A variable is a container. It can take on different values. For example x can be 0, 1, or 2.

A random variable takes on different values with probabilities.

If our variable x is a random variable then this means there are probabilities associated with the values it can take.

Since x can be 0, 1, or 2 in our example, each of these also has probabilities.

For a random variable x we can ask what is P(x=0) the probability that x has value 0?

For example I can make a random variable X that takes on the values 0, 1, 2, and 3 with the following probabilities:

P(X=0) = .2
P(X=1) = .3
P(X=2) = .4
P(X=3) = .1

The above set of probabilities for X is called the probability distribution of X

What happens if a random variable X takes on continuous values? Suppose we have X that can take on any value between 0 and 1. How do we model its distribution? Just like the sum of all probabilities for a discrete random variable (like given above) should sum to 1 we require that a continuous random variable has integral 1. In other words $\int_0^1 P(x)dx = 1$. For a continuous random variable the probability at a given point x is undefined. Only the cumulative probability is defined.

What is the cumulative? It's the sum of probabilities up to a certain point.

Suppose we toss a coin n times. Suppose probability of head is p.

Let's give exact numbers. Suppose we toss the coin 2 times and probability of head is 0.5.

Q1. What is the probability that I will see exactly one head in the first position?

Pr = Size of event space / size of sample space

What is the size of sample space? There are four possibilities if I toss the coin twice:

HH
HT
TH
TT

If I toss the coin 3 times what is the size of sample space? Is it 9?

HHH
HHT
HTH
HTT
THH
THT
TTH
TTT

It’s 8.

If I toss it 4 times what is the sample space size?

_ _ _ _

16

Now if I toss it n times what is the size of sample space? It’s 2^n

Suppose we toss the coin n times. What is the probability of getting exactly one head in position 1 and remainder all tails? The answer is $1/2^n$

What is the probability of getting exactly one head and remainder all tails?

Suppose we did 5 tosses. The sample space size is 2^5 = 32. The event space is tosses where we have exactly one head and remainder tails.

HTTTT
THTTT
TTHTT
TTTHT
TTTTH

For 5 tosses the answer is 5/32.

Instead of 5 tosses suppose we did n tosses. What is the probability of getting exactly one head in position 1 and remainder tails?

Pr = $n/2^n$

What is the probability of getting exactly two heads and remainder all tails?

HHTTT
HTHTT
HTTHT
HTTTH
THHTT
THTHT
THTTH
TTHHT
TTHTH
TTTHH

H _ _ _ _ 4 +
_H_ _ _ + 3
_ _ H _ _ + 2
_ _ _ H H + 1 = 10

The number of ways to select two unique objects from a group of n objects is n choose 2 = n(n-1)/2
The number of ways to select k unique objects from a group of n objects is

n choose k = $\frac{n!}{(n-k)!k!}$

In general we can ask what is the probability of getting exactly k heads in n tosses. The answer is (n choose k)/ 2^n

We can understand the formula for n choose k by starting with permutations. How many ways are there to permute n unique numbers? The answer is n!.

Suppose we I toss a fair coin n times. What is the probability of seeing at most 3 heads?

Answer: ((n choose 1) + (n choose 2) + (n choose 3))/2^n

What is a probability distribution?

Example of a Bernoulli random variable distribution: Single coin lands on head with probability 0.7 and tail with probability 0.3.

Example of binomial distribution: It's the sum of repeated independent Bernoulli trials.

Independent events:

Pr(Event X and Event Y) = Pr(Event X) * Pr(Event Y)

if and only if X and Y are independent.

Suppose we toss a coin three times. The probability of head is p.

What is the probability of seeing a head in each of the three tosses?

