EE230 HW1 SAMPLE SOLUTIONS

by Ogul Can YURDAKUL

Question 1

(a) $\Omega = \{H, TH, TTH, TTTH, ---\}$

Each outcome corresponds to either Ayse, Bora or Ceyda winning. If the number of tails in an outcome is equal to, for some $k \in \mathbb{N}$

* $3k$, Ayse wins.
* $3k+1$, Bora wins.
* $3k+2$, Ceyda wins.

Naturally, these three cases are mutually exclusive.

(b) $A = \{\omega \in \Omega \mid \text{the number of 'T's in } \omega \text{ is } 3k \text{ for some } k \in \mathbb{N}\}$

Ayse wins

$= \{H, TTTH, TTTTTTH, ---\}$

$B = \{\omega \in \Omega \mid \text{the number of 'T's in } \omega \text{ is } 3k+1 \text{ for some } k \in \mathbb{N}\}$

Bora wins

$= \{TH, TTTTH, TTTTTTTH, ---\}$

$A \cup B = \{\omega \in \Omega \mid \text{the number of 'T's in } \omega \text{ is } 3k \text{ or } 3k+1 \text{ for some } k \in \mathbb{N}\}$

Ayse or Bora wins

$= \{H, TH, TTTH, TTTTH, ---\}$

$(A \cup B)^c = \{\omega \in \Omega \mid \text{the number of 'T's in } \omega \text{ is NOT } 3k \text{ or } 3k+1 \text{ for any } k \in \mathbb{N}\}$

Ayse or Bora does not win

$\equiv$

Ceyda wins

$= \{\omega \in \Omega \mid \text{the number 'T's in } \omega \text{ is } 3k+2 \text{ for some } k \in \mathbb{N}\}$

$= \{TTH, TTTTTH, TTTTTTTTH, ---\}$

(c) For both P(A) and P(B) the calculation can be made via the sum of singleton event probabilities.

For a coin toss experiment, if the probability of getting a head is p, then that of getting a tail is $1-p$ for each trial.

$$P(A) = \underbrace{p}_{H} + \underbrace{(1-p)^3 p}_{TTTH} + \underbrace{(1-p)^6 p}_{TTTTTTH} + \cdots$$

$$= \sum_{k=0}^{\infty} p(1-p)^{3k}$$

$$= p \underbrace{\sum_{k=0}^{\infty} (1-p)^{3k}}$$ $\Rightarrow$ geometric series $\sum_{k=0}^{\infty} \alpha^k = \frac{1}{1-\alpha}$ for $|\alpha| < 1$. Here $\alpha = (1-p)^3 < 1$

$$= p \cdot \frac{1}{1-(1-p)^3}$$

$$P(B) = \underbrace{(1-p)p}_{TH} + \underbrace{(1-p)^4 p}_{TTTTH} + \underbrace{(1-p)^7 p}_{TTTTTTTH} + \cdots$$

$$= \sum_{k=0}^{\infty} p(1-p)^{3k+1} = p(1-p) \sum_{k=0}^{\infty} (1-p)^3$$

$$= p \cdot (1-p) \frac{1}{1-(1-p)^3}$$

(d) Events A & B are disjoint. Hence, by the additivity axiom

$$P(A \cup B) = P(A) + P(B) \tag{1}$$

$A \cup B$ & $(A \cup B)^c$ are also disjoint, whose union gives Ω. Therefore,

disjoint union symbol "$\sqcup$"

$$
\begin{aligned}
1 &= P(\Omega) && \text{(Normalization)} \\
&= P((A \cup B) \sqcup (A \cup B)^c) \\
&= P(A \cup B) + P((A \cup B)^c) && \text{(Additivity)} \\
&= P(A) + P(B) + P((A \cup B)^c) && \text{(By (1))}
\end{aligned}
$$

$$\Rightarrow \quad P((A\cup B)^c) = 1 - P(A) - P(B)$$

Then,

$$P(A\cup B) = \frac{p + p(1-p)}{1-(1-p)^3} = \frac{2p-p^2}{3p-3p^2+p^3}$$

$$P((A\cup B)^c) = 1 - P(A\cup B) = 1 - \frac{2p-p^2}{3p-3p^2+p^3}$$

$$= \frac{3p-3p^2+p^3-2p+p^2}{3p-3p^2+p^3} = \frac{p-2p^2+p^3}{3p-3p^2-p^3}$$

$$= \frac{p(1-p)^2}{1-(1-p)^3} \qquad (2)$$

rewrite the denominator as before

We can confirm this last result by computing $P((A\cup B)^c)$ using the second set description of $(A\cup B)^c$

$$P((A\cup B)^c) = \underbrace{(1-p)^2 p}_{TTH} + \underbrace{(1-p)^5 p}_{TTTTTH} + \underbrace{(1-p)^8 p}_{TTTTTTTTH} + \cdots$$

$$= \sum_{k=0}^{\infty} p(1-p)^{3k+2} = p(1-p)^2 \sum_{k=0}^{\infty} (1-p)^{3k}$$

$$= p(1-p)^2 \frac{1}{1-(1-p)^3} \qquad (3)$$

(2) & (3) match exactly.

