Bayesian Methods

LABORATORY

Lesson 1: Jan 24 2002

Software: R

The R Project for Statistical Computing

http://www.r-project.org/

- R is a language and environment for statistical computing and graphics.

The R Project for Statistical Computing

http://www.r-project.org/

- R is a language and environment for statistical computing and graphics.
  - R, like S, is designed around a computer language, and it allows users to add additional functionality by defining new functions.

The R Project for Statistical Computing

http://www.r-project.org/

- R is a language and environment for statistical computing and graphics.
  - R, like S, is designed around a computer language, and it allows users to add additional functionality by defining new functions.
  - The term "environment" is intended to characterize it as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools.

- It is a GNU project which is similar to the S language and environment.

- It is a GNU project which is similar to the S language and environment.
  - The GNU Project was launched in 1984 to develop a complete Unix-like operating system which is free software.

- It is a GNU project which is similar to the S language and environment.
  - The GNU Project was launched in 1984 to develop a complete Unix-like operating system which is free software.
  - Free software is a matter of the users' freedom to run, copy, distribute, study, change and improve the software. It is not a matter of price!

- It is a GNU project which is similar to the S language and environment.
  - The GNU Project was launched in 1984 to develop a complete Unix-like operating system which is free software.
  - Free software is a matter of the users' freedom to run, copy, distribute, study, change and improve the software. It is not a matter of price!
- R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.

ONE-DIMENSIONAL *parameter models*

1. The Binomial model and its coniugate Beta prior

in $Bin(n, \theta)$ there is a single parameter of interest (n is tipically assumed known), that is the probability θ of a certain outcome in each of the n trials considered.

Bayesian estimation of a probability from BINOMIAL *data*

Gelman book, pag. 39, sec. 2.5

R code placenta.r is in the Lab notes at the course web page

- Our interest focus on the proportion of female births in the so called maternal condition *placenta previa*

Bayesian estimation of a probability from BINOMIAL *data*

Gelman book, pag. 39, sec. 2.5

R code placenta.r is in the Lab notes at the course web page

- Our interest focus on the proportion of female births in the so called maternal condition *placenta previa*
- Our data consist in a early study in Germany: 437 females on 980 placenta previa births

Bayesian estimation of a probability from BINOMIAL *data*

Gelman book, pag. 39, sec. 2.5

R code placenta.r is in the Lab notes at the course web page

- Our interest focus on the proportion of female births in the so called maternal condition *placenta previa*
- Our data consist in a early study in Germany: 437 females on 980 placenta previa births
- How much evidence do they provide that the proportion of placenta previa female births is < 0.485, the proportion of the general population female births?

Analysis using a UNIFORM PRIOR

- Let the 1-parameter θ denote the proportion of *placenta previa* female births

Analysis using a UNIFORM PRIOR

- Let the 1-parameter θ denote the proportion of *placenta previa* female births
- We assume a $Bin(\theta, 980) \;\propto\; \theta^{437}\,(1-\theta)^{980-437}$ to be the model generating the data

