MSMS 408 : Practical 03

Ananda Biswas

Exam Roll No. : 24419STC053

February 17, 2026

Question 1

In a competitive examination, a student attempts 20 independent multiple-choice questions. Each question has four options, only one of which is correct, and the student answers each question randomly. The probability of answering any question correctly is therefore $1/3$. Let X denote the number of correct answers of the student.

(i) Assuming independence of questions, model X using an appropriate probability distribution.

(ii) Write down its probability mass function, compute the probability that the student answers exactly 6 questions correctly.

(iii) Find the expected number of correct answers.

(iv) Simulate observations of X using the inverse transform method to compare empirical and theoretical probabilities.

Theory

1 In the given set-up, X is said to have a Binomial distribution with parameters $n = 20$ and $p = \text{probability of success} = \dfrac{1}{3}$.

We write

$$X \sim \text{Binomial}\left(20, \frac{1}{3}\right).$$

2 The Probability Mass Function of X is given by

$$P[X = x] = \binom{20}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{20-x} \; ; \; x = 0, 1, 2, \ldots, 20.$$

Thus, the probability that the student answers exactly 6 questions correctly will be

$$\begin{aligned} P[X = 6] &= \binom{20}{6} \left(\frac{1}{3}\right)^6 \left(\frac{2}{3}\right)^{20-6} \\ &= \binom{20}{6} \left(\frac{1}{3}\right)^6 \left(\frac{2}{3}\right)^{14} \\ &\approx 0.182. \end{aligned}$$

3 If $X \sim \text{Binomial}\left(20, \dfrac{1}{3}\right)$, $E[X] = 20 \cdot \dfrac{1}{3} \approx 6.7$. So expected number of correct answers of the student is 6.

4 The inverse transform method to generate samples from a discrete random variable X is as follows.

I. Generate $u_i \sim U(0,1) \ \forall \ i = 1(1)n$.

II. Map $x_i = k$ whenever $F(k-1) < u_i \leq F(k) \ \forall \ i = 1(1)n$ where $F(\cdot)$ is the CDF of X. We are actually doing
$$X = min\{k : F(k) \geq U\}.$$

R Program

```
set.seed(22)
```

```
n <- 10000
```

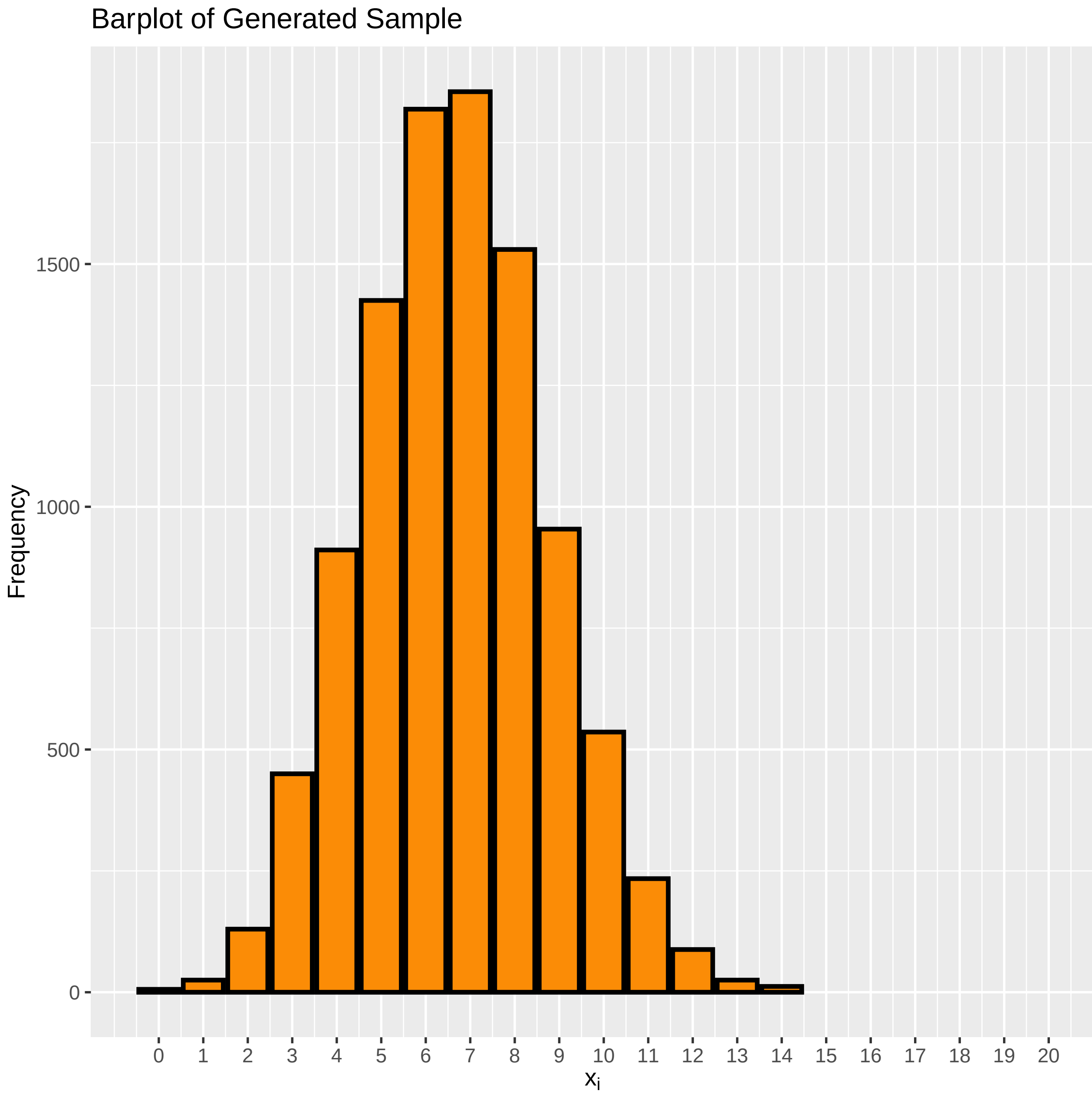
```
u <- runif(n)
```

```
CDF_F <- pbinom(q = 0:20, size = 20, prob = 1/3)
```

```
x <- findInterval(u, CDF_F, left.open = TRUE)
```

