⌄ ORIE 4580/5580/5581 Assignment 1

Students: Patric Huang (jh3223)   Kahei Lam (kl2235)

Instructions

- Due Thursday September 4, at 11.59pm on Gradescope.
- Assignment .ipynb files available for download on Canvas.
- Do all your work in provided notebook (text answers typeset in markdown; show all required code and generate plots inline), and then generate and submit a pdf.
- Ideally do assignments in groups of 2, and submit a single pdf with both names
- Please show your work and clearly mark your answers.
- You can use any code fragments given in class, found online (for example, on StackOverflow), or generated via Gemini/Claude/ChatGPT (you are encouraged to use these for first drafts) **with proper referencing**.
- You can also discuss with others (again, please reference them if you do so); but you must write your final answers on your own as a team.

Suggested reading

Chapters 1 and 2 of Introduction to Probability by Grinstead and Snell (This textbook tries to teach probability from a 'simulation-first' perspective. Some of you may also enjoy the historical remarks!)

⌄ Q1:

The probability density function of the random variable X is given by

$$f_X(x) = \begin{cases} \frac{1}{x} & \text{if } c \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

a. What is the value of the constant c?

∨ Ans.

Knowing

$$\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$$

so we can derive it as follow:

$$\int_{-\infty}^{+\infty} f_X(x) = \int_{c}^{1} \frac{1}{x}\, dx = ln1 - ln(c) = 0 - ln(c) = 1$$

so:

$$ln(c) = -1 \implies \mathbf{c = \frac{1}{e}}$$

---

b. Compute the cdf $F_X(x)$ and mean $\mathbb{E}[X]$ of X?

∨ Ans.

1. To derive the cdf $F_X(x)$

The CDF is defined as: $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt$

We analyze cases based on the support of $f_X(t)$:

Case 1: $x < c$

When $x < c$, the PDF $f_X(t) = 0$ for all $t \in (-\infty, x]$. Thus:

$$F_X(x) = \int_{-\infty}^{x} 0\, dt = 0$$

Case 2: $c \leq x \leq 1$

When $c \leq x \leq 1$, the PDF $f_X(t) = \frac{1}{t}$ (valid only on $[c, x]$, since $f_X(t) = 0$ outside $[c, 1]$). Integrating:

$$F_X(x) = \int_{c}^{x} \frac{1}{t}\, dt = \ln(x) - \ln(c)$$

Substitute $c = \frac{1}{e}$ (from part A):

$$\ln(c) = \ln\left(\frac{1}{e}\right) = -1$$

Thus:

$$F_X(x) = \ln(x) + 1$$

Case 3: $x > 1$

When $x > 1$, the integral covers the entire non-zero interval $[c, 1]$. By the normalization of the PDF:

$$F_X(x) = \int_c^1 \frac{1}{t}\, dt = 1$$

Final Result

Combining all cases, the CDF is:

$$F_X(x) = \begin{cases} 0 & \text{if } x < c, \\ \ln(x) + 1 & \text{if } c \le x \le 1, \\ 1 & \text{if } x > 1. \end{cases}$$

2. Compute the mean $E[x]$ of X

The expected value $E[x]$ of a continuous random variable X is defined as:

$$E[x] = \int_{-\infty}^{+\infty} x \cdot f_X(x)\, dx$$

Given the probability density function of X

$$f_X(x) = \begin{cases} \dfrac{1}{x} & \text{if } c \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

where $c = \frac{1}{e}$ (from part A)

Since $f_X(x) = 0$ outside $[c, 1]$, the integral simplifies to:

$$\mathbb{E}[X] = \int_c^1 x \cdot \frac{1}{x}\, dx = [x]_c^1 = 1 - c = 1 - \frac{1}{e}$$

---

c. Consider the random variable $Y = 1 + \ln(X)$. Compute the cdf $F_Y(y)$. Based on this, can you identify the distribution of Y?

∨ Ans.

The cumulative distribution function of Y is defined as:

$$F_Y(y) = P(Y \le y)$$

substitute

$$Y = 1 + ln(X)$$

so

$$F_Y(y) = P(1 + ln(X) \le y) = P\left(X \le e^{y-1}\right)$$

From previous results, X has the PDF:

$$f_X(x) = \begin{cases} \frac{1}{x} & \text{if } \frac{1}{e} \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The support of X is $\left[\frac{1}{e}, 1\right]$, which corresponds to a range for Y:

- When $X = \frac{1}{e} : Y = 1 + \ln\left(\frac{1}{e}\right) = 0$
- When $X = 1 : Y = 1 + \ln(1) = 1$

Thus, Y in [0, 1]. We analyze three cases for Y:

Case1: $y < 0$

For $y < 0$, $e^{y-1} < \frac{1}{e}$ (since e^x is increasing).
Since $X \ge \frac{1}{e}$:

$$P\left(X \le e^{y-1}\right) = 0 \implies F_Y(y) = 0$$

Case2: $0 \le y \le 1$

For $0 \le y \le 1$, $\frac{1}{e} \le e^{y-1} \le 1$
Using the PDF of

$$F_Y(y) = \int_{\frac{1}{e}}^{e^{y-1}} f_X(x)\, dx = \int_{\frac{1}{e}}^{e^{y-1}} \frac{1}{x}\, dx$$

Evaluate the integral:

$$\int_{\frac{1}{e}}^{e^{y-1}} \frac{1}{x}\, dx = [\ln(x)]_{\frac{1}{e}}^{e^{y-1}} = \ln\left(e^{y-1}\right) - \ln\left(\frac{1}{e}\right)$$

Simplify using logarithmic properties:

$$\ln\left(e^{y-1}\right) = y - 1 \quad \text{and} \quad \ln\left(\frac{1}{e}\right) = -1$$

Thus:

$$F_Y(y) = (y - 1) - (-1) = y$$

Case3: $y > 1$

For $y > 1$, $e^{y-1} > 1$. Since $X \le 1$:
$$P\left(X \le e^{y-1}\right) = 1 \implies F_Y(y) = 1$$

Result:

Combining all cases, the CDF of Y is :

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0, \\ y & \text{if } 0 \le y \le 1, \\ 1 & \text{if } y > 1. \end{cases}$$

The CDF $F_Y(y)$ matches the cdf of a uniform distribution over the interval $[0, 1]$, denotted $Y \sim \text{Uniform}(0, 1)$

A uniform distribution $\text{Uniform}(a, b)$ has the CDF:

$$F_Z(z) = \begin{cases} 0 & \text{if } z < a, \\ \frac{z-a}{b-a} & \text{if } a \le z \le b, \\ 1 & \text{if } z > b. \end{cases}$$

For $a = 0$ and $b = 1$, this simplifies to $F_Z(z) = z$, confirming $Y \sim \text{Uniform}(0, 1)$

---

d. Now let $U \sim U[0, 1]$ be a uniformly distributed random variable on $[0, 1]$, and consider the random variable $Z = e^{U-1}$. Compute the pdf $f_Z(z)$.