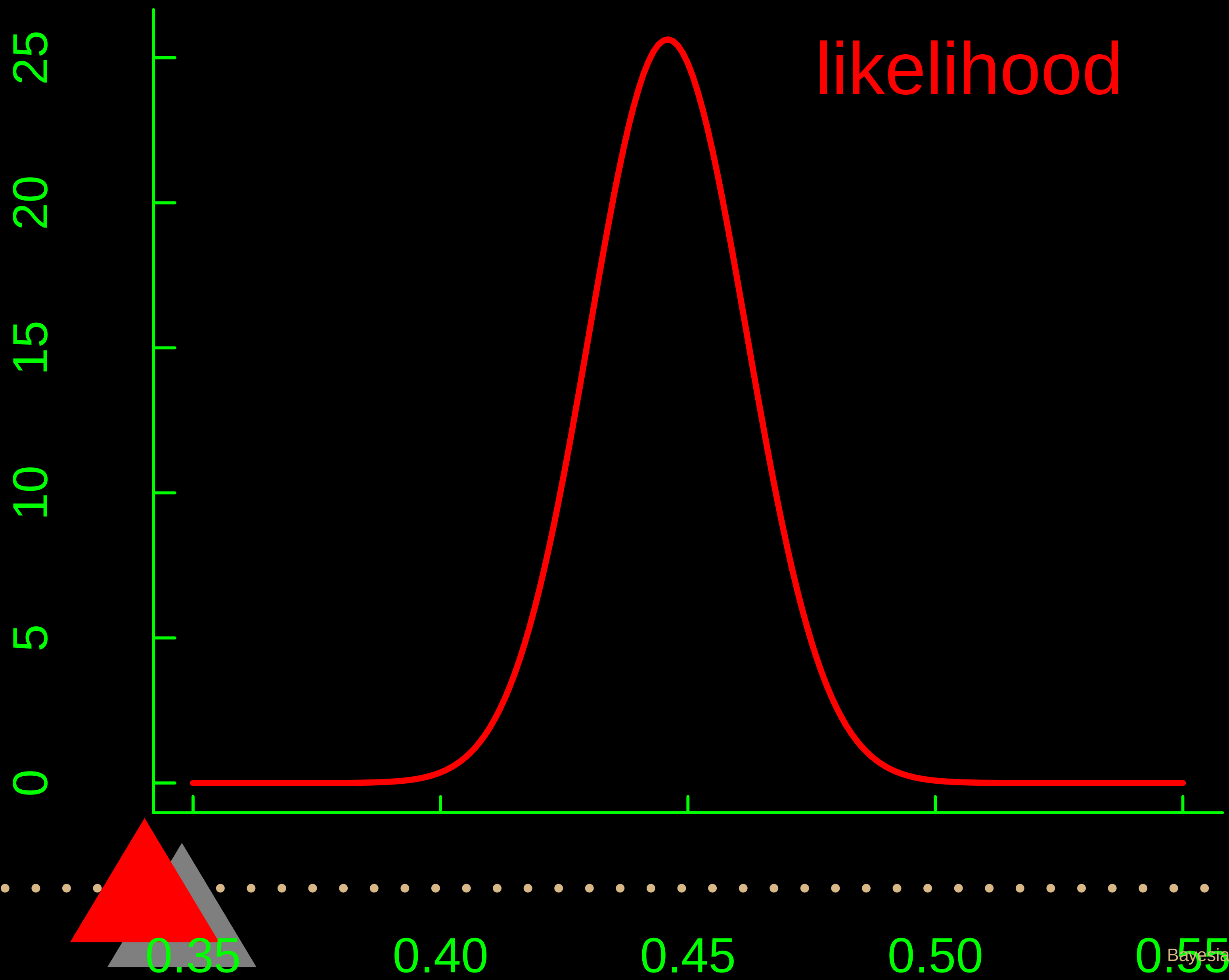
Analysis using a UNIFORM PRIOR

- Let the 1-parameter θ denote the proportion of *placenta previa* female births
- We assume a $Bin(\theta, 980) \;\propto\; \theta^{437}\,(1-\theta)^{980-437}$ to be the model generating the data
- We specify the *prior* for θ to be a $U[0,1]$

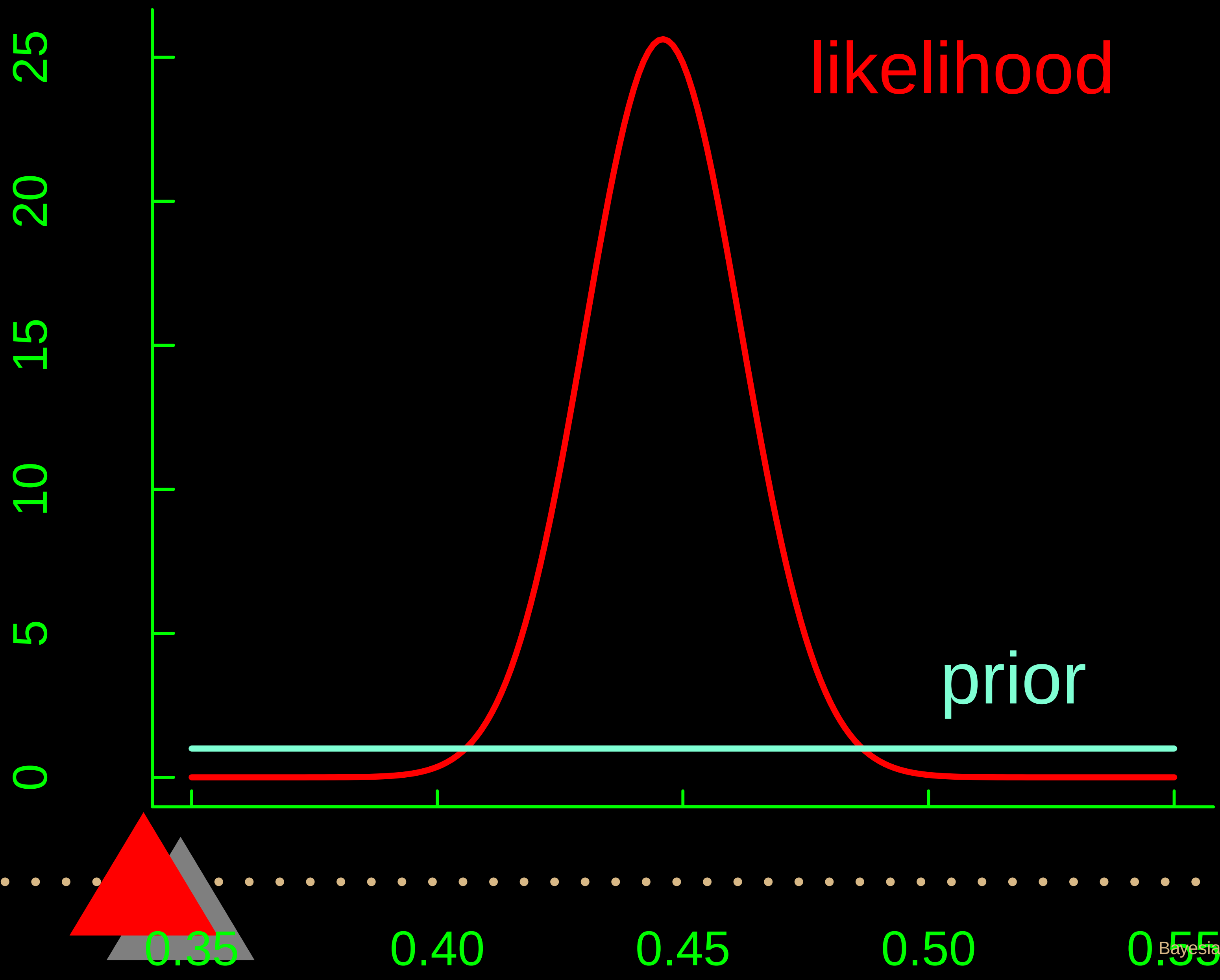
Analysis using a UNIFORM PRIOR

- Let the 1-parameter θ denote the proportion of *placenta previa* female births
- We assume a $Bin(\theta, 980) \;\propto\; \theta^{437}(1-\theta)^{980-437}$ to be the model generating the data
- We specify the *prior* for θ to be a $U[0,1]$
- The *posterior* for θ is, then, $\propto \theta^{437}(1-\theta)^{980-437}$, i.e., is a $Beta(437+1, 980-437+1)$

