Outline

Exercise: Solution

- **Q1: Calculate the probability of 13 heads in 30 independent bernoulli trials**
  - I directly do it in R:

```
x=13
dbinom(13, size = 30, prob = 0.5)

## [1] 0.1115
```

- **Q2: Calculate the binomial distribution of the above experiment**
  - To calculate binomial distribution, we need to calculate how the probability is distributed for each possible outcome. Since, we have 30 trials, our probability distribution ranges from getting zero heads to 30 heads

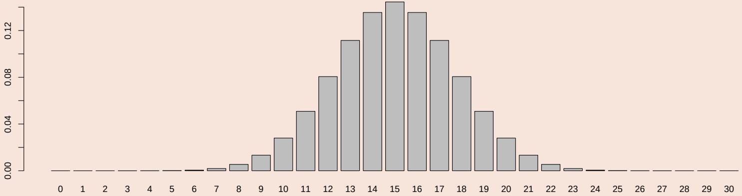
```
x= 0:30
y= dbinom(x, size = 30, prob = 0.5)
y

##  [1] 9.313e-10 2.794e-08 4.051e-07 3.781e-06 2.552e-05 1.327e-04
##  [7] 5.530e-04 1.896e-03 5.451e-03 1.332e-02 2.798e-02 5.088e-02
## [13] 8.055e-02 1.115e-01 1.354e-01 1.445e-01 1.354e-01 1.115e-01
## [19] 8.055e-02 5.088e-02 2.798e-02 1.332e-02 5.451e-03 1.896e-03
## [25] 5.530e-04 1.327e-04 2.552e-05 3.781e-06 4.051e-07 2.794e-08
## [31] 9.313e-10
```

Exercise: Solution

- Since, I have a vector of 30 probabilities, it is more meaningful to plot the probability dist.

```
barplot(y, names.arg = c(0:30))
```



- You can see the probability of getting 30 heads in 30 trials is extremely low
- You can also see the probability of getting 0 heads in 30 trials is extremely low (in other words getting 0 heads and 30 heads out of 30 trials are extremely rare but not statistically impossible)
- You can see that getting 15 heads in 30 trials has the highest probability
- The mid point of this dist. shows the mean value. Even without calculating the mean of this experiment, I have a good sense that the mean should be somewhere near 15 (we will check for it in the last question)

Exercise: Solution

- **Q3: Calculate the cummulative probability dist.**
  - Cummulative probability would simply be the sum of individual probabilities. It can be calculated in R in two ways:
    - **Method 1:** calculate the cummulative sum of the inidividual probabilities

```
x= 0:30
y= dbinom(x, size = 30, prob = 0.5)
z= cumsum(y)
z

##  [1] 9.313e-10 2.887e-08 4.340e-07 4.215e-06 2.974e-05 1.625e-04
##  [7] 7.155e-04 2.611e-03 8.062e-03 2.139e-02 4.937e-02 1.002e-01
## [13] 1.808e-01 2.923e-01 4.278e-01 5.722e-01 7.077e-01 8.192e-01
## [19] 8.998e-01 9.506e-01 9.786e-01 9.919e-01 9.974e-01 9.993e-01
## [25] 9.998e-01 1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
## [31] 1.000e+00
```

    - **Method 2:** pbinom function in R automatically computes cummulative probabilites

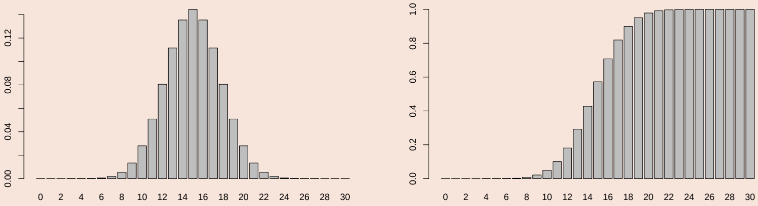
```
x=0:30
z=pbinom(x, size = 30, prob=0.5)
z

##  [1] 9.313e-10 2.887e-08 4.340e-07 4.215e-06 2.974e-05 1.625e-04
##  [7] 7.155e-04 2.611e-03 8.062e-03 2.139e-02 4.937e-02 1.002e-01
## [13] 1.808e-01 2.923e-01 4.278e-01 5.722e-01 7.077e-01 8.192e-01
## [19] 8.998e-01 9.506e-01 9.786e-01 9.919e-01 9.974e-01 9.993e-01
## [25] 9.998e-01 1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
## [31] 1.000e+00
```

Exercise: Solution

- Note the probabilities of cummulaties probability dist. sum to one as we reach 30
- It is more useful to plot cummulative dist.
- For fun, I plot probability dist. with cummulative dist. side by side:

```
par(mfrow=c(1,2))
barplot(y, names.arg = c(0:30))
barplot(z, names.arg = c(0:30))
```



Exercise: Solution

- **Q4: Calculate the mean and variance of heads in 30 bernoulli trials**
  - Mean for a binomial dist. is directly calcualted as:

$$E[X] = nP = 30(0.5) = 15$$

  - If you do not like the direct way of calculating mean, it can also be calculated as follows (as we did for other discrete variables earlier):

$$E[X] = \sum xP(x)$$

```
x=0:30
px= dbinom(x, size = 30, prob = 0.5)
sum(x*px)

## [1] 15
```

Note: Now look at the prob. dist figure again!

  - Variance is calculated as:

$$Var[X] = nP(1-P) = 30(0.5)[1-0.5] = 7.5$$

  - If you do not like the direct way, you can calculate it as follows (as we did for other discrete variables):

$$Var[X] = \sum (x-\mu)^2 P(x)$$

```
mean_x <- sum(x*px)
sum((x-mean_x)^2*px)

## [1] 7.5
```