⌄ Ans.

The PDF of a uniformly distributd random variable U on $[0, 1]$ is :

$$f_U[u] = \begin{cases} 1 & \text{if } 0 \le u \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The CDF of Z is defined as:

$$F_Z(z) = p(Z \le z) = P\left(e^{u-1} \le z\right) = P(u \le ln(z) + 1)$$

And now analyze cases for $F_Z(z)$ based on the rangge of z:

Case1: $z < e^{-1}$

If $z < e^{-1}$, then $ln(z) + 1 < 0$. Since $U \ge 0$:

$$P\left(U \le \ln(z) + 1\right) = 0 \implies F_Z(z) = 0$$

Case2: $e^{-1} \le z \le 1$

If $e^{-1} \le z \le 1$, then $0 \le \ln(z) + 1 \le 1$. Using the PDF of U:

$$F_Z(z) = \int_0^{\ln(z)+1} f_U(u) \, du = \int_0^{\ln(z)+1} 1 \, du$$

Evaluate the integral:

$$\int_0^{\ln(z)+1} 1 \, du = [u]_0^{\ln(z)+1} = \ln(z) + 1$$

Case 3: $z > 1$

If $z > 1$, then $\ln(z) + 1 > 1$. Since $U \le 1$:

$$P\left(U \le \ln(z) + 1\right) = P(U \le 1) = 1 \implies F_Z(z) = 1$$

Thus, the PDF of Z is:

$$f_Z(z) = \begin{cases} \dfrac{1}{z} & \text{if } e^{-1} \le z \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

From Part (a), the PDF of X is:

$$f_X(x) = \begin{cases} \dfrac{1}{x} & \text{if } e^{-1} \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

We observe that $f_Z(z) = f_X(z)$. Thus, Z and X follow the **same distribution**.

e. (OPTIONAL) Now consider a random variable X with cdf $F_X(x)$, and another random variable $Y = F_X(X)$. What is the distribution of Y? Next, let $F_X^{-1}(\cdot)$ denote the \emph{inverse} function of $F_X(\cdot)$, i.e., $F_X(x) = y \Leftrightarrow F_X^{-1}(y) = x$, and let $U \sim U[0, 1]$ be a uniformly distributed r.v. What is the distribution of $F_X^{-1}(U)$?

⌄ Ans.

Part1: Distribution of $Y = F_X(X)$

To find the distribution of $Y = F_X(X)$, we analyze its CDF of $F_Y(y)$

1. **Definition of $F_Y(y)$**
   The CDF of Y is:
   $$F_Y(y) = P(Y \le y) = P\left(F(X) \le y\right)$$

2. **Case Analysis for $F_Y(y)$**
   Since $F_X(x)$ is a CDF, its range is $[0, 1]$. We consider three cases:

   - Case1: $y < 0$
     $F_X(X) \ge 0$ for all X, so:
     $$F_Y(y) = P\left(F_X(X) \le y\right) = 0$$
   - Case2: $y \ge 1$
     $F_X(x) \le 1$ for all X, so:
     $$F_Y(y) = P\left(F_X(x) \le y\right) = 1$$
   - Case3: $0 \le y \le 1$
     Since $F_X(x)$ is non-decreasing, $F_X(X) \le y \iff X \le F_X^{-1}(y)$. Thus:
     $$F_Y(y) = P\left(X \le F_X^{-1}(y)\right) = F_X\left(F_X^{-1}(y)\right) = y$$

3. **Conclusion for Y**
   Combining cases, the CDF of Y is:
   $$F_Y(y) = \begin{cases} 0 & \text{if } y < 0, \\ y & \text{if } 0 \le y \le 1, \\ 1 & \text{if } y > 1. \end{cases}$$
   This matches the CDF of a uniform distribution on $[0, 1]$, so:
   $$Y \sim U[0, 1]$$

Part 2: Distribution of $F_X^{-1}(U)$ where $U \sim U[0, 1]$

Let $Z = F_X^{-1}(U)$. To find its distribution, analyze its CDF $ F_Z(z) :

1. **Definition of $F_Z(z)$**

   The CDF of Z is $F_Z(z) = P(Z \leq z) = P\left(F_X^{-1}(U) \leq z\right)$

2. **Simplify the Probability**

   Applying F_X to both sides of $F_X^{-1}(U) \leq z$ (using monotonicity of F_X:
   $$U \leq F_X(z)$$

   Since $U \sim U[0, 1]$, its CDF is $F_U(u) = u$ for $0 \leq u \leq 1$. Thus:
   $$F_Z(z) = P\left(U \leq F_X(z)\right) = F_U\left(F_X(z)\right) = F_X(z)$$

3. **Conclusion for (Z)**

   The CDF of $Z = F_X^{-1}(U)$ equals the CDF of X, so:
   $$F_X^{-1}(U) \sim X$$

Final Results

1. $Y = F_X(X)$ follows a uniform distribution: $\boxed{Y \sim U[0, 1]}$

2. $F_X^{-1}(U)$ has the same distribution as X: $\boxed{F_X^{-1}(U) \sim X}$

Q2:

Let X and Y be Bernoulli random variables with success probability p and q respectively, i.e.,

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases} \quad , \quad Y = \begin{cases} 1 & \text{with probability } q \\ 0 & \text{with probability } 1 - q \end{cases}$$

Let $\mathbb{P}(X = 1 \text{ and } Y = 1) = \frac{1}{3}$.

