From the six boxes
to the Bayes ‘billiard’

$\Rightarrow$ Introducing parametric inference

Inferring ‘proportions’

Let’s turn the toy experiment to a ‘serious’ physics case:

- Inferring H_j is the same as inferring the proportion of white balls:

$$H_j \quad \longleftrightarrow \quad j \quad \longleftrightarrow \quad p = \frac{j}{5}$$

Inferring 'proportions'

Let's turn the toy experiment to a 'serious' physics case:

- Inferring H_j is the same as inferring the proportion of white balls:
$$H_j \quad \longleftrightarrow \quad j \quad \longleftrightarrow \quad p = \frac{j}{5}$$
- Increase the number of balls
$$n: \quad 6 \to \infty$$

$\Rightarrow$ p continuous in $[0, 1]$

Inferring 'proportions'

Let's turn the toy experiment to a 'serious' physics case:

- Inferring H_j is the same as inferring the proportion of white balls:

$$H_j \quad \longleftrightarrow \quad j \quad \longleftrightarrow \quad p = \frac{j}{5}$$

- Increase the number of balls

$$n: \quad 6 \rightarrow \infty$$

$\Rightarrow$ p continuous in $[0, 1]$

- Generalize White/Black $\longrightarrow$ Success/Failure

$\Rightarrow$ efficiencies, branching ratios, ...

Bayes' *billiard*

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

Bayes' *billiard*

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

- imagine you roll a ball at random on a billiard;

Bayes' *billiard*

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

- imagine you roll a ball at random on a billiard;
- you mark the relative position of the ball along the billiard's length (l/L) and remove the ball

Bayes' *billiard*

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

- imagine you roll a ball at random on a billiard;
- you mark the relative position of the ball along the billiard's length (l/L) and remove the ball
- then you roll at random other balls

Bayes' *billiard*

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

- imagine you roll a ball at random on a billiard;
- you mark the relative position of the ball along the billiard's length (l/L) and remove the ball
- then you roll at random other balls
  - write down if it stopped left or right of the first ball;

Bayes' *billiard*

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

- imagine you roll a ball at random on a billiard;
- you mark the relative position of the ball along the billiard's length (l/L) and remove the ball
- then you roll at random other balls
  - write down if it stopped left or right of the first ball;
  - remove it and go on with n balls.

Bayes' *billiard*

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

- imagine you roll a ball at random on a billiard;
- you mark the relative position of the ball along the billiard's length (l/L) and remove the ball
- then you roll at random other balls
  - write down if it stopped left or right of the first ball;
  - remove it and go on with n balls.
- Somebody has to guess the position of the first ball knowing only how many balls stopped left and how many stopped right

Bayes' *billiard*

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

- imagine you roll a ball at random on a billiard;
- you mark the relative position of the ball along the billiard's length (l/L) and remove the ball
- then you roll at random other balls
  - write down if it stopped left or right of the first ball;
  - remove it and go on with n balls.
- Somebody has to guess the position of the first ball knowing only how many balls stopped left and how many stopped right

Bayes' *billiard*

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

- imagine you roll a ball at random on a billiard;
- you mark the relative position of the ball along the billiard's length (l/L) and remove the ball
- then you roll at random other balls
  - write down if it stopped left or right of the first ball;
  - remove it and go on with n balls.
- Somebody has to guess the position of the first ball knowing only how many balls stopped left and how many stopped right

Although Bayes never mentioned a *billiard*

Bayes' *billiard*

This is the original problem in the theory of chances solved by Thomas Bayes in late '700:

- imagine you roll a ball at random on a billiard;
- you mark the relative position of the ball along the billiard's length (l/L) and remove the ball
- then you roll at random other balls
  - write down if it stopped left or right of the first ball;
  - remove it and go on with n balls.
- Somebody has to guess the position of the first ball knowing only how many balls stopped left and how many stopped right

Although Bayes never mentioned a *billiard*

Bayes' billiard and Bernoulli trials

It is easy to recognize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Bayes' billiard and Bernoulli trials

It is easy to recognize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
    - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

Bayes' billiard and Bernoulli trials

It is easy to recognize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
    - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

$$f(p\,|\,S) \quad \propto \quad f(S\,|\,p) = p$$

Bayes' billiard and Bernoulli trials

It is easy to recognize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

$$\begin{aligned} f(p\,|\,S) &\propto f(S\,|\,p) = p \\ f(p\,|\,S,S) &\propto f(S\,|\,p)\cdot f(p\,|\,S) = p^2 \end{aligned}$$

