Chapter 2. Small Worlds and Large Worlds

Small world = our model of the problem, all behaviour is expected

Large world = real world, things behave differently

conjecture = belief based on incomplete information

To get the logic moving, we need to make assumptions, and these assumptions constitute the model. Designing a simple Bayesian model benefits from a design loop with three steps.

(1) Data story: Motivate the model by narrating how the data might arise.
(2) Update: Educate your model by feeding it the data.
(3) Evaluate: All statistical models require supervision, leading possibly to model revision.

The next sections walk through these steps, in the context of the globe tossing evidence.

The data story in this case is simply a restatement of the sampling process:

(1) The true proportion of water covering the globe is p.
(2) A single toss of the globe has a probability p of producing a water (W) observation. It has a probability $1 - p$ of producing a land (L) observation.
(3) Each toss of the globe is independent of the others.

Bayesian updating = updating based on new observations

Bayesian models make perfect decisions based on avaiable data. Small world is analtic, but is it right? "Bayesian model leearns in a demonstrably optimal way"

Graph of how bayesian updating learns :

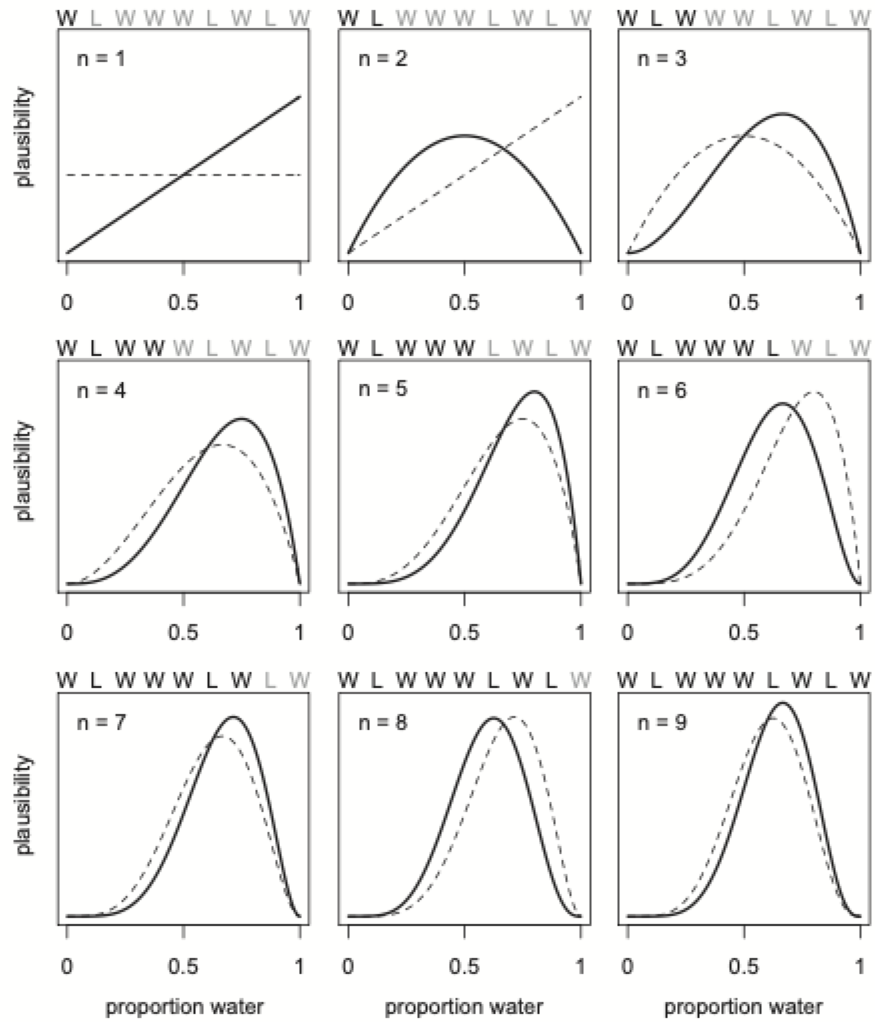

FIGURE 2.5. How a Bayesian model learns. Each toss of the globe produces an observation of water (W) or land (L). The model's estimate of the proportion of water on the globe is a plausibility for every possible value. The lines and curves in this figure are these collections of plausibilities. In each plot, previous plausibilities (dashed curve) are updated in light of the latest observation to produce a new set of plausibilities (solid curve).