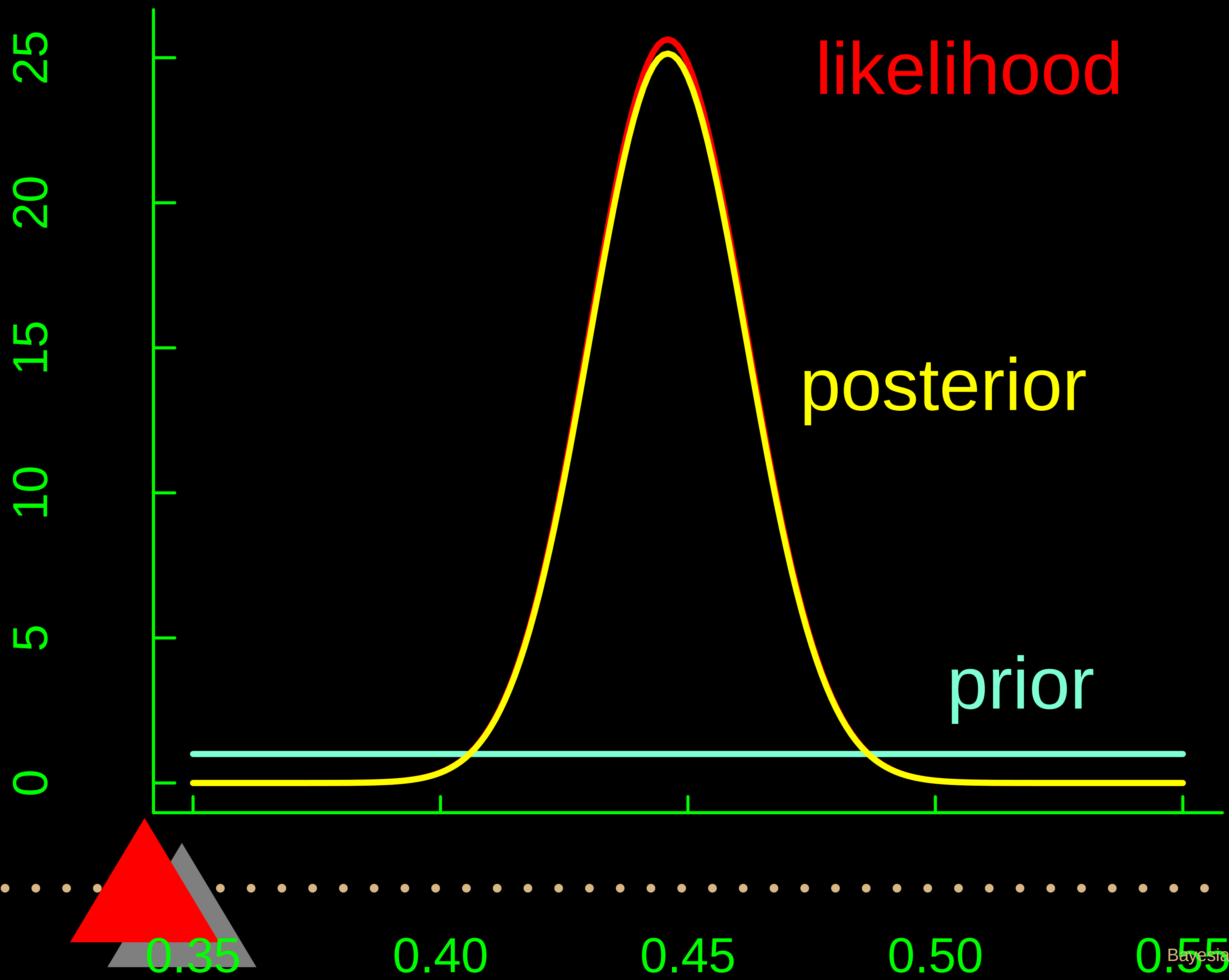
(Beta-)Uniform–Binomial



(Beta-)Uniform–Binomial



(Beta-)Uniform–Binomial



Analysis using different BETA PRIORS

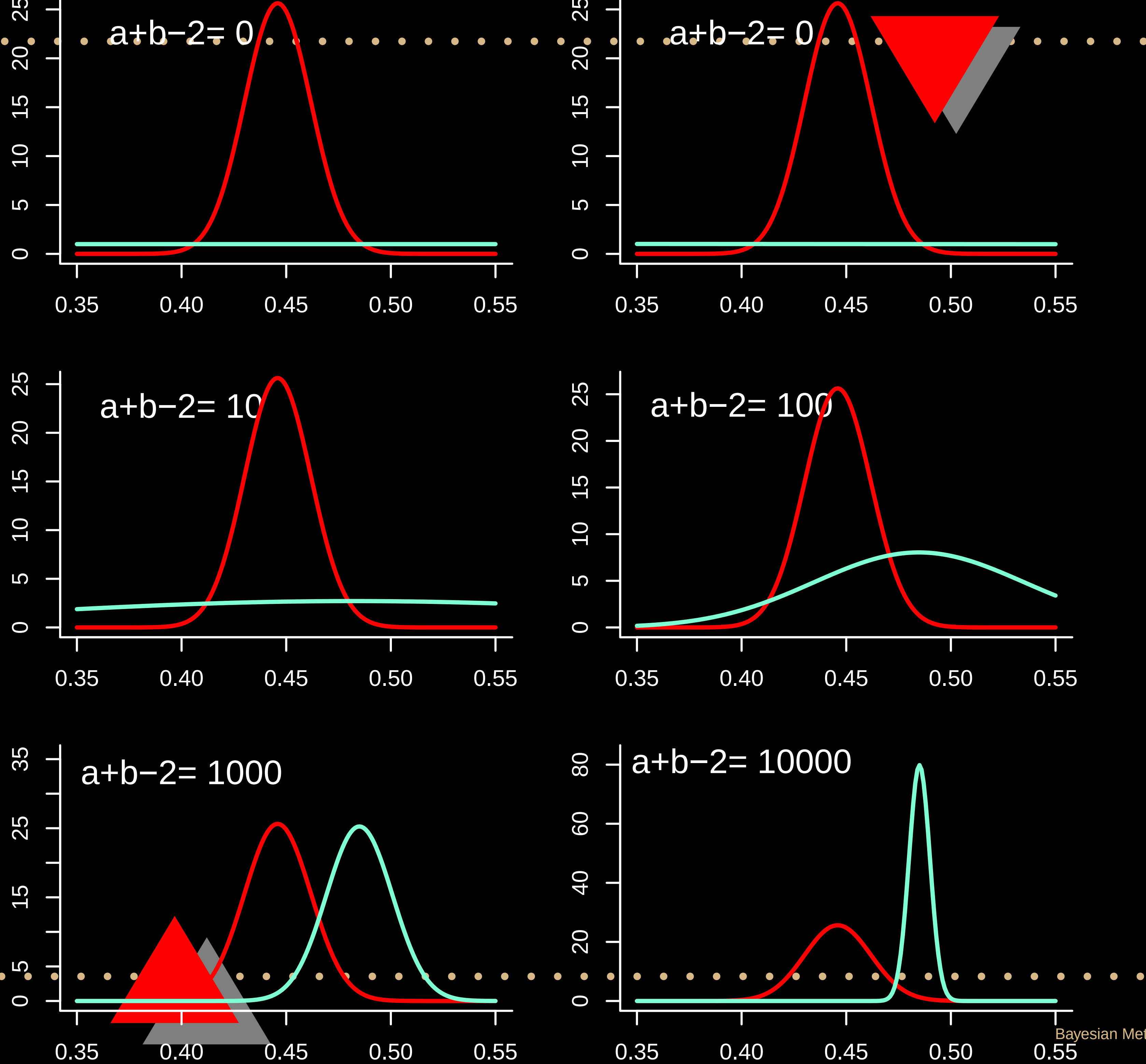
As the likelihood $p(y|\theta) \equiv L(\theta; y)$ is $\propto \theta^y \, (1-\theta)^{n-y}$