Bayes' billiard and Bernoulli trials

It is easy to recognize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

$$\begin{aligned} f(p\,|\,S) &\propto f(S\,|\,p) = p \\ f(p\,|\,S,S) &\propto f(S\,|\,p)\cdot f(p\,|\,S) = p^2 \\ f(p\,|\,S,S,F) &\propto f(F\,|\,p)\cdot f(p\,|\,S,S) = p^2(1-p) \end{aligned}$$

Bayes' billiard and Bernoulli trials

It is easy to recognize the analogy:

- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

$$\begin{aligned}
f(p\,|\,S) &\propto f(S\,|\,p) = p \\
f(p\,|\,S,S) &\propto f(S\,|\,p)\cdot f(p\,|\,S) = p^2 \\
f(p\,|\,S,S,F) &\propto f(F\,|\,p)\cdot f(p\,|\,S,S) = p^2(1-p) \\
\ldots &\quad\ \ldots \\
f(p\,|\,\#S,\#F) &\propto p^{\#S}(1-p)^{\#F} = p^{\#S}(1-p)^{(n-\#s)}
\end{aligned}$$

Bayes' billiard and Bernoulli trials

It is easy to recognize the analogy:

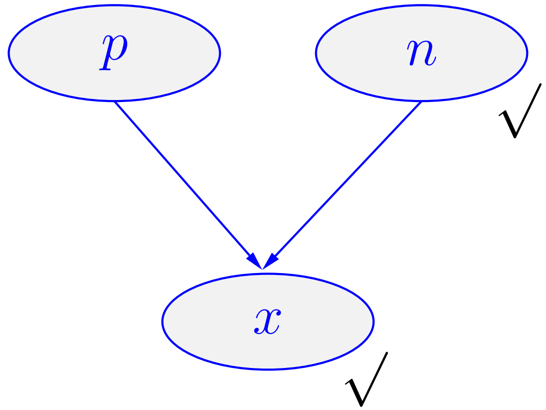
- Left/Right $\rightarrow$ Success/Failure
- if Left $\leftrightarrow$ Success:
  - $l/L \leftrightarrow p$ of binomial (Bernoulli trials)

Solution with modern notation:
Imagine a sequence $\{S, S, F, S, \ldots\}$ [f_0 is uniform]:

$$\begin{aligned}
f(p\,|\,S) &\propto f(S\,|\,p) = p \\
f(p\,|\,S,S) &\propto f(S\,|\,p)\cdot f(p\,|\,S) = p^2 \\
f(p\,|\,S,S,F) &\propto f(F\,|\,p)\cdot f(p\,|\,S,S) = p^2(1-p) \\
\ldots & \quad\ \ldots \\
f(p\,|\,\#S,\#F) &\propto p^{\#S}(1-p)^{\#F} = p^{\#S}(1-p)^{(n-\#s)} \\
f(p\,|\,x,n) &\propto p^x(1-p)^{(n-x)} \qquad\qquad [x=\#S]
\end{aligned}$$

