Mark Stanley: Dice Simulations, Histograms, Probability Density Functions

2024-03-09

Simulation of rolling a fair twelve-sided die 500 times in R:

```
diceSample <- sample(1:12, size=500, replace=TRUE)
diceSample
```

```
##   [1]  9 12  3  4  5  6  7  7  4  1  7  7 10 11  3  5  3  2  5 10  5  4  1 11  7
##  [26]  9  1 10  8  8 12  1 10  4  6  3  2  3  5  2  3  6  8 11  8  2 10  3  6  6
##  [51]  6  2  5 12 11  1  5  6  4 11  1  2  3  7  4 10 10 10  4  8 12 10  2  2  3
##  [76]  2 12  9 12  6  4  7  4 11  6  4 10  8  9  4  1 12  1  1 11  5  5  3  2  9
## [101] 10 12 12  4  7  1  3  5  7 11  7  3  9 11  3  8  5  9 11  9  2  9  3  1  3
## [126]  2  8  9  4  5  6  3  6 10 10  2  4 10  4  3 10  1  1 10  5  9  1  7  7  9
## [151]  2  2  8  7  5  2  6  1  2  4  7  7  1 12  9  7  6 11 11 10  4  6  3 12  7
## [176]  4  5  5  5  2  9  1  2  7  2  9  9  9  1  6  9 11  9  6  3  7  8  3  6  2
## [201]  8  9 10 12 11  7  9  8  7 10 11  3 12  8  1  7  9  1 11 10 10  3 12  1  9
## [226] 12 12 11 11  4  6  4  4  6  9  3 12  5 10  6  4  2  6 10  7  2  7  1  4  5
## [251]  7  8  2  6 11  2  7  6 12 12 11  3  9  5  5  6  8  9  2  6 12  9  8  3  1
## [276] 12  5  8  8  8  1  7  2  7  9  2  5  9 12  6 12  4  9  9 10  7 10  5  1  8
## [301]  1  8 12  3  5  4  6  9  3  7  8  4  7  9  9 10  6  9  2 11  4  5  5  9  4
## [326]  3 10  3 12  2  8  3  9  7  9  5  1  8  7  4  8  5  4  3  5  2  2  7 11  1
## [351]  9  5  4  5 10  6  6  9  8  2  1  6  4  9  6  7  1  4 10  9 11 12  3  2  4
## [376] 10  6  3 11 11  1  8  5  8 10  7  4  4  1  9  7  5  7  9  7  3  5  9  8  8
## [401] 12  9  1  6 11  4  8  8 12  4  7 11  3  4  9  5  5  4  2 12  7  9  8 12  5
## [426]  1  8  9  2  5  9  2  2  9  1  7  9  1 12  4 12  6  1  7  6  3  8  2 11  6
## [451] 12  5  5  1  5  1  5  1  7  5  2  4  8 10  6  8  6  5  3 10  9  8  8  8 11
## [476]  9  2 11  7  2  3  4  4  1  3 12  1  8 12  1  5 10  4  3  6  5  6  7 10  3
```

Use the sample function to achieve this.

Empirical probability of rolling a prime number or a multiple of 3 based on the simulation results:

```
primeCount <- 0
divis3Count <- 0

divis3Count <- length(diceSample[(diceSample%%3 == 0)])

primeCount <- length(diceSample[(diceSample %in% c(2,3,5,7,11))])

duplicates <- length(diceSample[(diceSample == 3)])

(divis3Count+primeCount - duplicates) / 500
```

```
## [1] 0.672
```

This result is accurate, as $8/12 = \sim 66\%$ of sampling units will be either prime, or divisible by 3.

Here is a bar graph showing the frequencies of each outcome (1 through 12) obtained from the simulation:

