**Supplementary Material S2**

The foundation of Bayesian statistics can be found in Bayes’ theorem:

**Equation 1**

To make it more explicit for our context, we will replace *A* with *H* for the hypothesis and *B* with *D* for the observed data:

**Equation 2**

* *P(H*|*D)* is the *posterior probability* distribution. This represent the updated knowledge about the hypothesis once our collected data are also considered.
* *P(D*|*H)* is the *likelihood*. This represents the evidence for the hypothesis provided by the data.
* *P(H)* is the *prior probability* distribution and represents the knowledge about the hypothesis that we bring, before our collected data are considered;
* *P(D)* is the *total probability* of the data, taking all possible hypotheses into account. It is the normalizing factor to ensure the posterior probability integrates to 1. While for simple discrete datasets the total probability is easy to calculate (see below), the total probability of continuous data requires numerical integration since there is an infinite number of possibilities within a continuous interval.

For example, suppose a colleague told you that it is difficult to hatch rainbow trout from eggs in captivity and that the success rate is less than 30%. You decide to test this assertion (because you believe it is actually not that difficult) and design an experiment. After placing 10 eggs into a hatching container, you notice a few weeks later that 4 out of 10 eggs hatched. While you can simply calculated the germination rate as 4/10 = 0.4 or 40%, you would also like to know what other hatching rates you could potentially expect, and which rates would be most frequently observed given the data you gathered from your experiment. In other words, you want to calculate a posterior probability distribution for rainbow trout hatching rates based on your sample and prior knowledge. For this example, we assume that the number of hatching successes can be modeled using the binomial distribution with parameter for the total number of trials and parameter for the hatching success probability:

**Equation 3**

The binomial distribution further assumes that each trial is independent of the other. In other words, the result of one trial must not influence the result of the next trial.

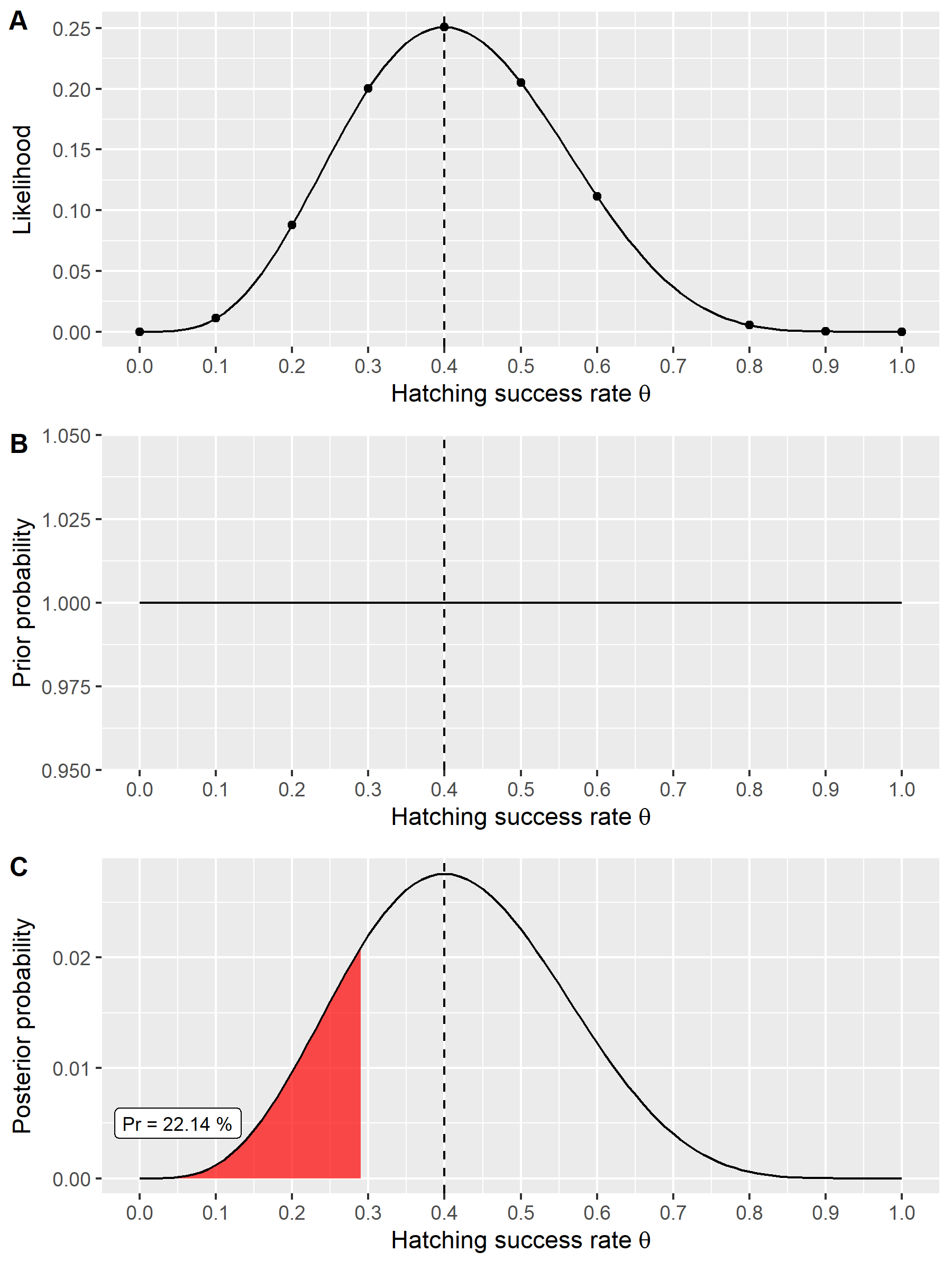
In order to calculate the posterior probability distribution, we need three ingredients:

