Analytical solution First, we need to define our likelihood function, which gives the probability for a certain outcome based on the parameter we want to fit. For the case of repeated coin tosses, a binomial model is a sensible choice. Let us in this section call the probability of heads Θ. Then we get for the number k of heads among n coin tosses the following likelihood function:

$$p(k|n,\Theta) = \binom{n}{k}\Theta^k(1-\Theta)^{n-k}$$

Now, we define our prior assumption about possible values of Θ and their respective probability. In the absence of any further knowledge, we assume that all possible values of $\Theta \in [0;1]$ are equally likely. In other words, our prior is the uniform distribution

$$p(\Theta) = \begin{cases} 1 & \Theta \in [0;1] \\ 0 & \text{else} \end{cases}.$$

Finally, we also need to calculate the so-called marginal likelihood $p(k)$ – this is just the probability of getting k heads in the first place. The marginal likelihood is not really important, as it is a constant term not depending on Θ, which is just used to rescale the posterior, so that the posterior sums up to 1.0. In other words, the marginal likelihood is only used to ensure that we end up with a valid probability distribution.

To get this marginal likelihood $p(k)$, we integrate over all possible values of Θ to get the overall probability of ending up with k heads after the experiment. Thus:

$$p(k) = \int_{\Theta=0}^{1} p(k|n,\Theta)p(\Theta)\, d\Theta$$

Now we are able to compute the posterior distribution of Θ using Bayes' rule, which is given by

$$\text{Posterior} = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Marginal Likelihood}}.$$

Specifically in our case, we compute

$$p(\Theta|n,k) = \frac{p(k|n,\Theta)p(\Theta)}{p(k)}$$

We get the following posterior over the interval $[0;1]$:

$$p(\Theta|n,k) = \frac{\binom{n}{k}\Theta^k(1-\Theta)^{n-k}}{p(k)}$$

It is easy to verify that $p(k)$ is a constant independent of Θ, and so is $\binom{n}{k}$. Thus, our posterior is proportional to the remaining part of the expression, i.e.:

$$p(\Theta|n,k) \propto \Theta^k(1-\Theta)^{n-k}$$

The RHS of this corresponds to a Beta distribution with $a = k+1$ and $b = n-k+1$ without the normalising constant that would force the distribution's integral to 1. (This "core part" of a distribution without the normalising constant is sometimes also called the "kernel" of a distribution.) Thus, our posterior distribution is a Beta distribution, with a, b determined by n, k.

For $n = 2, k = 1$ we get a Beta distribution with parameters $a = 2, b = 2$, and for $n = 1000, k = 500$ we get the Beta distribution with parameters $a = 500, b = 500$.