Posterior Distributions

In this lesson, we will introduce Posterior Distributions using an example.

Learn from Example

[It is] a spectacular vindication of the principle that each coin spun individually is as likely to come down heads as tails and therefore should cause no surprise that each time it does.

Thus Guildenstern (or is it Rosencrantz?) in Tom Stoppard's re-imagining of Hamlet "Rosencrantz and Guildenstern Are Dead". If you haven't seen it already, go watch the movie or at least the first scene.

It helps to know the plot of Hamlet: Before the play begins, prince Hamlet's uncle Claudius has murdered Hamlet's father, become king, and married Hamlet's mother, Gertrude. Hamlet is understandably perturbed by this sequence of events. Claudius and Gertrude have invited Hamlet's college friends Rosencrantz and Guildenstern to cheer him up, but SPOILER ALERT in a series of misadventures they end up dead, soon to be followed by everyone else; the plot, such as it is, of R&G Are Dead is this story told from their confused, uninformed, and amnesiac perspective.

.

.

.

Welcome back. As you now know, if you did not before, Rosencrantz and Guildenstern are flipping (or, as they say, spinning) a series of coins. In the film sometimes they are flipping a different coin each time, which Rosencrantz (or is it Guildenstern?) wins, and sometimes they're flipping the same coin several times in a row. Either way, the coin flipper (whatever his name is) hits an enormously unlikely sequence of heads.

Let's make up a variation on this scenario to see what we can learn about posterior distributions.

Introduction to Posterior Distributions

Suppose we have two kinds of coins: fair and double-headed. A fair coin has a $50 - 50$ probability of coming up heads or tails; a double-headed coin always comes up heads.

```
enum Coin { Fair, DoubleHeaded }
```

Now let's suppose we have a bag of coins. 999 are fair, and one is double-headed. Rosencrantz is going to pull a coin from the bag at random. Our prior is that the selection is random over the thousand coins in the bag, so:

```
var coins = new[] { Fair, DoubleHeaded };
var prior = coins.ToWeighted(999, 1);
```

Once he has his coin, Rosencrantz flips it, and of course, observes that it is *heads*.

The question we want to explore is: *what is the posterior probability that Rosencrantz just flipped the double-headed coin?*

That is, Rosencrantz draws a coin, flips it, it comes up heads, and now the question to you is: what would be fair odds for a bet on the question *"is it the double-headed coin or a regular coin that was just flipped?"* Before the observation, the fair odds are 1 to 999, but after we observe the flip, those odds change because we have more information.

Reasoning backward like this is a little tricky. (Of course, we call it the

posterior distribution because we are reasoning backward, in case that was unclear.) Again, let's break it down:

- Before flipping the coin, the probability of getting the double-headed coin is the aptly-named prior probability: 1 *in* 1000, which is odds of 1 *to* 999.

- If Rosencrantz flipped the coin and got tails, that would be solid evidence that he did not get the double-headed coin; the *posterior* probability of having the double-headed coin is zero. That is, the posterior probability is smaller: we've gone from 0.1% *to* 0%, a small decrease.

- If Rosencrantz got heads, that is very weak evidence that he got a double-headed coin, but it is not *zero* evidence. There should be some small increase over the prior in the posterior.

Of course, we don't need to do the math by hand; we know from earlier in this course that we can exactly solve for posterior probabilities by using a combination of `Where` and `SelectMany` query operators:

```
var results = new[] { Heads, Tails };
var prior = coins.ToWeighted(999, 1);

Func<Coin, IDiscreteDistribution<Result>> Likelihood = c =>
  results.ToWeighted(1, c == Fair ? 1 : 0);

var c1 = from c in prior
  from r in Likelihood(c)
  where r == Heads
  select c;
Console.WriteLine(c1.ShowWeights());
```

The output:

```
      Fair:999
DoubleHeaded:2
```

If Rosencrantz flipped heads, then the probability that he drew the double-headed coin is slightly increased over the prior. The prior probability of having the double-headed coin is 1 *in* 1000; the posterior probability is 2 *in* 1001. That's almost twice as likely, but still pretty unlikely.

Again, let's make sure that we understand what this means. It means that if we did millions of repetitions of this experiment where Rosencrantz picks a coin and flips it once, and if we *discarded* all the repetitions in which

Rosencrantz flipped tails, then in the remaining repetitions on average the ratio of double-headed coins to fair coins chosen would be about 2 *to* 999.