1. the likelihood of the data given,
2. the prior knowledge
3. the total probability of the data

For this example, we use grid approximation to calculate the posterior distribution. This is the easiest way to conceptualize Bayesian inference with a simple example. In reality, a much more sophisticated sampling technique is used to estimate parameter values, which is known as Markov Chain Monte Carlo (MCMC) ([van Ravenzwaaij et al., 2018](#_ENREF_2)). For our grid approximation example this means that we use an equally spaced sequence of hatching rates (parameter values) from zero to one labelled as to calculate the likelihood *P(D|)* for the observed rainbow trout hatching experiment with successes in trials:

**Equation 3**

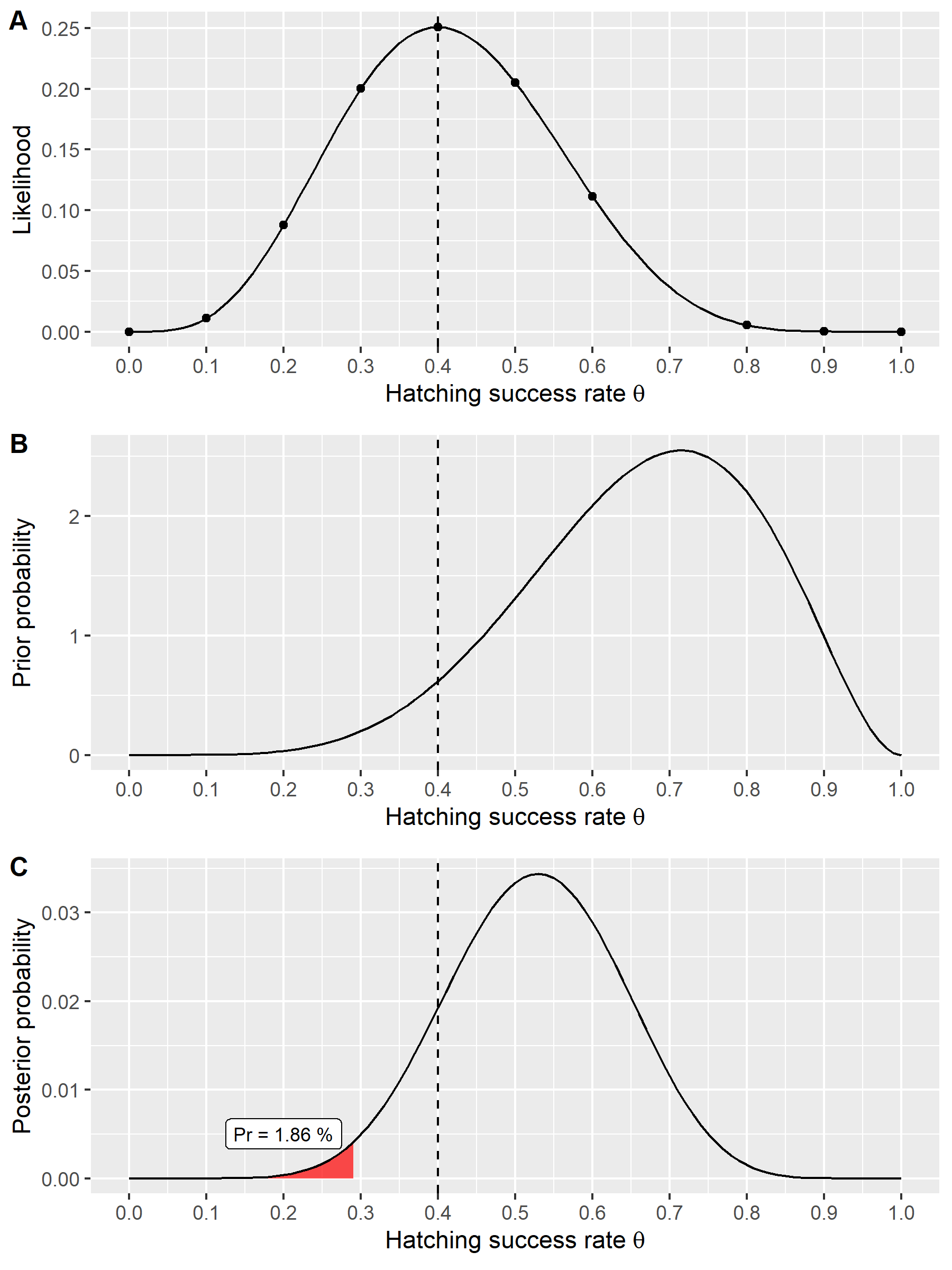
Figure 1a shows the resulting likelihood function. As expected, the most likely parameter for 4 successes out of 4 trials is 0.4. However, it also shows that other parameter values are certainly possible given our observation.



