estimator (UMVUE) of θ.

How does the mode of the posterior compare with $\widehat{\theta}$?

$$\begin{aligned} \text{mode} &= \frac{Y + a - 1}{n + a + b - 2} \\ &= w_n \widehat{\theta} + (1 - w_n) \left(\frac{a - 1}{a + b - 2} \right), \end{aligned}$$

where $w_n = n/(n + a + b - 2)$. This provides the following interesting interpretation:

> *The mode is a weighted average of the classical frequentist estimator and the mode of the prior distribution.*

The interpretation on the last page is not peculiar to this example. Typically, we have the following:

1. A Bayes point estimate is a weighted average of a common frequentist estimate and a parameter estimate obtained only from prior information.

2. The Bayes point estimate "shrinks" the frequentist estimate toward the prior estimate.

3. The weight on the frequentist estimate tends to 1 as n (the amount of data) tends to infinity.

The last property illustrates the second point on p. 12 of the notes (12N), namely, the effect of the prior becomes negligible as more and more data are obtained.

Another means of obtaining a Bayes point estimate is to use the Bayes decision principle. For example, suppose we use squared error loss. Then according to p. 37N, the Bayes action (or estimate) would be the value of a that minimizes

$$\int_{\Theta} (\boldsymbol{\theta} - a)^2 \pi(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta}.$$

If f is a density, the general solution of the problem

"Find a to minimize $\int (x-a)^2 f(x)\, dx$"

is $a = \int x f(x)\, dx$, i.e., the mean of the density.

Therefore,

the Bayes estimate with respect to squared error loss is always the mean of the posterior distribution (assuming that it exists).

The mean of a Beta(c, d) density is $c/(c+d)$, which is true for all $c > 0$ and $d > 0$. So, the mean of our Beta(a, b) prior is $a/(a+b)$, and the Bayes estimate of θ is

$$\begin{aligned}\frac{y+a}{n+a+b} &= \widehat{\theta}\left(\frac{n}{n+a+b}\right) \\ &\quad + \left(\frac{a}{a+b}\right)\left(1 - \frac{n}{n+a+b}\right).\end{aligned}$$

Note that this is a weighted average of the data mean and the mean of the prior density.

As did the posterior mode, this Bayes estimate has the three properties stated on p. 47N.

Example 6 Let θ be the proportion of all people suffering from a particular chronic illness who recover within one month when given a certain treatment. A clinical trial is to be done in which 50 persons suffering from this illness are given the treatment of interest.

Suppose that a beta prior is to be used. We'll look at the effect of choices for a and b on a Bayes estimate of θ.

Suppose that 18 of the 50 persons receiving the treatment recover within one month. The classical point estimate of θ is

$$\widehat{\theta} = \frac{18}{50} = 0.36.$$

Let's consider six different priors.

1. $\underline{a = 1,\ b = 1:}$ A uniform prior; gives equal weight to each value of θ.

$$\text{Bayes estimate } = \frac{19}{52} = 0.365$$

2. $\underline{a = 1/2,\ b = 1/2:}$ A particular type of non-informative prior.

$$\text{Bayes estimate } = \frac{18.5}{51} = 0.363$$

3. $\underline{a = 1,\ b = 2:}$ A fairly "vague" prior with mean 1/3.

$$\text{Bayes estimate } = \frac{19}{53} = 0.358$$

4. <u>$a = 20$, $b = 40$:</u> An informative prior with mean 1/3.

