Suppose I have two coins: one is double-headed and the other is fair. I pick one coin uniformly at random and flip it. Before you know the results of the flip, what is the probability that I picked the fair coin? 50%.

Now, consider the heads probability for the selected coin. Because the coin is random, this probability is actually a random variable! Let's call it $Y$. What is its distribution?

$$P(Y = 1/2) = \frac{1}{2} \qquad P(Y = 1) = \frac{1}{2}$$

Suppose I flip the coin and get a heads. Based on this information, do you still think there is a 50% chance that I had selected the fair coin? That is, do you still think that $P(Y = 1/2) = \frac{1}{2}$?

If now, how do we calculate the the probability that the coin is fair, in light of this new evidence. Let $X$ be the result of the coin flip, with $X = 1$ indicating a heads. We want$P(Y = 1/2|X = 1)$, the **posterior probability**.

We can calculate this using Bayes' Rule:

$$P(Y = 1/2|X = 1) = \frac{P(X = 1|Y = 1/2)}{P(X = 1)} P(Y = 1/2)$$

The first term in this expression is called **the likelihood ratio** and the second term is the **prior probability**. Observe by LoTP:

$$\begin{aligned}
P(Y = 1/2|X = 1) =& \frac{P(X = 1|Y = 1/2)}{P(X = 1)} P(Y = 1/2) \\
=& \frac{P(X = 1|Y = 1/2)}{P(X = 1|Y = 1/2)P(Y = 1/2) + P(X = 1|Y = 1P(Y = 1)} P(Y = 1/2) \\
=& \frac{\frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}} \frac{1}{2} \\
=& \frac{1/4}{3/4} = \frac{1}{3}
\end{aligned}$$

Now, suppose that instead of two coins, I have a very large number of coins (10 billion), each with a different probability of heads. What is a possible histogram for the distribution of heads probabilities for my coins?

Since I have so many coins, it might be more reasonable to model the heads probability as a continuous distribution, instead of a discrete one. A reasonable choice for this distribution is $Y \sim \text{Beta}(a, b)$, with density

$$f(y) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1 - y)^{b-1} \qquad 0 < y < 1$$

Again, I will choose 1 coin at random from among my very large number of coins. What is the prior distribution on the heads probability $Y$ of my chosen coin?

Now, suppose that the coin lands heads. What is the posterior distribution on the heads probability $Y$ of my chosen coin? Again, we use Bayes Rule (this time, the continuous version), along with continuous LoTP

$$f(y|X = 1) = \frac{P(X = 1|Y = y)f(y)}{P(X = 1)} = \frac{P(X = 1|Y = y)f(y)}{\int_0^1 P(X = 1|Y = t)f(t)\,dt}$$

The numerator is

$$P(X = 1|Y = y)f(y) = y\frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1 - y)^{b-1} = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} y^{a}(1 - y)^{b-1}$$

The denominator is

$$\int_0^1 P(X = 1|Y = t)f(t)\,dt = \int_0^1 t\frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1 - t)^{b-1}\,dt = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \int_0^1 t^{a}(1 - t)^{b-1}\,dt$$

How do we solve this integral? Notice that it looks very, very similar to the PMF of a Beta distribution! (We are just missing the coefficients). We know that from the Beta$(a+1, b)$ distribution

$$\int_0^1 \frac{\Gamma(a+1+b)}{\Gamma(a+1)\Gamma(b)} t^a (1-t)^{b-1}\, dt = 1$$

and so

$$\int_0^1 t^a (1-t)^{b-1}\, dt = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)}$$

And therefore, our original integral is

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 t^a (1-t)^{b-1}\, dt = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)}$$

Note that this is just a number (it does not depend on $y$).

Putting this together, our posterior probability is

$$f(y|X=1) = \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^a (1-y)^{b-1}}{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)}}$$

Can this unweildy expression simplify? (absolutely!) One way to do so is to use properties of the Gamma function. But we can also be sneaky!

Note that everything involving the gamma function is just a number. Call this number $c$:

$$c = \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}}{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)}}$$

so that

$$f(y|X=1) = c \cdot y^a (1-y)^{b-1}$$

Now, observe that $f(y|X=1)$ must be a probability density. So what happens if we integrate $f(y|X=1)$ across its support?

$$1 = \int_0^1 c \cdot y^a (1-y)^{b-1}\, dy$$

But we've already done this integral once before. $c \cdot y^a (1-y)^{b-1}$ is the density of Beta$(a+1, b)$ and so $c = \frac{\Gamma(a+1,b)}{\Gamma(a+1)\Gamma(b)}$.

In other words, we've learned that the posterior distribution of $Y|X=1$ is Beta$(a+1, b)$.

In this case, we started with a prior distribution of $Y$ which was some form of Beta distribution, and we ended up with a posterior distribution which was also some form of the Beta distribution (albeit with different parameters). This is an example of a **conjugate prior** relationship.