

What's the probability a baby will be a boy?

Jonathan Auerbach, Owen Ward

October 31, 2019

How do scientists estimate the probability of a theory?

```
library("knitr")  
library("HistData")  
library("tidyverse")  
theme_set(theme_bw())
```

# What's the probability a baby will be a boy?

- ▶ Laplace (1786) knew the number of male and female births in London and Paris.
- ▶ Arbuthnot previously had analysed rate in London, shown more males born than females.
- ▶ Laplace wanted to know:
  - ▶ Probability baby a boy
  - ▶ Whether this probability was different in London and Paris

Recall that Arbuthnot computed the probability of observing more boys in each of 82 years if boy/girl equally likely...

```
Arbuthnot %>%  
select(Year, Males, Females) %>%  
filter(Year < 1633 | Year > 1708) %>%  
kable()
```

Year	Males	Females
1629	5218	4683
1630	4858	4457
1631	4422	4102
1632	4994	4590
1709	7840	7380
1710	7640	7288

... that's equivalent to flipping a coin 82 times, landing heads every time

```
dbinom(82,82,prob=0.5)
```

```
## [1] 2.067952e-25
```

- ▶ This is called a “p-value”: the probability of such strong evidence if  $p = 0.5$  was essentially 0.
- ▶ Arbuthnot inferred the probability of male and female births different—but could not tell by how much.
- ▶ Laplace used Bayes statistics to answer this question. (Although he developed his solution independently of Bayes.)

## Bayes Statistics

- ▶ Thomas Bayes (1763), a British minister who published the theorem (posthumously) that bears his name.



$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

- ▶ Interpretation: Have some prior knowledge/belief about an event “A”, update this based on observing another event “B”.
- ▶ In Laplace’s problem, “A” is the likelihood a birth is a boy is some proportion, say .51, and “B” is the number of boy and girl birth observed.
- ▶ Laplace argued  $P(A) = P(A^c) = .5$  by the “principle of insufficient reasoning”

## Bayes and Laplace



## Laplace's solution

- ▶ Laplace provided methods to numerically estimate  $P(A|B)$  for every possible proportion with high precision several years later.
- ▶ This allowed him to compute the probability the rate was higher in Paris than London.
- ▶ Today, we can also approximate this integral by computing the numerator on a grid of points, divide by the sum of all values.



## MÉMOIRE SUR LES PROBABILITÉS\*

P. S. Laplace<sup>†</sup>

*Mémoires de l'Académie royale des Sciences de Paris*, 1778 (1781)  
*Oeuvres complètes* 9, pp. 227–332

### XVII.

When we have nothing given *a priori* on the possibility of an event, it is necessary to assume all the possibilities, from zero to unity, equally probable; thus, observation can alone instruct us on the ratio of the births of boys and of girls, we must, considering the thing only in itself and setting aside the events, to assume the law of possibility of the births of a boy or of a girl constant from zero to unity, and to start from this hypothesis into the different problems that we can propose on this object.

We suppose, for example, that we have observed that, out of  $p + q$  infants, there is born  $p$  boys and  $q$  girls, and that we seek the probability  $P$  that, out of  $m + n$  infants who must be born, there will be  $m$  boys and  $n$  girls; if we name  $x$  the probability that an infant who must be born will be a boy, and  $1 - x$  that it will be a girl, by designating

$$\frac{1.2.3 \dots (p+q)}{1.2.3 \dots p.1.2.3 \dots q}$$

by  $\lambda$ , we will have

The posterior probability of more boys in 82 out of 82 years (Arburthnot)

```
tibble(p      = seq(0, 1, 0.01),  
       post = dbinom(82,82,p)/sum(dbinom(82,82,p))) %>%  
  ggplot() +  
  aes(p, post) +  
  geom_point()
```