---

a. What is $Cov(X, Y)$?

∨ Ans.

The covariance of X and Y is defined as:

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

For Bernoulli random variables, the expectation equals the success probability:

$$E[X] = P(X = 1) = p, \quad E[Y] = P(Y = 1) = q$$

The product XY is 1 only if both $X = 1$ and $Y = 1$; otherwise, it is 0:

$$E[XY] = P(XY = 1) = P(X = 1, Y = 1) = \frac{1}{3}$$

Substitute into the covariance formula:

$$\mathrm{Cov}(X, Y) = \frac{1}{3} - pq$$

---

b. For what values of p and q are X and Y uncorrelated?

∨ Ans.

Random variables are uncorrelated if their covariance is 0:

$$\mathrm{Cov}(X, Y) = 0 \implies \frac{1}{3} - pq = 0$$

Thus:

$$pq = \frac{1}{3}$$

---

c. Are X and Y independent for the values in (b)?

⌄ Ans.

From Part (b), $pq = \frac{1}{3}$. We verify all cases:

1. For $x = 1, y = 1$:

$$P(X = 1, Y = 1) = \frac{1}{3} = pq = P(X = 1)P(Y = 1)$$

2. For $x = 1, y = 0$:

$$P(X = 1, Y = 0) = P(X = 1) - P(X = 1, Y = 1) = p - \frac{1}{3}$$

$$P(X = 1)P(Y = 0) = p(1 - q) = p - pq = p - \frac{1}{3}$$

They are equal.

3. For $x = 0, y = 1$ and $x = 0, y = 0$, similar calculations show equality.

Thus, X and Y are independent when $pq = \frac{1}{3}$.

---

d. Suppose you have a function *Bernoulli*(r) that simulates the outcomes of a Bernoulli random variable for any success probability r, i.e., it will generate a 1 with probability r and 0 with probability $1 - r$.

Explain how you can use this function to generate a realization of X and Y above for $p = \frac{1}{2}, q = \frac{2}{3}$.

∨ Ans.

Using the definition of conditional probability:
$$P(X = 1, Y = 1) = P(X = 1) \cdot P(Y = 1 \mid X = 1)$$
Substitute known values:
$$\frac{1}{3} = \frac{1}{2} \cdot P(Y = 1 \mid X = 1) \implies P(Y = 1 \mid X = 1) = \frac{2}{3}$$
By the law of total probability:
$$P(Y = 1) = P(Y = 1 \mid X = 1)P(X = 1) + P(Y = 1 \mid X = 0)P(X = 0)$$
Substitute known values:
$$\frac{2}{3} = \frac{2}{3} \cdot \frac{1}{2} + P(Y = 1 \mid X = 0) \cdot \frac{1}{2}$$
Solve for $P(Y = 1 \mid X = 0)$:
$$P(Y = 1 \mid X = 0) = \frac{2}{3}$$

1. Generate $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$:
$$X = \text{Bernoulli}\left(\frac{1}{2}\right)$$

2. Generate $Y \sim \text{Bernoulli}\left(\frac{2}{3}\right)$ (same distribution regardless of X):
$$Y = \text{Bernoulli}\left(\frac{2}{3}\right)$$

---

e. Could you also use this function to get a realization of X and Y above if $p = \frac{1}{2}$, $q = \frac{1}{2}$ (and $\mathbb{P}(X = 1 \text{ and } Y = 1) = \frac{1}{3}$)? Explain how, or why not.

⌄ Ans.

Using the joint probability formula:

$$\frac{1}{3} = \frac{1}{2} \cdot P(Y = 1 \mid X = 1) \implies P(Y = 1 \mid X = 1) = \frac{2}{3}$$

Apply the law of total probability:

$$\frac{1}{2} = \frac{2}{3} \cdot \frac{1}{2} + P(Y = 1 \mid X = 0) \cdot \frac{1}{2}$$

Solve for $P(Y = 1 \mid X = 0)$:

$$P(Y = 1 \mid X = 0) = \frac{1}{3}$$

1. Generate $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$:

$$X = \text{Bernoulli}\left(\frac{1}{2}\right)$$

2. Generate Y based on X:

$$Y = \begin{cases} \text{Bernoulli}\left(\frac{2}{3}\right) & \text{if } X = 1, \\ \text{Bernoulli}\left(\frac{1}{3}\right) & \text{if } X = 0. \end{cases}$$

For both cases, the joint probability satisfies:

$$P(X = 1, Y = 1) = \frac{1}{3}$$

as required.

⌄ Q3:

In Chapter 1 example 1.4, Grinstead and Snell talk about measuring two natural quantities associated with the so-called *simple random walk* $S_0 = 0$, $S_t = S_{t-1} + X_t$, where $X_t = \pm 1$ with probability $\frac{1}{2}$ (they describe this as a simple betting game between Peter and Paul, where at each time t they toss a fair coin, and if it comes up heads, then Peter wins a penny from Paul, else Paul wins a penny from Peter). We will now replicate their experiments.

a. Generate $n = 1000$ copies of a simple random walk $S^i = (S_0^i, S_1^i, \ldots, S_T^i)$ (with $T = 100$ and $i \in \{1, 2, \ldots, n\}$). For each walk, compute $Z_T^i = \frac{S_T^i}{T}$ (i.e., the per-round winnings of Paul after 100 games), as well as $W_T^i = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{\{S_t^i > 0\}}$ (i.e., the fraction of rounds in $[1, T]$ where Paul is in the lead). Plot a histogram of both these quantities (using `plt.hist(data, density=True,bins='auto')`).

