Modelli a parametro singolo

Analisi Bayesiana di dati binomiali

Matilde Trevisani

Lab 1 del corso di Inf Stat Bayesiana

Il modello Binomiale

In $Bin(n, \theta)$ c'è un solo parametro di interesse (tipicamente, n si assume noto), i.e. la **probabilità** θ di un certo risultato in ognuna delle n prove considerate.

Stima Bayesiana di una probabilità da dati binomiali

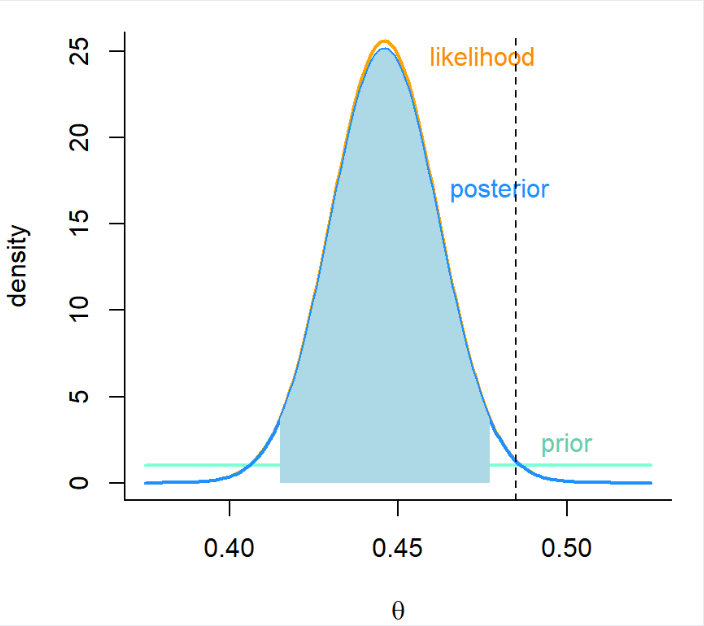
BDA3: pag. 37, sez. 2.4, codice R `placenta.r` nelle note del Lab

- interesse nella proporzione di nascite femminili nella cd condizione materna "placenta previa"
- i dati consistono in un precedente studio in Germania: 437 femmine su 980 nascite "placenta previa"
- con che evidenza i dati ci dicono che la proporzioe di nascite femminili "placenta previa" è minore della proporzione di nascite femminili nella popolazione generale pari a 0.485 ?

Analisi con una a priori uniforme

Il parametro 1-dimensionale θ denoti la proporzione di nascite femminili "placenta previa"

- Assumiamo che $Bin(\theta, 980) \; \propto \; \theta^{437} \, (1-\theta)^{980-437}$ sia il modello che abbia generato i dati
- Specifichiamo la a priori per θ essere una uniforme $U[0,1]$
- La a posteriori per θ è, quindi, $\propto \theta^{437} \, (1-\theta)^{980-437}$, i.e., è una $Beta(437+1, 980-437+1)$.