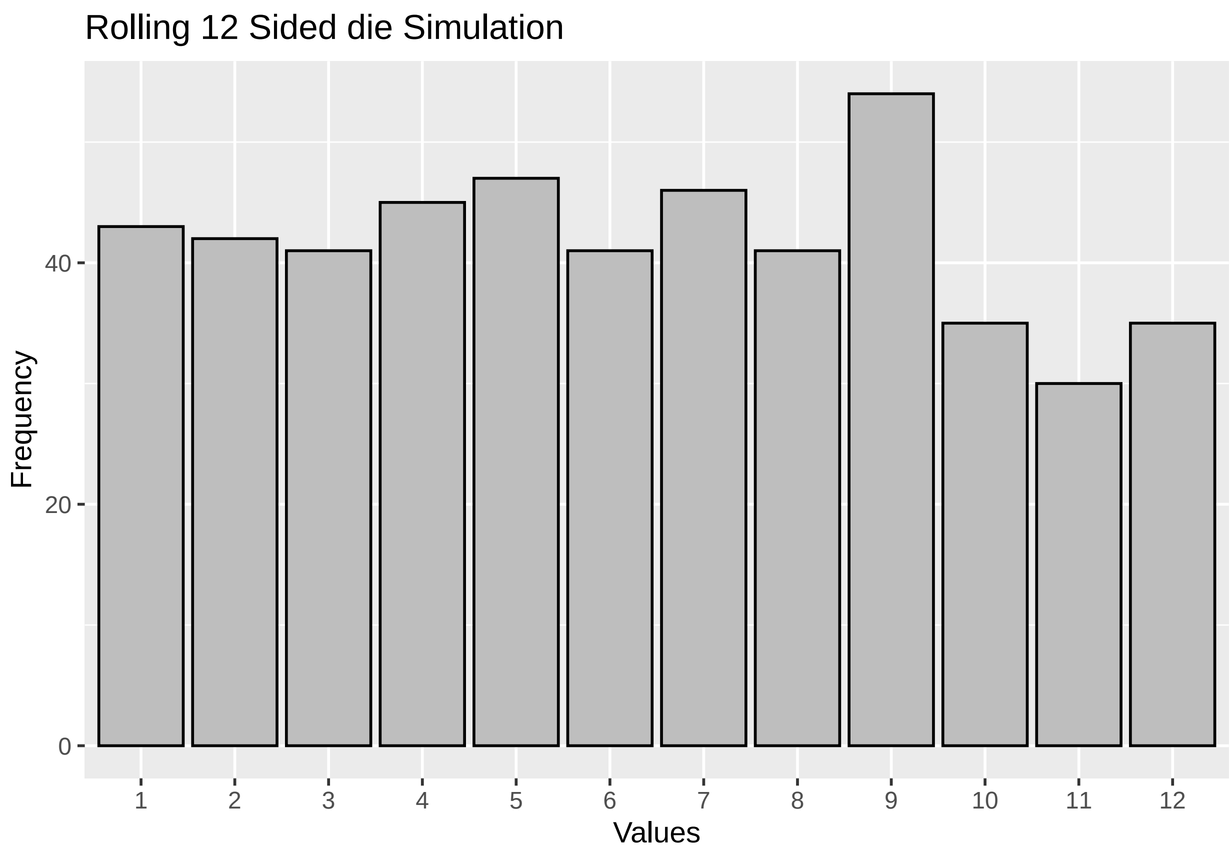
```r
library(ggplot2)

freq <- table(diceSample)

frame <- data.frame(freq)
names(frame) <- c("rolls", "frequency")

bar_chart <- ggplot(data = frame, aes(x = rolls, y = frequency)) +
    geom_bar(color = "black",fill = "gray", stat = "identity") +
    labs(title = "Rolling 12 Sided die Simulation",x = "Values",y = "Frequency")

print(bar_chart)
```



Here we use the ggplot2 library to make a bar graph.

Here is a random sample of size 100 from a negative binomial distribution with parameters r = 5 and p = 0.2:

```r
r <- 5
p <- 0.2
sample_size <- 100
```

```r
random_sample <- rnbinom(sample_size, size = r, prob = p)

random_sample
```