⌄ Ans.

```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(111)

n = 1000
T = 100

steps = np.random.choice([-1, 1], size=(n, T))
S = np.cumsum(steps, axis=1)
S = np.hstack([np.zeros((n, 1)), S])

Z_T = S[:, -1] / T

W_T = np.mean(S[:, 1:] > 0, axis=1)

plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
plt.hist(Z_T, bins='auto', density=True)
plt.title(r'Histogram of $Z_T^i = S_T^i / T$')
plt.xlabel(r'$Z_T^i$')
plt.ylabel('Density')
plt.subplot(1, 2, 2)
plt.hist(W_T, bins='auto', density=True)
plt.title(r'Histogram of $W_T^i = \frac{1}{T} \sum_{t=1}^T 1_{\{\S_t^i > 0\}}$')
plt.xlabel(r'$W_T^i$')
plt.ylabel('Density')
plt.show()
```

Histogram of $Z_T^i = S_T^i/T$

Histogram of $W_T^i = \frac{1}{T} \sum\limits_{t=1}^{T} 1_{\{S_t^i > 0\}}$
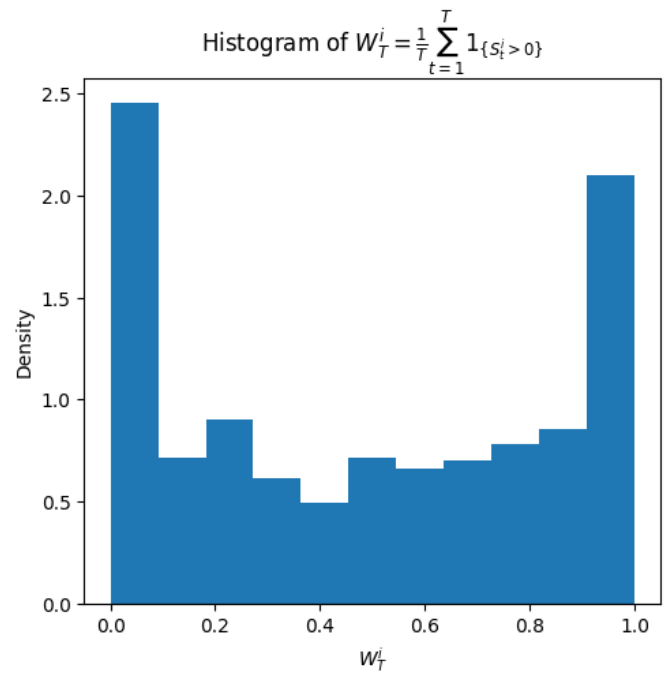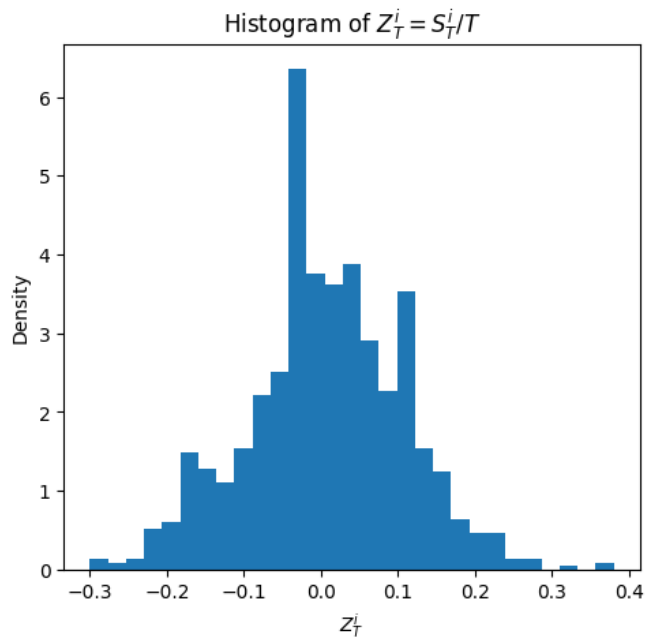
b. Repeat for $T = 1000$ and $T = 10000$.
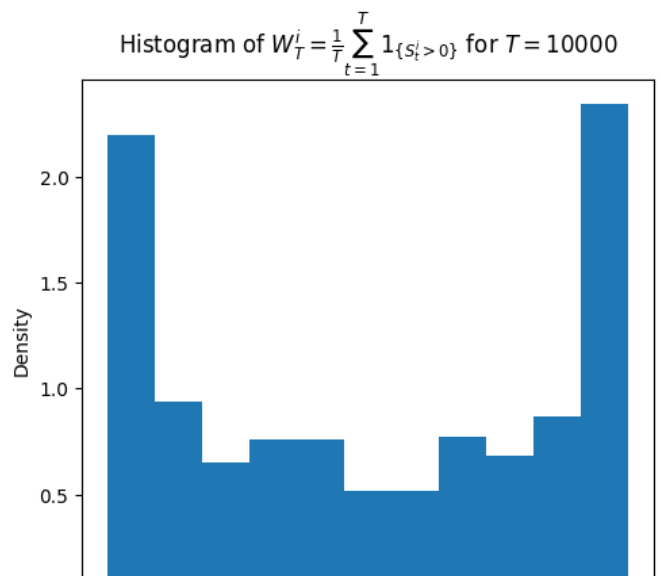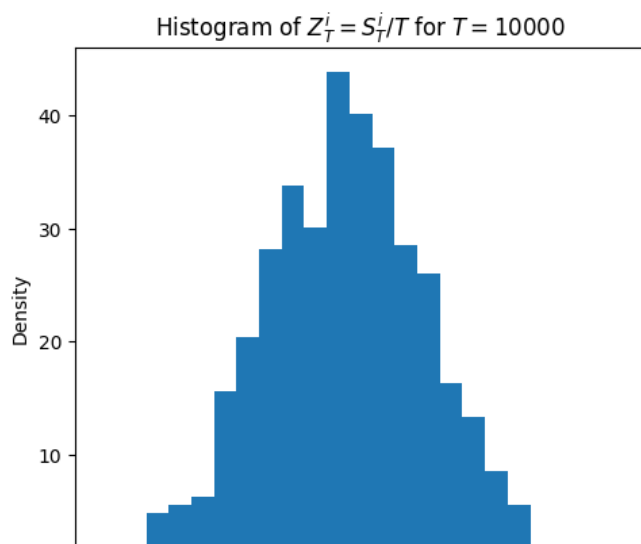
Ans.

```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(111)

n = 1000
T_values = [1000, 10000]

for T in T_values:
    steps = np.random.choice([-1, 1], size=(n, T))
    S = np.cumsum(steps, axis=1)
    S = np.hstack([np.zeros((n, 1)), S])
```
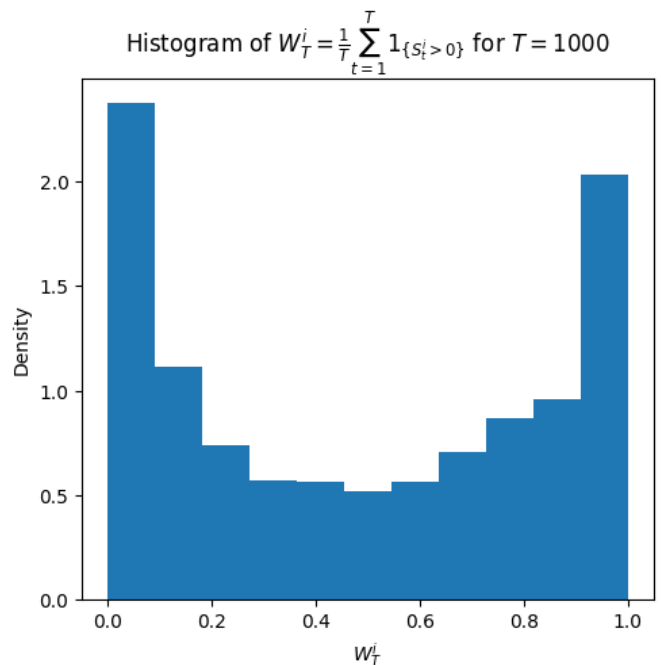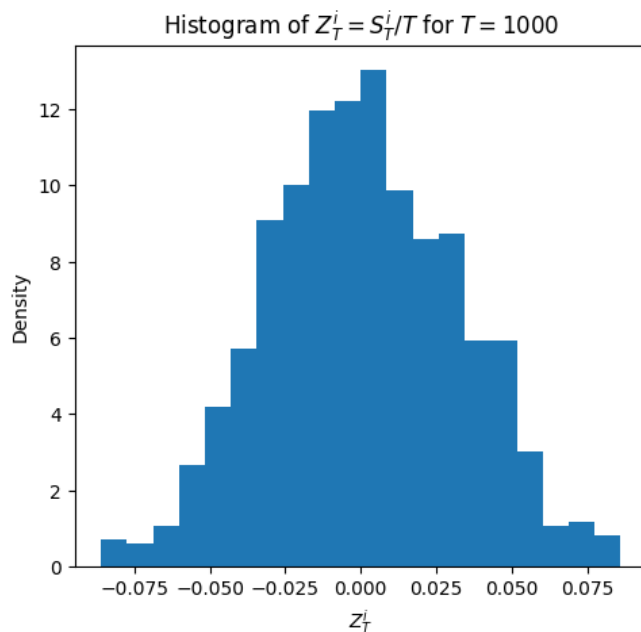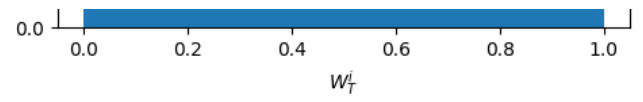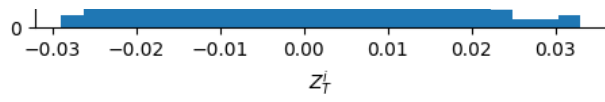
```python
Z_T = S[:, -1] / T

W_T = np.mean(S[:, 1:] > 0, axis=1)

plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
plt.hist(Z_T, bins='auto', density=True)
plt.title(rf'Histogram of $Z_T^i = S_T^i / T$ for $T={T}$')
plt.xlabel(r'$Z_T^i$')
plt.ylabel('Density')
plt.subplot(1, 2, 2)
plt.hist(W_T, bins='auto', density=True)
plt.title(rf'Histogram of $W_T^i = \frac{{1}}{{T}} \sum_{{t=1}}^{{T}} 1_{{\
plt.xlabel(r'$W_T^i$')
plt.ylabel('Density')
plt.show()
```