Analisi con a priori Beta

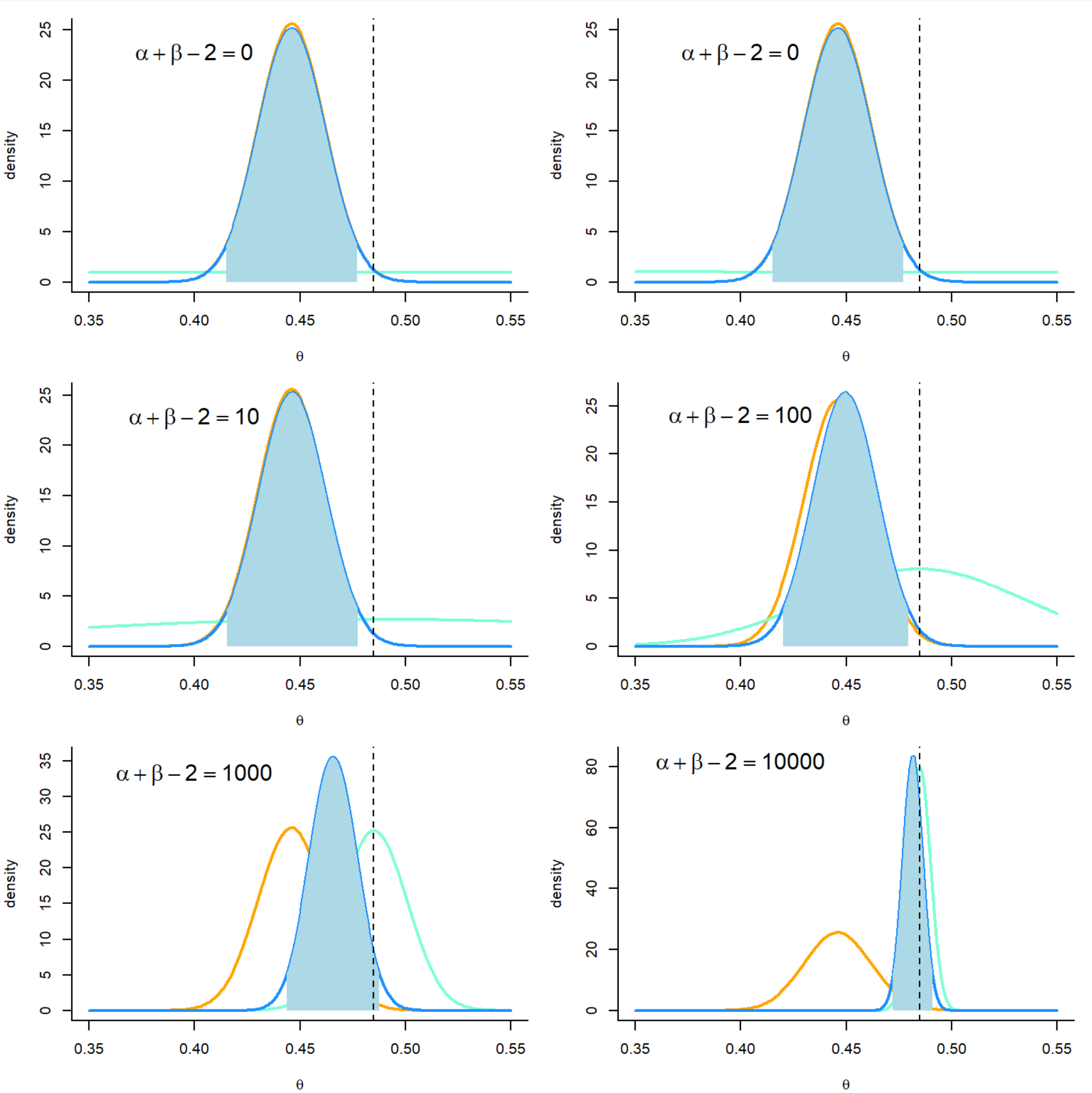
Poichè la verosimiglianza $p(y|\theta) \equiv L(\theta; y)$ è $\propto \theta^y \, (1-\theta)^{n-y}$, se la priori è della stessa forma, e.g., $p(\theta)$ è $\propto$

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

allora la posteriori sarà anche di questa forma. Infatti, $p(\theta|y)$ è $\propto$

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

La distribuzione a priori Beta è una **famiglia coniugata** per la verosimiglianza binomiale.



α + β − 2 = 0
α + β − 2 = 0
α + β − 2 = 10
α + β − 2 = 100
α + β − 2 = 1000
α + β − 2 = 10000
density
θ

Quale compromesso la a posteriori realizza tra la a priori e i dati?

- Il compromesso dipende da quanto peso (**informativa**) la a priori ha (è) rispetto ai dati disponibili

- i.e., nel caso binomiale, dipende dal peso relativo di

  $$\alpha + \beta - 2$$

  $\approx$ il numero delle "osservazioni a priori" ($\sim$ **precisione** della a priori) rispetto a

  $$n$$

  la grandezza campionaria.

