Probability Distribution

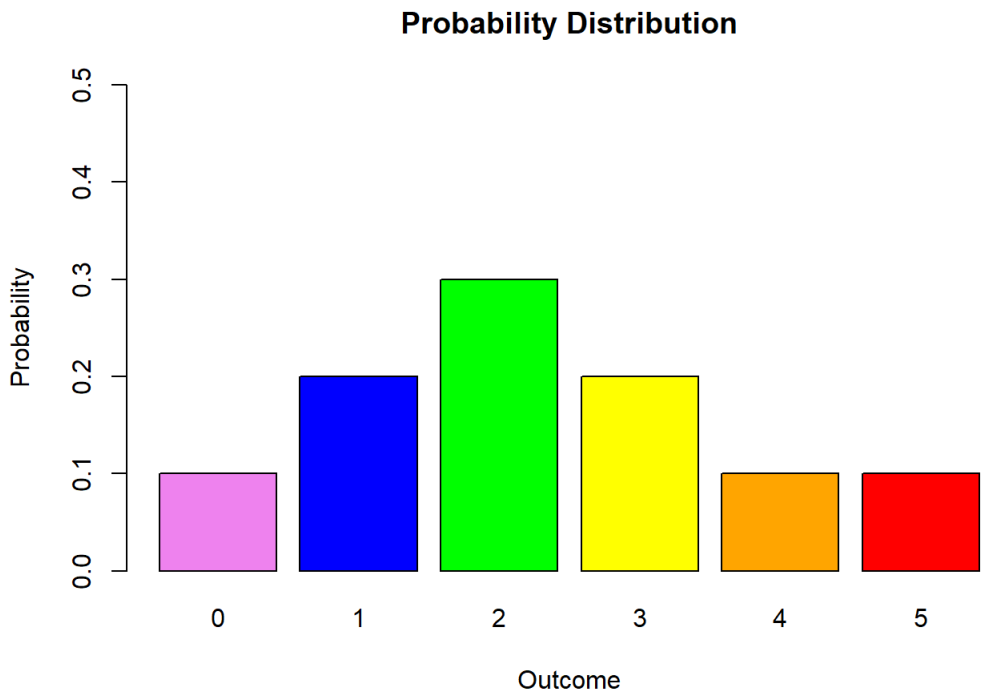
Chirantan Ganguly

28/04/2020

Probability and Mass Density function

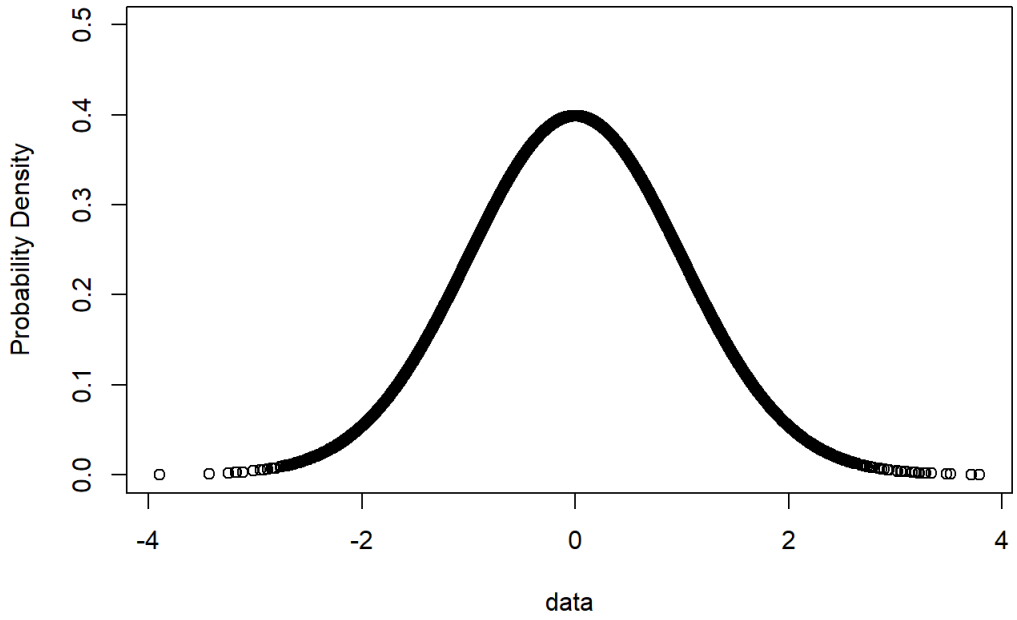
For Discrete random Variables

```
probability_distribution <- data.frame(outcome = 0:5, probs = c(0.1, 0.2, 0.3, 0.2, 0.1, 0.1))
barplot(height=probability_distribution$probs, names.arg=probability_distribution$outcome,
        col=c("violet","blue","green","yellow","orange","red"),
        main="Probability Distribution", ylab="Probability", xlab="Outcome",
        ylim=c(0,0.50), width=0.00001)
```



For Continuous random Variables

```
set.seed(11225)
data <- rnorm(10000)
density<-dnorm(data)
plot(data,density,ylim=c(0,0.5), ylab="Probability Density")
```