c. How do you think the distributions for Z_T and W_T behave as $T \to \infty$? Can you give some intuition why this is the case?

Ans.

As $T \to \infty$, we have: $Z_T^i \to 0$ in probability.

Since each X_t has mean 0 and variance 1, we have:
$$\mathbb{E}[S_T] = T \cdot \mathbb{E}[X_t] = 0,$$
$$\mathrm{Var}(S_T) = T \cdot \mathrm{Var}(X_t) = T.$$

So, for large T, we can approximate:
$$S_T \sim N(0, T).$$

so $Z_T = \frac{S_T^i}{T} \sim N\left(0, \frac{1}{T}\right)$, which concentrates around 0 as T increases.

As $T \to \infty$, the distribution of W_T^i converges to the arcsine distribution, showned by a U-shaped curve: $W_T^i \xrightarrow{d} \mathrm{Arcsine}(0, 1)$. This means most walks spend almost all or almost none of the time above zero and switching leads is rare. This is because a random walk has momentum — once it drifts above 0, it tends to stay positive for a while, and vice versa.

Q4:

As we discuss in class, one major application of simulation is to estimate difficult integrals. In this question, we will perform a simple simulation to estimate the area of a unit circle.

a. Generate $10,000$ points uniformly distributed in $[-1, 1]$ and plot a histogram of these points using 50 uniform-sized bins.

Hint: See the histogram function in matplotlib; in particular, if your data is stored in an array X, then you can use `<plt.hist(X,bins = 50,density = True)>`

