

# binomial distribution experiment

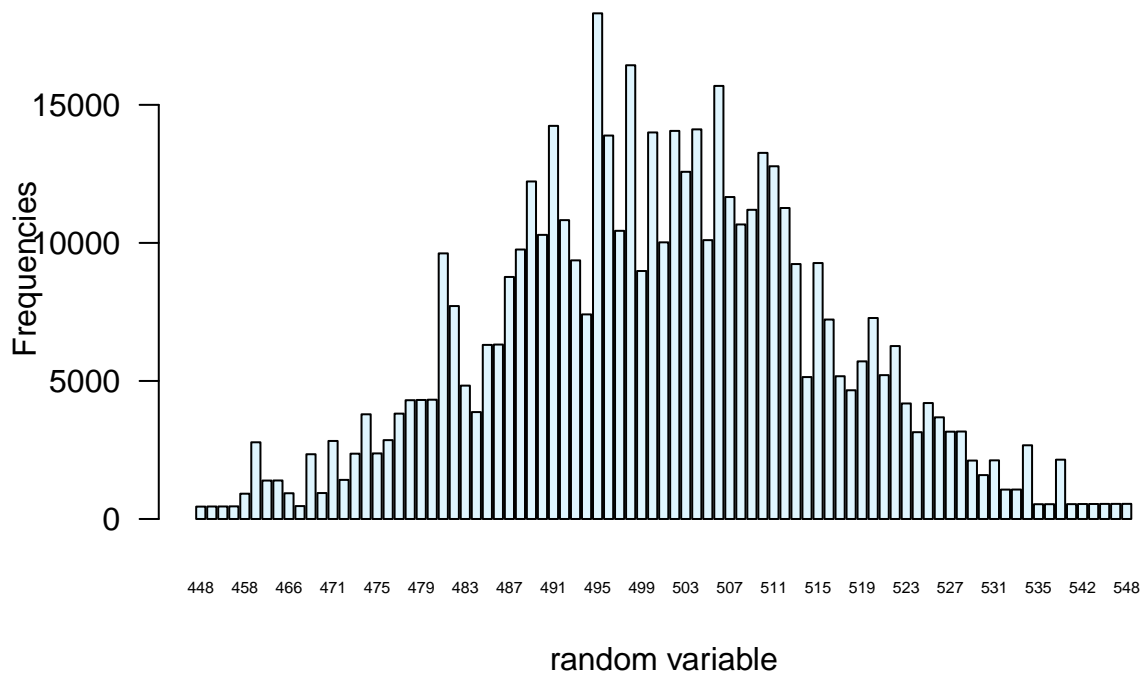
## How to know a population is binomial distribution

We can know by sampling and drawing the distribution. Here, I am going to sample from the binomial population, and then draw the distribution, if the distribution shape is bell curve, it is usually a binomial distribution

```
sampl <- rbinom(1000, 1000, 0.5)
sampl.df <- data.frame(sampl)

sampl.agg.df <- aggregate(sampl.df$sampl, by = list(rv = sampl.df$sampl), FUN = sum)

barplot(sampl.agg.df$x~sampl.agg.df$rv, col=c("#ccf0fe9f"), horiz=FALSE, cex.names=0.5, las=1, width=1,
        xlab="random variable", ylab="Frequencies")
```