The cumulative probability distribution

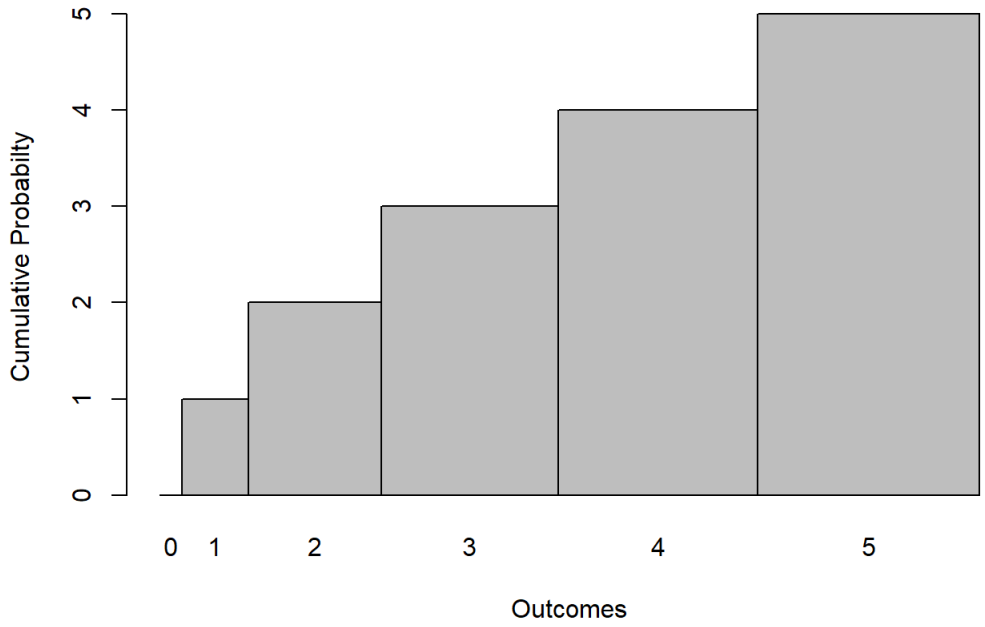
A cumulative probability refers to the probability that the value of a random variable falls within a specified range. Frequently, cumulative probabilities refer to the probability that a random variable is less than or equal to a specified value.

Lets find Cumulative Probability and stor it in the dataset as a new variable

```
probability_distribution$cum_probs<-cumsum(probability_distribution$probs)
probability_distribution
```

```
##   outcome probs cum_probs
## 1       0   0.1       0.1
## 2       1   0.2       0.3
## 3       2   0.3       0.6
## 4       3   0.2       0.8
## 5       4   0.1       0.9
## 6       5   0.1       1.0
```

```
barplot(probability_distribution$outcome,probability_distribution$cum_probs,space=0,names.arg = probability_
distribution$outcome,ylab="Cumulative Probabilty",xlab="Outcomes")
```



Expected Score, Variance and Standard Deviation of a Probability distribution

```
expected_score <- sum(probability_distribution$outcome * probability_distribution$probs)
expected_score
```

```
## [1] 2.3
```

```
variance<-sum((probability_distribution$outcome-expected_score)^2*probability_distribution$probs)
variance
```

```
## [1] 2.01
```

```
std<-variance^0.5
std
```

```
## [1] 1.417745
```

The normal distribution and cumulative probability

Hair length is considered to be normally distributed with a mean of 25 centimeters and a standard deviation of 5. Imagine we wanted to know the probability that a woman's hair length is less than 30. We can do this in R using the pnorm() function. This function calculates the cumultative probability. We can use it the following way: pnorm(30, mean = 25, sd = 5). If you wanted to calculate the probability of a woman having a hair length larger or equal to 30 centimers, you can set the lower.tail argument to FALSE. For instance, pnorm(30, mean = 25, sd = 5, lower.tail = FALSE)

```
pnorm(30, mean = 25, sd = 5)
```

```
## [1] 0.8413447
```

```
pnorm(30, mean = 25, sd = 5, lower.tail = FALSE)
```

```
## [1] 0.1586553
```

The normal distribution and quantiles

The normal distribution and quantiles

Sometimes we have a probability that we want to associate with a value.Say we want the value of a woman's hair length that corresponds with the 0.2 quantile (=20th percentile).If our variable is normally distributed, in R we can use the function qnorm() to do so. We can specify the probability as the first parameter, then specify the mean and then specify the standard deviation.

```
qnorm(.85,mean=25,sd=5)
```

```
## [1] 30.18217
```

Calculating probabilities of binomial distributions in R

Let's consider the example: we had an exam with 25 questions and 0.2 probability of guessing a question correctly. Calculate the probability of exactly answering say 5 questions correctly. When we want to calculate the probability of answering 5 questions correctly, we can use the dbinom function. This function calculates an exact probability. If we would like to calculate an interval of probabilities, say the probability of answer 5 or more questions correctly, we can use the pbinom function.

```
five_correct <- dbinom(5, size = 25, prob = 0.2)
five_correct
```

```
## [1] 0.1960151
```

```
atleast_five_correct <- pbinom(4, size = 25, prob = 0.2, lower.tail = FALSE)
atleast_five_correct
```

```
## [1] 0.5793257
```

Quantiles and the binomial distribution

Quantiles are used when you have a probability and you want to associate this probability with a value. In our last example we had 25 questions and the probability of guessing a question correctly was 0.2. When we want to calculate the value that is associated with for instance the 0.2 quantile (=20th percentile). In case we are working with a binomial distribution, we can use the function qbinom() for this.

```
qbinom(0.6, size=25, prob=0.2)
```

```
## [1] 5
```