Let us have a barplot of the generated sample.

```
data.frame(x) %>%
  ggplot(aes(x = x)) +
  geom_bar(fill = "#FB8C06", color = "black", linewidth = 1) +
  scale_x_continuous(limits = c(-0.5,20), breaks = 0:20) +
  labs(x = expression(x[i]), y = "Frequency",
       title = "Barplot of Generated Sample")
```



✎ Most of the sample values are in between 1 and 13. Values in the tail have a very small probability mass, so with n = 10000 only, it is unlikely to observe them and exactly that has happened to be the case.

The empirical probabilities are obtained as

$$\hat{P}(X = x) = \frac{\text{count}(x)}{n}; \ \forall x = 0(1)20.$$

```
emp_prob <- table(factor(x, levels = 0:20)) / n
emp_prob <- as.numeric(emp_prob)
```

```
theo_prob <- dbinom(x = 0:20, size = 20, prob = 1/3)
```

```
df <- data.frame(x = 0:20,
                 empirical.probability = emp_prob,
                 theoretical.probability = theo_prob)
```

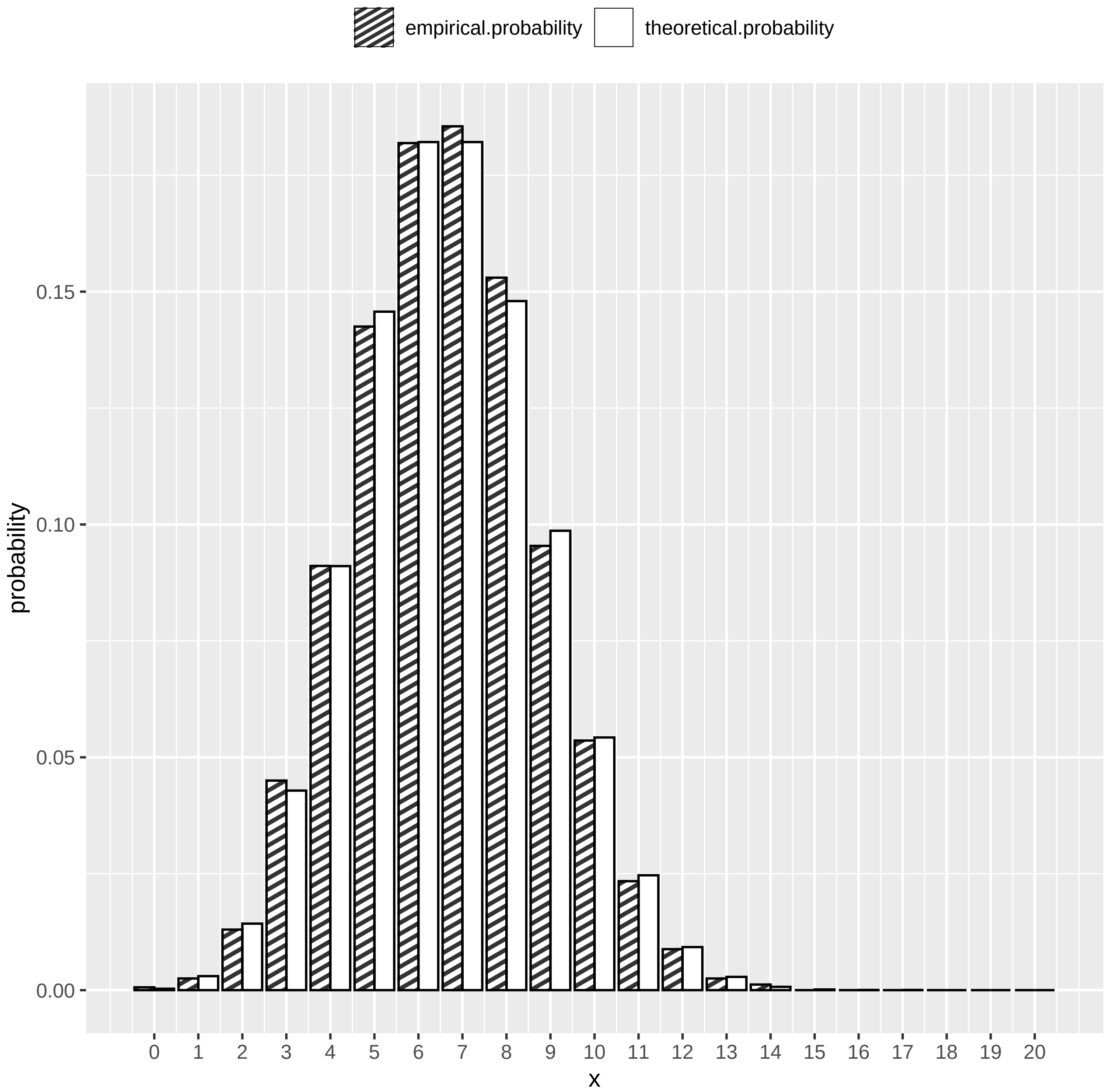
Table 1:

x	empirical.probability	theoretical.probability
0	0.00060	0.00030
1	0.00250	0.00301
2	0.01300	0.01428
3	0.04500	0.04285
4	0.09110	0.09106
5	0.14250	0.14570
6	0.18190	0.18213
7	0.18550	0.18213
8	0.15300	0.14798
9	0.09540	0.09865
10	0.05360	0.05426
11	0.02340	0.02466
12	0.00880	0.00925
13	0.00250	0.00285
14	0.00120	0.00071
15	0	0.00014
16	0	0.00002
17	0	0.000003
18	0	0.0000002
19	0	0
20	0	0

```
df_melted <- df %>%
                pivot_longer(cols = c("empirical.probability", "theoretical.probability"),
                             names_to = "probability_type",
                             values_to = "probability")
```

Let us have a grouped barplot from comparison of empirical and theoretical probabilities.

```
df_melted %>%
  ggplot(aes(x = x, y = probability, pattern = probability_type)) +
  geom_col_pattern(position = "dodge",
                   pattern_spacing = 0.01,
                   fill = "white", color = "black") +
  scale_pattern_manual(values = c("stripe", "none")) +
  scale_x_continuous(limits = c(-0.5, 20.5), breaks = 0:20) +
  theme(legend.position = "top",
        legend.title = element_blank())
```



Conclusion

✎ Empirical probabilities very much agree with theoretical probabilities.

Question 2

A fair die is rolled 100 times and let X denote the number appearing on the top face in each roll. Generate the 100 outcomes using a suitable random number method, compute the frequency of each face (1–6), and compare the observed relative frequencies with the theoretical probability 1/6 for a fair die.

