

A Computer Scientist in a Business School

Random thoughts of a computer scientist who is working behind the enemy lines; and lately turned into a double agent.

Tuesday, January 15, 2008

Are you a Bayesian or a Frequentist? (Or Bayesian Statistics 101)

OK, the previous post was actually a brain teaser given to me by Roy Radner back in 2004, when I joined Stern, in order to teach me the difference between Bayesian and Frequentist statistics. It actually illustrates nicely how the two techniques lead to different conclusions. For completeness, let me list the question again:

You have a coin that when flipped ends up head with probability p and ends up tail with probability $1 - p$. (The value of p is unknown.)

Trying to estimate p , you flip the coin **14** times. It ends up head **10** times.

Then you have to decide on the following event: "*In the next two tosses we will get two heads in a row.*"

Would you bet that the event will happen or that it will not happen?

One of the basic differences of Bayesian and Frequentists is how they treat the parameters. Using frequentist statistics (the statistics that I learned in school :-)) we would say that the best (maximum likelihood) estimate for p is $p = \frac{10}{14}$, i.e., $p \approx 0.714$.

In this case, the probability of two heads is $0.714^2 \approx 0.51$ and it makes sense to bet for the event.

Bayesian approach: p is not a value, it's a distribution

A Bayesian approach, instead of considering only the ML estimate for p , it would treat p as a random variable with its own distribution of possible values. The distribution can be defined by the existing evidence. The logic goes as follows. What is the probability of a given value of p , given the data?

$$Pr\{p|data\} = \frac{Pr\{data|p\} \cdot Pr\{p\}}{Pr\{data\}}$$

Since the term $Pr\{data\}$ is going to be the same for all values of p , we can for now ignore it. We will see that we can infer it later, as the probability distribution when integrated over all values of p will need to be equal to 1.

The value $Pr\{data|p\}$ is easy to compute using simple statistics. The result of the experiment is a binomial distribution and we have

$$Pr\{data|p\} = \binom{14}{10} \cdot p^{10} \cdot (1 - p)^4$$

Again, ignoring the binomial coefficient $\binom{14}{10}$, we have:

$$Pr\{p|data\} \propto p^{10} \cdot (1 - p)^4 \cdot Pr\{p\}$$

where the \propto symbol means proportional.

Now, the factor $Pr\{p\}$, the prior distribution, comes into play. A **very convenient** prior distribution for this scenario (also known as the **conjugate prior** of the binomial) is the **Beta distribution**, $Beta(p; a, b)$ defined as:

$$Beta(p; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot p^{a-1} \cdot (1 - p)^{b-1}$$

In this case, we have:

About Me

Panos Ipeirotis

[View my complete profile](#)

Subscribe

 [Blog Posts](#)

 [Subscribe by Email](#)

Twitter

[Tweets by @Ipeirotis](#)

Total Pageviews

Popular Posts

[Why was my Amazon Mechanical Turk registration denied?](#)

[Identify Verification \(and how to bypass it\)](#)

[Are you a Bayesian or a Frequentist? \(Or Bayesian Statistics 101\)](#)

[How many Mechanical Turk workers are there?](#)

[The Google attack: How I attacked myself using Google Spreadsheets and I ramped up a \\$1000 bandwidth bill](#)

[Mechanical Turk account verification: Why Amazon disables so many accounts](#)

[The New Demographics of Mechanical Turk](#)

[Why I will never pursue cheating again](#)

[How big is Mechanical Turk?](#)

[Uncovering an advertising fraud scheme. Or "the Internet is for porn"](#)

Blog Archive

► [2018](#) (2)

► [2017](#) (1)

► [2016](#) (3)

► [2015](#) (4)

► [2014](#) (4)

► [2013](#) (10)

► [2012](#) (30)

► [2011](#) (29)

► [2010](#) (40)

$$Pr\{p\} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot p^{a-1} \cdot (1-p)^{b-1}$$

Dropping once again the constant:

$$Pr\{p|data\} \propto p^{10} \cdot (1-p)^4 \cdot p^{a-1} \cdot (1-p)^{b-1}$$

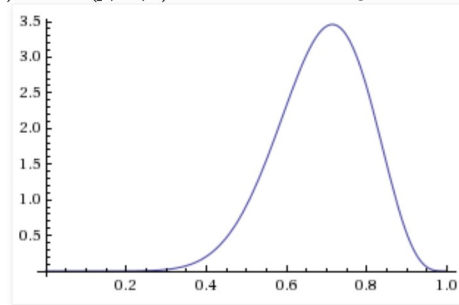
$$Pr\{p|data\} \propto p^{(10+a-1)} \cdot (1-p)^{(4+b-1)}$$

Now, what distribution is that? We need the normalizing factor to get the proper distribution. Looking at the form of the Beta distribution above, we can see that this is again a Beta distribution, and the normalizing factor is

$$\frac{\Gamma((10+a)+(4+b))}{\Gamma(10+a)\Gamma(4+b)} = \frac{1}{B(10+a, 4+b)}$$

where $B(x,y)$ is the [Beta function](#) (not to be confused with the [Beta distribution](#)!)

