INFSCI 2595
Quiz Review Group 4 - Week 4

Bayesian Formation and Beta distribution
Yixin Zhou

Since the posterior probability is proportional to the product of the prior probability and the probability that the posterior event will occur given the prior event, we can estimate the probability μ in the same way with Bayesian formulation:

$$p(\mu|m, N) \propto f(\mu) \cdot p(\mu) \tag{1}$$

Then we can use BETA distribution to encode the prior probability p(μ). BETA distribution is a probability density function for continuous variables between 0 and 1, whose shape is controlled by the hyper-parameters α and β:

$$p(\mu|a, b) = Beta(\mu|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)+\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1} \tag{2}$$

As $\frac{\Gamma(a+b)}{\Gamma(a)+\Gamma(b)}$ is a coefficient, we can focus on $Beta(\mu|a, b) \propto \mu^{a-1}(1-\mu)^{b-1}$, which has the same functional form as the Binomial distribution: $Binomial(m|\mu, N) \propto \mu^m(1-\mu)^{N-m}$. Beta distribution is the conjugate prior of the binomial likelihood. Then we can find that the posterior distribution on μ is:

$$p(\mu|m, N) = Beta(\mu|a+m, b+(N-m)) \tag{3}$$

We can view $(a+m)$ as a_{new} and $(b+(N-m))$ as b_{new}, then we get:

$$p(\mu|m, N) = \mathrm{Beta}(\mathrm{a_{new}}, \mathrm{b_{new}}) \tag{4}$$

Finally we will update our belief about μ under this circumstances: The posterior distribution is Beta distribution.

Q&A

Q1: How to choose the prior distribution?
A: Based on the observation. The larger the size, the less sensitive the data is.

Q2: In what circumstance we should change our belief about μ ?
A: When the observation is really large and useful.

Probability
Jinghong Zhang

The Probability is proportion (fraction) of times an event occurs out of the total number of trials. It reflects the probability of random events appearing. Random events are events that may or may not occur under the same conditions. For example, if you randomly select one item from a batch of genuine and defective products, "the one you get is genuine" is a random event. Suppose that n times of experiments and observations are made on a random phenomenon, in which event A occurs m times, that is, the frequency of its occurrence is m/n. After a lot of trial and error, m/n often gets closer and closer to a certain constant (see Bernoulli's law of large numbers for the proof of this assertion). This constant is the probability of the occurrence of event A, commonly expressed as P (A).

Questions:

1. Given three boxes A.B.C. A has two red balls .B has two blue balls .C has one red ball and one blue ball. Pick a box at random, pick a ball from the box at random, find that it is a red ball, and find the probability that the other ball in the box is also a red ball.
   (2/3)

2. How many dice rolls does it take on average to get a 6?
   (6)

Independence is a fundamental notion in probability theory, as in statistics and the theory of stochastic processes. Two events are independent, statistically independent, or stochastically independent[1] if, informally speaking, the occurrence of one does not affect the probability of occurrence of the other or, equivalently, does not affect the odds. Similarly, two random variables are independent if the realization of one does not affect the probability distribution of the other.

When dealing with collections of more than two events, two notions of independence need to be distinguished. The events are called pairwise independent if any two events in the collection are independent of each other, while mutual independence (or collective independence) of events means, informally speaking, that each event is independent of any combination of other events in the collection. A similar notion exists for collections of random variables. Mutual independence implies pairwise independence, but not the other way around. In the standard literature of probability theory, statistics, and stochastic processes, independence without further qualification usually refers to mutual independence.

1. There are two white balls and one black ball in the bag. Now I have put it back and take the ball from the bag, one at a time, A is "the first time to get the white ball", B is "the second time to get the black ball", what is the P(AB)?
   (2/9)

2. (True/False)If $A$and B are unrelated to each other, they are independent.
   (False)

Binomial Distributions
Vishruth Reddy

Real life example:
Suppose you are looking at the ratings of objects on Amazon. You come across this product that is sold by different sellers

Seller 1: 10 reviews 100% positive rating Seller2: 50 reviews and 96% positive rating Seller3: 200 reviews 93% positive rating

Intuition: The more data we see gives us more confidence in data rating. As the rating can go haywire even if change one review out of just 10 observations. How do we put this quantitatively?

What are we modeling?

Every seller is producing random experiences and each seller has a constant underlying success rate (S) associated with it. Ex: for a seller ‘x’ the P(Positive experience) = 0.75

Note: for first seller, the ratings are 10/10 that does not imply that P(positive experience) for seller 1 is not 1, can 0.95 or 0.90. This we don’t know.