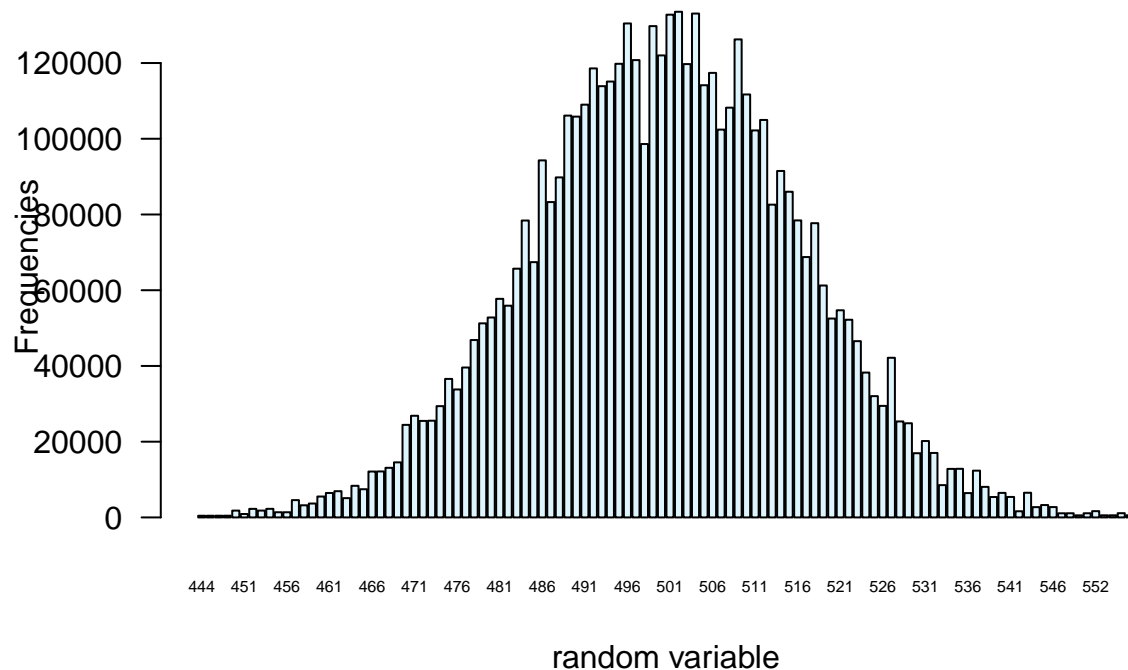


What if I increase the sample? If the sample size is increased, the bell curve becomes more and more what?

```
sampl <- rbinom(10000, 1000, 0.5)
sampl.df <- data.frame(sampl)

sampl.agg.df <- aggregate(sampl.df$sampl, by = list(rv = sampl.df$sampl), FUN = sum)

barplot(sampl.agg.df$x~sampl.agg.df$rv, col=c("#ccf0fe9f"), horiz=FALSE, cex.names=0.5, las=1, width=1,
        xlab="random variable", ylab="Frequencies")
```



## What happens when $n$ is going to be infinitive in binominal distribution

When  $n$  is going to infinitive, the distribution converges to normal distribution.

I set the  $n$  to 5000, and smaple size is 50000, the distribution is almost the normal distribution.

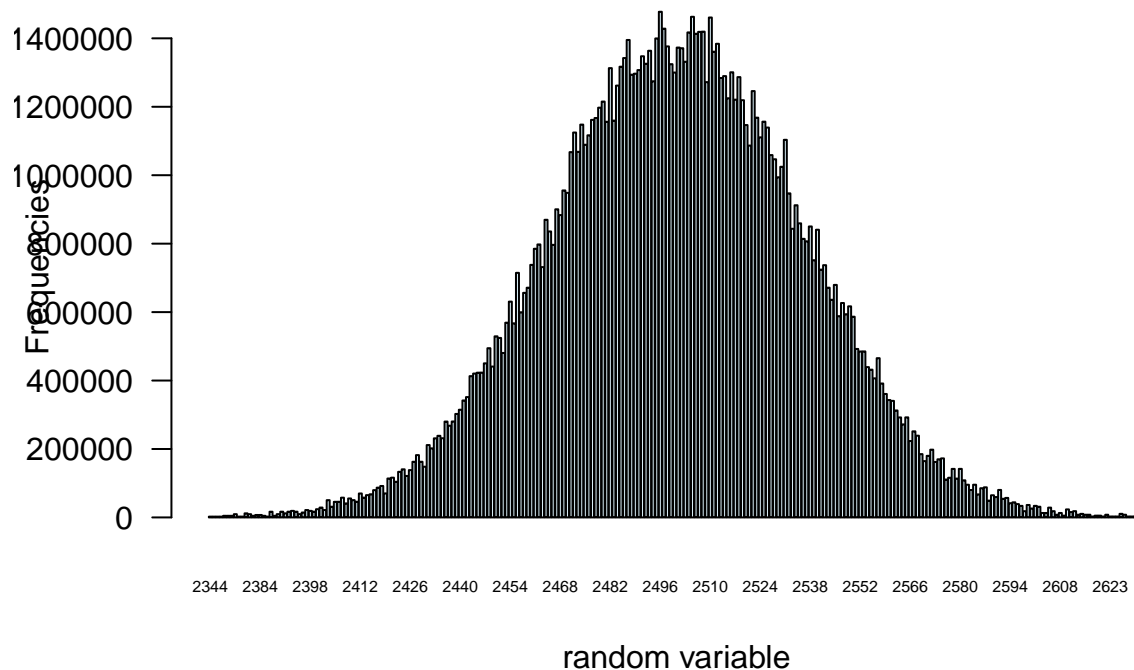
So when  $n$  is going to infinitive, we can guess that its distribution converges to normal distribution, what a amazing coincidence!