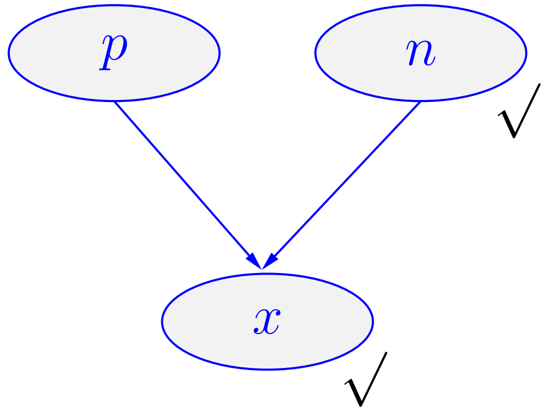
n independent Bernoulli processes

Inferring p



n independent Bernoulli processes

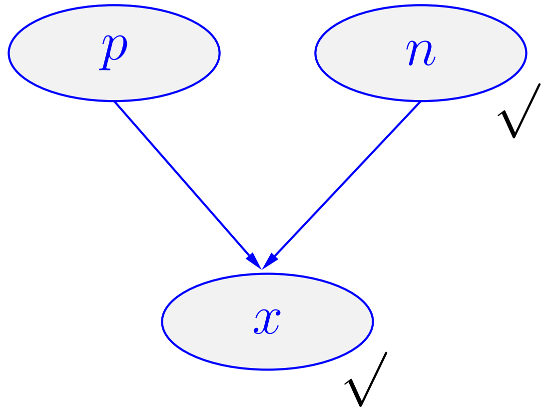
Inferring *p*



$$f(p \mid x, n) = \frac{f(p, x \mid n)}{f(x \mid n)}$$

n independent Bernoulli processes

Inferring p

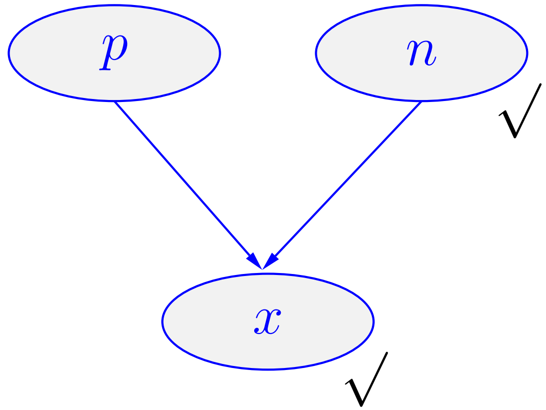


$$f(p\,|\,x,n) = \frac{f(p,x\,|\,n)}{f(x\,|\,n)}$$

$$= \frac{f(x\,|\,n,p)\cdot f_0(p)}{f(x\,|\,n)}$$

n independent Bernoulli processes

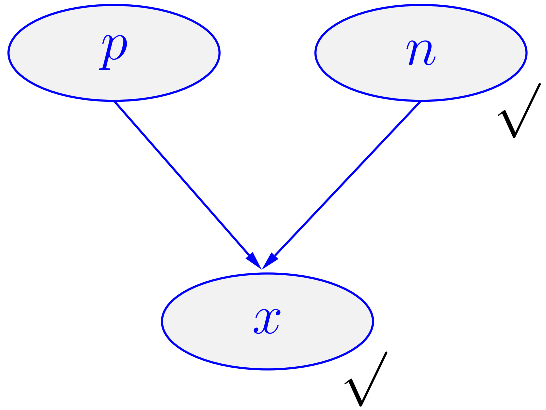
Inferring p



$$
\begin{aligned}
f(p\,|\,x,n) &= \frac{f(p,x\,|\,n)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{\int_0^1 f(x\,|\,n,p)\cdot f_0(p)\,dp}
\end{aligned}
$$

n independent Bernoulli processes

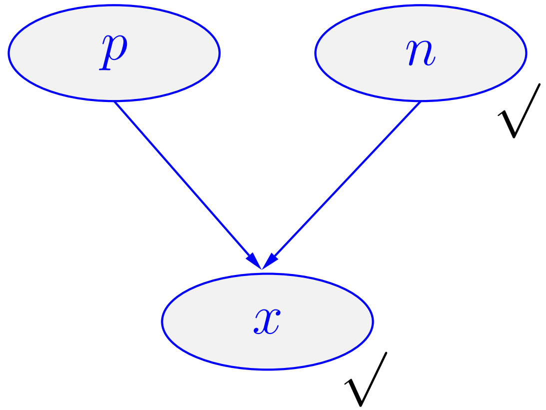
Inferring p



$$
\begin{aligned}
f(p \,|\, x, n) &= \frac{f(p, x \,|\, n)}{f(x \,|\, n)} \\
&= \frac{f(x \,|\, n, p) \cdot f_0(p)}{f(x \,|\, n)} \\
&= \frac{f(x \,|\, n, p) \cdot f_0(p)}{\int_0^1 f(x \,|\, n, p) \cdot f_0(p)\, dp} \\
&\propto f(x \,|\, n, p) \cdot f_0(p)
\end{aligned}
$$

n independent Bernoulli processes

Inferring p



$$
\begin{aligned}
f(p\,|\,x,n) &= \frac{f(p,x\,|\,n)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{f(x\,|\,n)} \\
&= \frac{f(x\,|\,n,p)\cdot f_0(p)}{\int_0^1 f(x\,|\,n,p)\cdot f_0(p)\,dp} \\
&\propto f(x\,|\,n,p)\cdot f_0(p)
\end{aligned}
$$

(denominator just normalization!)

Inferring the “Bernoulli's p”

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas (“Bayesian”) has always been a severe problem!

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented (like what we have called the "Gaussian trick").

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented (like what we have called the "Gaussian trick").

Here is a very elegant one, particularly suitable useful to infer Bernoulli's p.

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$\rightarrow$ **Computational barrier**

Some tricks have been invented (like what we have called the "Gaussian trick").
Here is a very elegant one, particularly suitable useful to infer Bernoulli's p.

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a (1-p)^b$$

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$$\rightarrow \textbf{Computational barrier}$$

Some tricks have been invented (like what we have called the "Gaussian trick").

Here is a very elegant one, particularly suitable useful to infer Bernoulli's p.