$$\text{Bayes estimate } = \frac{38}{110} = 0.345.$$

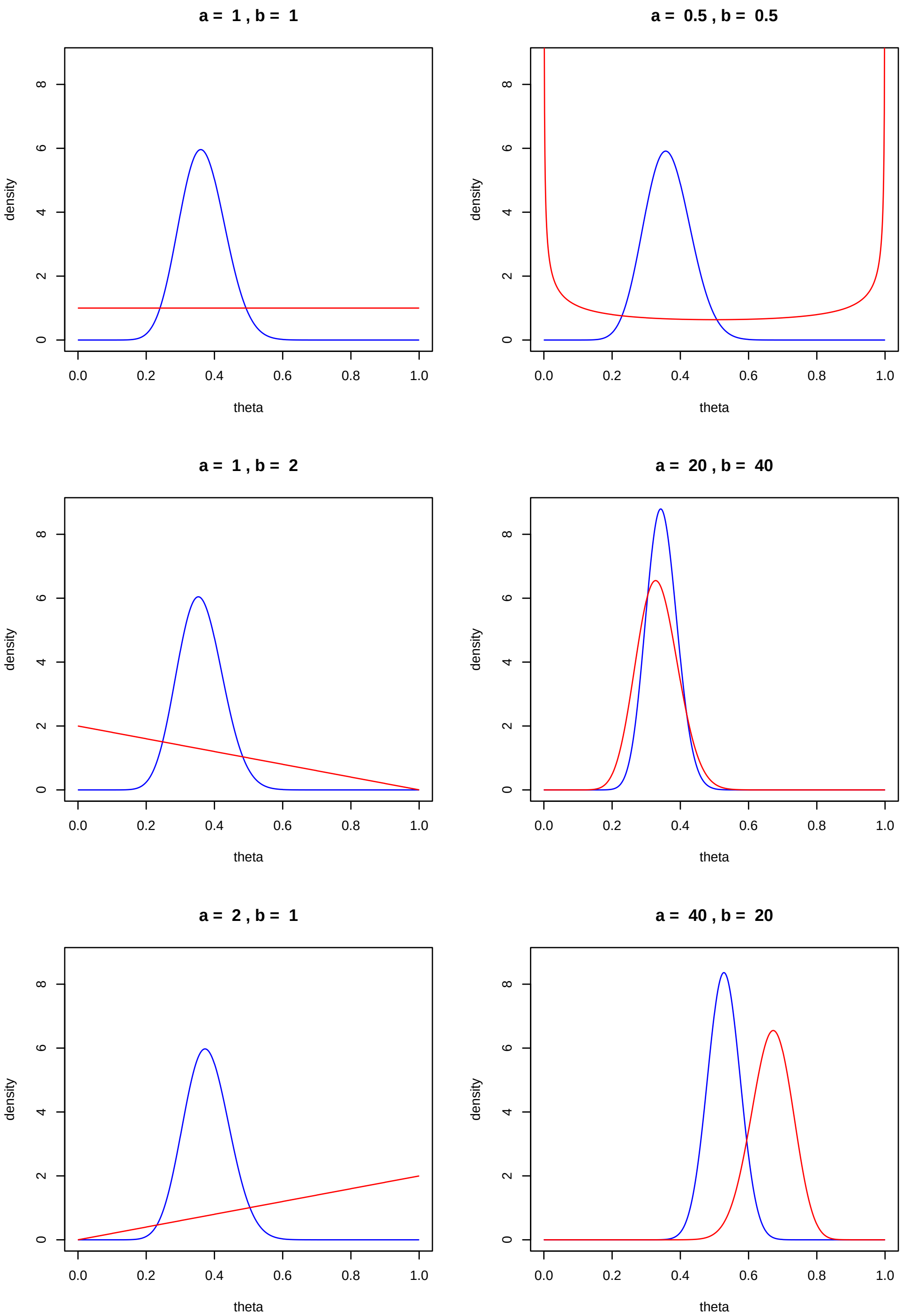
5. <u>$a = 2$, $b = 1$:</u> A vague prior with mean 2/3.

$$\text{Bayes estimate } = \frac{20}{53} = 0.377$$

6. <u>$a = 40$, $b = 20$:</u> An informative prior with mean 2/3.

$$\text{Bayes estimate } = \frac{58}{110} = 0.527$$

Priors and posteriors for Example 6



Prior: —— Posterior: ——