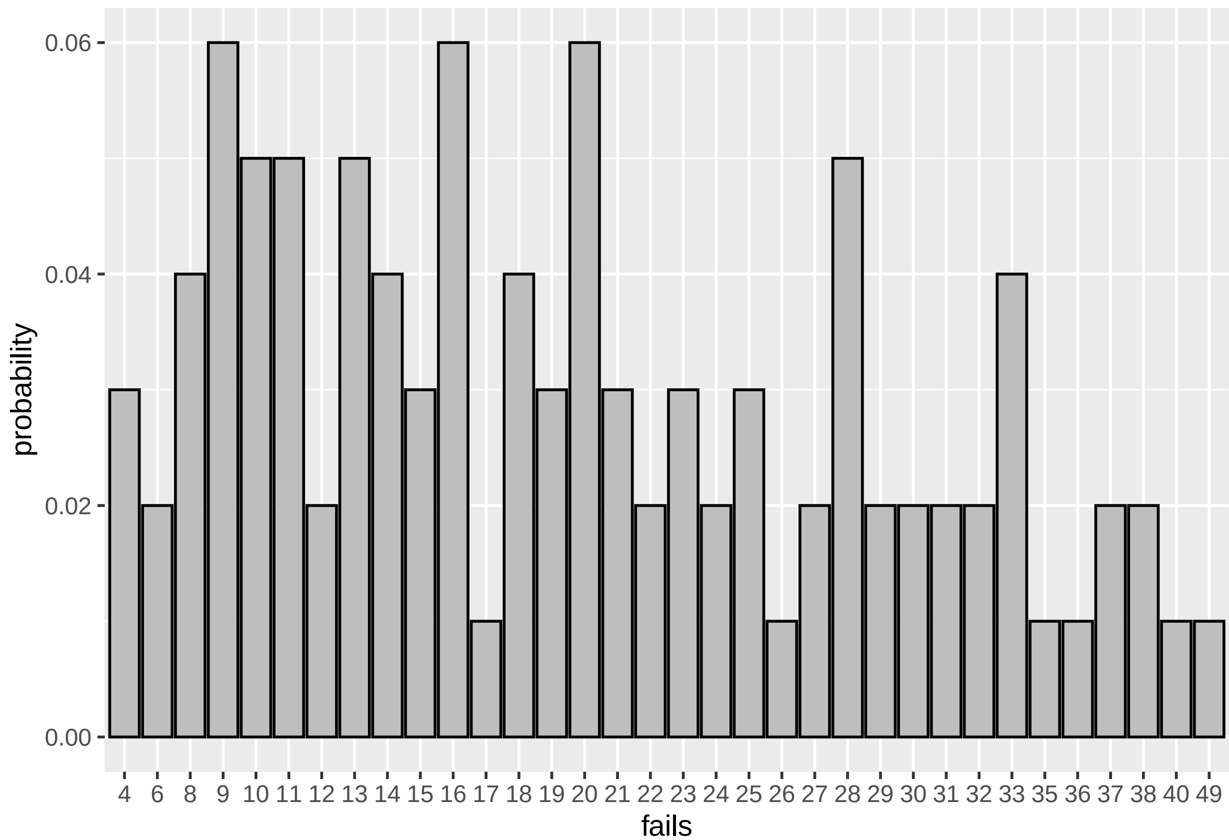
```
##   [1] 11 18 36 23  9 22 12 24 29 16 37  8 13 30 20  8 26  9 32 11 15 20 38 30  9
##  [26] 17 35 37 10 28 13 33  9 12 33 10 28 14 16 32 11 16 18 13 21 40 18  4 23 31
##  [51] 31 22 19 19 27 21 28 33 18 19 49 13 11 16 16 14 27 15 28 10 29  6 20 20 11
##  [76] 20 13 38  6 23 21 10 28 25 25 24  9 10 15  4  9 14  8 16 33  4  8 25 14 20
```

Here is a bar chart representing the PMF:

```r
PMF <- table(random_sample) /100

PMF_frame <- data.frame(PMF)

names(PMF_frame) <- c("fails", "probability")
ggplot(PMF_frame, aes(x = fails, y = probability)) +
  geom_bar(stat = "identity", color = "black", fill = "gray")
```



Here we use the ggplot2 library to make the bar chart.

Here is code that calculates the probability of getting at least 4 heads in 8 coin tosses, assuming p = 0.6:

```r
n <- 8
p <- 0.6  # prob of success
x <- 3

probability_at_least_4_heads <- 1 - pbinom(x, size = n, prob = p)
# prob will be 1 minus prob of 3 or less heads

print(probability_at_least_4_heads)
```

```
## [1] 0.8263296
```