Nota: $\text{precisione} = \frac{1}{\text{varianza}}$ dove $\text{var} = \frac{\mu(1-\mu)}{\alpha+\beta+1}$

Una prima analisi di sensitività

Concetto di sensitività: **sensitività** o **robustezza** dell'inferenza rispetto alla scelta della a priori

A former sensitivity analysis

Prior information		Posterior information			
alpha + beta - 2	**mean**	**mean**	**median**	**95% interval**	**length**
0	0.5	0.446	0.446	[0.415 , 0.477]	0.062
0	0.485	0.446	0.446	[0.415 , 0.477]	0.062
10	0.485	0.446	0.446	[0.416 , 0.477]	0.062
100	0.485	0.45	0.45	[0.42 , 0.479]	0.059
1000	0.485	0.466	0.466	[0.444 , 0.488]	0.044
10000	0.485	0.482	0.482	[0.472 , 0.491]	0.019

Si noti che nello studio $\hat{p} = .446$ e $n \approx 1000$.

Un approccio alla stima basato sulla simulazione

- L'approccio moderno alla stima bayesiana è diventato strettamente legato ai **metodi di stima basati sulla simulazione**.
- In effetti, la stima bayesiana si concentra sulla **stima dell'intera densità di un parametro**.
- Questa stima della densità si basa sulla **generazione di campioni dalla densità a posteriori** dei parametri stessi o di funzioni dei parametri.

Nel modello BETA-BINOMIALE, la coniugatezza ci permette di conoscere la densità a posteriori in forma chiusa.

- Quindi, è fattibile sia il calcolo diretto che la **simulazione diretta** dalla densità a posteriori.
- Tuttavia, anche se la densità a posteriori non può essere integrata in modo esplicito, i metodi di **simulazione iterativa** (o **MCMC**) vengono in alternativa utilizzati.

Una prima simulazione (diretta)

Congdon, pag. 77, ex. 3.4, Program 2.11

- Wilcox (1996) presenta i dati di un sondaggio d'opinione (fatto da *Gallup Poll*, agenzia statunitense pubblica per sondaggi d'opinione) sulla moralità del presidente Bush che non aiutò i gruppi ribelli iracheni dopo la fine formale della Guerra del Golfo. Dei 751 adulti che risposero, 150 ritennero che le azioni del presidente non furono etiche.

- Ci interessa valutare la probabilità che un adulto campionato a caso giudichi "immorale" il presidente.

- Nell'inferenza potremmo usare le evidenze di precedenti sondaggi sulla proporzione della popolazione che in generale può considerare le azioni di un presidente immorali.

Beta: alcune note

A priori $\pi \sim \mathcal{B}eta(\alpha, \beta)$

Interpretazione dei parametri: assumiamo che si siano osservati α successi e β fallimenti in un campione di $s = \alpha + \beta$ prove, con

$$E(\pi) = \mu = \frac{\alpha}{\alpha + \beta} \quad \mathrm{var}(\pi) = V = \frac{\mu(1-\mu)}{\alpha + \beta + 1}$$

Beta può anche essere espressa come $\mathcal{B}eta(s\mu, s(1-\mu))$.

Si noti come non sia del tutto chiaro cosa significhi definire una a priori non-informativa.

Ad es, l'a priori uniforme $\mathcal{B}eta(1,1)$ comporta una media a posteriori $(y+1)/(n+2)$, come anche una $\mathcal{B}eta(\epsilon, \epsilon)$ con $\epsilon \to 0$ comporta una media a posteriori che tende alla MLE y/n.

Tuttavia, una $\mathcal{B}eta(0,0)$ può essere vista come informativa dato che si riduce a due punti massa 0 e 1 come per i casi meno estremi di bimodalità a priori, e.g., $\mathcal{B}eta(0.5, 0.5)$ ove $p(\pi) \propto \frac{1}{\sqrt{\pi(1-\pi)}}$