What are we optimizing?
Basically want to maximize your probability of having a good experience. (Still not sure about the success rate).

So we have many probabilities associated with a seller. We want to know the the likelihood of each of these probabilities for the particular seller. In cases like coin flips and dice rolls we know that the underlying probability of getting a heads (in case of coin flips) is 0.5 and getting a number from 1 to 6 in the case of a dice roll is 1/6. But this is not the case here. We have an uncertainty associated with the longterm performance of the seller itself.

Say we knew the true success rate for a seller,
How would we compute the probability of 10 positive reviews and 0 negative reviews for seller 1 P(10 positive, 0 negative, given that s = 0.95)? Or P(48 positive, 2 negative, given that s = 0.95) for seller 2? i.e what is the Probability of data given an assumed success rate?

We can write this as,

$P(data|s)$

Probability of data given as assumed success rate?

One way we can do this is by plotting the frequency of obtaining 48 positive reviews and 2 negative reviews out of 50 reviews by randomly sampling again and again.

1-> positive review and 0-> negative review We can write the different samples as

00011111111111111111111111111111111111111111111111 - 3 negatives 47 positives
11101110111111111111111111111111111111111111111111 - 2 negatives 48 positives
11101110111111011111111111111111111111111111111111 - 3 negatives 47 positives
11111111111111111111111011111111111111111110111111 - 2 negatives 48 positives
11111111111111111111111111111110111111111111111111 - 1 negatives 49 positives
11101111111111110111111111111111111111111111111111 - 2 negatives and 48 positives and so on...

And then find a sense how probable it is to obtain a sample which has 2 negative reviews and 48 positive reviews based on the frequency. But that's not practical right?

We can model this using an exact formula that looks like this,

$P(48 \text{ positive}, 2 \text{ negative} \mid s = 0.95) = (^{50}C_{48}) (0.95)^{48} (1-0.95)^{2} = 0.26$

