3. Approccio bayesiano alla stima di parametri

Massimiliano Pastore
Università di Padova

Contents

The Bayesian approach

... when you have eliminated the impossible,
whatever remains, however improbable, must be the truth ...
(Doyle, 1890)

- The general principles of Bayesian analysis are easy to understand.
- First, we model uncertainty or *degree of belief* by using probability (*prior*, $p(\theta)$).
- Second, we use observed data to update the *prior* information or beliefs to become *posterior* information or belief, $p(\theta|D)$.

(Lee & Wagenmakers, 2013)

Conditional probabilities

- Bayesian statistics are based on Bayes' principle of conditional probabilities:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

- This equation is often verbalized as

Conditional probabilities

- Bayesian statistics are based on Bayes' principle of conditional probabilities:
$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$
- This equation is often verbalized as
$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

Major points

- **More can be learned about parameter estimates and model fit.**
  - ML assumes that the distribution of the parameter estimate is normal.
  - In contrast, Bayes provides the whole distribution, referred to as a posterior distribution, not assuming it is normal.
  - The ML confidence interval assumes a symmetric distribution, whereas the Bayesian credibility interval allows for a strongly skewed distribution.

Major points

- **Better small-sample performance can be obtained and large-sample theory is not needed.**
  - This point is illustrated by better Bayesian small-sample performance for factor analyses prone to Heywood cases and better performance when a small number of clusters are analyzed in multilevel models.
  - This, however, requires a judicious choice of the prior.

van de Schoot, R., & Miocevic, M. (2020). *Small Sample Size Solutions: A Guide for Applied Researchers and Practitioners*. Routledge, London and New York.

Major points

- **Analyses can be made less computationally demanding.**
  - Many models are computationally cumbersome or impossible using ML, such as with categorical outcomes and many latent variables resulting in many dimensions of numerical integration.
  - Such an analyst may view the Bayesian analysis simply as a computational tool for getting estimates that are analogous to what would have been obtained by ML had it been feasible.

Major points

- **New types of models can be analyzed.**
  - For example models with a very large number of parameters or where ML does not provide a natural approach.

Frequentist approach

- In the traditional frequentist approach to statistical inference, the probability of an event is interpreted as the relative frequency of an event given an infinite sequence of samples from an identical (i.e. fixed) probability distribution.
- In the frequentist approach the model parameters are assumed as fixed, e.g. in the form of a Null-Hypothesis that fixes $\theta = 0$.

Bayesian inference

- In contrast, the Bayesian approach focuses directly on the probability of an effect, i.e. on the probability of observing the estimated parameters given the data, i.e. on $p(\theta|D)$.
- Further, in addition to the sampling uncertainty of data, the Bayesian approach also treats the model parameters as uncertain, i.e. assumed as following a probability distribution, namely the prior distribution $p(\theta)$.

Example

Let us consider a sample of subjects for whom there is a suspicion of a cognitive deficit and $\mu = 90$ indicates the average value of the score for normal subjects in a specific test measuring cognitive abilities.

Example

Let us consider a sample of subjects for whom there is a suspicion of a cognitive deficit and $\mu = 90$ indicates the average value of the score for normal subjects in a specific test measuring cognitive abilities.

NHST approach:

$$\begin{cases} H_0 : \mu = 90 \\ H_1 : \mu < 90 \end{cases}$$

Example

Let us consider a sample of subjects for whom there is a suspicion of a cognitive deficit and $\mu = 90$ indicates the average value of the score for normal subjects in a specific test measuring cognitive abilities.

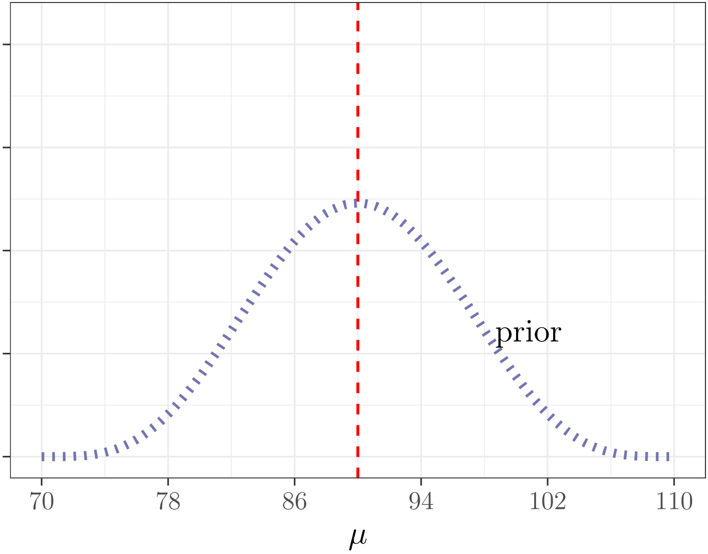
NHST approach:

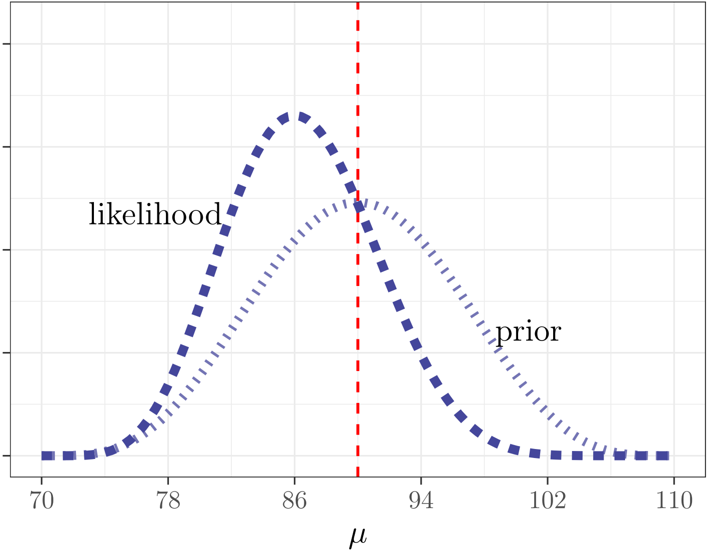
$$\begin{cases} H_0 : \mu = 90 \\ H_1 : \mu < 90 \end{cases}$$

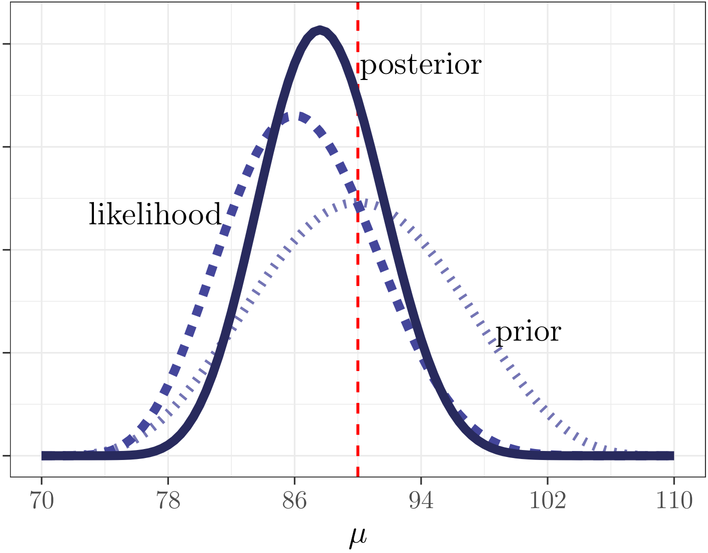
Bayes approach:

$$\mu \sim \mathcal{N}(90, \tau)$$

With Bayesian way it's possible, for example, to estimate $P(\mu < 90)$ while in NHST this probability can be only 0 (H_1 is false) or 1 (H_1 is true).


prior
70
78
86
94
102
110
μ


likelihood
prior
70
78
86
94
102
110
μ


posterior
likelihood
prior
70
78
86
94
102
110
μ

Example

- An ornithologist wants to estimate the proportion (θ) of male kiwi in a specified area of the NZ.
- Without any prior information, any proportion could be equally plausible.
- He observes a sample of 9 animals in the following sequence:

$$F, F, F, F, M, F, M, F, F$$

  2 are male and 7 female.
- After observing these data, what can he think?

Model of observations

- We can make the following assumptions about the sampling process:
  1. The true proportion of male individuals is θ
  2. By randomly pulling out an individual this will be male (with probability θ) or female (with probability 1-θ)
  3. The outcome (male or female) of any observation is independent of the outcome of any other observation
- This collection of assumptions about the kiwi sampling process is our model of the observations.

Model of observations

The model of observations is the model that describes the probabilities of observable events.

- We can have a formula that describes the probability that an observed animal will be male.
- Or a formula that tells us what is the probability of observing k male on n animals.

Model of beliefs

- We could made a second set of assumptions about our beliefs regarding the proportion of males.
- For example, assuming that we believe most strongly in the proportion being close to 50%, but we also allowing for the possibility that the proportion could be different.
- Thus, we have a set of assumptions about how likely it is for the proportion to be 50% or to be another value to different amounts.

Model of beliefs

The model of beliefs describe the extent to which we believe in various underlying possibilities.

- e.g. we can have a formula that describes how much we believe in each possible proportion of males.

Priors and Posteriors

- We define *Priors* all the beliefs that we have before collecting any kind of information.
- We define *Posteriors* all beliefs that take into account a particular set of information collected.

Priors and Posteriors

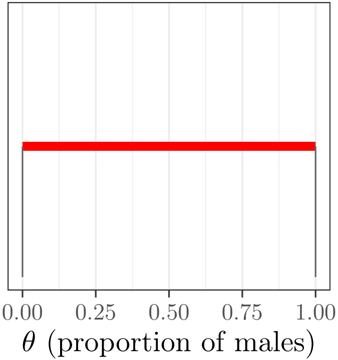
- We define *Priors* all the beliefs that we have before collecting any kind of information.
- We define *Posteriors* all beliefs that take into account a particular set of information collected.

For example:

- Before observing a kiwi we assume to have a probability of 50% that it is male.
- After observing the animals of our sample we change our belief because the result of 2 males (out of 9 individuals observed) can lead us to reduce the probability that we assign to the fact that the number of males and females is equal.