if the prior is of the same form, e.g., $p(\theta)$ is $\propto$

$$\theta^{\alpha-1} \, (1-\theta)^{\beta-1}$$

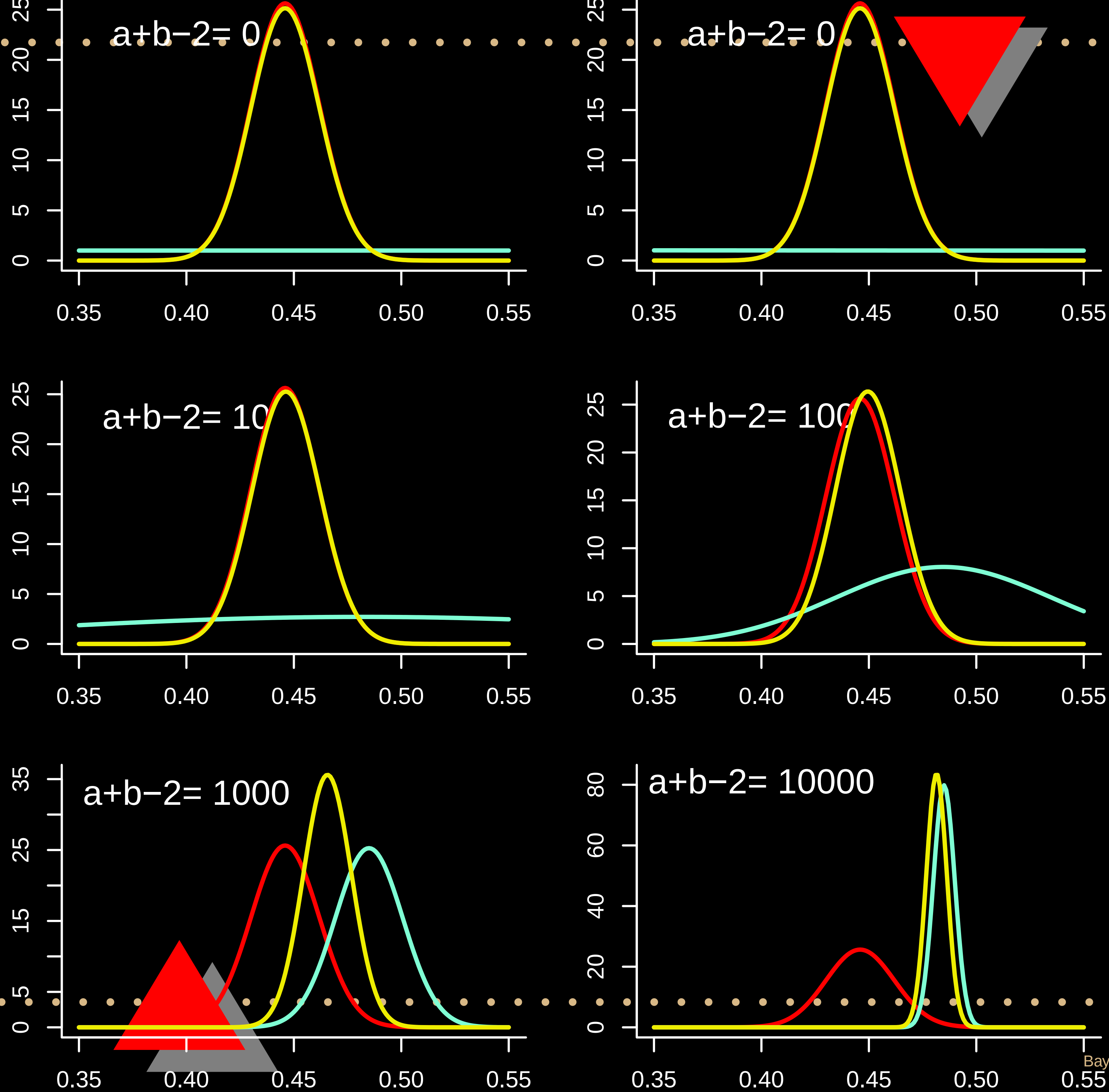
then the posterior will also be of this form. In fact, $p(\theta|y)$ is