**Fig. 1.** The likelihood function for an equally spaced vector of hatching success rates in a binomial experiment with four successes in ten independent trials (**A**). The prior probability, which is assumed to be one over all hatching success rates (**B**). The posterior probability describing the probability of the parameters given the observed data and prior knowledge (**C**).

Now since we have the likelihood *P(D|H)*, we can multiply it with the prior probability *P(H)*. The first prior we use is a flat or uninformative prior simply being a vector of 1s over all hatching rates (Fig 1b). This means that all parameter values are equally likely to occur. Hence the shape of the likelihood function will not change after it is being multiplied with the likelihood (which is 1 in for all cases). However, once we normalize it by the total probability of the data *P(D)*, which is the sum of the likelihood and the prior, the resulting posterior integrates to 1 (see change on y axis) (Fig 1c). We now have estimated the posterior probability of rainbow trout hatching success rates given our observed data and prior knowledge. Going back to our experiment, we can now calculate the probability of observing germination rates of 30 % or less than that, as suggested by our colleague. This can be done by adding up all individual probabilities of the posterior probability distribution for the hatching parameter values from 0 to 0.3. This turns out to be a probability of 22.14 % of observing hatching rates less than or equal to 30 % given our data (see R script).

However, since the key feature of Bayesian statistics is to incorporate prior knowledge, let us reimagine the experiment. Instead of using a flat and uninformative prior, we conducted a literature review on rainbow trout hatching success rates. We found that on average hatching success was 68 % and that the data can described using a beta distribution with shape parameters α = 6 and β = 3 (Fig. 2b). This distribution represents our informative prior. Now we collect data and run the hatching experiment as described before. For the sake of argument let us assume that again four out of ten eggs hatched and that the likelihood is hence the same. Now we can calculate the posterior probability as before by multiplying the likelihood with the prior and normalizing by the total probability of the data. Running the analysis with the informative beta prior shows a drastic change in the posterior distribution compared to the posterior distribution based on the flat or uninformative prior (Fig 2c). Now the probability of observing germination rates equal or less than 30 % is only 1.86 %.



**Fig 2.** The likelihood function for an equally spaced vector of hatching success rates in a binomial experiment with four successes in ten independent trials (A). The prior probability, which is assumed to be beta distributed with parameters α = 6 and β = 3 over all hatching success rates (B). The posterior probability describing the probability of the parameters given the observed data and prior knowledge (C).

In summary, Bayesian statistics allows us to make probabilistic statements about hypotheses given the sampled data. It also requires us to include prior knowledge in from of a prior probability distribution. If no previous knowledge is available, then it can be represented by an uninformative or flat prior. Such a prior could come from a uniform distribution over a given interval, or from a normal distribution with a large standard deviation, hence exerting no influence on the calculation of the posterior probability. However, the use of non-informative priors has become increasingly criticized in favour for weakly-informative priors ([Lemoine, 2019](#_ENREF_1)). Furthermore, the new posterior distribution could also be used iteratively as the new prior for subsequent analyses of the same type. This highlights the benefit of the Bayesian data analysis approach as it allows to update the evidence based on (new) prior knowledge.

**References**

Lemoine, N.P., 2019. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. Oikos, 128(7): 912-928.

van Ravenzwaaij, D., Cassey, P., Brown, S.D., 2018. A simple introduction to Markov Chain Monte–Carlo sampling. Psychonomic Bulletin & Review, 25(1): 143-154.