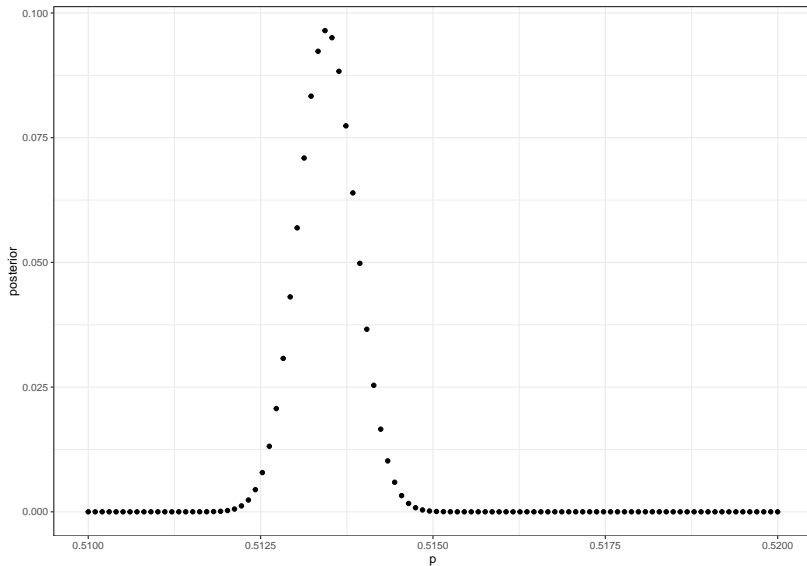
## The posterior probability of more boys in London (Laplace)

```
tibble(values = c(251527, 241945, 738629, 698958),
        city   = c("Paris", "Paris", "London", "London"),
        sex    = c("Male", "Female", "Male", "Female"))
```

```
## # A tibble: 4 x 3
##   values city   sex
##   <dbl> <chr> <chr>
## 1 251527 Paris  Male
## 2 241945 Paris  Female
## 3 738629 London Male
## 4 698958 London Female
```

```
lik <- function(p)
  dbinom(x = 737629, size = 737629 + 698958, prob = p)
tibble(p = seq(.51, .52, length.out = 100)) %>%
  mutate(likelihood = map_dbl(p, lik),
         posterior   = likelihood/sum(likelihood)) %>%
  qplot(data = ., x = p, y = posterior)
```

# ## The posterior probability of more boys in London (Laplace)



The posterior probability Parisian boy more likely than a London boy

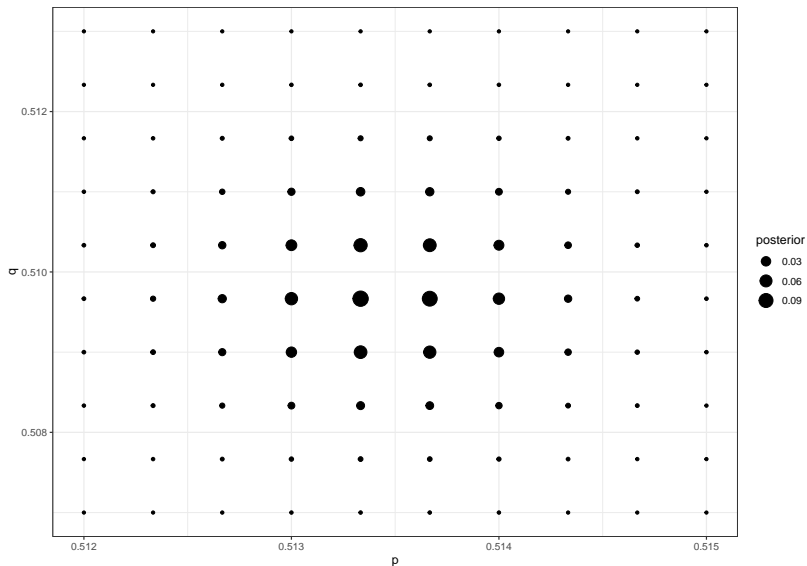
```
lik2 <- function(p, q)
  dbinom(x = 737629, size = 737629 + 698958, prob = p) *
  dbinom(x = 251527, size = 241945 + 251527, prob = q)
```

## The posterior probability Parisian boy more likely than a London boy

```
tibble(p = seq(.512, .515, length.out = 10),
       q = seq(.507, .513, length.out = 10)) %>%
  expand(p, q) %>%
  mutate(likelihood = map2_dbl(p, q, lik2),
         posterior = likelihood/sum(likelihood)) %>%
  qplot(data = ., x = p, y = q, size = posterior)

tibble(p = seq(.512, .515, length.out = 50),
       q = seq(.507, .513, length.out = 50)) %>%
  expand(p, q) %>%
  mutate(likelihood = map2_dbl(p, q, lik2),
         posterior = likelihood/sum(likelihood)) %>%
  qplot(data = ., x = p, y = q, fill = posterior,
        geom = "tile") +
  geom_abline(intercept = 0, slope = 1, color = "red")
```

# The posterior probability Parisian boy more likely than a London boy



The posterior probability Parisian boy more likely than a London boy

```
tibble(p = seq(.512, .515, length.out = 50),
       q = seq(.507, .513, length.out = 50)) %>%
  expand(p, q) %>%
  mutate(likelihood = map2_dbl(p, q, lik2),
         posterior = likelihood/sum(likelihood)) %>%
  qplot(data = ., x = p, y = q, fill = posterior,
        geom = "tile") +
  geom_polygon(data=data.frame(x = c(0.512, 0.512, 0.513),
                              y = c(0.512, 0.513, 0.513)),
              aes(x=x,y=y), alpha=0.5, fill="red")
```



The posterior probability Parisian boy more likely than a London boy

