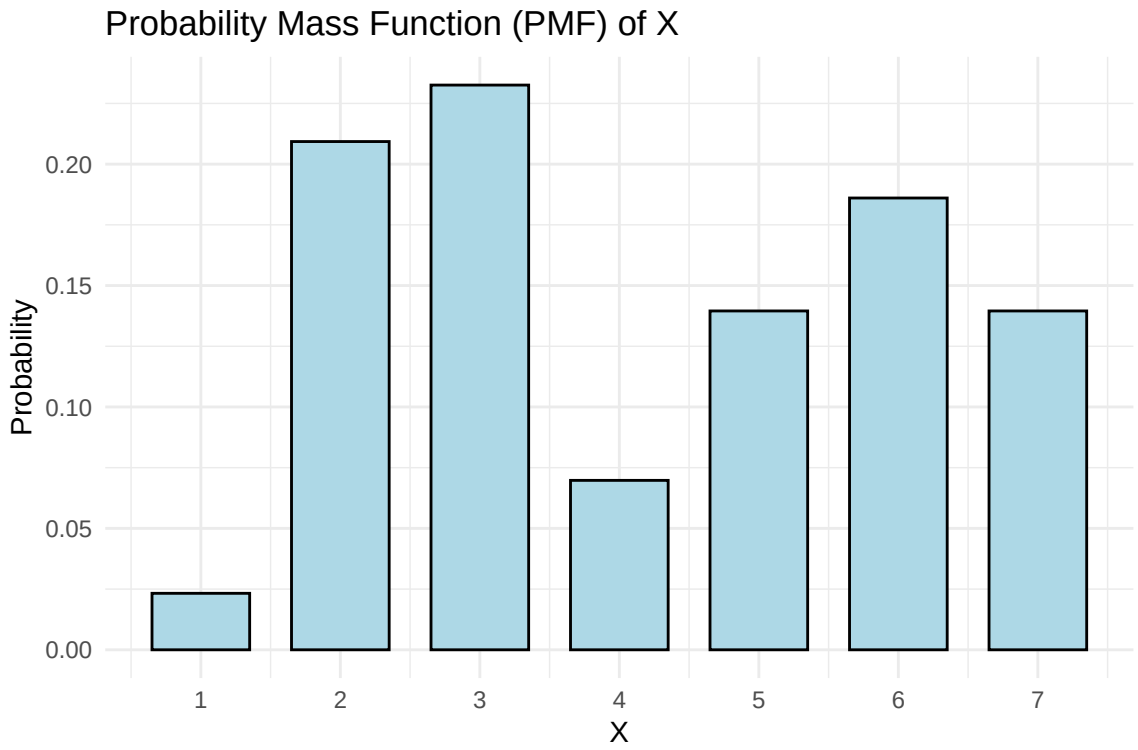
STATS 7022 - Data Science PG Assignment 1

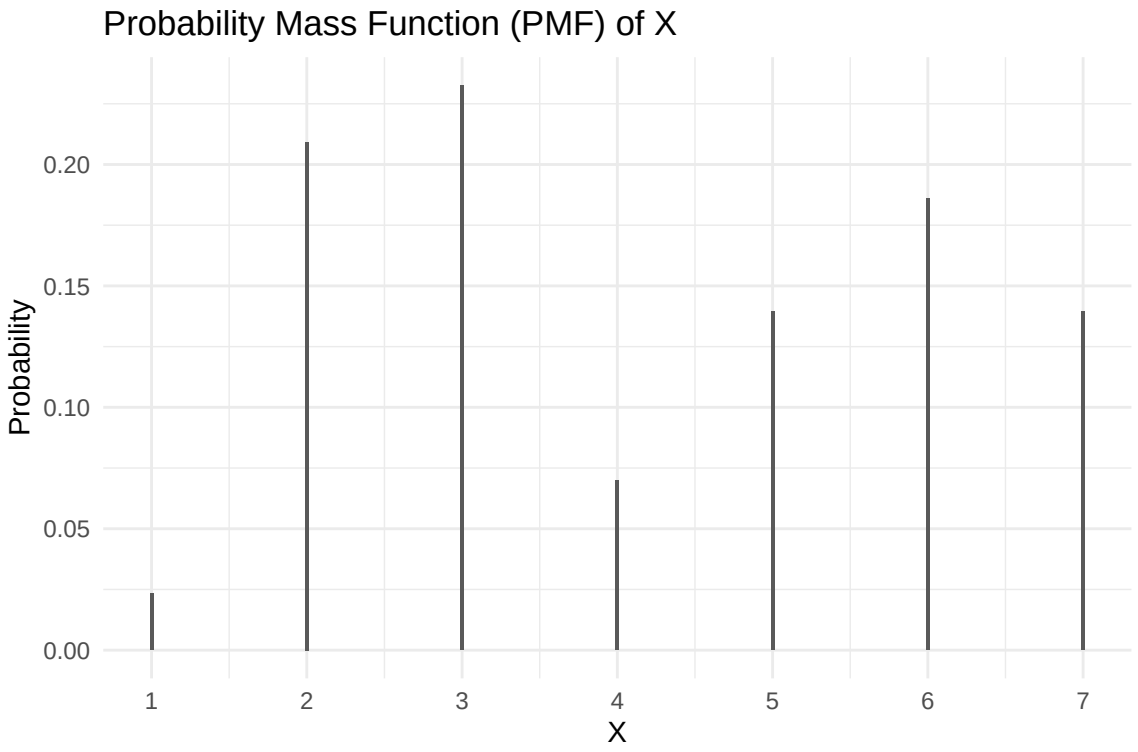
Dang Thinh Nguyen

2024-06-24

```
q2 <- tibble(
  X = c(1:7),
  Pr = c(1/43, 9/43, 10/43, 3/43, 6/43, 8/43, 6/43)
)

q2 %>%
  ggplot(aes(x = X, y = Pr)) +
  geom_bar(stat="identity", width=0.7, color='black', fill='lightblue') +
  labs(x = 'X',
       y = 'Probability',
       title="Probability Mass Function (PMF) of X",) +
  scale_x_continuous(breaks = 1:7) +
  theme(text = element_text(size = 40)) +