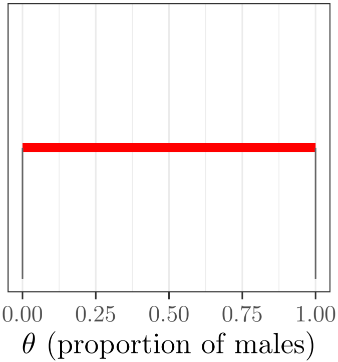
Priors and Posteriors

Before observing animals, we assume that any θ value is equally plausible.

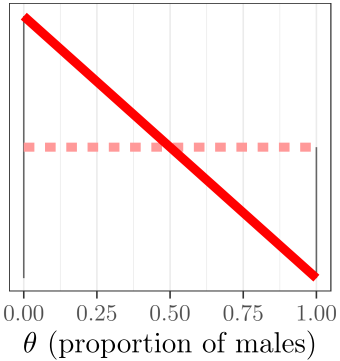


Priors and Posteriors

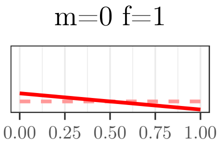
Before observing animals, we assume that any θ value is equally plausible.



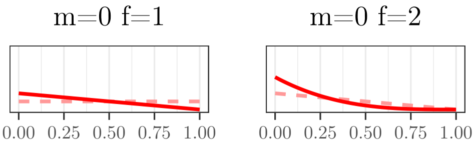
After observing the first animal (F) we change the plausibility associated with θ values.



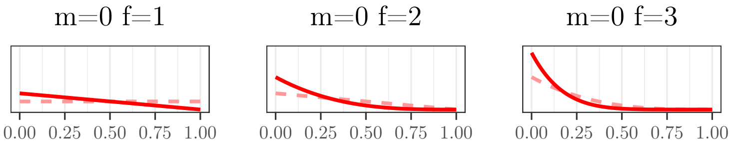
Priors and Posteriors



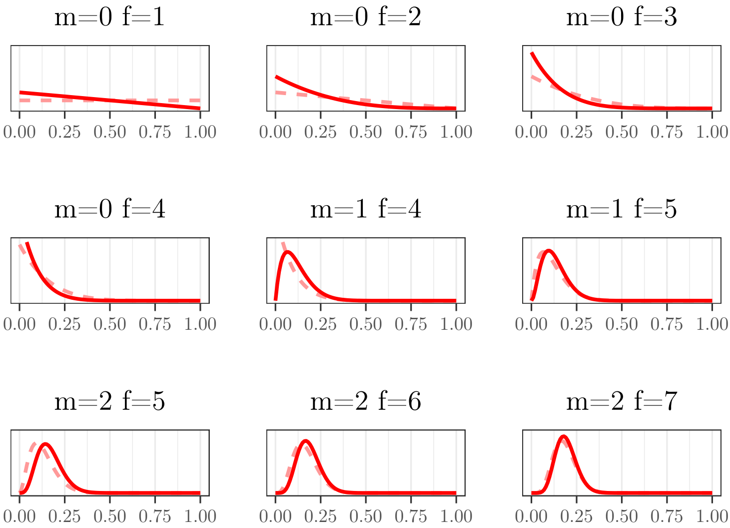
Priors and Posteriors



Priors and Posteriors



Priors and Posteriors



Goals for bayesian inference

When we make observation of data, we typically have one of these three goals in mind (like in the frequentist approach):

1. Estimation of Parameter Values.
2. Prediction of Data Values.
3. Model Comparison.

Estimation of Parameter Values

- It means deciding the extent to which we should believe in each of the possible values of an underlying parameter.
- Let θ be true (unknown) value of males proportion in the kiwi-observation scenario.
- Because the observation of an animal is a random process, we cannot be certain of the underlying true proportion of males; so our posterior beliefs are an estimate.
- The process of shifting our beliefs in various parameter values is called **estimation of parameter values**.

Prediction of Data Values

- Based on our beliefs, i.e. the probability of a male, we may predict other values: for example the outcome of a new observation or the number of males in ten animals.
- Prediction simply means inferring the values of some missing data based on some other included data, regardless of the actual temporal relationship of the included and missing data.

Model Comparison

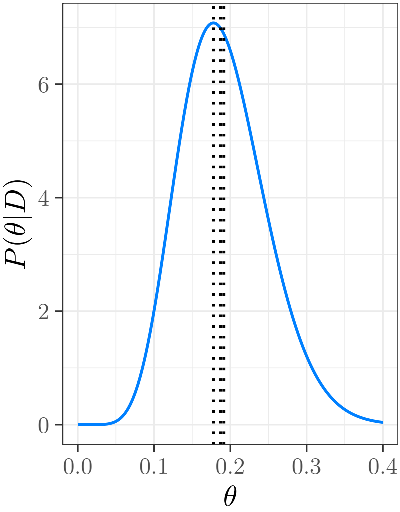
- If we have two different models of how something happens, then an observation of what really does happen can influence which model we believe in most.
- For example, suppose we have two different models for the male/female proportion. One model assumes that the proportion of males could be 20%, 50% or 80%. The second model assumes that the proportion is either a perfectly balanced proportion or else a gender is completely missing: 0%, 50% or 100%.
- After observing 2 males out of 9 animals, which model do we believe in more? The mathematics of Bayesian inference **can tell exactly how much more to believe in the first model than in the second**.