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a\,(1-p)^b$$

- Then the inference becomes

$$f(p\,|\,x,n) \;\;\propto\;\; p^x\,(1-p)^{n-x}\cdot p^a\,(1-p)^b$$

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$$\rightarrow \textbf{Computational barrier}$$

Some tricks have been invented (like what we have called the "Gaussian trick").

Here is a very elegant one, particularly suitable useful to infer Bernoulli's p.

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a \, (1-p)^b$$

- Then the inference becomes

$$\begin{aligned} f(p \,|\, x, n) &\propto p^x \, (1-p)^{n-x} \cdot p^a \, (1-p)^b \\ &\propto p^{a+x} \, (1-p)^{b+(n-x)} \end{aligned}$$

Inferring the "Bernoulli's p"

Mathematically convenient priors

Before the advent of powerful computers, applying Laplace' ideas ("Bayesian") has always been a severe problem!

$$\rightarrow \textbf{Computational barrier}$$

Some tricks have been invented (like what we have called the "Gaussian trick").

Here is a very elegant one, particularly suitable useful to infer Bernoulli's p.

- imagine that we could express $f_0(p)$ in the following form

$$f_0(p) \propto p^a\,(1-p)^b$$

- Then the inference becomes

$$\begin{aligned} f(p\,|\,x,n) &\propto p^x\,(1-p)^{n-x}\cdot p^a\,(1-p)^b \\ &\propto p^{a+x}\,(1-p)^{b+(n-x)} \\ &\propto p^{a'}\,(1-p)^{b'} \end{aligned}$$

Beta distribution

Indeed, such a pdf exists ($a = r - 1$; $b = s - 1$).

Beta distribution

Indeed, such a pdf exists ($a = r - 1$; $b = s - 1$).
In general, given the generic uncertain number X,

$$f(x \,|\, \text{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \leq x \leq 1 \end{cases}$$

Beta distribution

Indeed, such a pdf exists ($a = r - 1$; $b = s - 1$).
In general, given the generic uncertain number X,

$$f(x \,|\, \mathrm{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \le x \le 1 \end{cases}$$

- The denominator is just for normalization, i.e.

$$\beta(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\,\mathrm{d}x$$

Beta distribution

Indeed, such a pdf exists ($a = r - 1$; $b = s - 1$).
In general, given the generic uncertain number X,

$$f(x \,|\, \text{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \le x \le 1 \end{cases}$$

- The denominator is just for normalization, i.e.

$$\beta(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\,\mathrm{d}x$$

Indeed this integral defines the beta function, resulting in

$$\beta(r,s) = \frac{\Gamma(r)\,\Gamma(s)}{\Gamma(r+s)}$$

Beta distribution

Indeed, such a pdf exists ($a = r - 1$; $b = s - 1$).
In general, given the generic uncertain number X,

$$f(x \,|\, \text{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\ s > 0 \\ 0 \le x \le 1 \end{cases}$$

- The denominator is just for normalization, i.e.