Notice how the hieght increases with more samples, a denser probability mass with more confidence

2.3.1. Likelihood. The first and most influential component of a Bayesian model is the **LIKELIHOOD**. The likelihood is a mathematical formula that specifies the plausibility of the data.

Parameters are things like mean and variance of gaussian

Rethinking: Datum or parameter? It is typical to conceive of data and parameters as completely different kinds of entities. Data are measured and known; parameters are unknown and must be estimated from data. Usefully, in the Bayesian framework the distinction between a datum and a parameter is fuzzy. A datum can be recast as a very narrow probability density for a parameter, and a parameter as a datum with uncertainty. Much later in the book (Chapter 14), you'll see how to exploit this continuity between certainty (data) and uncertainty (parameters) to incorporate measurement error and missing data into your modeling.

The prior is the initial set of plausabilities

subjective bayesian = priors from knowledge of scientist

Posterior is combination of prior and likelihood (then normalise)

Good graphic of posterior from prior and likelihood:

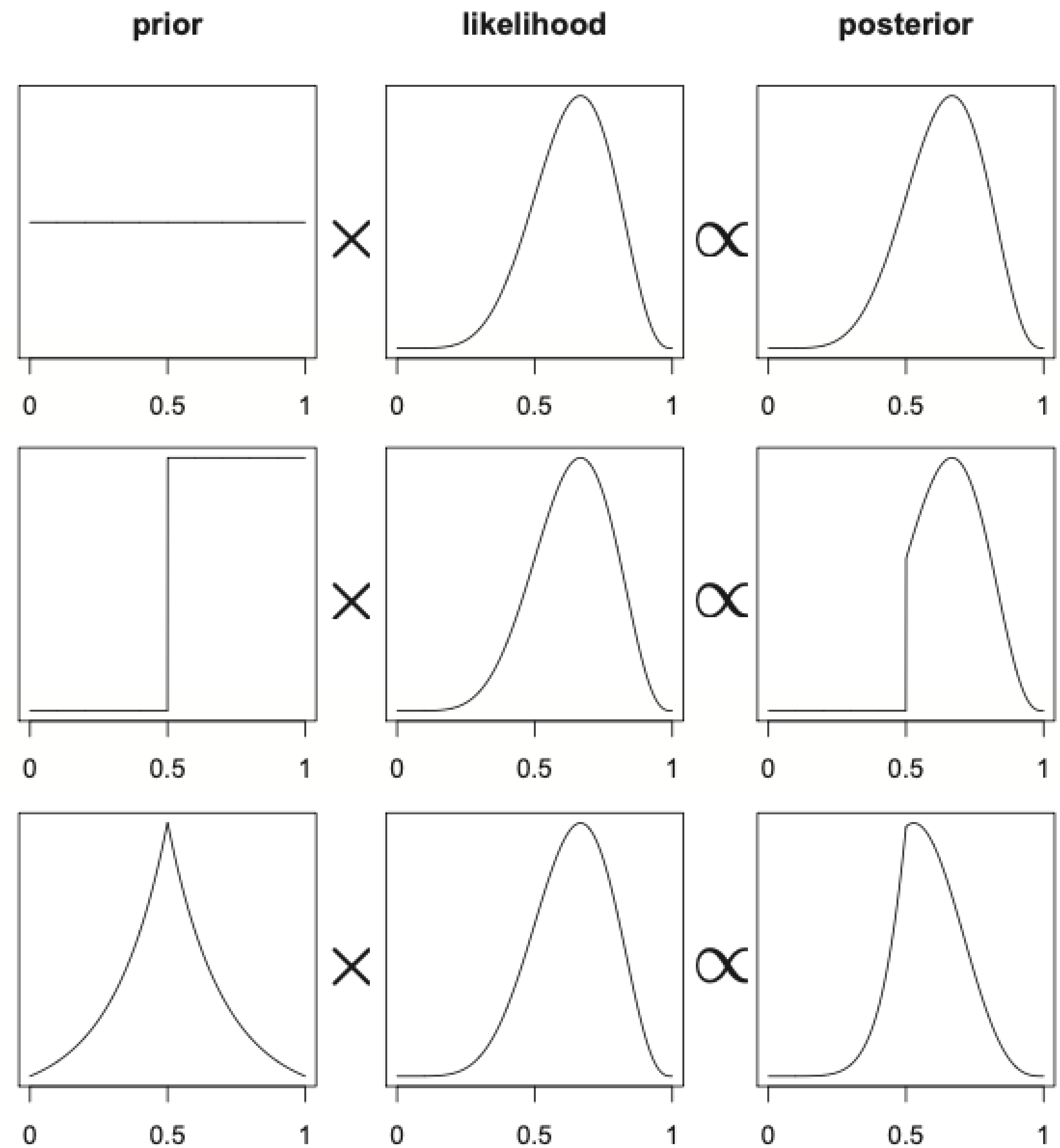