(e) <u>For the events A & B given above</u>, we have $A\cap B = \emptyset$, so

$$P(A\cap B) = P(\emptyset) = 0 \qquad \text{(Proven in Class)}$$

We also have

$$\begin{aligned} 0 &\leq P((A \cup B)^c) && \text{(Non-negativity)} \\ &= 1 - P(A) - P(B) && \text{(Above)} \end{aligned}$$

$$\Rightarrow \quad P(A) + P(B) - 1 \leq 0$$

Combining the two results, we obtain

$$P(A) + P(B) - 1 \leq 0 = P(A \cap B)$$

However, this line of argumentation is not a proof for any two A & B as it relies on the extra assumption of $A \cap B = \emptyset$. Let's now make a proof for any $C, D \subseteq \Omega$.

<u>Claim:</u> For any $C, D \subseteq \Omega$, $P(C \cap D) \geq P(C) + P(D) - 1$

<u>Proof:</u> We will first show the following:

$$P(C \cup D) = P(C) + P(D) - P(C \cap D)$$

To do so, write

disjoint union: implies that the constituent sets have $\emptyset$ intersection

$$C \cup D = (C - D) \sqcup (C \cap D) \sqcup (D - C)$$

$$\begin{aligned} \Rightarrow P(C \cup D) &= P(C-D) + P(C \cap D) + P(D-C) && \text{(Additivity)} \\ &= \underbrace{P(C-D) + P(C \cap D)}_{=} + \underbrace{P(D-C) + P(C \cap D)}_{=} - P(C \cap D) \\ &= P(C) + P(D) - P(C \cap D) && \text{(Additivity, twice)} \end{aligned}$$

And so, because $\Omega \supseteq C \cup D$,

$$\begin{aligned} 1 &= P(\Omega) && \text{(Normalization)} \\ &= P(C \cup D) + \underbrace{P((C \cup D)^c)}_{\geq 0} && \text{(Additivity)} \\ & && \text{(Positivity)} \end{aligned}$$

$$\Rightarrow \quad 1 \geq P(C \cup D) = P(C) + P(D) - P(C \cap D)$$

$$\Rightarrow \quad P(C \cap D) \geq P(C) + P(D) - 1 \qquad \square$$

Question 2

(a) Because B is the event that the sampled number is divisible by 3 (A_3) or divisible by 5 (A_5), we have

$$B = A_3 \cup A_5$$

The given event is the set A_2. So,

$$P(B \mid A_2) = \frac{P(B \cap A_2)}{P(A_2)} = \frac{P((A_3 \cup A_5) \cap A_2)}{P(A_2)}$$

(b) We will calculate probabilities through finite set cardinalities.

$$|(A_3 \cup A_5) \cap A_2| = |(A_3 \cap A_2) \cup (A_5 \cap A_2)|$$
$$= \underbrace{|A_3 \cap A_2|}_{:= A_6} + \underbrace{|A_5 \cap A_2|}_{:= A_{10}} - \underbrace{|A_3 \cap A_5 \cap A_2|}_{:= A_{30}}$$

$$|A_6| = |\{6, \ldots, 96\}| = \frac{96-6}{6} + 1 = 16$$

$$|A_{10}| = |\{10, \ldots, 100\}| = \frac{100-10}{10} + 1 = 10$$

$$|A_{30}| = |\{30, 60, 90\}| = 3$$

$$|(A_3 \cup A_5) \cap A_2| = 16 + 10 - 3 = 23$$

$$|A_2| = |\{2, \ldots, 100\}| = \frac{100-2}{2} + 1 = 50$$

$$|\{1, \ldots, 100\}| = 100$$

$$\Rightarrow P(B \mid A_2) = \frac{23/100}{50/100} = \frac{23}{50} = 0.46$$

(c) Let a, b be prime numbers with $a, b \leq N$. Then

$$P(A_a \cup A_b \mid A_2)$$
$$= P(A_a \mid A_2) + P(A_b \mid A_2) - P(A_a \cap A_b \mid A_2)$$
$$= \frac{P(A_a \cap A_2)}{P(A_2)} + \frac{P(A_b \cap A_2)}{P(A_2)} - \frac{P(A_a \cap A_b \cap A_2)}{P(A_2)}$$
$$= \frac{P(A_a \cap A_2) + P(A_b \cap A_2) - P(A_a \cap A_b \cap A_2)}{P(A_2)}$$

Declaring $A_a \cap A_b = A_{ab}$ or $A_a \cap A_b \cap A_2 = A_{2ab}$ may not always be correct, as we may have $a = b$ or a or b equal to 2.