$$\beta(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\,\mathrm{d}x$$

Indeed this integral defines the beta function, resulting in

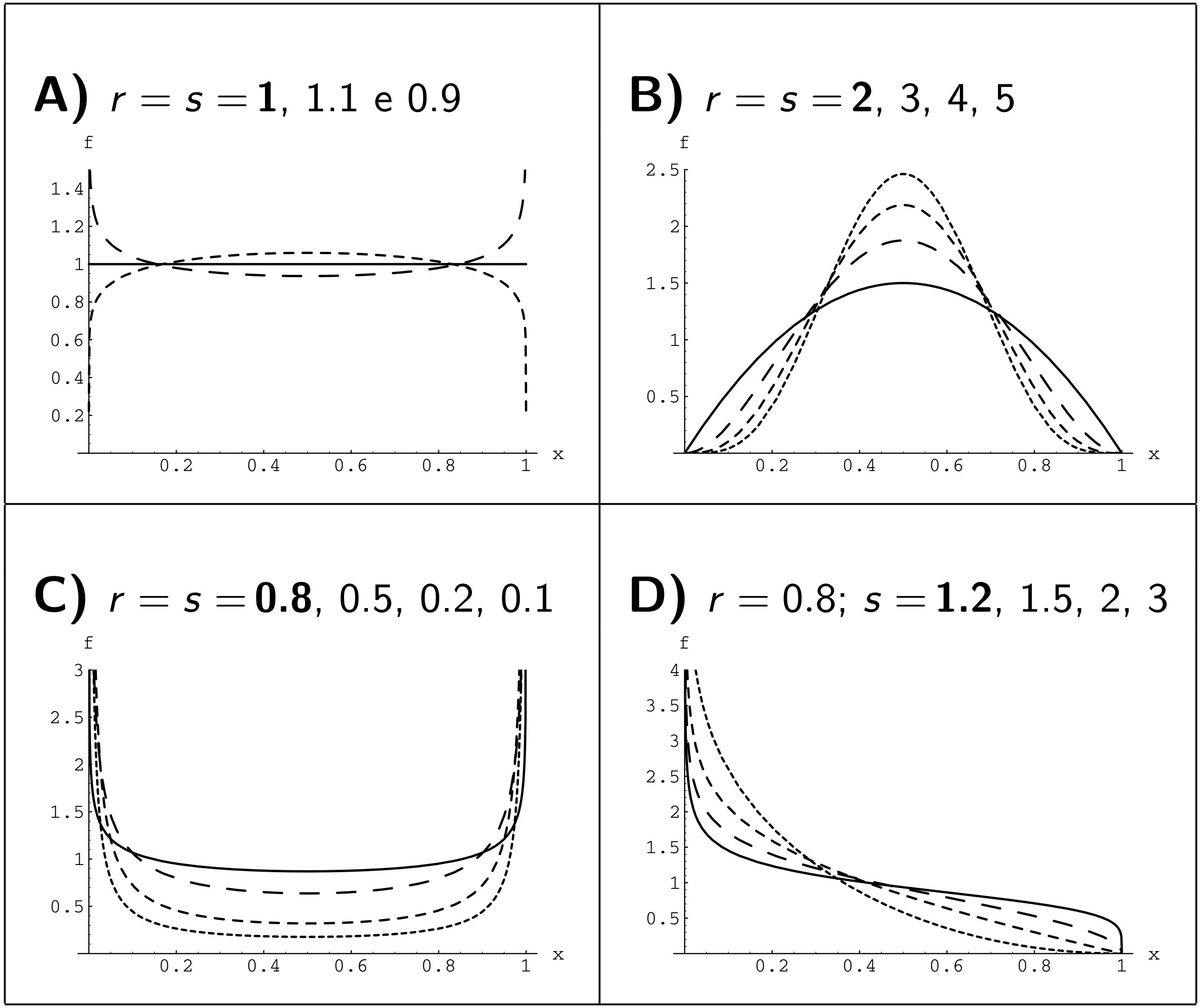
$$\beta(r,s) = \frac{\Gamma(r)\,\Gamma(s)}{\Gamma(r+s)}$$

Try e.g.

```
> p<-seq(0,1,by=0.01)
> plot(p, dbeta(p, 3, 5), ty='l', col='blue')
```