$$\propto \; \theta^{y+\alpha-1} \, (1-\theta)^{n-y+\beta-1} = Beta(\alpha+y, \beta+n-y)$$

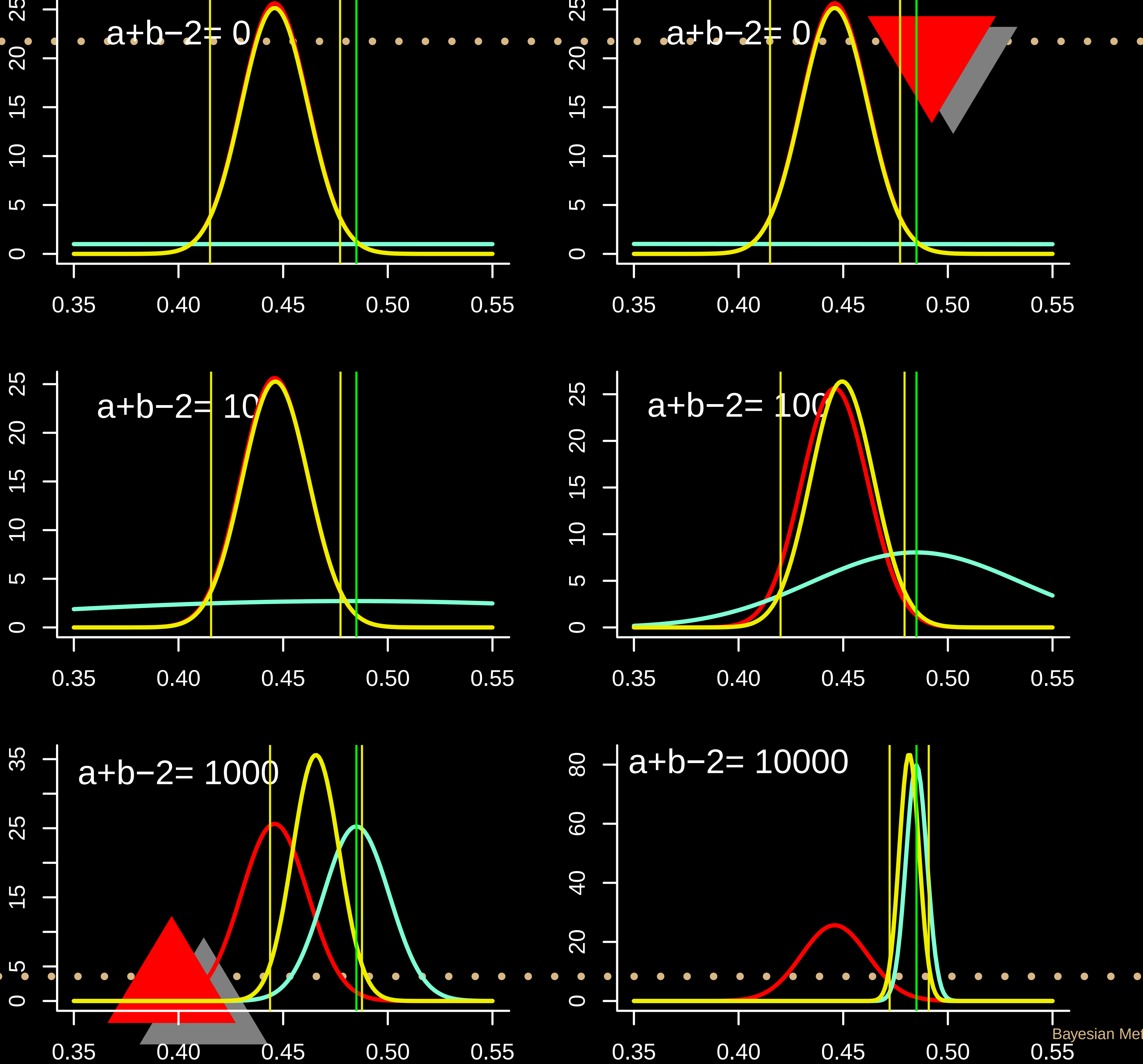
-> the BETA prior distribution is a coniugate family for the BINOMIAL likelihood



a+b−2= 0
a+b−2= 0
a+b−2= 10
a+b−2= 100
a+b−2= 1000
a+b−2= 10000
0.35
0.40
0.45
0.50
0.55



a+b−2= 0
a+b−2= 0
a+b−2= 10
a+b−2= 100
a+b−2= 1000
a+b−2= 10000
0.35
0.40
0.45
0.50
0.55


a+b−2= 0
a+b−2= 0
a+b−2= 10
a+b−2= 100
a+b−2= 1000
a+b−2= 10000
0.35
0.40
0.45
0.50
0.55

How does posterior COMPROMISE *between prior and the data?*

- The compromise depends on how much weight prior has (or how much informative it is) w.r.t. the data at hand