Diversi gradi di "informatività" della priori Beta

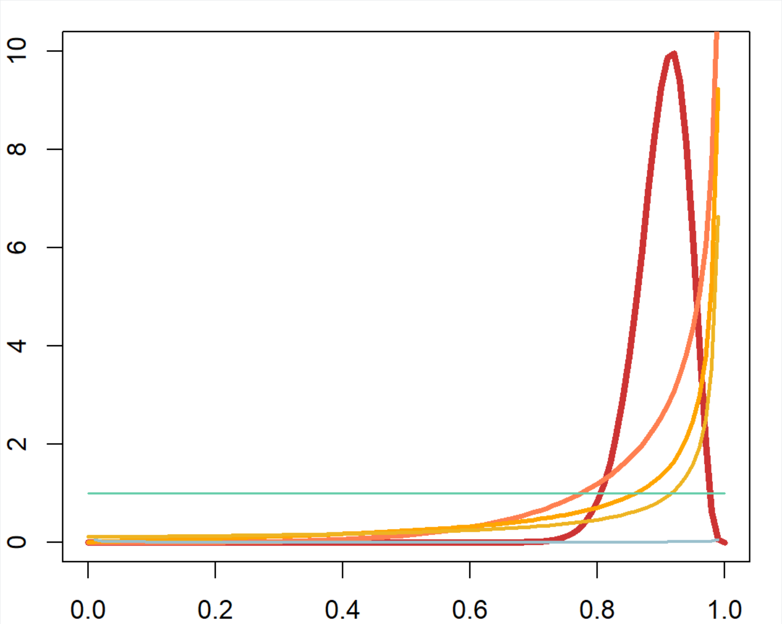
Presentiamo l'inferenza bayesiana sulla probabilità che un adulto risponda "immorale" assumendo diverse a priori Beta:

1. $\alpha = \beta = 1$ prior information ~ 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 ~ 0 $E = 0.9$
5. $\alpha = 4.5,\ \beta = 0.5$ prior information ~ 3 $E = 0.9$
6. $\alpha = 45,\ \beta = 5$ prior information ~ 48 $E = 0.9$

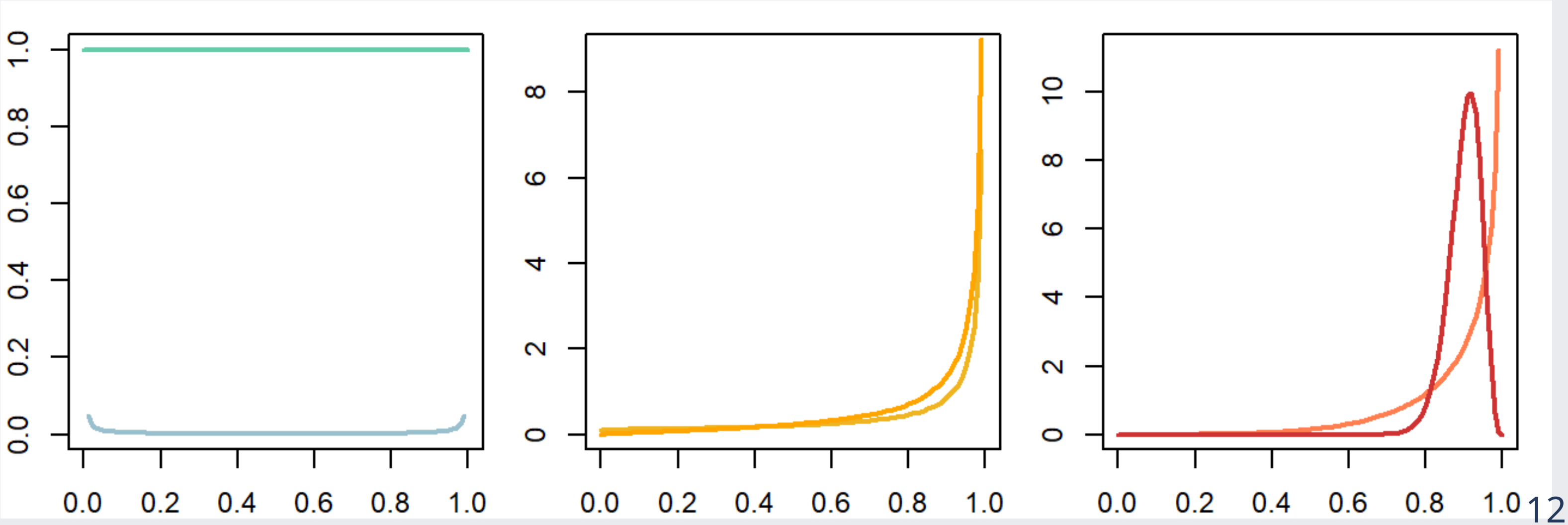
Priori 1. e 2. sono non-informative, sebbene 2. possa essere più specifica per eventi *one-off* (o correlati)