Ans.

```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(111)
X = np.random.uniform(-1, 1, size=10000)

plt.hist(X, bins=50, density=True, edgecolor='black')
plt.title("Histogram of 10,000 Uniform Samples in [-1, 1]")
plt.xlabel("Value")
plt.ylabel("Density")
plt.grid(True)
plt.show()
```
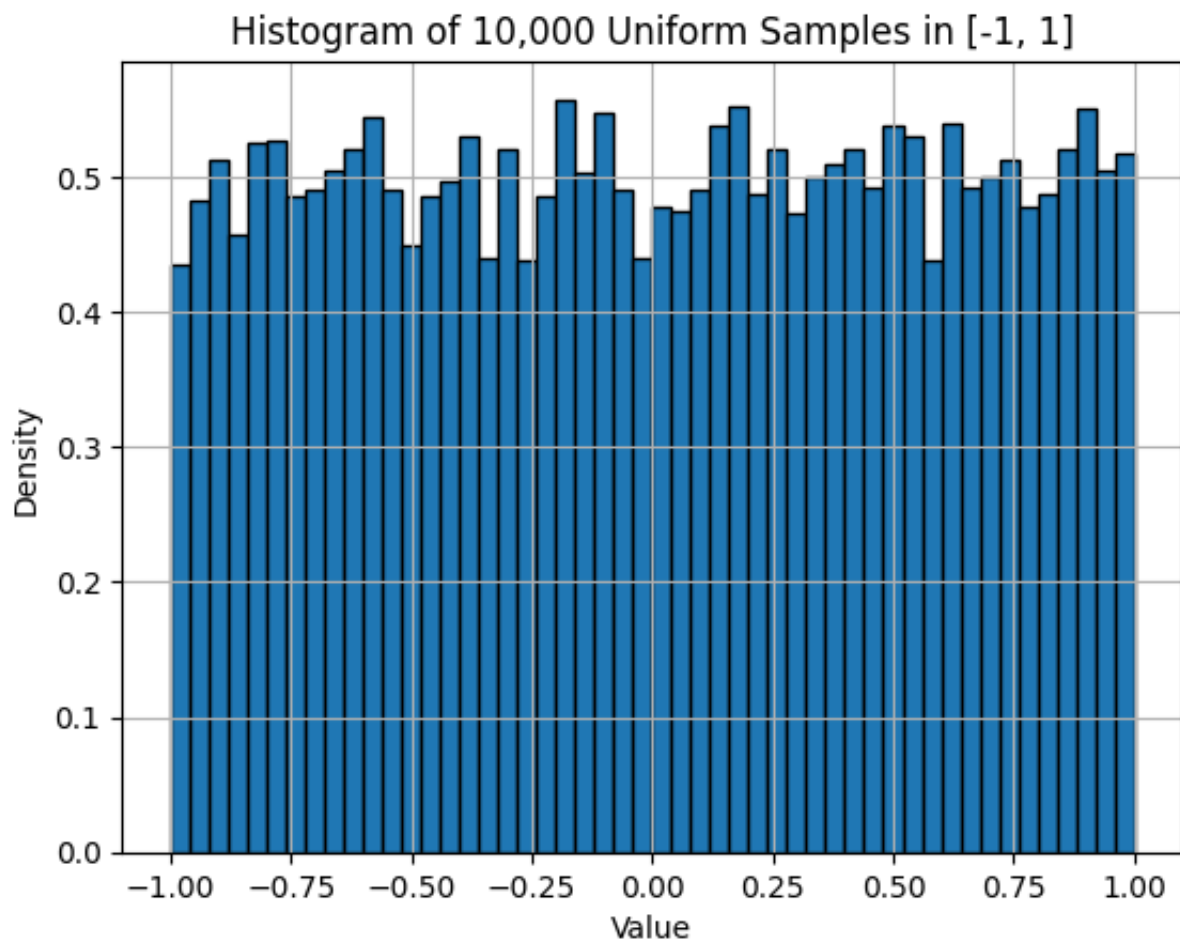
b. Generate $10,000$ points that are uniformly distributed in a square around the origin with corners $(-1, -1), (-1, 1), (1, -1)$ and $(1, 1)$. To do this, independently generate the x- and y-coordinates of a point to be uniformly distributed in $(-1, 1)$. Create an additional array Z of length $10,000$, which indicates whether each random point lies within a circle of radius 1 centered at the origin. Plot these points using a scatter plot with different colors for points within the circle and points outside.

Hint: If you store your coordinate vectors as X and Y and the vector of indicators as Z, then use <plt.scatter(X,Y, c = Z, s = 2.0, cmap=plt.cm.Paired)>
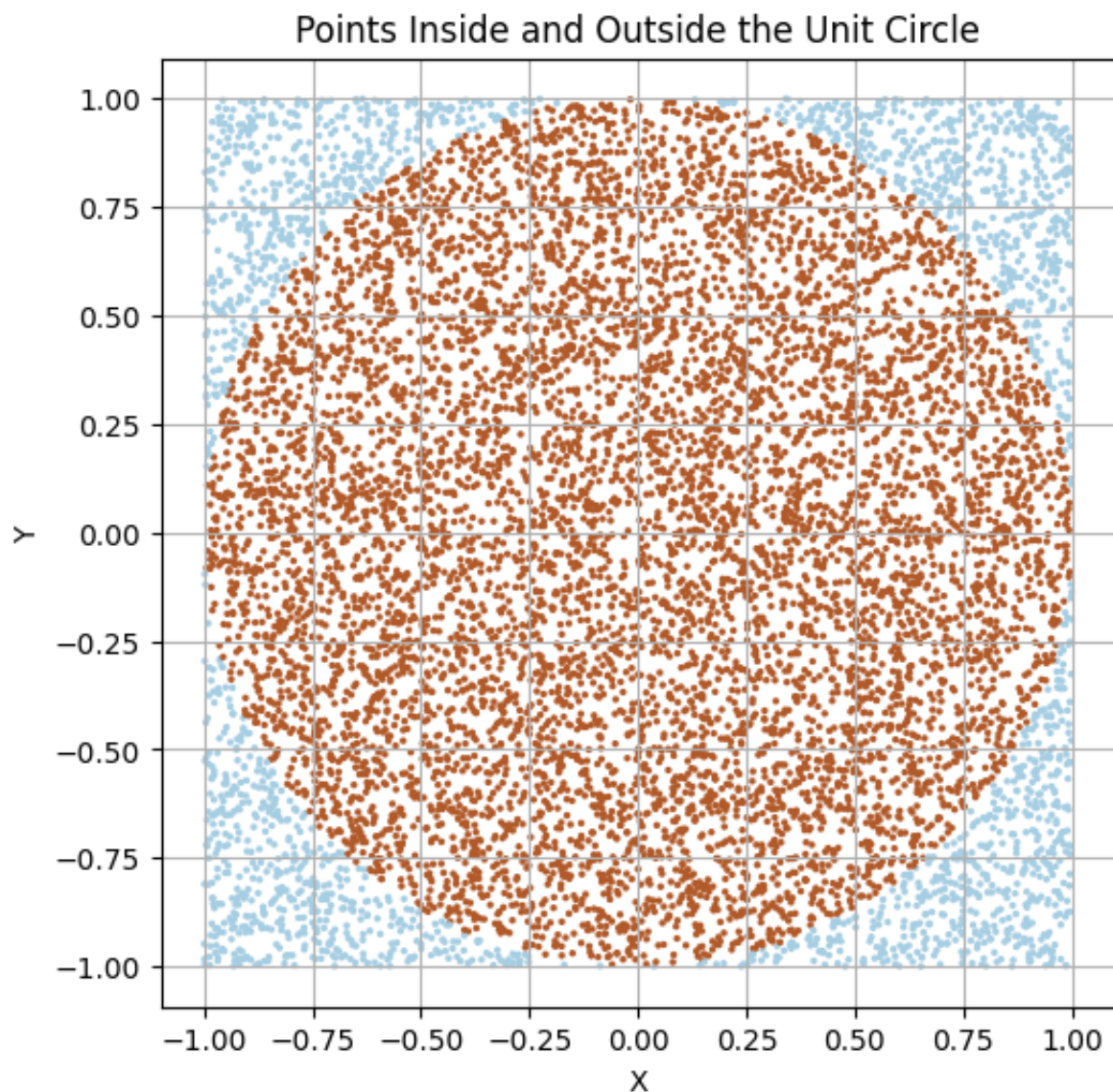
⌄ Ans.