Pr(Firsttoss = H **and** secondtoss = H **and** thirdtoss = H)=
Pr(Firsttoss=H) * Pr(secondtoss = H) * Pr(thirdtoss = H) = p * p * p = p^3

Suppose we toss a coin four times and the probability of head is p.
What is

Pr(Firsttoss = H **and** secondtoss = T **and** thirdtoss = H **and** fourthtoss = T) =
Pr(Firsttoss = H) * Pr(secondtoss = T) * Pr(thirdtoss = H) * Pr(fourthtoss = T) = p*(1-p)*p*(1-p) = p^2 * (1-p)^2

Expected value of random variable X = $\sum_{i=1}^{n}$ Pr(X=i)i

If X is Bernoulli with probability of success p what is E(X) = 0*(1-p) + 1*(p) = p

Define X = $\sum_{i=1}^{n} X_i$ where each Xi is a Bernoulli variable with probability of success p. In other words X is a binomal random variable. Recall that a Bernoulli random variable has the distribution:
P(X=0) = 1-p
P(X=1) = p

What is $E(X) = E(\sum_{i=1}^{n} X_i)$?

We apply the distributive property of expectation: E(X+Y) = E(X) + E(Y)

What is $E(X) = \sum_{i=1}^{n} E(X_i) = np$?

So this means if you flip a fair coin 100 times we can expect to see 100*(0.5) = 50 heads. Suppose the coin is biased and p = .1. Now if we flip it 100 times I expect 100*.1 = 10 heads.

Probability distribution of a Binomial random variable which sums three Bernoullis. I can model this as three coin flips. The number of outcomes are given below. To complete the distribution I need to assign a probability to each outcome. Suppose the probability of head P(X=1) is p.

Outcome	Probability
HHH	P(H and H and H) = P(H)P(H)P(H) = p*p*p
HHT	p*p*(1-p)
HTH	p*(1-p)*p
HTT	p*(1-p)*(1-p)
THH	(1-p)*p*p
THT	(1-p)*p*(1-p)
TTH	(1-p)*(1-p)*p
TTT	(1-p)*(1-p)*(1-p)

Outcome	
0 Heads	(1-p)*(1-p)*(1-p)
1 Head	p*(1-p)*(1-p) + (1-p)*p*(1-p) + (1-p)*(1-p)*p
2 Heads	p*p*(1-p) + p*(1-p)*p + (1-p)*p*p
3 Heads	p*p*p

Basic hypothesis testing:

We are given some data and we want to know if the data "agrees" with some distribution. The basic idea is to determine the probability of the data under a null distribution. If the probability is small we can reject the null hypothesis.

Variance of random variable

$Var(X) = E(X - E(X))^2$

$= E(X^2 + (E(X))^2 - 2*X*E(X))$

$= E(X^2) + E(X)^2 - 2E(X)^2$

$= E(X^2) - E(X)^2$

What is the variance of a Binomial random variable? In other words what is $Var(X) = Var(\sum_{i=1}^{n} X_i)$

To solve this you need to first know what is Var(X+Y)=?

$Var(X+Y) = E((X + Y)^2) - (E(X+Y) * E(X+Y))$

$= E ((X^2+Y^2+2XY)) - ((E(X) + E(Y))*(E(X) + E(Y))$

$= E(X^2) + E(Y^2) + E(2XY) - (E(X))^2 + 2E(X)E(Y) + (E(Y))^2$

$= E(X^2) - (E(X))^2 + E(Y^2) - (E(Y))^2 + 2(E(XY) - E(X)E(Y))$

$= Var(X) + Var(Y) - 2Cov(X,Y)$

Covariance between variables X and Y is defined as

$Cov(X,Y) = E((X - E(X)) * (Y - E(Y))$

$= E(XY - XE(Y) - YE(X) + E(X)E(Y))$

$= E(XY) - E(Y)E(X) - E(Y)E(X) + E(X)(Y)$

$= E(XY) - E(X)E(Y)$

Conditional probability:

P(A and B) = P(B) * P(A|B)

I can also write P(A and B) as P(B and A)

P(B and A) = P(A) * P(B|A)

But P(A and B) is the same as P(B and A) therefore

P(B) * P(A|B) = P(A) * P(B|A)

and rearranging the above gives us Bayes rule:

P(A|B) = P(B|A)P(A)/P(B)
