

```
{:title "Python - Prior, Posterior Probability and Normalization" :layout :post
:date "2017-1-10" :tags ["Python" "Machine Learning" "Guide"]}
```

When learning about Bayes's theorem there is often talk about Priors, Posterior Probability, and Normalization. Those are fancy words when I first heard them but really they are simple concepts so I will try to explain them in the simplest terms I can. So let's start with trying to give a basic idea of a prior. A way to think about a prior is as your belief in a certain event or information. If a person believes that there's equal chances for either event to occur such as maybe a coin flip then that person has a prior of 0.5 for heads and 0.5 for tails. One thing to note about the prior is that this is a value assigned before any evidence and before the coin is even flipped. Now let's say we flip the coin a thousand of times and it seems to favor heads slightly more than tails for whatever reason then your original prior can be updated to favor the evidence and so now your prior could be 0.60 for heads and 0.40 for tails. Normalization is done by first finding the probability of being heads and not being heads which is done by multiplying by the prior and then adding up the probabilities. Adding up the probability will give you the total probability which does not usually add up to one. In that case you have to normalize it by adding both probabilities and then dividing the probability of the event over the total probability. The posterior probability is then determined by calculating the probability of the event by multiplying by the prior but this time dividing by the total probability so that the probability of not occurring will equal to 1. So you can think of the posterior probability as your updated probability after examining the data. So let's look at an example of putting all this stuff into practice. Let's say we have two guys Jake and Tony who both like Emily. Let's say our belief/prior is that they both have a fifty-fifty chance of hooking up with Emily. Now the next part is a bit unrealistic as we are going to quantify personality, appearance, and intelligence which are things that are quite difficult to do so. However for now we'll just be using this example to demonstrate and look at a more realistic example later on. So let's say Tony has 0.5 appearance, 0.3 intelligence, and 0.2 personality. Notice we are making the character traits into a probability adding to 1. Now the same for Jake but he may not be as appearance inclined as Tony but has slightly more intelligence. So Jake has 0.3 appearance, 0.5 intelligence, 0.2 personality. Now let's say Emily doesn't care about appearance since you know it's overrated now a days. So she values intelligence and personality so how do we know who is more likely to hook up with Emily? From just common sense you could probably figure it out since it isn't really a close call but let's do the math. So let's write a quick python script to figure it out or you could just do the math on a calculator if you want but I would rather create a small script so if you wanted to change the values you would do it rather quickly and easily.

```
def main():

    prior = 0.5
    tony = {"Appearance": 0.5, "Intelligence": 0.3, "Personality": 0.2}
```

```

jake = {"Appearance": 0.3, "Intelligence": 0.5, "Personality": 0.2}

total_probability = (tony["Intelligence"] * tony["Personality"] * prior) + (jake["Intelligence"] * jake["Personality"] * prior)

print(total_probability)

tony_posterior_probability = (tony["Intelligence"] * tony["Personality"] * prior) / total_probability
jake_posterior_probability = (jake["Intelligence"] * jake["Personality"] * prior) / total_probability

print("Tony's Posterior Probability: ", tony_posterior_probability)
print("Jake's Posterior Probability: ", jake_posterior_probability)

if __name__=="__main__":
    main()

```

After running the calculations you'll find that the results are that Jake has a higher posterior probability and that makes sense since you put into consideration that Emily doesn't care about appearance. So you would get that Tony has a posterior probability of 0.375 and Jake has a posterior probability of 0.625. Sad news for Tony unfortunately.