Result of a Bayesian Analysis

- The *work* of a bayesian model is to produce the *posterior distribution* by sampling[1].
- After having produced the posterior, our work will consist in summarizing and interpreting this distribution.
- Depending on our goals, we can have various ways to summarize the posterior: (1) point estimates, (2) intervals based on defined boundaries, (3) intervals based on defined probability densities.

[1] This does not exclude particular, generally simpler cases where posterior distribution is analytically defined.

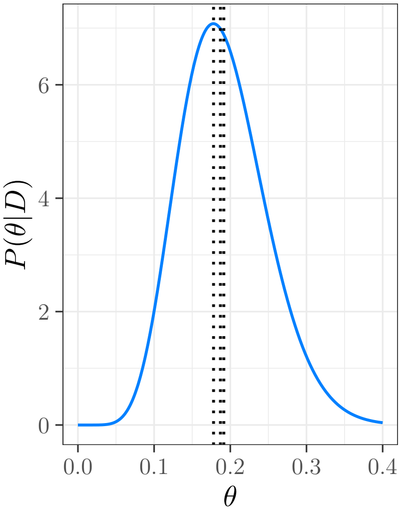
Point estimates



Let's consider three alternative point estimates:

- the *maximum a posteriori*, MAP = 0.178
- the posterior mean: 0.191
- the posterior median: 0.187

Point estimates



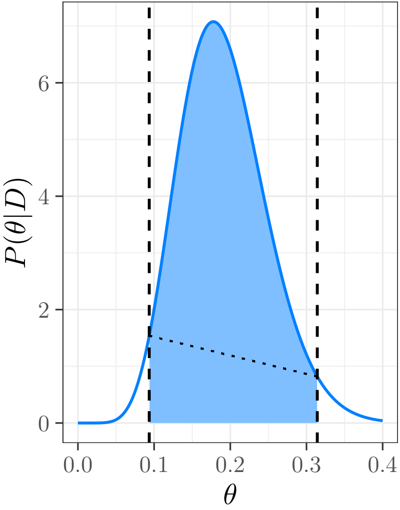
Let's consider three alternative point estimates:

- the *maximum a posteriori*, MAP $= 0.178$
- the posterior mean: 0.191
- the posterior median: 0.187

Problem:

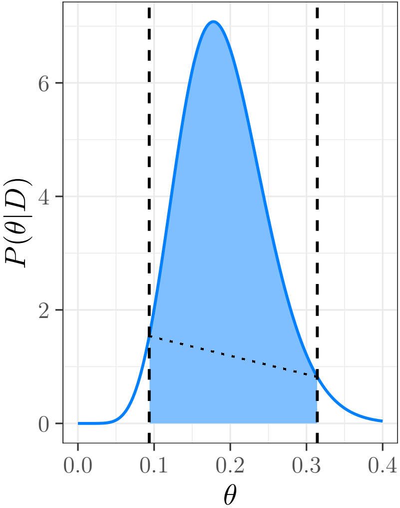
- If the distribution is far from symmetry these measures will tend to be different.

Intervals based on quantiles



We can use a quantile-based interval, for example the classicals 2.5% and 97.5%.

Intervals based on quantiles

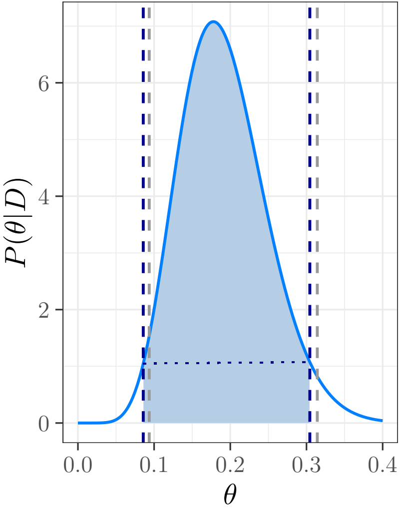


We can use a quantile-based interval, for example the classicals 2.5% and 97.5%.

Problem:

- If the distribution is skewed we will have within the interval less likely values than those outside the interval.

HPD intervals



This is an interval $[l(y), u(y)]$ such that:

$$\Pr[l(y) < \theta < u(y)|D] = 1 - \alpha$$

i.e. the probability that $l(y) < \theta < u(y)$ after observing data.

The binomial model

From Bernoulli to binomial model

- When we observe an animal, the result can be a male (M) or a female (F).
- We will denote the result by y, with $y = 1$ for M and $y = 0$ for F.
- Let θ be the probability of M:

From Bernoulli to binomial model

- When we observe an animal, the result can be a male (M) or a female (F).
- We will denote the result by y, with $y = 1$ for M and $y = 0$ for F.
- Let θ be the probability of M:

$$\begin{cases} p(y = 1|\theta) = \theta \\ p(y = 0|\theta) = 1 - \theta \end{cases}$$