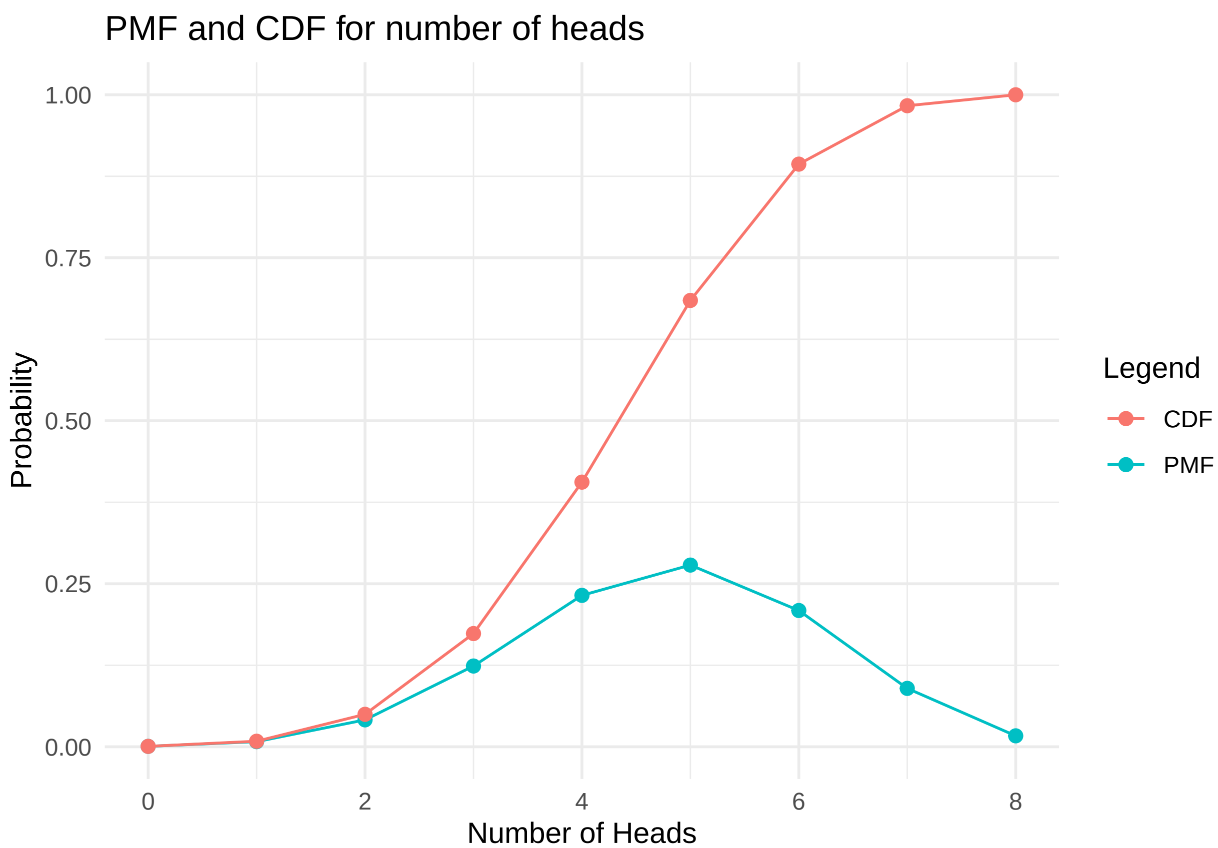
So there is an around 82 percent chance that from 8 trials, at least 4 will be heads

Here we plot the probability mass function (PMF) and cumulative distribution function (CDF) together:

```r
# Values for PMF and CDF

PMF <- dbinom(0:8, size = 8, prob = 0.6)
CDF <- pbinom(0:8, size = 8, prob = 0.6)


df <- data.frame(Heads = 0:8, PMF,
                 CDF)
# Plotting
ggplot(df, aes(x = Heads)) +
  geom_point(aes(y = PMF, color = "PMF"), size = 2) +
  geom_line(aes(y = PMF, color = "PMF")) +
  geom_point(aes(y = CDF, color = "CDF"), size = 2) +
  geom_line(aes(y = CDF, color = "CDF")) +
  labs(title = "PMF and CDF for number of heads",
       x = "Number of Heads", y = "Probability",
       color = "Legend") +
  theme_minimal()
```



Here we use R to compute the probability of observing exactly 3 events in a given time interval, assuming lambda = 2.5 events per hour (using a poisson distribution):

```
lambda <- 2.5  # average events per hour
k <- 3  # num of events

# Probability of exactly 3 events
probability_3_events <- dpois(k, lambda)


print(probability_3_events)
```

```
## [1] 0.213763
```

The cumulative probability of observing 3 or fewer events:

```
lambda <- 2.5  # average events per hour
k <- 3  # num of events

# Cumulative probability of observing 3 or fewer events using ppois
cumulative_probability_3_or_fewer <- ppois(k, lambda)

print(cumulative_probability_3_or_fewer)
```