See plot below.

```python
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(111)
X = np.random.uniform(-1, 1, 10000)
Y = np.random.uniform(-1, 1, 10000)
Z = (X**2 + Y**2 <= 1).astype(int)

plt.figure(figsize=(6, 6))
plt.scatter(X, Y, c=Z, s=2.0, cmap=plt.cm.Paired)
plt.title("Points Inside and Outside the Unit Circle")
plt.xlabel("X")
plt.ylabel("Y")
plt.axis("equal")
plt.grid(True)
plt.show()
```

c. Define N_{circle} to be the (random) number of points which lie within the circle. What is $\mathbb{E}[N_{circle}]$ and $Var(N_{circle})$ (express your answer in terms of π)?

Suppose also that you did not know the value of π; can you still argue that $Var(N_{circle}) \leq \frac{n}{4}$?

Ans.

The area of the square is 4, and the area of the unit circle is π, so the probability that a randomly chosen point falls inside the circle is:

$$p = \frac{\pi}{4}$$

Since each point falls inside the circle independently with probability p, the number of points inside the circle follows a binomial distribution:

$$N_{\text{circle}} \sim \text{Binomial}(n, p)$$

The expected number of points inside the circle is:

$$\mathbb{E}[N_{\text{circle}}] = n \cdot p = n \cdot \frac{\pi}{4}$$

The variance of the number of points inside the circle is:

$$\text{Var}(N_{\text{circle}}) = n \cdot p \cdot (1 - p) = n \cdot \frac{\pi}{4}\left(1 - \frac{\pi}{4}\right)$$

The function $p(1 - p)$ achieves its maximum at $p = \frac{1}{2}$, where $p(1 - p) = \frac{1}{4}$. Therefore, even if we do not know the value of π, we can bound the variance as:

$$\text{Var}(N_{\text{circle}}) \leq n \cdot \frac{1}{4}$$

---

d. Given n random points generated as in part (b), we can estimate π as $\pi_n = \frac{4N_{circle}}{n}$ for π (where as before N_{circle} is the random number of points that lie inside the circle). For n between 1 and 1000, plot the relative error $|\pi_n - \pi|/\pi$. Also plot the function $1/\sqrt{n}$ on the same plot. How do the relative errors compare to this function?

∨ Ans.

The relative error stays below to $1/\sqrt{n}$ from $n = 1$ to $n = 1000$, which is the bound on the error of Monte Carlo methods.

```python
import numpy as np
import matplotlib.pyplot as plt

pi_true = np.pi
n_vals = np.arange(1, 1001)
pi_estimates = np.zeros_like(n_vals, dtype=float)
rel_errors = np.zeros_like(n_vals, dtype=float)

np.random.seed(111)
X = np.random.uniform(-1, 1, size=1000)
Y = np.random.uniform(-1, 1, size=1000)
inside_circle = (X**2 + Y**2 <= 1).astype(int)
cumulative_inside = np.cumsum(inside_circle)

for i, n in enumerate(n_vals):
    pi_n = 4 * cumulative_inside[i] / n
    pi_estimates[i] = pi_n
    rel_errors[i] = np.abs(pi_n - pi_true) / pi_true

plt.figure(figsize=(10, 6))
plt.plot(n_vals, rel_errors, label='Relative Error $|\\pi_n - \\pi| / \\pi$', c
plt.plot(n_vals, 1 / np.sqrt(n_vals), label='$1/\\sqrt{n}$', color='red')
plt.xlabel("n (number of samples)")
plt.ylabel("Error")
plt.legend()
plt.grid(True)
plt.show()
```