What if we increase the number of flips of that same coin, and he got two heads?

```
var c2 = from c in prior
         from r1 in Likelihood(c)
         where r1 == Heads
         from r2 in Likelihood(c)
         where r2 == Heads
         select c;
Console.WriteLine(c2.ShowWeights());
```

The output:

```
    Fair:999
DoubleHeaded:4
```

The posterior probability of having the double-headed coin goes up again, this time to 4 *in* 1003.

What if we increased it to ten flips (again, of the same coin), and they're all heads? By the time we get to ten heads, there are two possibilities: the *exactly* one-in-a-thousand chance that Rosencrantz got the double-headed coin, or the *approximately* one-in-a-thousand chance that he got a fair coin and flipped ten heads in a row; the ratio of those two probabilities is the posterior odds. At this point, it becomes pretty much even money whether he has the double-headed coin or a fair coin. (As you might expect, the exact odds are 999 "fair" to 1024 "double-headed", which is pretty close to $50 - 50$.)

If we observe eleven or more heads, it becomes much more likely that Rosencrantz has drawn the one-in-a-thousand double-headed coin than that he is lucky enough to observe that sequence of heads from a fair coin, and of course by the time we get to 156 heads, it is absurdly more likely to be the double-headed coin than a fair coin.

Continuous vs. Discrete Version

This is a cute problem and all, but it seems a little out of place at this point in our course; we stopped talking about discrete distributions that we could solve exactly for posteriors a while back. We are discussing it now because all this is in the way of introducing a variation on the above problem:

- We have a mint with really poor quality control; it produces coins that are out of balance so that some of them favor heads and some of them favor tails.
- To model this, let's say that the mint samples from some continuous distribution that produces values strictly between 0.0 and 1.0. Let's say that we know this distribution.
- Each coin that is minted is modeled as a Bernoulli distribution of heads and tails, with the given probability of heads.
- Rosencrantz flips a random coin that was obtained from that mint; this is our prior.
- We observe that Rosencrantz gets heads.
- What is the posterior distribution on the fairness of that coin?

At first glance, you might find this continuous version of the problem a bit harder to figure out, but fortunately, it is not fundamentally different than the discrete version we gave above. In both cases, we have a prior distribution of coin fairness that we sample from to get a coin. In both cases, we then obtain an observation of "heads". And in both cases, we can use that observation to deduce something about the likely provenance of the coin. Whether we're in the continuous case or the discrete case, flipping heads is evidence that the coin is more likely to be on the "heads heavy" side of the prior distribution. But how much more likely?

In the all-discrete case, we could compute that posterior exactly. The question before us now is: given the *continuous* prior, the discrete likelihood function and an observation of a discrete value, can we compute the weight function of the continuous posterior? That is, can we answer the question "*what is the probable fairness level of this coin given that we've observed one head?*" And of course, given all that information, can we also produce a distribution object that we can sample from?

Give it some thought.

Implementation

Let's have a look at the code:

Program.cs
Bernoulli.cs
BetterRandom.cs
Distribution.cs
DistributionBuilder.cs
Empty.cs
Episode29.cs
Extensions.cs
Flip.cs
IDiscreteDistribution.cs
IDistribution.cs
IWeightedDistribution.cs
Markov.cs
MarkovBuilder.cs
Metropolis.cs

```csharp
using System;

namespace Probability
{
    using static Coin;
    using static Result;

    enum Coin { Fair, DoubleHeaded }

    enum Result { Heads, Tails }

    static class Episode29
    {
        public static void DoIt()
        {
            Console.WriteLine("Episode 29 -- Rosencrantz flips a coin");
            // 999 fair coins, one double-headed coin :
            var coins = new[] { Fair, DoubleHeaded };
            var results = new[] { Heads, Tails };
            var prior = coins.ToWeighted(999, 1);
```

```
        Func<Coin, IDiscreteDistribution<Result>> Likelihood = c =>
            results.ToWeighted(1, c == Fair ? 1 : 0);


        var c1 = from c in prior
                from r in Likelihood(c)
                where r == Heads
                select c;
        Console.WriteLine(c1.ShowWeights());

        var c2 = from c in prior
                from r1 in Likelihood(c)
                where r1 == Heads
                from r2 in Likelihood(c)
                where r2 == Heads
                select c;
        Console.WriteLine(c2.ShowWeights());
    }
  }
}
```

In the next lesson, we will attempt to produce a distribution object that we can sample from, using the tools we've developed thus far.