Bernoulli

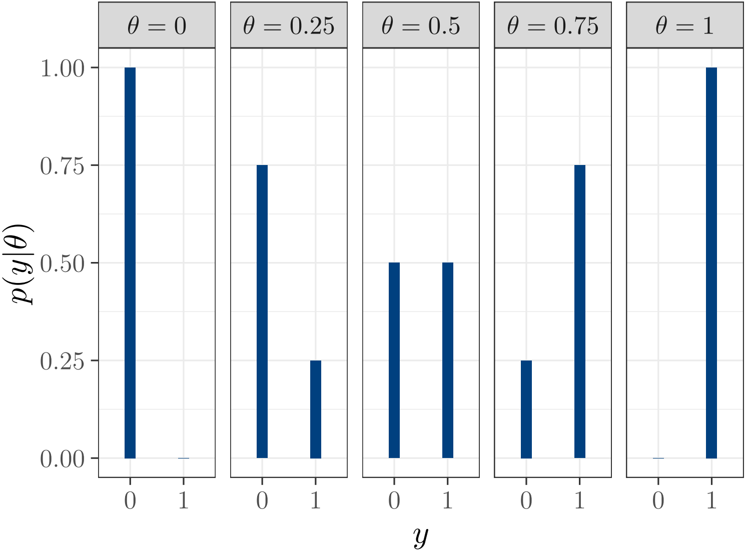
- The two equations can be combined into a single expressions as follows:

$$p(y|\theta) = \theta^{y}(1-\theta)^{(1-y)}$$

which expresses the Bernoulli distribution with $y \in \{0, 1\}$ and $\theta \in [0, 1]$.

- This distribution represents the probability to obtain 0 or 1 given fixed θ.
- If, for example, $\theta = \{0, 0.25, 0.5, 0.75, 1\}$, we will obtain 5 different probability distributions.

Bernoulli mass function

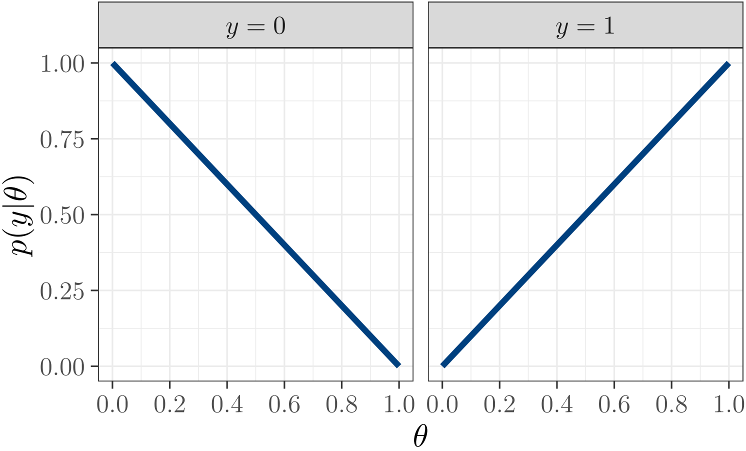


Bernoulli likelihood

- Contrariwise, by fixing y and varying θ we obtain the *likelihood function*.

Bernoulli likelihood

- Contrariwise, by fixing y and varying θ we obtain the *likelihood function*.



Important points:

- The same function $p(y|\theta)$ can be two different things: a Bernoulli distribution or a *likelihood* function.

Important points:

- The same function $p(y|\theta)$ can be two different things: a Bernoulli distribution or a *likelihood* function.
- The Bernoulli is a discrete distribution: it give us the probability of y given fixed θ.
- The *likelihood* is a continuous function: it specifies a probability at each value of θ given y, but is not a probability distribution.

Important points:

- The same function $p(y|\theta)$ can be two different things: a Bernoulli distribution or a *likelihood* function.
- The Bernoulli is a discrete distribution: it give us the probability of y given fixed θ.
- The *likelihood* is a continuous function: it specifies a probability at each value of θ given y, but is not a probability distribution.
- In the NHST approach we consider only θ values under H_0.
- In the bayesian approach, θ is considered and treated as a random variable.

Binomial model

- When we observe N animals, we have a set of data, $D = \{y_1, \ldots, y_N\}$, where each y_i is 0 (F) or 1 (M).
- By assuming each observation as independent from the others, the probability of getting the set of N animals D is the product of the individual outcome probabilities:

$$p(\{y_1, \ldots, y_N\}) =$$

Binomial model

- When we observe N animals, we have a set of data, $D = \{y_1, \ldots, y_N\}$, where each y_i is 0 (F) or 1 (M).
- By assuming each observation as independent from the others, the probability of getting the set of N animals D is the product of the individual outcome probabilities:

$$p(\{y_1, \ldots, y_N\}) = \prod_i p(y_i|\theta) =$$

Binomial model

- When we observe N animals, we have a set of data, $D = \{y_1, \dots, y_N\}$, where each y_i is 0 (F) or 1 (M).
- By assuming each observation as independent from the others, the probability of getting the set of N animals D is the product of the individual outcome probabilities:

$$p(\{y_1, \dots, y_N\}) = \prod_i p(y_i|\theta) = \prod_i \theta^{y_i}(1-\theta)^{(1-y_i)}$$

- If the numbers of males in the set of animals is denoted $z = \sum_i^N y_i$then we can write:

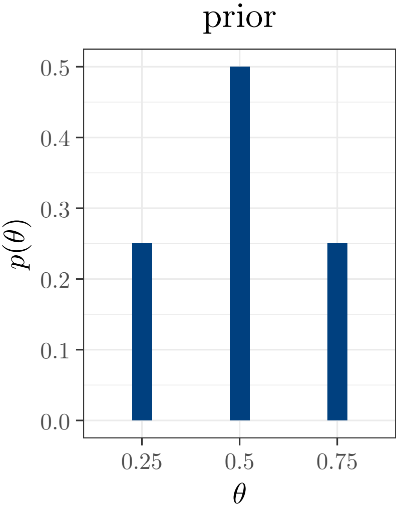
$$p(z, N|\theta) = \theta^z(1-\theta)^{(N-z)}$$

Prior definition

- Let us suppose to be interested in assessing whether proportion male/female is quite similar.
- First, we specify our *prior beliefs*.
- We denote the proportion of males as $\theta = p(\mathrm{M})$.
- Suppose that we believe there are only 3 possible values for this proportion: $\theta = 0.25, 0.5, 0.75$.