Priori 3. e 4. possono essere assunte sulla base di precedenti indagini, ma, sebbene $E = 0.9$, rimangono *diffuse*, cioè sono poco informative

Priori 5. e 6. sono crescentemente informative, Le a priori



distinte per grado di "informatività"



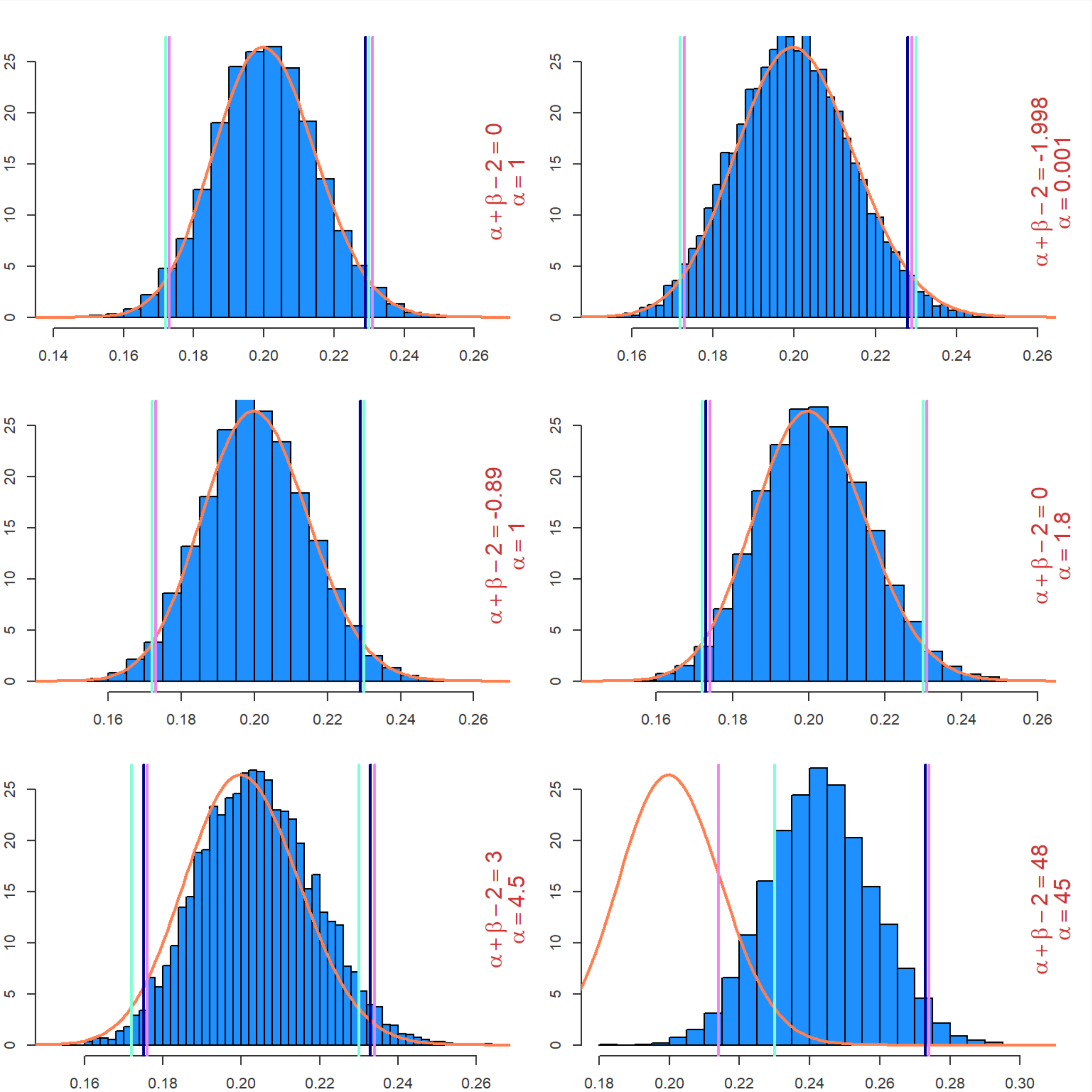
Inferenza Bayesiana sulla base della simulazione diretta