How does posterior COMPROMISE *between prior and the data?*

- The compromise depends on how much weight prior has (or how much informative it is) w.r.t. the data at hand
- i.e., in the binomial case, depends on the relative weight of

$$\alpha + \beta - 2$$

$\approx$ number of *prior observations* ($\sim$ prior precision)

Note: precision=1/variance, var=$\frac{\theta(1-\theta)}{\alpha+\beta+1}$

w.r.t. n, the sample size

A first SENSITIVITY ANALYSIS

concept of sensitivity: sensitivity or robustness of the inferences to the choice of the prior

Prior information		Posterior information	
$\alpha + \beta - 2$	mean	mean	95% interval
0	0.500	0.446	[0.415 , 0.477]
0	0.485	0.446	[0.415 , 0.477]
10	0.485	0.446	[0.416 , 0.477]
100	0.485	0.450	[0.420 , 0.479]
1000	0.485	0.466	[0.444 , 0.488]
10000	0.485	0.482	[0.472 , 0.491]

NOTE: in placenta previa example n $\approx$ 1000 and $\bar{y} = 0.446$

The SIMULATION*-based estimation approach*

- The modern approach to Bayesian estimation has become closely linked to **simulation-based estimation methods**.

The SIMULATION*-based estimation approach*

- The modern approach to Bayesian estimation has become closely linked to **simulation-based estimation methods**.
- In fact, Bayesian estimation focuses on estimating the entire density of a parameter.

The SIMULATION*-based estimation approach*

- The modern approach to Bayesian estimation has become closely linked to **simulation-based estimation methods**.
- In fact, Bayesian estimation focuses on estimating the entire density of a parameter.
- This density estimation is based on generating samples from the posterior density of the parameters themselves or of functions of parameters.

- In the BETA-BINOMIAL model, the coniugacy allows us knowing the posterior density in closed form.

- In the BETA-BINOMIAL model, the coniugacy allows us knowing the posterior density in closed form.
- Then, direct calculations are feasible or **direct simulation** from it can be performed.

- In the BETA-BINOMIAL model, the coniugacy allows us knowing the posterior density in closed form.
- Then, direct calculations are feasible or **direct simulation** from it can be performed.
- However, even if posterior density cannot be explicitly integrated, **iterative simulation** methods (or MCMC) are alternatively used. We will see them in future lab's.

a first (direct) simulation

Congdon book, pag. 31, sec. 2.11

- Wilcox (1996) presents data from a 1991 gallup opinion poll about the morality of President Bush's not helping Iraqi rebel groups after the formal end of the gulf war. Of the 751 adults responding, 150 thought the president's actions were not moral.

a first (direct) simulation

Congdon book, pag. 31, sec. 2.11

- Wilcox (1996) presents data from a 1991 gallup opinion poll about the morality of President Bush's not helping Iraqi rebel groups after the formal end of the gulf war. Of the 751 adults responding, 150 thought the president's actions were not moral.
- We are interested in assessing the probability that a randomly sampled adult would respond 'immoral'.

a first (direct) simulation

Congdon book, pag. 31, sec. 2.11

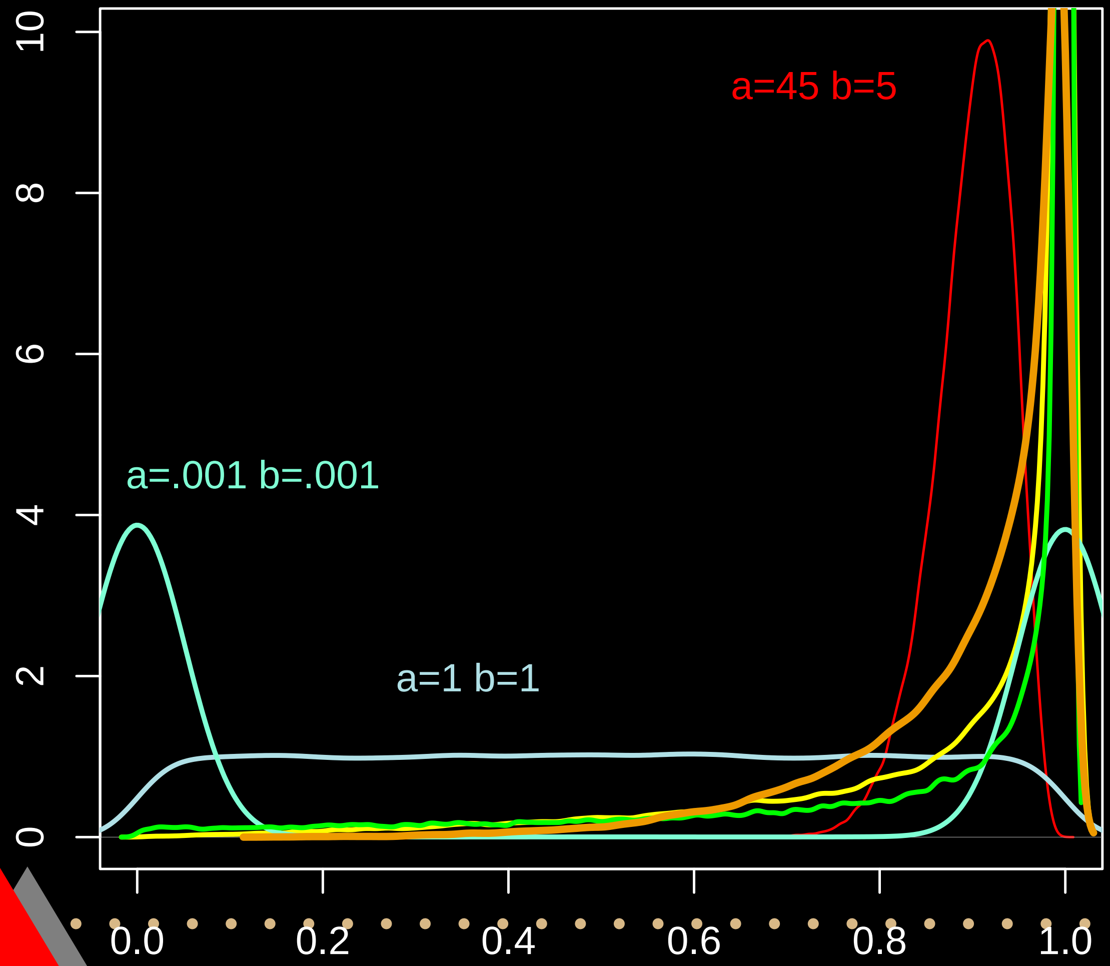
- Wilcox (1996) presents data from a 1991 gallup opinion poll about the morality of President Bush's not helping Iraqi rebel groups after the formal end of the gulf war. Of the 751 adults responding, 150 thought the president's actions were not moral.
- We are interested in assessing the probability that a randomly sampled adult would respond 'immoral'.
- In the inference we might use evidence from previous polls on the proportion of the population generally likely to consider a President's actions immoral.