Therefore the distribution for the parameter p is $Beta(p; a+10, b+4)$. If we assume that we know nothing about p , we can assume that the prior is a uniform distribution, and the uniform distribution is a special case of Beta, with $a = b = 1$. In this case the distribution for p is $Beta(p; 1+10, 1+4) = Beta(p; 11, 5)$, which has the following form:



Probability of "Two Heads in a Row"

Now, we need to compute the probability of the event "two heads in a row", given the data that we have, which is $Pr\{HH|data\}$. To simplify:

- We do not update p between the two tosses.
- We assume conditional independence that, given p , the two tosses are independent.

Given that we will not update the value of p , we can break the process into two steps. First, we compute the possible values of p as described above; then, for each value of p we compute the probability of two heads in a row. These steps are encoded into the following integral computation:

$$Pr\{HH|data\} = \int_0^1 Pr\{HH|p\} \cdot Pr\{p|data\} dp$$

Given the conditional independence of the two tosses:

$$Pr\{HH|p\} = [Pr\{H|p\}]^2 = p^2$$

So, we have:

$$Pr\{HH|data\} = \int_0^1 p^2 \cdot Pr\{p|data\} dp$$

From above, we know that:

$$Pr\{p|data\} = \frac{1}{B(10+a, 4+b)} \cdot p^{10+a-1} \cdot (1-p)^{4+b-1}$$

By replacing this value, we get:

$$Pr\{HH|data\} = \frac{1}{B(10+a, 4+b)} \int_0^1 p^{(10+a-1)+2} \cdot (1-p)^{4+b-1} dp$$

Solving the integral, we have:

$$Pr\{HH|data\} = \frac{B(10+a+2, 4+b)}{B(10+a, 4+b)}$$

If we assume that we know nothing about p , we can assume that the prior is a uniform distribution, and the uniform distribution is a special case of Beta, with $a = b = 1$. In this case, the probability of the event "two heads in a row" is $\frac{B(13,5)}{B(11,5)} = 0.485$ and it makes sense to bet against the event. So, the Frequentist approach gives probability **51%** and the Bayesian approach with uniform prior gives **48.5%**. Kudos to Roy

► [2009](#) (41)

▼ [2008](#) (47)

► [November](#) (5)

► [October](#) (6)

► [September](#) (4)

► [August](#) (1)

► [July](#) (4)

► [June](#) (2)

► [May](#) (3)

► [April](#) (3)

► [March](#) (8)

► [February](#) (6)

▼ [January](#) (5)

[Who is Reviewing the Reviewers?](#)

[Misunderstandings of Power-Law Distributions](#)

[Are you a Bayesian or a Frequentist? \(Or Bayesian ...](#)

[Are you a Bayesian or a Frequentist?](#)

[Defining Probability in Prediction Markets](#)

► [2007](#) (32)

for coming up with example, and shame on me for screwing up the initial posting!

(**Update** based on Foster's comment below: instead of using the uniform distribution as a prior, we can be even more agnostic. In this case, we can use the Beta(0,0) distribution as a prior. Such a distribution corresponds to the case where *any mean* of the distribution is equally likely. In this case, the two approaches, Bayesian and frequentist give the same results.)

(**Update 2** based on multiple comments below: I changed the example to be more "fully Bayesian". Initially, I was estimating the full distribution for p , but then, in order to compute the "two heads in a row" probability, I was collapsing p into a point estimate, using its mean value. Now, with the help of the commenters, the example follows closer the Bayesian paradigm. I had to change the number of heads and tails for the example to still be valid. [WolframAlpha was surprisingly easy to use for identifying numbers that will work for this example](#). The values (3 heads, 1 tails), (5 heads, 2 tails), (8 heads, 3 tails), (try your own value of tails)... also work.)

Posted by [Panos Ipeirotis](#) at 7:03 PM 

[53 Comments](#)

Labels: [bayesian](#), [frequentist](#), [statistics](#)