  theme_minimal()
```

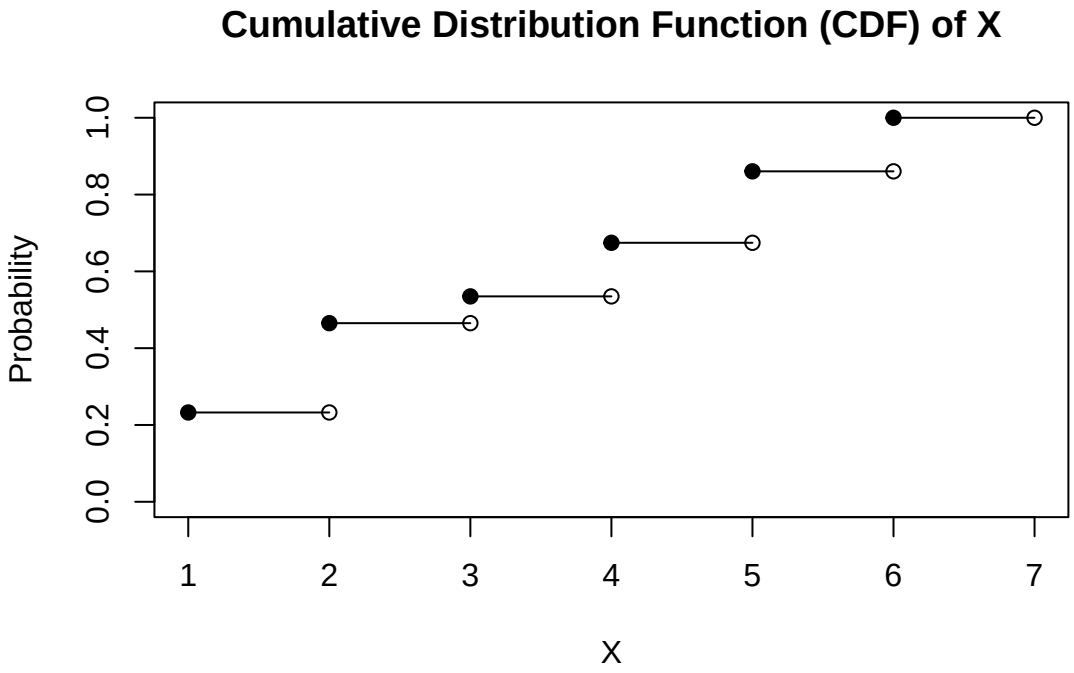


```r
q2 <- tibble(
  X = c(1:7),
  Pr = c(1/43, 9/43, 10/43, 3/43, 6/43, 8/43, 6/43)
)

q2 %>%
  ggplot(aes(x = X)) +
  geom_bar(aes(y = Pr), stat = 'identity', alpha = 1, width = 0.02) +
  labs(x = 'X',
       y = 'Probability',
       title="Probability Mass Function (PMF) of X",) +
  scale_x_continuous(breaks = 1:7) +
  theme(text = element_text(size = 40)) +
  theme_minimal()
```

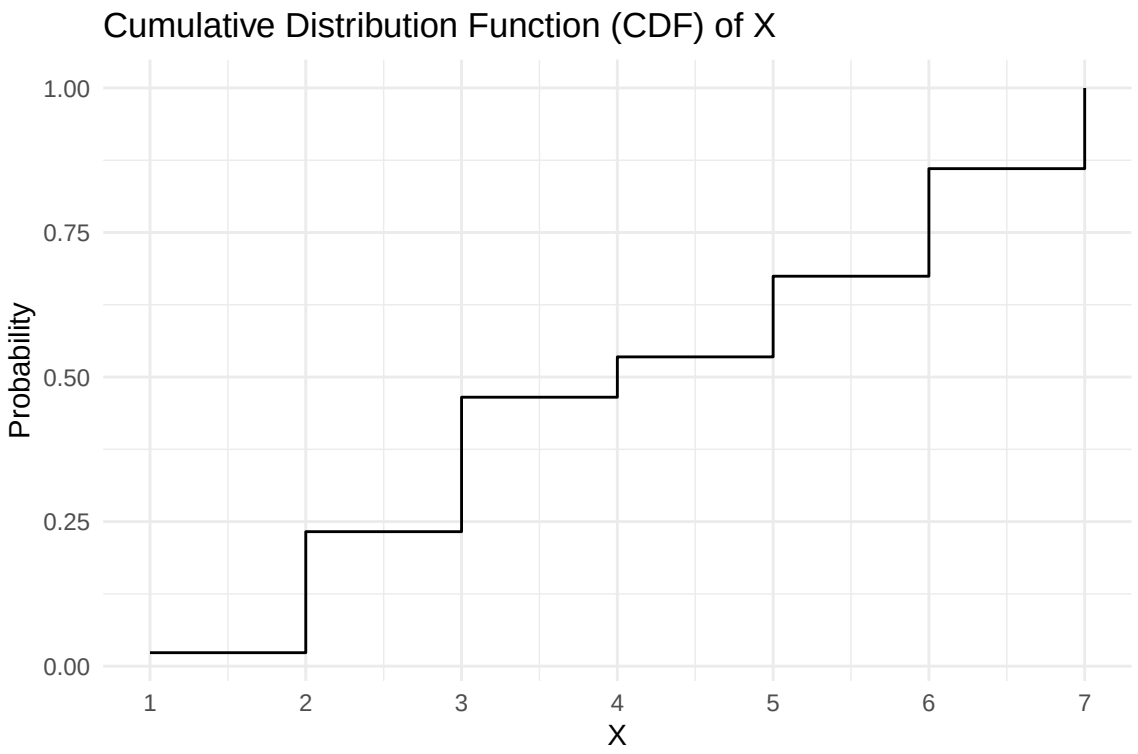



```r
x <- 1:7