Theory

In the said set-up, X is said to have a Discrete Uniform distribution with support $\{1, 2, 3, 4, 5, 6\}$. The Probability Mass Function of X is given by

$$P[X = x] = \frac{1}{6};\ x = 1(1)6.$$

We shall use Inverse Transform method to generate sample from X. Recall that, the CDF of X at $\{1, 2, 3, 4, 5, 6\}$ is

$$F(x) = \frac{x}{6};\ x = 1(1)6.$$

R Program

```
n <- 10000; u <- runif(n); CDF_F <- (1:6) / 6
```

```
x <- findInterval(u, CDF_F, left.open = TRUE) + 1
```

```
df1 <- data.frame(x = 1:6, count = as.numeric(table(factor(x, levels = 1:6))))
```

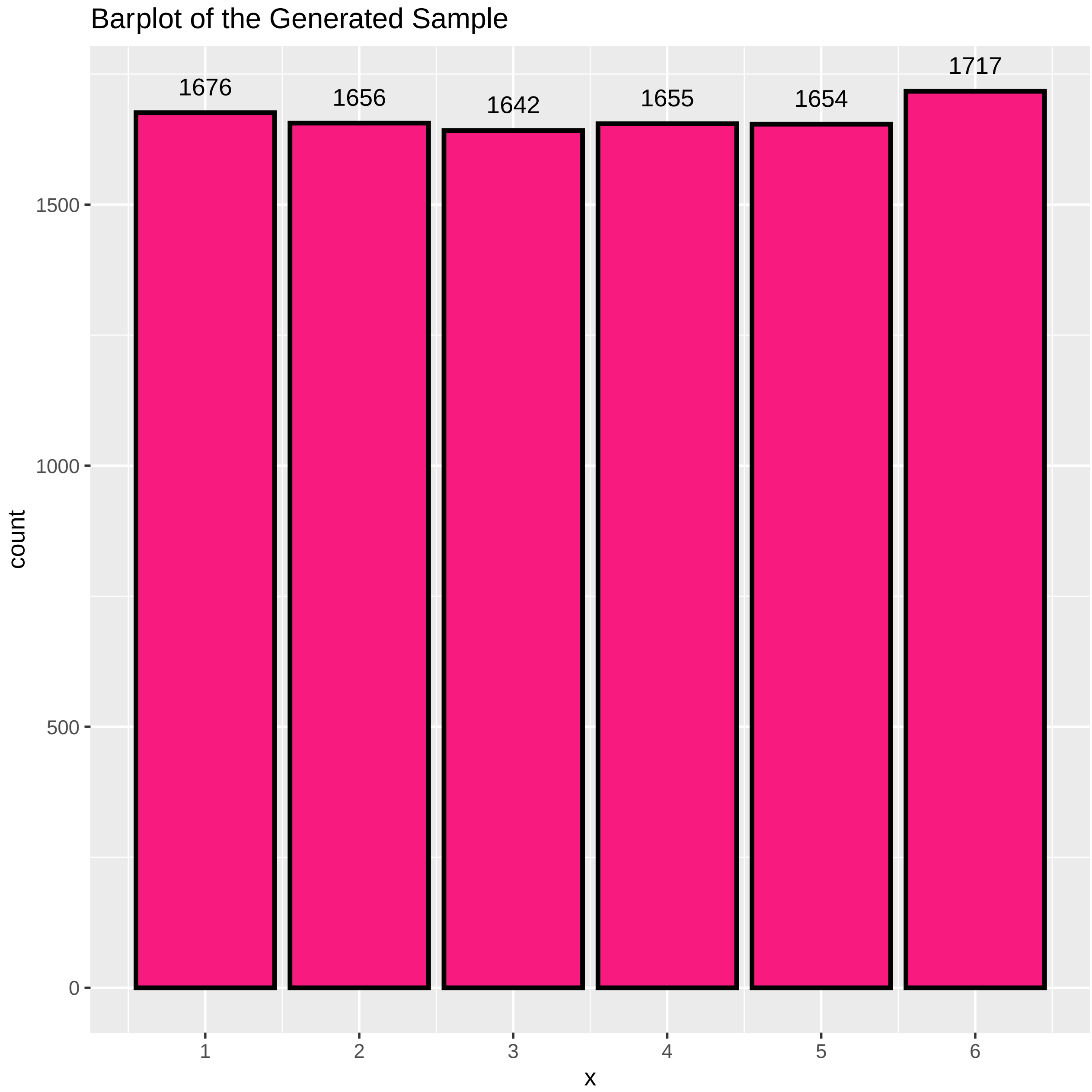
```
stargazer(df1, summary = FALSE, rownames = FALSE, label = "Table 2")
```