In ogni figura, curve:

- likelihood $\propto \mathcal{B}in(150, 751)$
- posteriori $\mathcal{B}eta(\alpha + 150, \beta + 601)$: istogramma di 10,000 **estrazioni** (dirette)

intervalli a posteriori 95%

- del modello Uniforme-Binomiale (come riferimento a non-informatività)
- del modello Beta-Binomiale (calcolo esatto)
- utilizzando un'approssimazione normale sulle **estrazioni** della simulazione
- invertendo l'intervallo costruito su scala logit sulle **estrazioni** della simulazione


α + β − 2 = 0
α = 1
α + β − 2 = -1.998
α = 0.001
α + β − 2 = -0.89
α = 1
α + β − 2 = 0
α = 1.8
α + β − 2 = 3
α = 4.5
α + β − 2 = 48
α = 45

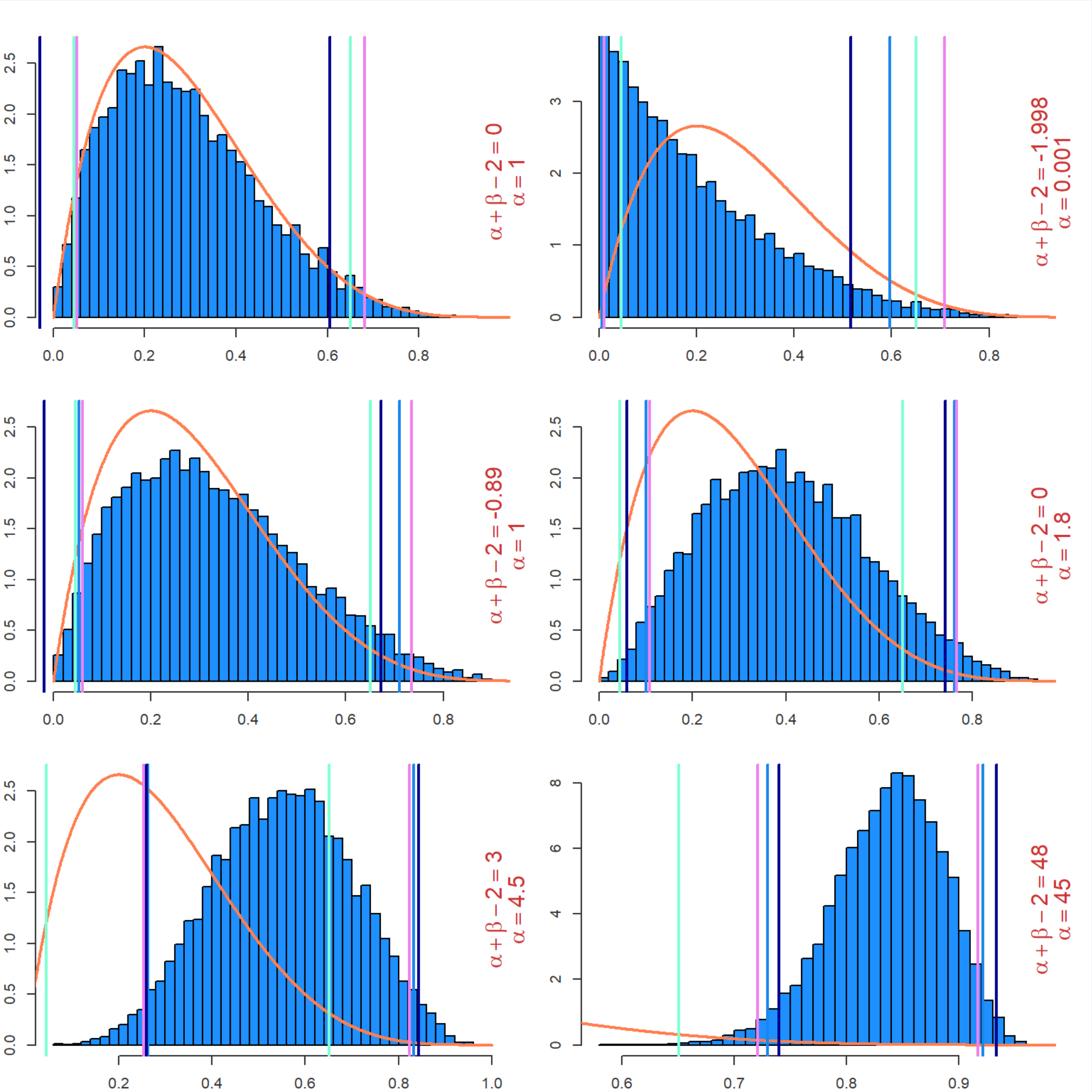
Campione grande

Sebbene θ sia non lontano da 0 (~ 0.2), poichè il campione è grande ($n = 751$), l'approssimazione normale è buona così come le inferenze a posteriori sono robuste rispetto alla priori scelta (anche se discorde rispetto ai dati), almeno per un'informazione a priori ≤ 0

Campione piccolo

Cosa accade se $n = 5$ e $y = 1$ adulti che ritengono immorali le azioni del presidente?

Si noti che la proporzione empirica è sempre 0.2.


α + β − 2 = 0
α = 1
α + β − 2 = -1.998
α = 0.001
α + β − 2 = -0.89
α = 1
α + β − 2 = 0
α = 1.8
α + β − 2 = 3
α = 4.5
α + β − 2 = 48
α = 45

Large sanple