Figure 2.6. The posterior distribution, as a product of the prior distribution and likelihood. Top row: A flat prior constructs a posterior that is simply proportional to the likelihood. Middle row: A step prior, assigning zero probability to all values less than 0.5, resulting in a truncated posterior. Bottom row: A peaked prior that shifts and skews the posterior, relative to the likelihood.

Three techniques for conditioning the prior on the data.... finding posterior over parameters

(1) Grid approximation
(2) Quadratic approximation
(3) Markov chain Monte Carlo (MCMC)

Grid approximation. Bad scalability.

Set grid of parameters, calculate posterior for each, normalise

Quadratic approximation. (Gaussian apprxoimation)

Called quadratic approximation because log of gaussian is parabola (quadratic function).

Works because the peak of the posterior is generally close to gaussian

Use optimisation (MAP) to find mode of posterior, estimate curivture with gaussian

How more samples (n) improves approximation:

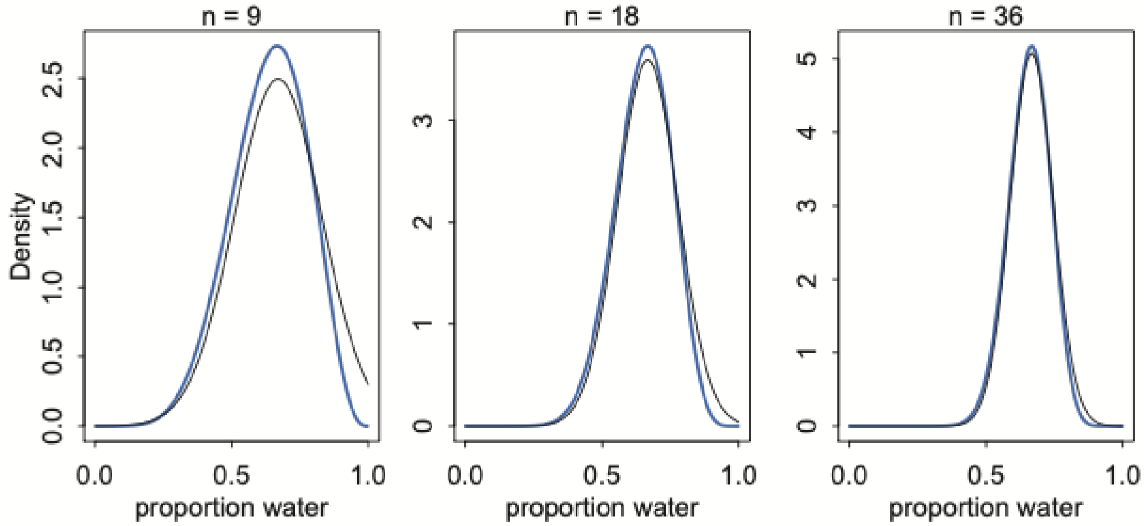

FIGURE 2.8. Accuracy of the quadratic approximation. In each plot, the exact posterior distribution is plotted in blue, and the quadratic approximation is plotted as the black curve. Left: The globe tossing data with $n = 9$ tosses and $w = 6$ waters. Middle: Double the amount of data, with the same fraction of water, $n = 18$ and $w = 12$. Right: Four times as much data, $n = 36$ and $w = 24$.

MCMC : good for when quadratic doesnt fit

Sample from posterior, use to estimate posterior