fx <- c(1/43, 9/43, 10/43, 3/43, 6/43, 8/43, 6/43)

Fx <- cumsum(fx)
n <- length(x)

plot(x = NA, y = NA, pch = NA,
     xlim = c(1, max(x)),
     ylim = c(0, 1),
     xlab = "X",
     ylab = "Probability",
     main = "Cumulative Distribution Function (CDF) of X")
points(x = x[-n], y = Fx[-1], pch=19)
```

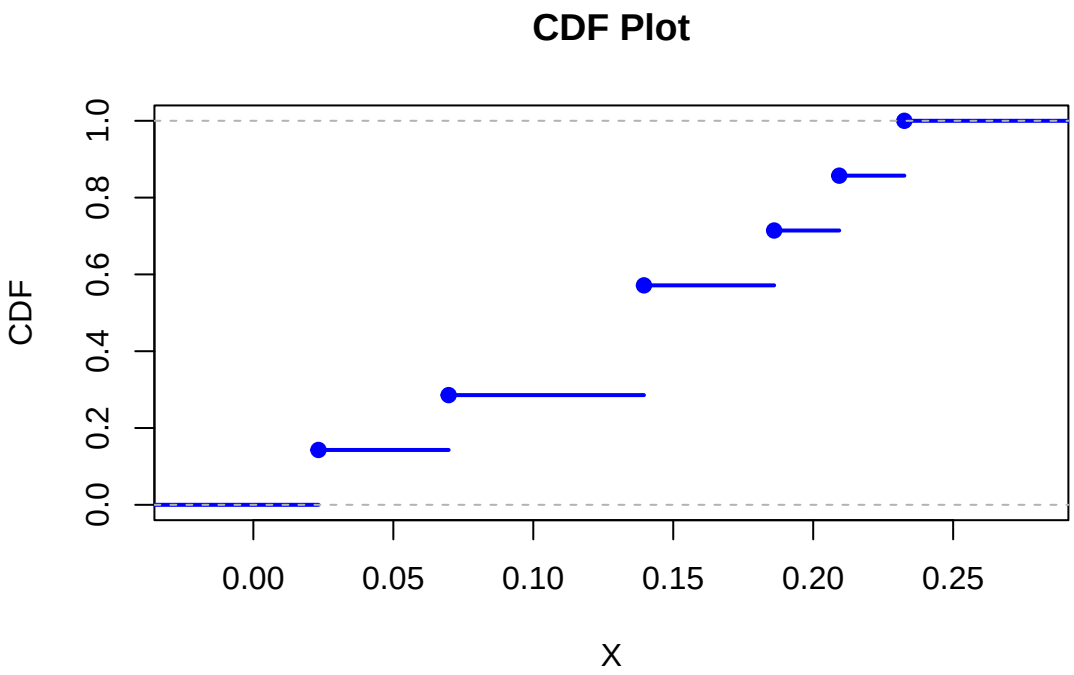
```r
points(x = x[-1], y = Fx[-1], pch=1)
for(i in 1:(n-1)) points(x=x[i+0:1], y=Fx[c(i,i)+1], type="l")
```




```r
q2 %>%
  ggplot(aes(x = X, y = cumsum(Pr))) +
  geom_step() +
  labs(x = 'X',
       y = 'Probability',
       title="Cumulative Distribution Function (CDF) of X",) +
  scale_x_continuous(breaks = 0:7) +
  theme(text = element_text(size = 40)) +
  theme_minimal()
```




```r
q2 <- tibble(
  X = c(1:7),
  Pr = c(1/43, 9/43, 10/43, 3/43, 6/43, 8/43, 6/43)
)

# calculate CDF
CDF <- ecdf(q2$Pr)

# draw the cdf plot
plot(CDF, main = "CDF Plot", xlab = "X", ylab = "CDF", col = "blue", lwd = 2)
```


CDF Plot
CDF
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20
0.25
X