---
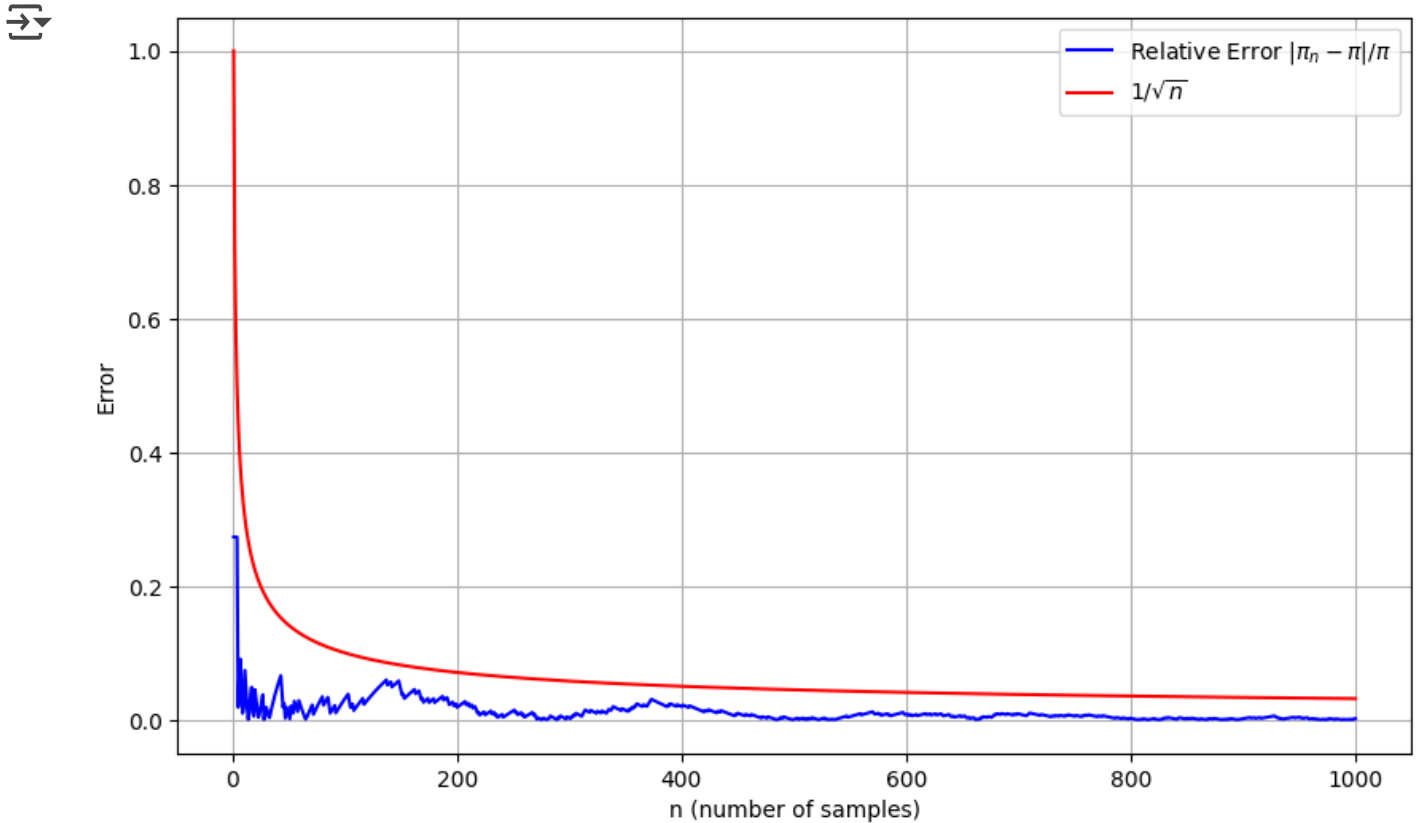
e. Recall that a 95% confidence interval for an estimate is given by $\pi_n \pm \dfrac{1.96\sqrt{Var(\pi_n)}}{\sqrt{n}}$ (the relative error plot in previous question shows why this is the case). Using your upper bound for $Var(\pi_n)$ from part (c), compute what value of n is required to ensure that the width of the confidence interval is less than 0.001, and then, simulate using this value n and report an estimate with a 95% confidence interval for π that is accurate up to the second digit after the decimal point.

∨   Ans.

```python
import numpy as np
import matplotlib.pyplot as plt

max_width = 0.001
z = 1.96
numerator = 2 * z * 2
required_n = int(np.ceil(numerator / max_width))
print(f"Required sample size n to get CI width < {max_width}: {required_n}")

np.random.seed(111)
X = np.random.uniform(-1, 1, size=required_n)
Y = np.random.uniform(-1, 1, size=required_n)
inside_circle = (X**2 + Y**2 <= 1)
N_circle = np.sum(inside_circle)

pi_hat = 4 * N_circle / required_n

std_error_bound = np.sqrt(4 / required_n)
ci_lower = pi_hat - z * std_error_bound
ci_upper = pi_hat + z * std_error_bound

print(f"\nEstimated π: {pi_hat:.2f}")
print(f"95% Confidence Interval: [{ci_lower:.2f}, {ci_upper:.2f}]")
print(f"Confidence Interval Width: {ci_upper - ci_lower:.2f}")
```

```
Required sample size n to get CI width < 0.001: 7840

Estimated π: 3.15
95% Confidence Interval: [3.11, 3.20]
Confidence Interval Width: 0.09
```

Q5: (For 5580/5581 students; OPTIONAL for 4580)

Above we used simulation to compute the area of a unit circle. Using a similar process, we can find the volume of a 'sphere' in n dimensions. Estimate this volume for dimensions $d \in \{2, 3, \dots, 20\}$, using $100,000$ samples in each dimension. How does your estimate of the volume of the sphere behave as the dimension increases? Also, do your estimates seem smooth or noisy (and if the latter, why)?

⌄ Ans.

As d increases, the volume shrinks rapidly. The estimates become noisier in higher dimensions due to fewer points landing inside the ball. This might because most of the cube's volume lies far from the origin in high dimensions.

```python
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(111)

num_samples = 100000
dimensions = range(2, 21)
volumes = []

for d in dimensions:
    X = np.random.uniform(-1, 1, size=(num_samples, d))
    squared_norms = np.sum(X**2, axis=1)
    inside_ball = np.sum(squared_norms <= 1)

    cube_volume = 2 ** d
    ball_volume = (inside_ball / num_samples) * cube_volume
    volumes.append(ball_volume)

plt.figure(figsize=(10, 6))
plt.plot(dimensions, volumes)
plt.title("Estimated Volume of the Unit Ball in $\mathbb{R}^d$")
plt.xlabel("Dimension $d$")
plt.ylabel("Estimated Volume")
plt.grid(True)
plt.show()
```

```
<>:21: SyntaxWarning: invalid escape sequence '\m'
<>:21: SyntaxWarning: invalid escape sequence '\m'
/tmp/ipython-input-436234323.py:21: SyntaxWarning: invalid escape sequence
  plt.title("Estimated Volume of the Unit Ball in $\mathbb{R}^d$")
```

Estimated Volume of the Unit Ball in $\mathbb{R}^d$