Prior information		Posterior information				
mean	**sd**	**mean**	**sd**	**95% L**	**95% U**	**length**
0.500	0.289	0.201	0.015	0.173	0.230	0.057
0.500	0.500	0.200	0.015	0.171	0.229	0.058
0.901	0.206	0.201	0.015	0.173	0.230	0.057
0.900	0.173	0.202	0.015	0.173	0.231	0.058
0.900	0.122	0.204	0.015	0.177	0.234	0.057
0.900	0.042	0.243	0.015	0.214	0.275	0.061

Tiny sanple

Prior information		Posterior information				
mean	**sd**	**mean**	**sd**	**95% L**	**95% U**	**length**
0.500	0.289	0.286	0.160	0.041	0.646	0.605
0.500	0.500	0.200	0.163	0.007	0.608	0.601
0.901	0.206	0.327	0.176	0.053	0.710	0.657
0.900	0.173	0.400	0.173	0.100	0.753	0.653
0.900	0.122	0.550	0.150	0.255	0.827	0.572
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

In WinBUGS

Program 2.11 Presidential Actions

BUGS code for Example 3.4 Presidential Actions in *Bayesian Statistical Modelling* (Congdon, 2nd ed.)

Blith (1986) suggerisce che quando $y = 0$, il limite superiore di un CI al 95% dovrebbe essere $1 - \alpha^{1/n}$ invece che vicino a 0 (come risulta se si usa l'approssimazione usuale).

Per $n = 8$ e $\alpha = .05$ il limite estremo di in CI al 95% risulterebbe quindi 0.312344.

Lo si confronti con i risultati nell'inferenza Bayesiana, considerando l'effetto delle due a priori diffuse.