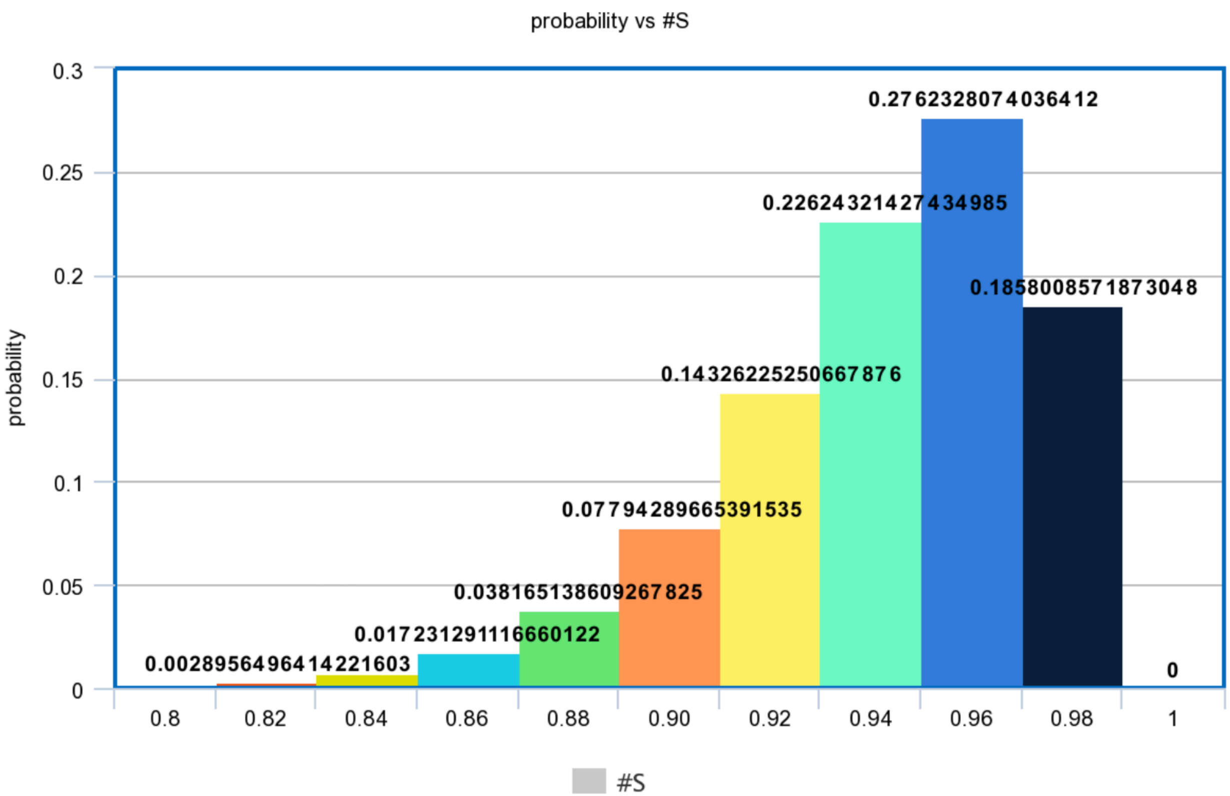
Lets break it down term by term,

$(^{50}C_{48})$ represents the number of ways you can 'choose' 48 slots out of 50 i.e 48 good reviews out of 50 total reviews. It is around 1225 ways.

We multiply it by the probability of seeing a single positive review 48 times i.e (0.95)^48 And multiply it by the probability of seeing a single negative review 2 times ie (1-0.95)^2

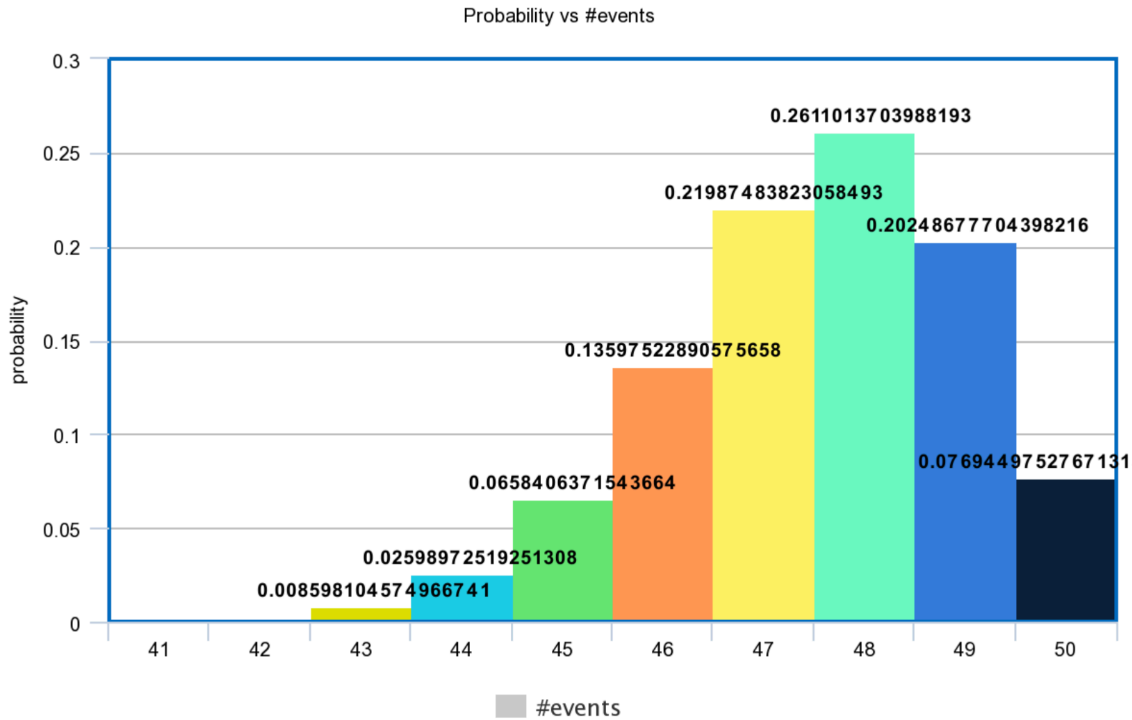
Important thing to note: Each review is independent of the last hence we are able to multiply the probabilities.

Lets plot for the probabilities of different values of positive and negative reviews given S = 0.95



This is called a binomial distribution where given an assumed success rate we are calculating the probability of one of the trials.

Now, let’s tinker around with different values of S.



We can observe that the value of S for which 48 positive reviews and 2 negative reviews is most likely to happen around S = 0.96, which is what we started out with! If we decrease the value of S the probability of getting 48 positive reviews decreases and similarly of we increase the value of S the probability again decreases. Which is pretty intuitive as the success rate reaches 1 the chance of getting 2 negative reviews is almost 0.

We can generalize the above formula by,
P(data|s) = c . s^(no of positive reviews) . (1-s)^(no of negative reviews)

If we had 500 reviews and wanted to see the probability of 480 positive reviews and 20 negative reviews the plot would be still centered around 0.96 but would more concentrated.