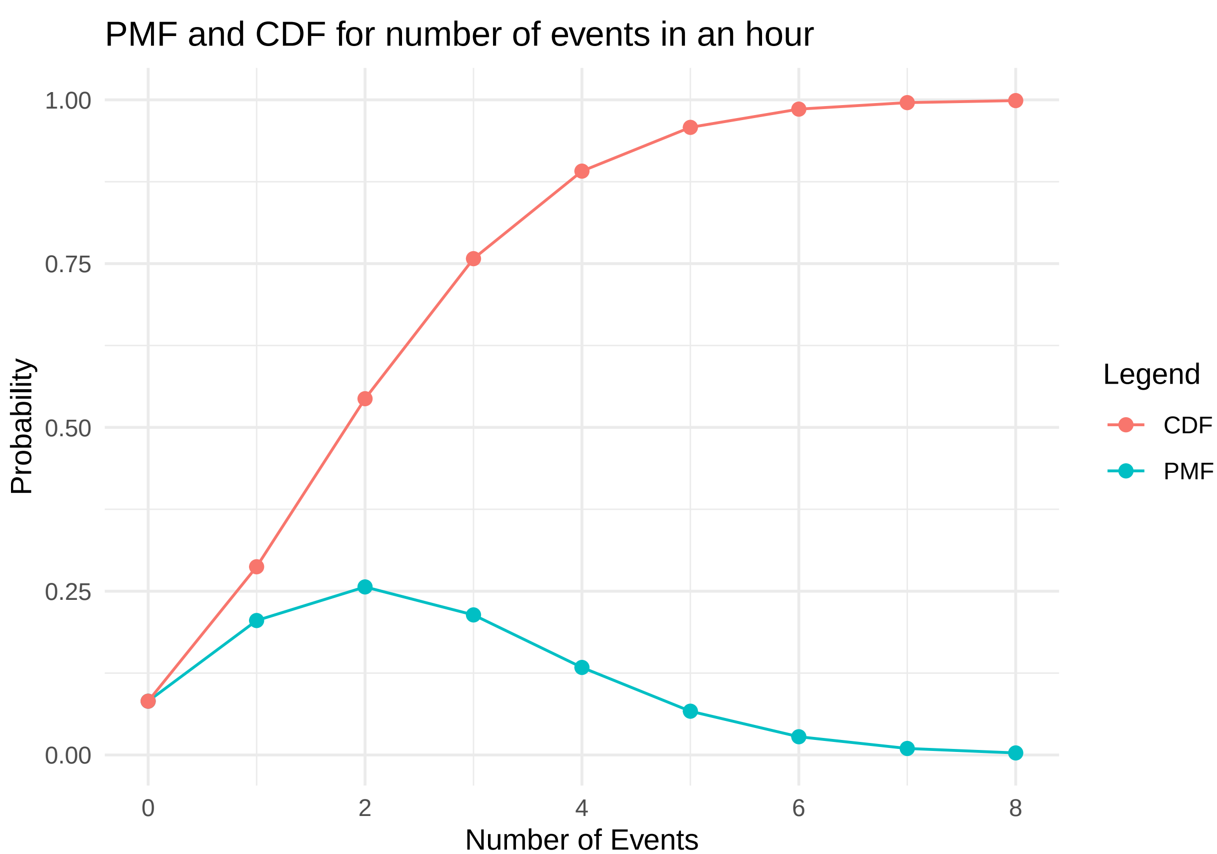
```
## [1] 0.7575761
```

The probability function (PMF) and cumulative distribution function (CDF):

```
library(ggplot2)

# Calculate PMF and CDF using Poisson distribution
PMF <- dpois(0:8, lambda = 2.5)
CDF <- ppois(0:8, lambda = 2.5)
df <- data.frame(events = 0:8, PMF, CDF)

# Plotting
ggplot(df, aes(x = events)) +
  geom_point(aes(y = PMF, color = "PMF"), size = 2) +
  geom_line(aes(y = PMF, color = "PMF")) +
  geom_point(aes(y = CDF, color = "CDF"), size = 2) +
  geom_line(aes(y = CDF, color = "CDF")) +
  labs(title = "PMF and CDF for number of events in an hour",
       x = "Number of Events",
       y = "Probability",
       color = "Legend") +
  theme_minimal()
```



Simulation of rolling three fair six-sided dice 1000 times:

```
num_dice <- 3
num_sides <- 6
num_simulations <- 1000
```

```r
First_Roll = sample(1:num_sides, num_simulations, replace = TRUE)
Second_Roll = sample(1:num_sides, num_simulations, replace = TRUE)
Third_Roll = sample(1:num_sides, num_simulations, replace = TRUE)

dice_df <- data.frame( First_Roll,Second_Roll,Third_Roll)

print(head(dice_df))
```

```
##   First_Roll Second_Roll Third_Roll
## 1          4           2          2
## 2          2           1          5
## 3          2           4          5
## 4          2           1          2
## 5          4           4          3
## 6          1           6          5
```