a first (direct) simulation

The R code is in betabin.r at the course web page

- We present Bayesian inference about the probability of an adult responding ‘immoral’ assuming different Beta priors:
  1. $\alpha = \beta = 1$ prior information $\sim$ 0 $E = 1/2$
  2. $\alpha = \beta = 0.001$ prior information $<$ 0 $E = 1/2$
  3. $\alpha = 1$ $\beta = 0.11$ prior information $<$ 0 $E = 0.9$
  4. $\alpha = 1.8$ $\beta = 0.2$ prior information $\sim$ 0 $E = 0.9$
  5. $\alpha = 4.5,\ 45$ $\beta = 0.5,\ 5$ prior information $\sim$ 5,50 $E = 0.9$

1., 2. are both non informative, but 2. is a reasonable choice for ‘one-off’ events (or for correlated data) 3., 4. may be assumed on the basis of previous polls. Although E=0.9 they still are diffuse. 5., 6 are increasingly informative.



Legend for the next figure –>

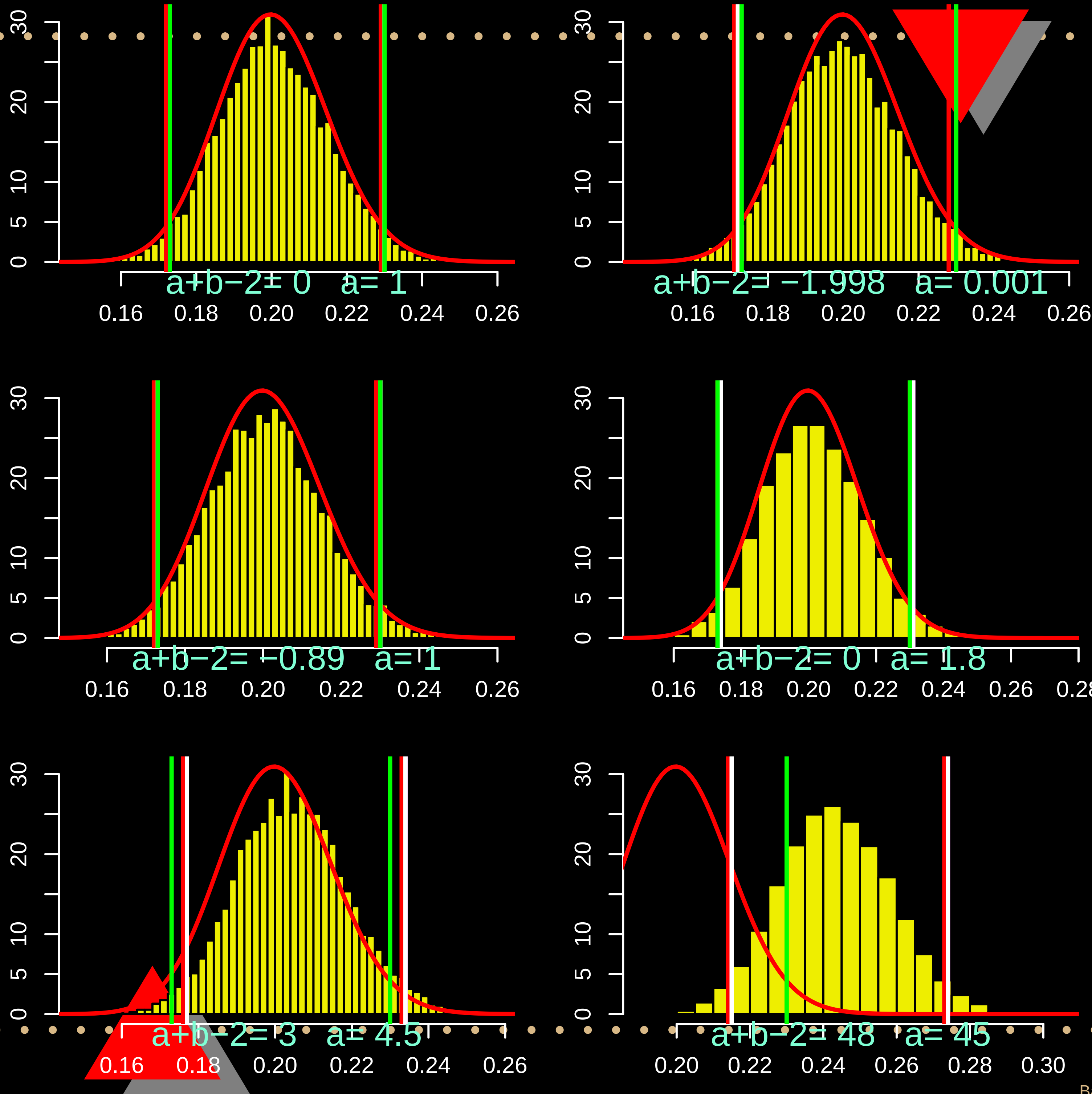
in each figure:

- curves: histogram of 10,000 draws from the posterior Beta(150+α,601+β); likelihood $\propto$ Bin(150,751). intervals: Unif-Bin 95% posterior interval; 95% (Beta(150+α,601+β)) posterior interval; Normal approximation of the 95% posterior interval; Inverted 95% posterior interval on the logit scale.

