Bayesian Analysis

- The Bayesian world-view interprets probability as measure of *believability in an event*, that is, how confident we are in an event occurring.
- Frequentists, known as more classical version of statistics, assume that probability is the long-run frequency of events. This makes sense for many probabilities of events, but becomes more difficult to understand when events have no long-term frequency of occurrences.
- Bayesians have more intuitive approach.

Bayes Theorem

Conjoint probability is the probability that two things are true.

P(*A* and *B*) is the probability that *A* and *B* are both true.

For independent events P(*A* and *B*) = P(*A*) p(*B*).

In general, p(*A* and *B)* = P(*B*|*A*) p(*A*).

A and B are interchangeable P(*B* and *A*) = P(*A*|*B*) p(*B*)

Bayes theorem:

$$P(B|A)\,p(A) = P(A|B)\,p(B)$$

$$P(A|B) = P(B|A)\,p(A)\,/p(B)$$

The Cookie problem:

- Bowl 1 contains 30 vanilla cookies and 10 chocolate cookies.
  Bowl 2 contains 20 vanilla cookies and 20 chocolate cookies.
- Suppose you choose one of the bowls at random and select a cookie at random. The cookie is vanilla. What is the probability that it came from Bowl 1?
- We want to find p(Bowl 1 | vanilla), but it is not obvious to compute.

- It is easy to find p(vanilla | Bowl 1) = 30/40 = ¾
- Using Bayes theorem
  p(Bowl 1| Vanilla) = p(Vanilla| Bowl 1) p(Bowl 1)/p(Vanilla)
- p(Bowl 1) = ½, the probability that we chose Bowl 1
- p(Vanilla| Bowl 1) is the probability of getting a vanilla cookie from Bowl 1, ¾
- P(vanilla) is the probability of drawing a vanilla cookie from either bowl, 50/80 or 5/8.
- p(Bowl 1| Vanilla) = (3/4) * (1/2) / (5/8) = 3/5

Bayesian framework

- Bayes's theorem gives us a way to update the probability to update the probability of a hypothesis, H, in light of some body of data, D.
    - Rewriting with H and D yields:
      $P(H|D) = P(H)\ p(D|H)/p(D)$
- p(H) is the probability of the hypothesis before we see the data, called the prior probability
- p(H|D) is the probability of the hypothesis after we see the data, called the posterior
- p(D|H) is the probability of the data under the hypothesis H, called the likelihood
- p(D) is the probability of the data under any hypothesis, called the evidence or normalizing constant

Monty Hall problem

- The object of the game is to guess which door has the car. If you guess right, you get to keep the car.
- You pick a door, which we will call Door A. We’ll call the other doors B and C.
- Before opening the door you chose, Monty increases the suspense by opening either Door B or C, whichever does not have the car. Monty opens door B.
- Then Monty offers you the option to stick with your original choice or switch to the one remaining unopened door.

Bayesian Solution of Monty Hall problem

- You picked door A and Monty opens door B.
- In this problem likelihood is the probability to open door B (our data) under different hypothesis.

Hypothesis	Prior p(H)	Likelihood P(D|H)	P(D|H)p(H)	Posterior p(H|D)
A	1/3	1/2	1/6	1/3
B	1/3	0	0	0
C	1/3	1	1/3	2/3

M & M problem

- In 1995, they introduced blue M & M’s.
- Before then, the color mix was 30% brown, 20% yellow, 20 % red, 10% Green, 10% orange, 10% tan.
- Afterward it was 24% blue, 20 % green, 16% orange, 14% yellow, 13% red, 13% brown.
- You have 2 bags, one is from 1994 and one from 1996. you draw one from each bag. One is yellow and one is green. What is the probability that the yellow one came from the 1994 bag?
  Hypothesis A: bag 1 is from 1994 and bag 2 is from 1996.
  Hypothesis B: Bag 1 is from 1996 and bag 2 from 1994.

M & M problem

Hypothesis	Prior P(H)	Likelihood p(D|H)	p(D|H)p(H)	Posterior p(H|D)
A	1/2	(20)(20)	20*20*1/2=200	20/27
B	1/2	(14)(10)	14*10*1/2=70	7/27

Normalizing constant isp(D) = p(D|A)p(A) + p(D|B)p(B) = 200 + 70 = 270

Posterior p(H|D) = p(D|H) * p(H) / p(D)

Conjugate prior - Bernoulli distribution

- $P(H|D) = P(H)\ p(D|H)/p(D)$
- It would be convenient if the product of $p(D|H)$ and $p(H)$ results in a function of the same form as $p(H)$.
- The prior and posterior are described using the same form of function
- When the forms of $p(D|H)$ and $P(H)$ combine so that the posterior distribution has the same form as prior, then $p(H)$ is called a conjugate prior
- Binomial: Coin flipping (head or tail),
  - the proportion of free throws hit by a player,
  - The proportion of babies born that are girl
  - the proportion of people who agree with a statement on a survey

Flipping coin - Conjugate prior.

- For binomial distribution we need a mathematical formula that describes the prior belief for each value of the bias θ in the interval [0,1].
  - The probability of the coin up $p(y=1|\theta)$. Generally, $p(y|\theta) = \theta^{y}(1-\theta)^{1-y}$
  - When we flip coin N times and the number of heads $z=\sum y_i$ the likelihood $p(z, N \mid \theta) = \theta^{z}(1-\theta)^{N-z}$
  - If the prior is of the form $\theta^{a}(1-\theta)^{b}$ then our posterior will be a function of the same form $\theta^{(z+a)}(1-\theta)^{(N-z+b)}$
  - A probability density of that form is called a beta distribution.

Beta distribution

- Formally, a beta distribution has two parameters a and b, and the density itself is defined as
  $p(\theta|a, b) = \text{beta}(\theta; a,b) = \theta^{(a-1)}(1 - \theta)^{(b-1)} / B(a,b)$
- The mean of the beta(θ; a,b) is $\langle\theta\rangle = a/(a+b)$. For a = b, mean $\langle\theta\rangle = ½$.
- The standard deviation of the beta distribution is sqrt{$\langle\theta\rangle(1 - \langle\theta\rangle)/(a + b + 1)$}. Notice that the standard deviation gets smaller when a + b gets larger.