The joint probability of obtaining a sum greater than 10 on the first two dice rolls and a sum less than 5 on the third roll:

```r
# Use desired outcomes and total outcomes (Based on the data we created)
desired_outcomes <- subset(dice_df, (First_Roll + Second_Roll > 10) & (Third_Roll < 5))

probability <- nrow(desired_outcomes) / num_simulations

print(probability)
```

```
## [1] 0.05
```

This figure is calculated using the data. The theoretical probability should be 1/12*2/3 = 1/18 = about 0.055 as the events are independent.

The conditional probability of obtaining a sum greater than 12 on the third roll given that the sum of the first two rolls is 9:

```r
desired_outcomes2 <- subset(dice_df, (First_Roll+Second_Roll == 9)&(Third_Roll > 3))

available_outcomes <- subset(dice_df, (First_Roll + Second_Roll) == 9)

probability <- nrow(desired_outcomes2) / nrow(available_outcomes)

print(probability)
```

```
## [1] 0.3925234
```

Again this is an experimental probability, the theoretical probability would simply be 50 percent, as the first two events are independent of the third and there is a 50 percent chance that you roll over a 3.

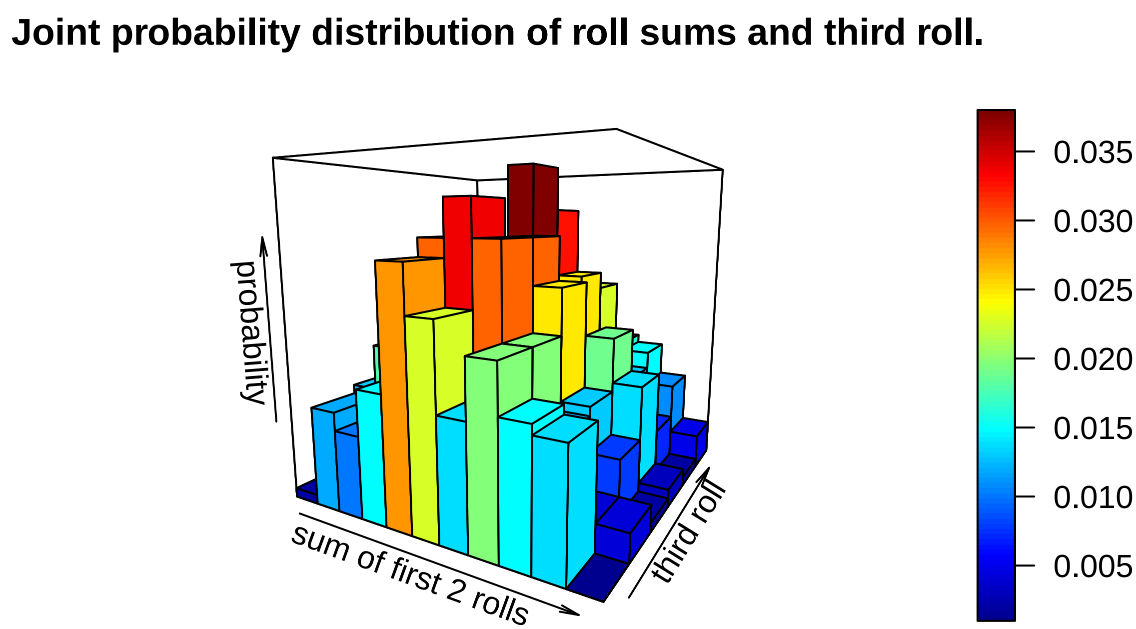
Here is a 3D histogram showing the joint probabilities of all possible outcomes:

```r
library(plot3D)

sum_first2rolls <- data.frame(First_Roll+Second_Roll,Third_Roll)

joint_prob <- table(sum_first2rolls) / 1000

hist3D(z = joint_prob, border = "black", theta = 30, phi = 10,
       xlab = "sum of first 2 rolls",
       ylab = "third roll",
       zlab = "probability",
       main = "Joint probability distribution of roll sums and third roll.")
```



This 3D histogram demonstrates the probability of joint outcomes. One axis represents the sum of the first 2 rolls, which is distributed as expected, with more 6 and 7 sums (near the middle), as that is a more likely outcome. The other horizontal axis represents the third roll, which will be very uniform when facing that direction, as the events are independent.