Prior definition

- Let us suppose to be interested in assessing whether proportion male/female is quite similar.
- First, we specify our *prior beliefs*.
- We denote the proportion of males as $\theta = p(\mathrm{M})$.
- Suppose that we believe there are only 3 possible values for this proportion: $\theta = 0.25, 0.5, 0.75$.



Likelihood definition

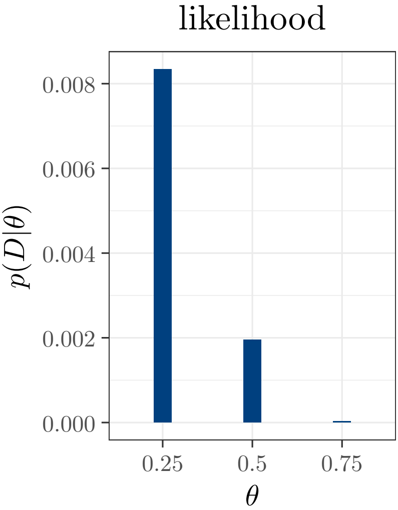
- Next, we observe a sample of animals to get some data D and determine the likelihood $p(D|\theta)$.
- Suppose we observe 9 animals and it comes up 2 M and 7 F.
- Consequently:

$$p(D|\theta) = \theta^2(1-\theta)^7$$

Likelihood definition

- Next, we observe a sample of animals to get some data D and determine the likelihood $p(D|\theta)$.
- Suppose we observe 9 animals and it comes up 2 M and 7 F.
- Consequently:

$$p(D|\theta) = \theta^2(1-\theta)^7$$



Posterior computing

- For computing the posterior distribution we use the Bayes' rule:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

- In the binomial case the computation of $p(D)$ is quite easy:

$$p(D) = \sum_{\theta} p(D|\theta)p(\theta)$$

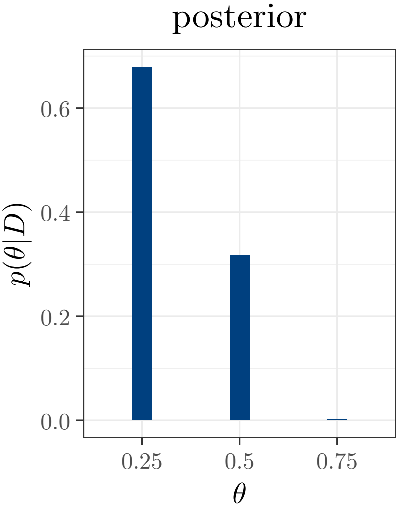
Posterior computing

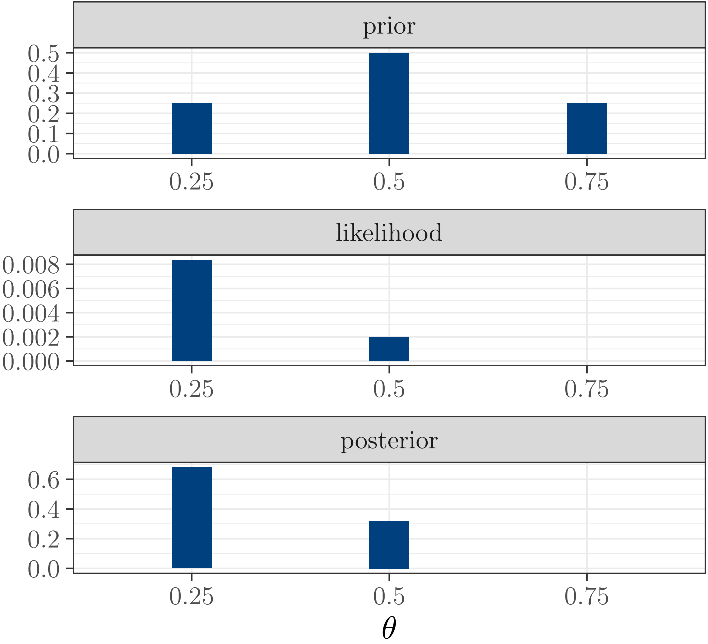
- For computing the posterior distribution we use the Bayes' rule:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

- In the binomial case the computation of $p(D)$ is quite easy:

$$p(D) = \sum_{\theta} p(D|\theta)p(\theta)$$




prior
0.5
0.4
0.3
0.2
0.1
0.0
0.25
0.5
0.75
likelihood
0.008
0.006
0.004
0.002
0.000
0.25
0.5
0.75
posterior
0.6
0.4
0.2
0.0
0.25
0.5
0.75
θ

```
theta <- c(.25,.5,.75) # parameter values
(prior <- pmin(theta,1-theta)) # prior distribution

[1] 0.25 0.50 0.25

data <- c(1,1,rep(0,7)) # observed data
n <- length(data)
males <- sum(data) # number of males
(like <- theta^males*(1-theta)^(n-males)) # likelihood

[1] 8.342743e-03 1.953125e-03 3.433228e-05

pData <- sum(prior*like)
(post <- prior*like/pData) # posterior distribution

[1] 0.679192547 0.318012422 0.002795031
```

