# Binomial Distribution

The binomial distribution is a important distribution in the probability theory and statistics, it gives the the number of successes *k* in a sequence of *n independent Yes/No* experiments with **p** as the probability of success of each yield

Lets understand the distribution  
If *p* is the probability that an event will happen in a single trial.  
*q =1-p* is the probability that it will fail to happen,

then the probability that the event will happen exactly *k* in *n* trial is given

p(k) = where, k = 0, 1, 2, ..., n,

is the binomial coefficient, hence the name binomial distributions.

Lets understand this distribution more using some simple examples

**Example 1** :  
Flip a fair coin 8 times and Probability of getting Tail X number of times.

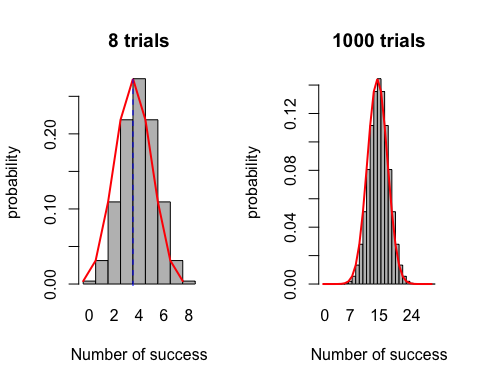
**solution:**  
Outcome from first flip can be a head or tail, so total out come 2 outcomes; Probability of a fair coin = 0.5. .

Using Binomial Distribution formula ,  
The probability of getting *0* tails in 8 flips == 1/256,

In similar ways,  
Probability of getting 0 tail in 8 flips = = 0.0039062  
Probability of getting 1 tail in 8 flips = = 0.03125  
Probability of getting 2 tail in 8 flips = = 0.109375  
Probability of getting 3 tail in 8 flips = = 0.21875  
Probability of getting 4 tail in 8 flips = = 0.2734375  
Probability of getting 5 tail in 8 flips = = 0.21875  
Probability of getting 6 tail in 8 flips = = 0.109375  
Probability of getting 7 tail in 8 flips = = 0.03125  
Probability of getting 8 tail in 8 flips = = 0.0039062

The Binomial distribution is very important, when the sample size is huge, like flipping a coin 1000 times or more. Lets plot the curves and compare when the number of trial increase from 8 trials to 1000 trials. You can use the r syntax below

n=8 # sample size i.e. 8 flips  
k <- seq(0,8, by = 1) # Number of success 0, 1.....8  
y <- c(dbinom(k, size = n ,prob = 0.5)) #dbinom gives the binomial dist.  
par(mfrow=c(1,2))   
barplot( y,k, width= 1,space =0,  
 main ="8 trials" , xlab = "Number of success" ,  
 ylab="probability")  
lines(k, (dbinom(k, n,0.5)), col = "red", lwd = 2)  
abline(v=4, lty=2, col ="blue")  
  
# 30 flips  
n= 30  
k <- seq(0,30, by = 1)   
y <- c(dbinom(k, size = n ,prob = 0.5))  
barplot( y,k, width= 1,space =0,   
 main ="1000 trials", xlab = "Number of success" ,  
 ylab="probability")  
lines(k, (dbinom(k, n,0.5)), col = "red", lwd = 2)



**Example 2**  
If each gender has an independent 50% probability for each birth.What's the probability of getting 3 or more girl in 7 trials ?

probability of getting a girl *p* = 0.5  
probability of not getting a girl *q* = 1- 0.5= 0.5

*p(3)* = = 0.2734375  
*p(4)* = = 0.2734375  
*p(5)*= = 0.1640625  
*p(6)*= = 0.0546875  
*p(7)* = = 0.0078125

**The probability of 3 or more girl** (sum up above probabilities) = 0.7734375

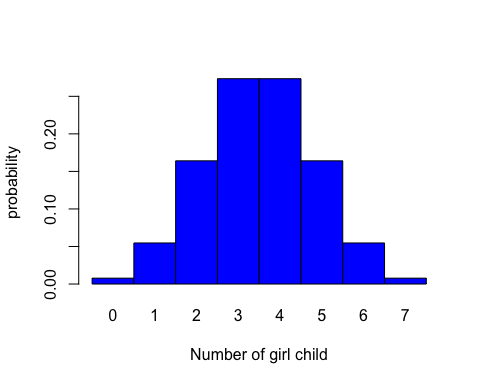
We can use dbinom or pdinom to get the probability of 3 or more girls: dbinom(3, size =7,prob =0.5) +  
dbinom(4, size =7,prob =0.5) +  
dbinom(5, size =7,prob =0.5) +  
dbinom(6, size =7,prob =0.5) +  
dbinom(7, size =7,prob =0.5) = 0.7734375  
OR  
 pbinom(2, size = 7, prob = 0.5, lower.tail = FALSE) =0.7734375.

Similarly,

* The probability of getting 3 or less girls dbinom(3, size =7,prob =0.5) + dbinom(2, size =7,prob =0.5) + dbinom(1, size =7,prob =0.5) + dbinom(0, size =7,prob =0.5) = 0.5.
* using **pbinom**, pbinom(3,size = 7, prob = 0.5, lower.tail =TRUE) = 0.5
* The probability of getting 1 girl or less = **d**binom(0, size =7,prob =0.5) + dbinom(1, size =7,prob =0.5)\* = 0.0625
* The probability of getting 1 or less girls =**p**binom(1 ,size = 7, prob = 0.5, lower.tail = TRUE)\* = 0.0625
* The probability of getting 1 or more girls= *pbinom(0,size = 7, prob = 0.5, lower.tail =FALSE)* = 0.9921875

Lets see graphically

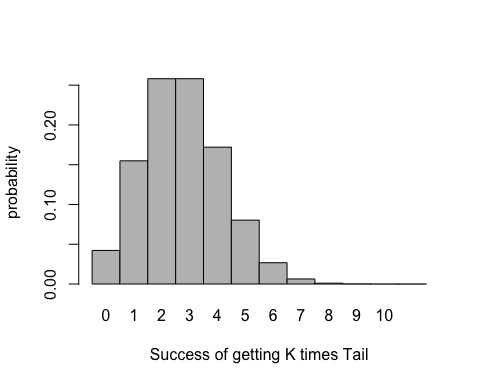
n=7  
k <- seq(0,7, by = 1) # Number of success   
y <- c(dbinom(k, size = n ,prob = 0.5))  
barplot( y,k, width= 1,  
 space =0,col = "blue",  
 xlab = "Number of girl child" , ylab="probability")



**Example 3** Suppose there are 10 multiple choice questions in an Math quiz. Each question has 4 possible answers, and only one of them is correct. Find the probability of having 5 or less correct answers if a student attempts to answer every question at random.

**Solution**  
There are 4 option, but only one correct answer , so probability = 1/4 = 0.25  
Total size = 10

n=11  
k <- seq(0,11, by = 1) # Number of success 0, 1.....8  
y <- c(dbinom(k, size = n ,prob = 0.25))  
barplot( y,k, width= 1,space =0,  
 xlab = "Success of getting K times Tail" , ylab="probability")



The probability of 0 correct answer = dbinom(0,size =10, prob = 1/4) = 0.0563135

The probability that , only 5 or less correct answer = pbinom(5, size = 10, prob =0.25,lower.tail = TRUE) = 0.9802723

The probability, Six or more correct answers =pbinom(5, size = 10, prob =0.25.lower.tail= FALSE) = 0.0197277