Bayesian statistics

So far we have thought of probabilities as the long term "success frequency": #successes / #trails $\rightarrow$ P(success).

In **Bayesian statistics probabilities are subjective**!
Examples

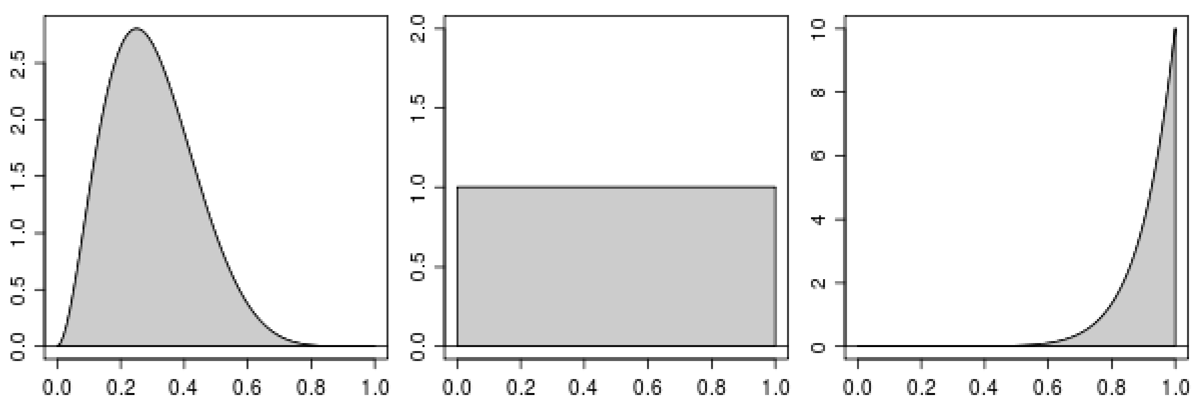
* Probability that two companies merge
* Probability that a stock goes up
* Probability that it rains tomorrow

We typically want to make inference for a parameter θ, for example μ, σ^2 or π. How is this done using subjective probabilities?

Bayesian statistics

Bayesian idea: We describe our “knowledge” about the parameter of interest, θ, in terms of a distribution $\pi(\theta)$. This is known as the **prior distribution** (or just prior) – as it describes the situation *before* we see any data.

Example: Assume θ is the probability of success.
Prior distributions describing what value we *think* θ has:



Bayesian statistics
Posterior

Let x denote our **data**. The conditional distribution of θ given data x is denoted the **posterior distribution**:

$$\pi(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{g(x)}$$

Here $f(x|\theta)$ tells how data is specified conditional on θ.

Example:

Let *x* denote the number of successes in *n* trail.

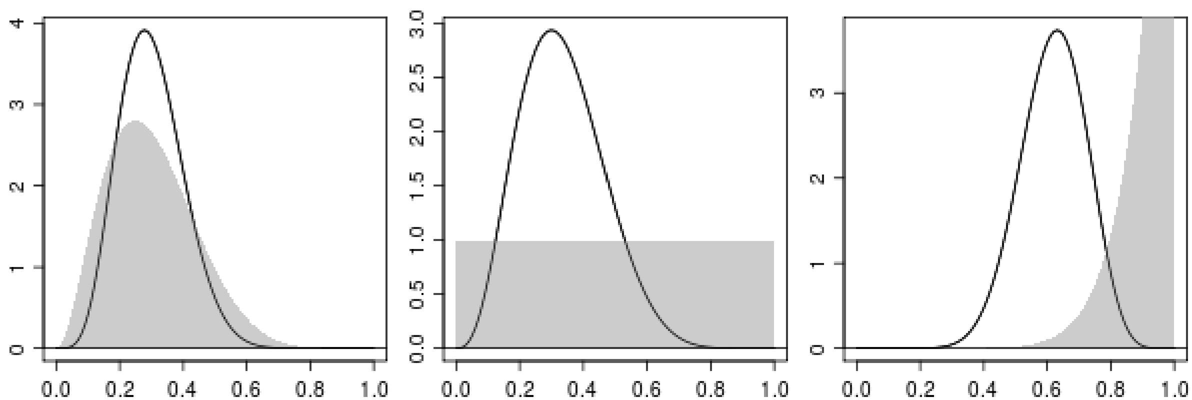
Conditional on θ, *x* follows a binomial distribution:

$$f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Bayesian statistics
Posterior – some data

We now observe n=10 experiment with x=3 successes, i.e. x/n=0.3

Posterior distributions – our “knowledge” after having seen data.



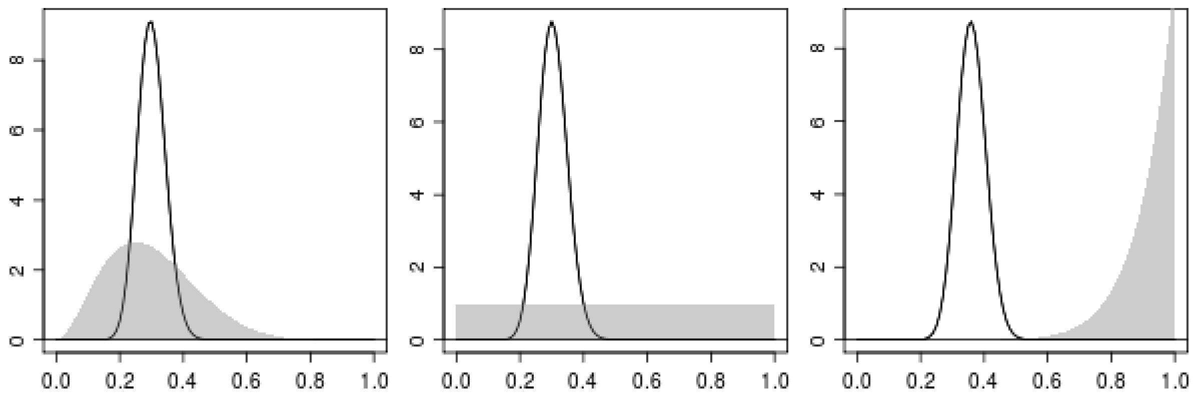
Shaded area: Prior distribution
Solid line: Posterior distribution

Notice that the posteriors are moving towards 0.3.

Bayesian statistics
Posterior – some data

We now observe $n = 100$ experiment with $x = 30$ successes, i.e. $x/n = 0.3$

Posterior distributions – our "knowledge" after having seen data.



Shaded area: Prior distribution
Solid line: Posterior distribution

Notice that the posteriors are almost identical.

Bayesian statistics
Mathematical details

A prior often used with the binomial is given by a so-called **Beta distribution** with parameters $\alpha > 0$ and $\beta > 0$:

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \quad \text{for} \quad 0 \le \theta \le 1$$

The posterior then becomes

$$\pi(\theta \mid x) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)}\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}$$

a Beta distribution with parameters $\alpha+x$ and $\beta+n-x$.