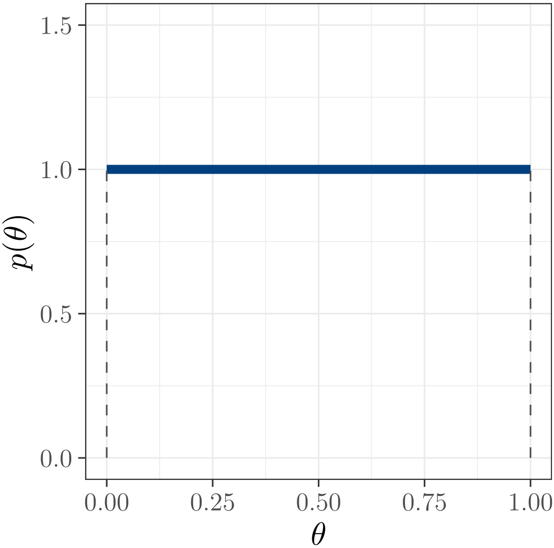
The logic of Bayesian analysis:

- Similarly, we can define other *priors*, or we could use the obtained *posterior* as a new *prior* and then collect other data.
- The main objective of the analysis is to obtain the posterior distribution.
- Given the posterior distribution, we can proceed depending on our final goal: parameter estimation, prediction or model comparison.

Prior definition

- Actually, the proportion is a continuous variable, so it is better to model our prior beliefs with a continuous distribution.
- All we know, before observing the animals, is that the proportion of males will certainly be in the [0,1] range.
- Consequently, in order to formalise our prior hypothesis we need an appropriate probability distribution.
- Any idea?

Uniform Prior



Beta Prior

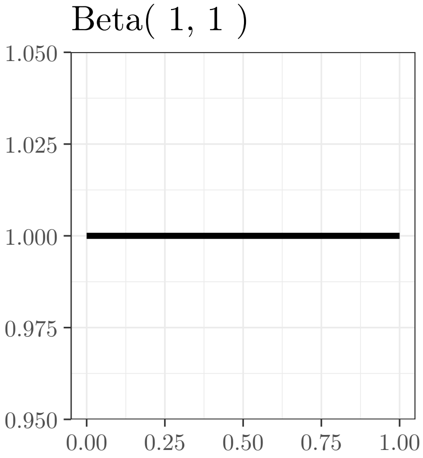
- The Uniform distribution is actually a special case of the Beta distribution.
- The Beta is a continuous density distribution ranging from 0 to 1 and is therefore suitable for dealing with proportions or parameters that fall within this range.
- This distribution is defined as

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

and can easily be used to formalise hypotheses about the binomial case.

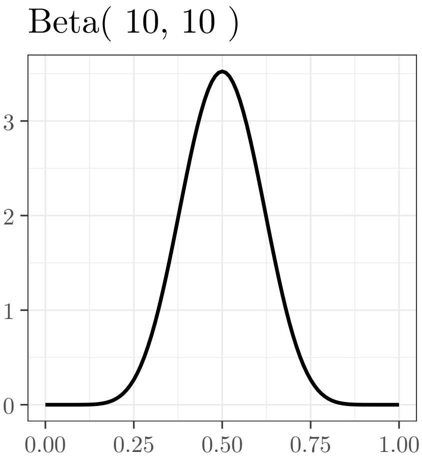
- Assuming that we know nothing about the proportion of males (maximum uncertainty), how many males can we expect to see when observing two birds at random?

- Assuming that we know nothing about the proportion of males (maximum uncertainty), how many males can we expect to see when observing two birds at random?



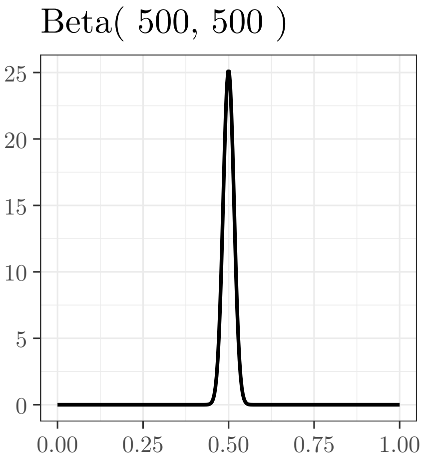
- Assuming that we believe that the proportion of males is about 50%, how many males can we expect to see when observing 20 birds at random?

- Assuming that we believe that the proportion of males is about 50%, how many males can we expect to see when observing 20 birds at random?