Confidence interval around binomial estimate of 0 or 1

What is the best technique to calculate a confidence interval of a binomial experiment, if your estimate is that $p = 0$ (or similarly $p = 1$) and sample size is relatively small

A general advice is to never use the normal approximation (i.e., the asymptotic / Wald confidence interval),

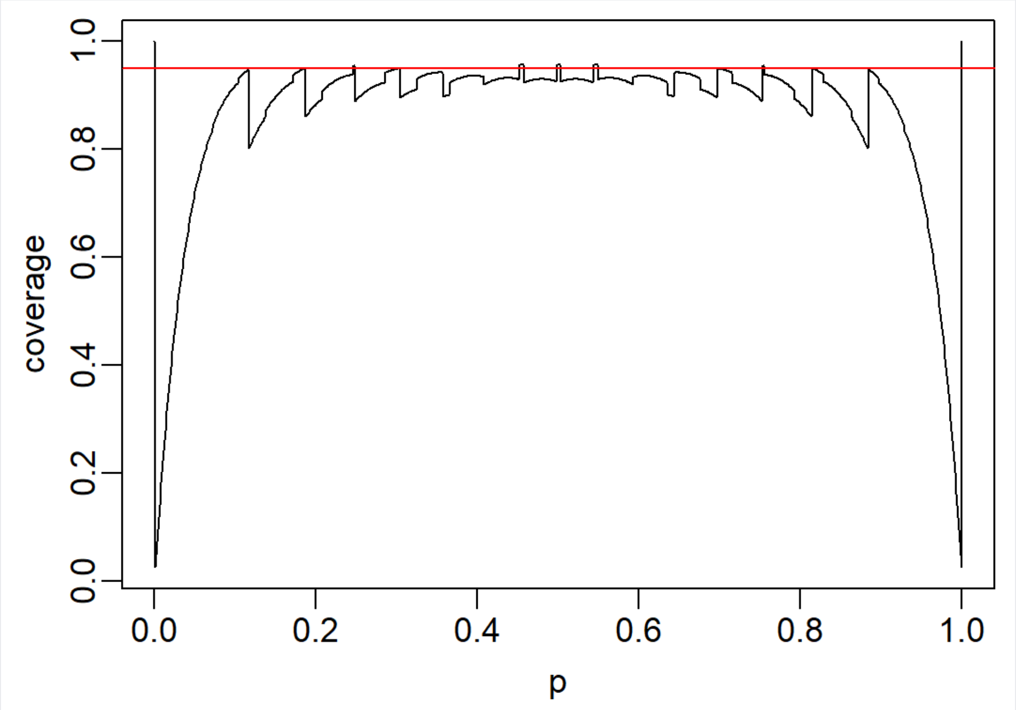
$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

as it has terrible coverage properties.

The standard interval is often presented with the caveat that it should be usedonly when $n \cdot \min(p, 1-p)$ is at least 5 (or 10).

See the erratic behavior of the coverage probability of the standard Wald confidence interval

```
library(binom)
p <- seq(0,1,.001)
coverage <- binom.coverage(p, n=25, method="asymptotic")$coverage
plot(p, coverage, type="l") ; abline(h=.95, col="red")
```



For small success probabilities, you might ask for a 95% confidence interval, but actually get, say, a 10% confidence interval!

Different methods to obtain a confidence interval on the binomial probability.

```
binom.confint(x=0, n=8)
```

```
##           method x n       mean       lower     upper
## 1  agresti-coull 0 8 0.00000000 -0.04776038 0.3721679
## 2     asymptotic 0 8 0.00000000  0.00000000 0.0000000
## 3          bayes 0 8 0.05555556  0.00000000 0.2075080
## 4        cloglog 0 8 0.00000000  0.00000000 0.3694166
## 5          exact 0 8 0.00000000  0.00000000 0.3694166
## 6          logit 0 8 0.00000000  0.00000000 0.3694166
## 7         probit 0 8 0.00000000  0.00000000 0.3694166
## 8        profile 0 8 0.00000000  0.00000000 0.3203933
## 9            lrt 0 8 0.00000000  0.00000000 0.2134440
## 10     prop.test 0 8 0.00000000  0.00000000 0.4022967
## 11        wilson 0 8 0.00000000  0.00000000 0.3244076
```

Brown et al. (in *Statistical Science*, 2001) recommend the Wilson interval or the equal-tailed Jeffreys prior interval for small n and the interval suggested in Agresti and Coull for larger n.