Actually, this guess is true.

```
sampl <- rbinom(50000, 5000, 0.5)
sampl.df <- data.frame(sampl)

sampl.agg.df <- aggregate(sampl.df$sampl, by = list(rv = sampl.df$sampl) , FUN = sum)

barplot(sampl.agg.df$x~sampl.agg.df$rv, col=c("#ccf0fe9f"), horiz=FALSE, cex.names=0.5, las=1, width=1,
        xlab="random variable", ylab="Frequencies")
```



## How to get the propability of binominal distribution

When we know a population is binominal distribution, how we get the probability.

Actually, we can get the probability from the sample. When the sample size is very big, the estimated probability is getting closed to the population probability.

It is also very simple to compute the probability by counting the yes of outcome from the sample, which divides the sample size.

According to the Law of Large Number, with the sample size increasing, the probability is getting close to the population probability.

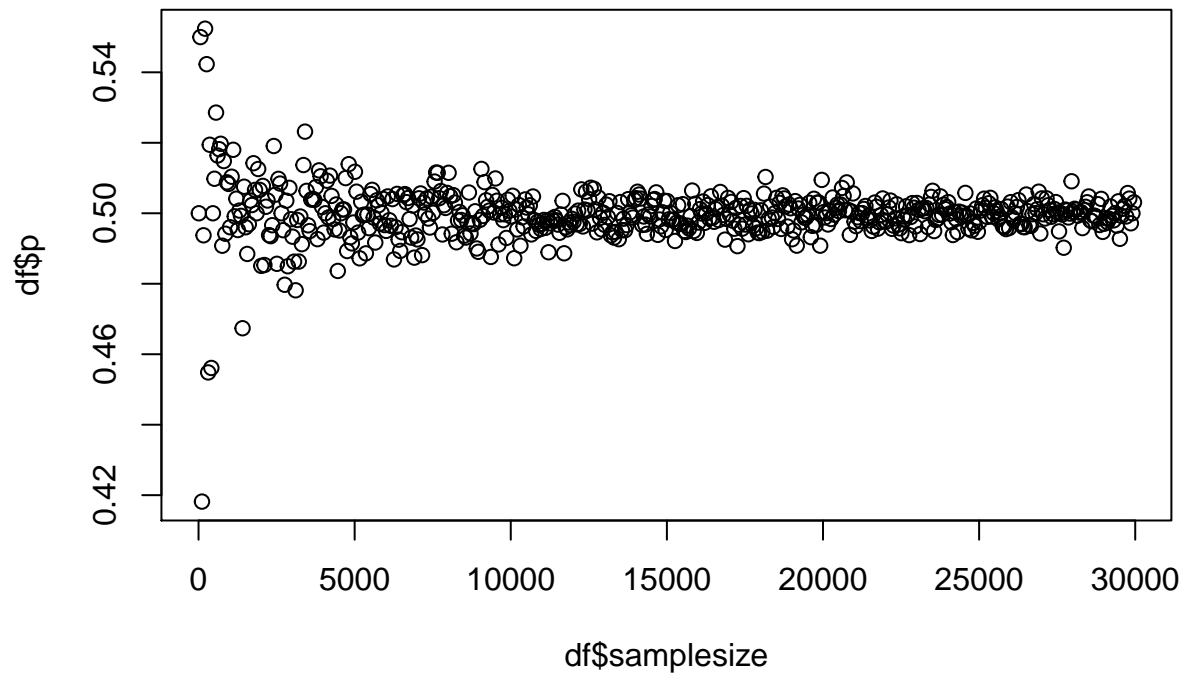
Let's experiment here.

```

samplesize = 10
df <- data.frame(samplesize= numeric(0), p= numeric(0))
for (i in seq(10, 30000, by = 50) ){
  sampl <- rbinom(i, 1, 0.5)
  sampl.df <- data.frame(sampl)
  df[i, ] <- c(i, mean(sampl.df$sampl))
}

plot(df$samplesize, df$p)

```



From the graph above, we know that if when the sample size increase , it converges to 0.5, which the probability of population mean