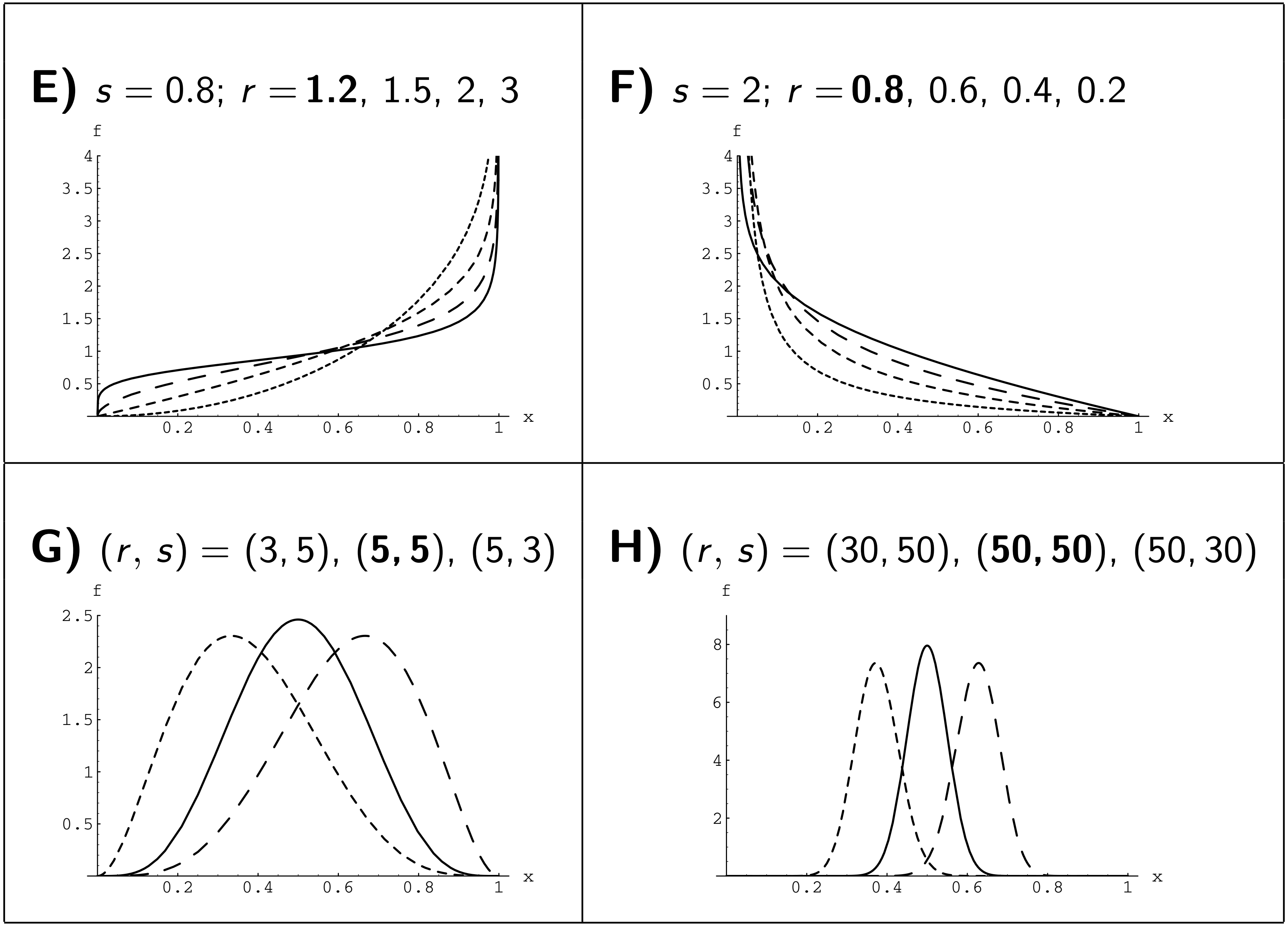
Beta distribution

Some examples



Beta distribution

Some examples



Beta distribution

Summaries

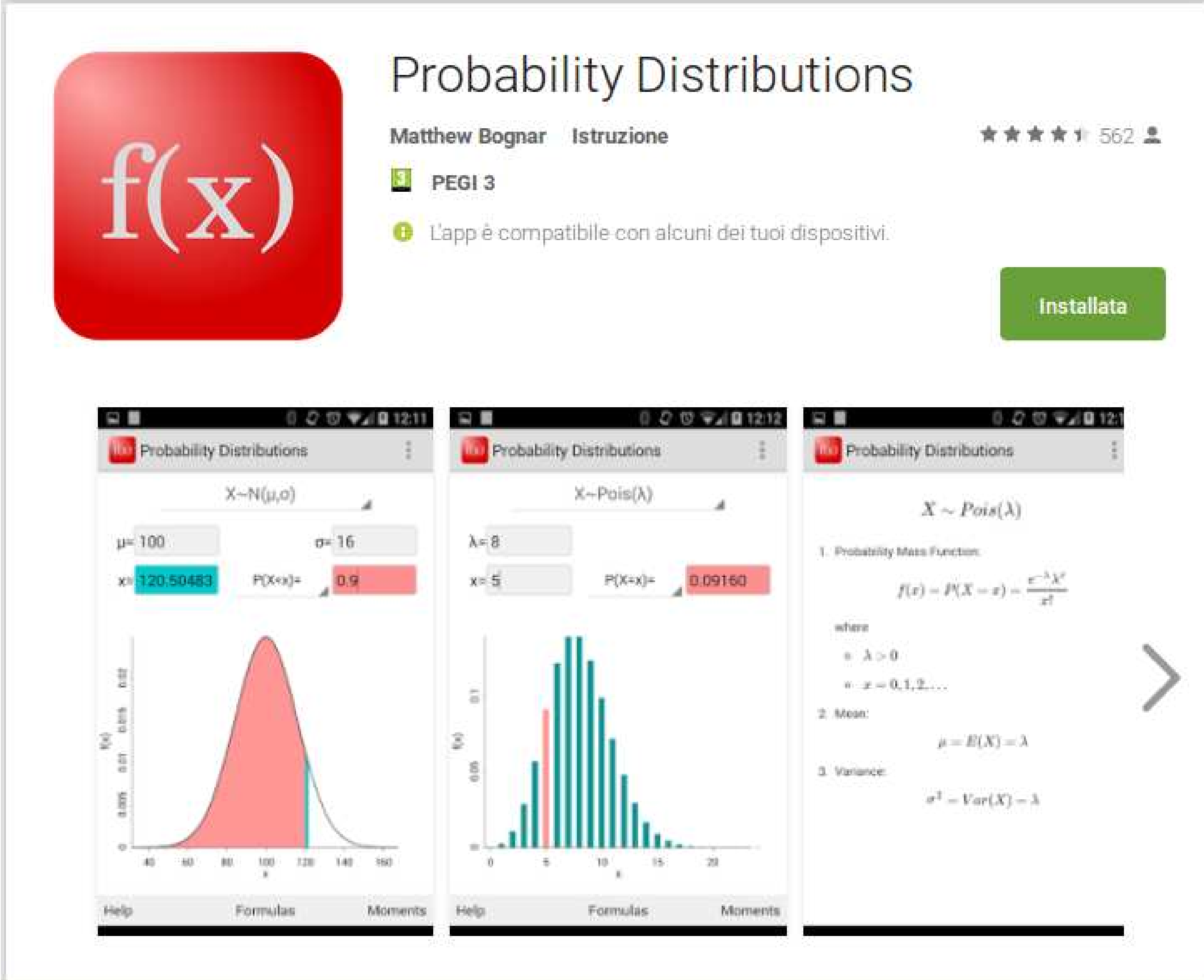
$$
\begin{aligned}
\mathrm{E}(X) &= \frac{r}{r+s} \\
\mathrm{Var}(X) &= \frac{rs}{(r+s+1)\,(r+s)^2}\,.
\end{aligned}
$$