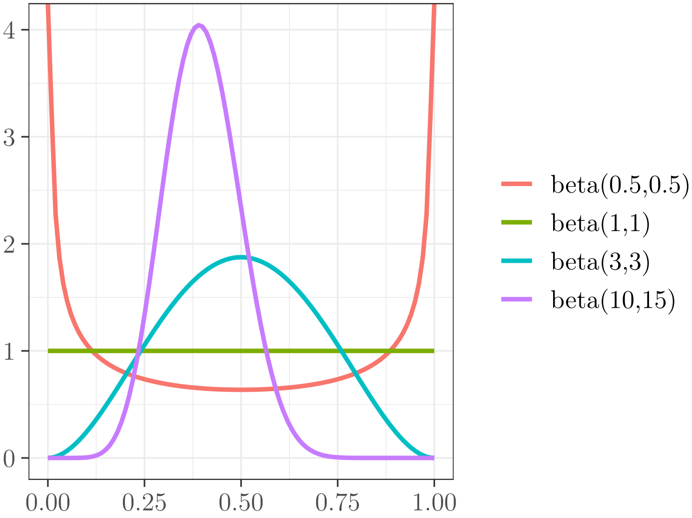


- Assuming that we strongly believe that the proportion of males is about 50%, how many males can we expect to see when observing 1000 birds at random?

- Assuming that we strongly believe that the proportion of males is about 50%, how many males can we expect to see when observing 1000 birds at random?



Beta distribution



Binomial inference

Returning to the kiwi example, we want to estimate the proportion of males in the population (θ) after observing 2 males and 7 females (D):

- The prior will be $p(\theta) = Beta(1, 1)$
- The likelihood will be $p(D|\theta) = \theta^2(1-\theta)^7$
- Using Bayes' theorem, the posterior will be

$$p(\theta|D) = \frac{p(\theta) \times p(D|\theta)}{p(D)}$$

The interesting thing is that the posterior distribution will again be a Beta distribution with parameters $a = 1 + 2$ and $b = 1 + 7$, $Beta(3, 8)$.

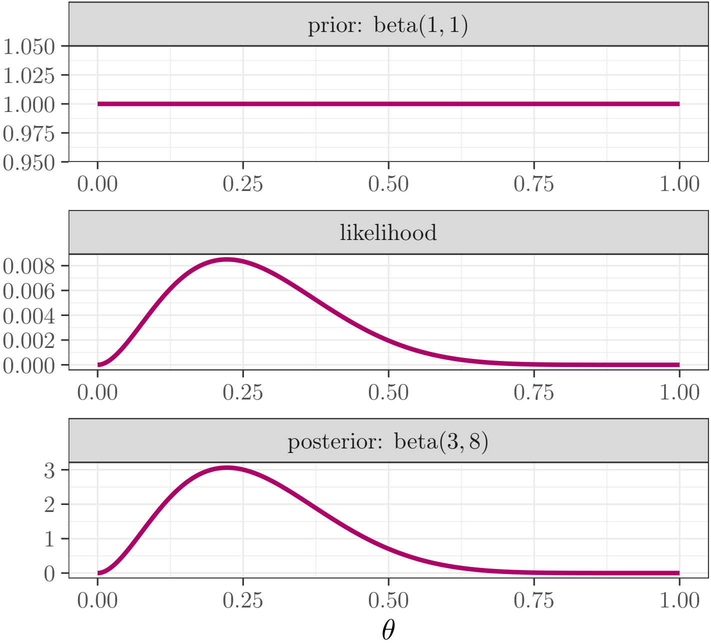
Binomial inference

Returning to the kiwi example, we want to estimate the proportion of males in the population (θ) after observing 2 males and 7 females (D):

- The prior will be $p(\theta) = Beta(1, 1)$
- The likelihood will be $p(D|\theta) = \theta^2(1-\theta)^7$
- Using Bayes' theorem, the posterior will be

$$p(\theta|D) = \frac{p(\theta) \times p(D|\theta)}{p(D)} = Beta(1+2, 1+7)$$

The interesting thing is that the posterior distribution will again be a Beta distribution with parameters $a = 1+2$ and $b = 1+7$, $Beta(3, 8)$.


prior: beta(1, 1)
1.050
1.025
1.000
0.975
0.950
0.00
0.25
0.50
0.75
1.00
likelihood
0.008
0.006
0.004
0.002
0.000
posterior: beta(3, 8)
3
2
1
0
θ

Used R packages

- `bayesplot`. Gabry J, Mahr T (2025).
- `brms`. Bürkner P (2017).
- `cmdstanr`. Gabry J, Češnovar R, Johnson A, Bronder S (2025).
- `coda`. Plummer M, Best N, Cowles K, Vines K (2006).
- `cowplot`. Wilke C (2024).
- `ggplot2`. Wickham H (2016).
- `knitr`. Xie Y (2025).
- `posterior`. Bürkner P, Gabry J, Kay M, Vehtari A (2025).
- `R`. R Core Team (2025).
- `Rcpp`. Eddelbuettel D, Francois R, Allaire J, Ushey K, Kou Q, Russell N, Ucar I, Bates D, Chambers J (2025).
- `report`. Makowski D, Lüdecke D, Patil I, Thériault R, Ben-Shachar M, Wiernik B (2023).
- `rethinking`. McElreath R (2024).
- `rstan`. Stan Development Team (2025).
- `rstanarm`. Goodrich B, Gabry J, Ali I, Brilleman S (2024).
- `StanHeaders`. Stan Development Team (2020).

massimiliano.pastore@unipd.it
https://psicostat.dpss.psy.unipd.it/