Table 2:

x	count
1	1,676
2	1,656
3	1,642
4	1,655
5	1,654
6	1,717

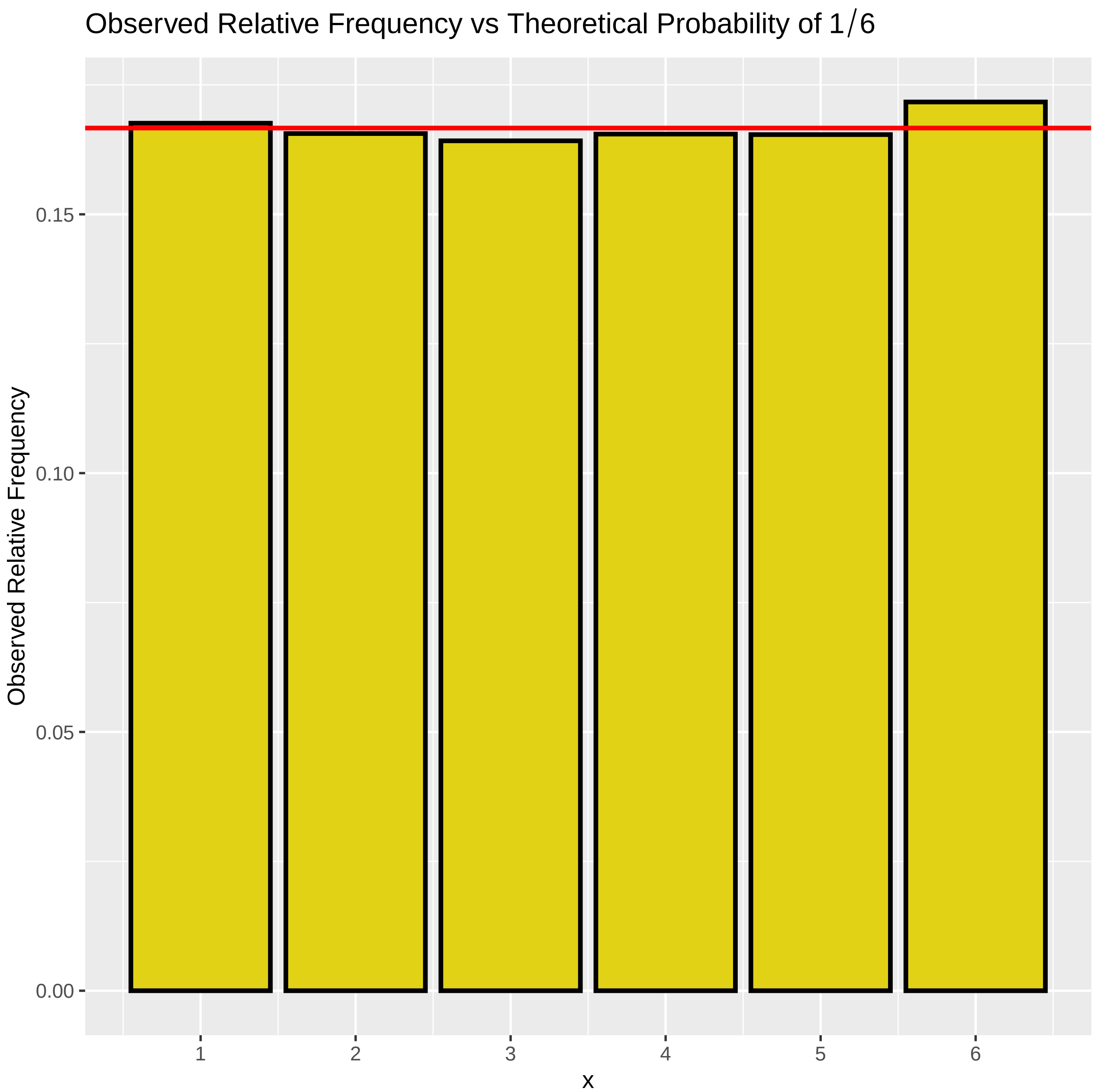
Let us have a barplot of the generated sample.

```
df1 %>%
  ggplot(aes(x = x, y = count)) +
  geom_col(fill = "#F81A7F", col = "black", linewidth = 1) +
  scale_x_continuous(breaks = 1:6) +
  geom_text(aes(label = count), vjust = -1, size = 4) +
  labs(title = "Barplot of the Generated Sample")
```



Here we have a plot to compare the observed relative frequencies and theoretical probability of $\frac{1}{6}$.

```
df1 %>%
  ggplot(aes(x = x, y = count / n)) +
  geom_col(fill = "#E1D216", col = "black", linewidth = 1) +
  scale_x_continuous(breaks = 1:6) +
  geom_hline(yintercept = 1/6, col = "red", linewidth = 1) +
  labs(y = "Observed Relative Frequency",
       title = expression("Observed Relative Frequency vs Theoretical Probability of" ~ 1/6))
```



Conclusion

✎ Observed relative frequencies are consistent with theoretical probability.