Legend for the next figure –>

in each figure:

- curves: histogram of 10,000 draws from the posterior Beta(150+α,601+β); likelihood $\propto$ Bin(150,751). intervals: Unif-Bin 95% posterior interval; 95% (Beta(150+α,601+β)) posterior interval; Normal approximation of the 95% posterior interval; Inverted 95% posterior interval on the logit scale.
- Though θ is close to 0, because of the large sample size (751), the normal approximation is good as well as posterior inferences are *insensitive* to prior choice (even if discordant to data), at least for prior information $\leq$ 0.

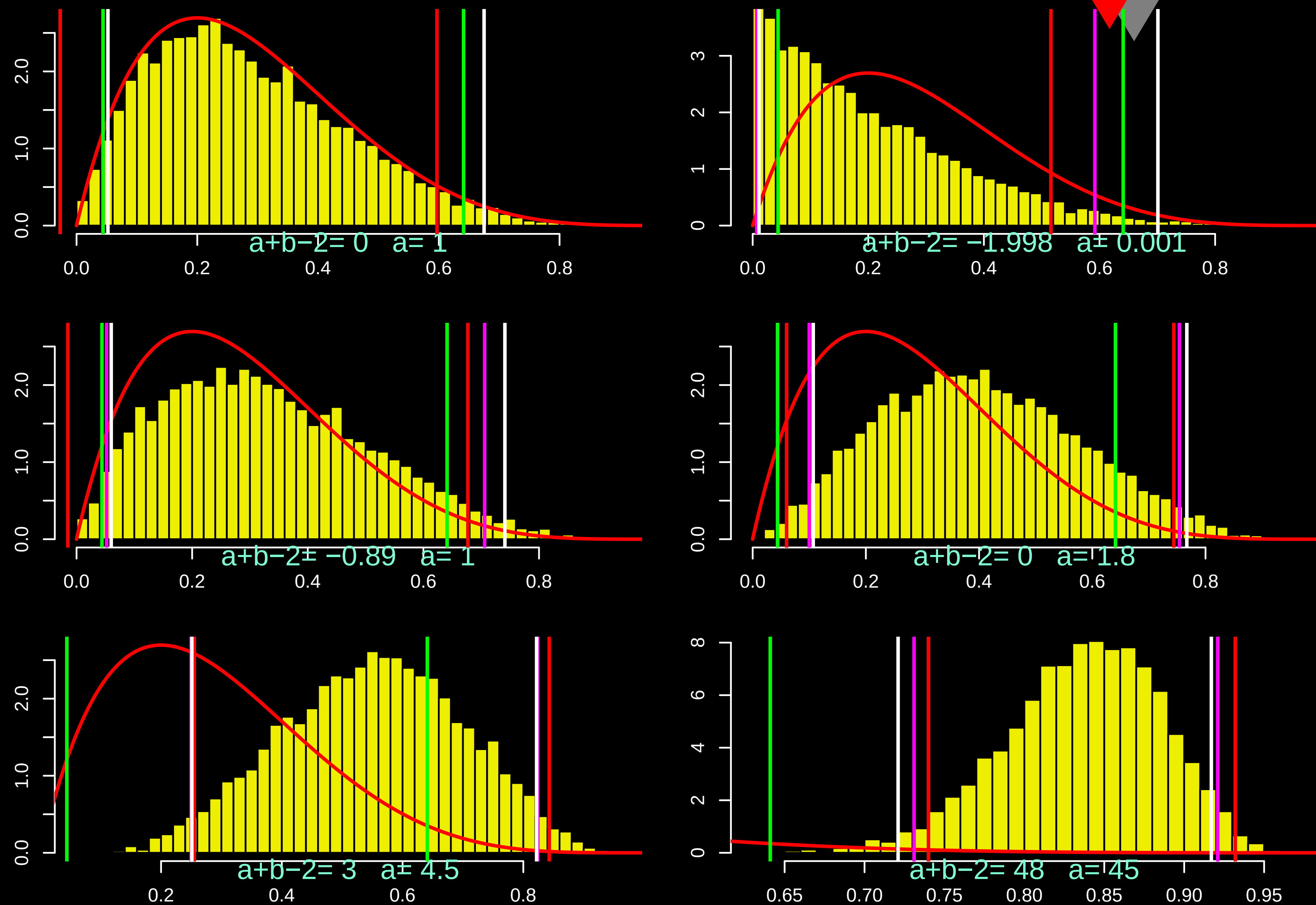


a+b−2= 0 a= 1
a+b−2= −1.998 a= 0.001
a+b−2= −0.89 a= 1
a+b−2= 0 a= 1.8
a+b−2= 3 a= 4.5
a+b−2= 48 a= 45

And what about if our sample size was only

$n = 5$, *with* $y = 1$ *adults considering immoral the President's actions? -*$>$

NOTE: the empirical mean still is $y/n = 0.2$


post
a+b−2= 0 a= 1
a+b−2= −1.998 a= 0.001
a+b−2= −0.89 a= 1
a+b−2= 0 a= 1.8
a+b−2= 3 a= 4.5
a+b−2= 48 a= 45