Mode, unique if $r > 1$ and $s > 1$:

$$
\frac{r-1}{r+s-2}
$$

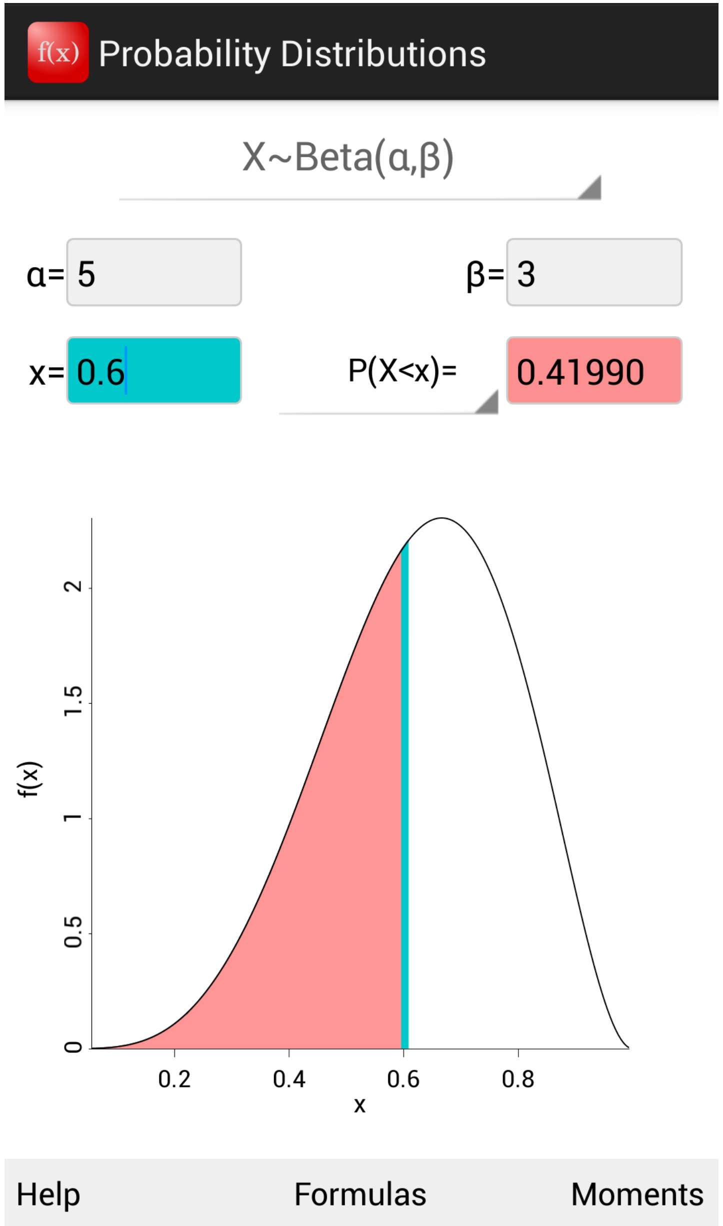
A useful app

https://play.google.com/store/apps/details?id=com.mbognar.probdist



A useful app

An example



Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$
\begin{aligned}
f(p \mid n, x, \text{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\
&\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1} \,.
\end{aligned}
$$

Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$
\begin{aligned}
f(p \,|\, n, x, \text{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i - 1}(1-p)^{s_i - 1}\right] \\
&\propto p^{x + r_i - 1}(1-p)^{n - x + s_i - 1}\,.
\end{aligned}
$$

Simple updating rule:

$$
\begin{aligned}
r_f &= r_i + x \\
s_f &= s_i + (n - x)
\end{aligned}
$$

Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$\begin{aligned} f(p \,|\, n, x, \text{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i - 1}(1-p)^{s_i - 1}\right] \\ &\propto p^{x + r_i - 1}(1-p)^{n - x + s_i - 1}\,. \end{aligned}$$

Simple updating rule:

$$\begin{aligned} r_f &= r_i + x \\ s_f &= s_i + (n - x) \end{aligned}$$

Check the case of **uniform prior** ($r_i = s_i = 1$)

$$\text{E}(X) = \frac{r_f}{r_f + s_f} = \frac{x+1}{n+2}$$

Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$
\begin{aligned}
f(p \,|\, n, x, \mathrm{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\
&\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1}\,.
\end{aligned}
$$

Simple updating rule:

$$
\begin{aligned}
r_f &= r_i + x \\
s_f &= s_i + (n-x)
\end{aligned}
$$

Check the case of **uniform prior** ($r_i = s_i = 1$)

$$
\begin{aligned}
\mathrm{E}(X) &= \frac{r_f}{r_f + s_f} = \frac{x+1}{n+2} \\
\mathrm{Var}(X) &= \frac{r_f\, s_f}{(r_f + s_f + 1)\,(r_f + s_f)^2} = \frac{(x+1)(n-x+1)}{(n+3)\,(n+2)^2}
\end{aligned}
$$

Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$
\begin{aligned}
f(p \,|\, n, x, \text{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\
&\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1}\,.
\end{aligned}
$$

Simple updating rule:

$$
\begin{aligned}
r_f &= r_i + x \\
s_f &= s_i + (n-x)
\end{aligned}
$$

Check the case of **uniform prior** ($r_i = s_i = 1$)

$$
\begin{aligned}
\text{E}(X) &= \frac{r_f}{r_f+s_f} = \frac{x+1}{n+2} \\
\text{Var}(X) &= \frac{r_f\, s_f}{(r_f+s_f+1)\,(r_f+s_f)^2} = \frac{(x+1)(n-x+1)}{(n+3)\,(n+2)^2} \\
\text{mode}(X) &= \frac{r_f-1}{r_f+s_f-2} = \frac{x}{n}
\end{aligned}
$$

Beta distribution as prior

Let us finally apply it to infer the Bernoulli's p

$$
\begin{aligned}
f(p \,|\, n, x, \text{Beta}(r_i, s_i)) &\propto \left[p^x (1-p)^{n-x}\right] \times \left[p^{r_i-1}(1-p)^{s_i-1}\right] \\
&\propto p^{x+r_i-1}(1-p)^{n-x+s_i-1}\,.
\end{aligned}
$$

Simple updating rule:

$$
\begin{aligned}
r_f &= r_i + x \\
s_f &= s_i + (n-x)
\end{aligned}
$$

Check the case of **uniform prior** ($r_i = s_i = 1$)

$$
\begin{aligned}
\text{E}(X) &= \frac{r_f}{r_f+s_f} = \frac{x+1}{n+2} \\
\text{Var}(X) &= \frac{r_f\, s_f}{(r_f+s_f+1)\,(r_f+s_f)^2} = \frac{(x+1)(n-x+1)}{(n+3)\,(n+2)^2} \\
\text{mode}(X) &= \frac{r_f-1}{r_f+s_f-2} = \frac{x}{n} \qquad \surd
\end{aligned}
$$

Conjugate priors

The Beta distribution is an example of conjugate prior:

Conjugate priors

The Beta distribution is an example of conjugate prior:

- a pdf such that prior and posterior belong to the same family;
- its parameters are updated by the the ‘likelihood’.

Conjugate priors

The Beta distribution is an example of conjugate prior:

- a pdf such that prior and posterior belong to the same family;
- its parameters are updated by the the 'likelihood'.

Note:

- not all *conjugate priors* are as flexible as the Beta.

Conjugate priors

The Beta distribution is an example of conjugate prior:

- a pdf such that prior and posterior belong to the same family;
- its parameters are updated by the the 'likelihood'.

Note:

- not all *conjugate priors* are as flexible as the Beta.

  (In particular, the Gaussian is self-conjugate,
  which is not so great...)