Question 3

Let,

- A_k denote the event where a family has k children for $k = 1, ..., 4$;
- C_y denote the event where the selected child is the youngest in the family, and
- C_o denote the event where the selected child is the oldest in the family.

Now that we have our events, we can talk of their probabilities. We will make often use of the Bayes' rule. Let us first calculate a useful probability that will come up often.

Because by the problem formulation, the events A_k, $k = 1, ..., 4$ form a partition of the sample space, we can use the Total Probability law to calculate $P(C_y)$, as follows:

$$P(C_y) = \sum_{k=1}^{k} \underbrace{P(C_y | A_k)}_{= \frac{1}{k}} \underbrace{P(A_k)}_{p_k} \quad \text{(Total Probability)}$$

p_k as given in the question.

$= \frac{1}{k}$ because all children are equally probable

$$= 1 p_1 + \frac{1}{2} p_2 + \frac{1}{3} p_3 + \frac{1}{4} p_4 = 0.5667$$

(a) $$P(A_1 | C_y) = \frac{\overbrace{P(C_y | A_1)}^{1} \overbrace{P(A_1)}^{p_1 = 0.25}}{\underbrace{P(C_y)}_{\text{above}}} = 0.4412$$

(b) $$P(A_3 \mid C_y) = \frac{\overset{\frac{1}{3}}{P(C_y \mid A_3)}\,\overset{p_3 = 0.2}{P(A_3)}}{\underset{\text{above}}{P(C_y)}} = 0.1176$$

(c) Because the children are chosen uniformly from any family, the probability that the selected child is the oldest in the family is equal to the probability of them being the youngest. In short, because $P(C_y \mid A_k) = P(C_o \mid A_k)$ for all $k = 1, \dots, 4$ we have $P(C_o) = P(C_y) = 0.5667$.

Question 4

(a) There are $\frac{5!}{3!\,2!}$ different outcomes where we have 2 w_q's & 3 w_p's,

$5!$ → total length

$3!$ → 3 repeated w_q

$2!$ → 2 repeated w_q

all of which have $p^3 q^2$ singleton probabilities. So, the given event has probability

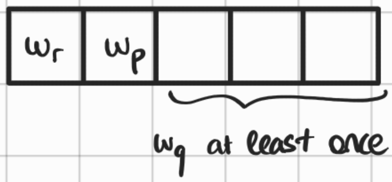
$$\frac{5!}{3!\,2!}\,p^3 q^2$$

(b) Solution 1: Assuming w_q must occur in 5 trials.

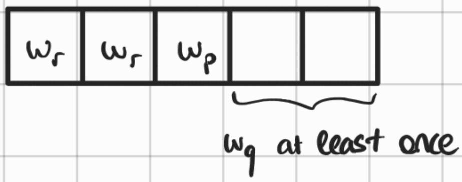
We can partition this event into 4 sets:

* A_1: w_p occurs on the first try & w_q occurs at least once in the following 4.

* A_2: w_r occurs on the first try, w_p the second, and w_q at least once in the following 3.



* A_3: w_r occurs twice first, w_p the third, and w_q at least once in the following 2.



* A_4: w_r occurs thrice first, w_p the fourth, and w_q the fifth.

Then, because $A_i \cap A_j = \phi$ whenever $i \neq j$, we have

$$P(A) = P\left(\bigsqcup_{i=1}^{4} A_i\right) = \sum_{i=1}^{4} P(A_i)$$

$$= p\left(1-(1-q)^4\right) + r\,p\,\underbrace{\left(1-(1-q)^3\right)}_{\to P\left(\substack{w_q \text{ occurs at} \\ \text{least once}}\right) = 1 - P\left(\substack{w_q \text{ never} \\ \text{occurs}}\right)}$$

$$+ r^2 p\left(1-(1-q)^2\right) + r^3 p\,\underbrace{\left(1-(1-q)\right)}_{=q}$$

Solution 2: Assuming ω_q need not occur in these 5 trials. Under this assumption, the solution is quite similar to the previous one, with an additional (singleton) partition set added to our overall event set. The new partition is as follows:

* A_1 :

ω_p				

ω_q at least once

Same as before

* A_2 :

ω_r	ω_p			

ω_q at least once

* A_3 :

ω_r	ω_r	ω_p		

ω_q at least once

* A_4 :

ω_r	ω_r	ω_r	ω_p	ω_q

* A_5: ω_q did not occur at all.

ω_r	ω_r	ω_r	ω_r	ω_p

Then we write

$$P(A) = P\left(\bigsqcup_{i=1}^{5} A_i\right) = \sum_{i=1}^{5} P(A_i)$$

$$= p\left(1-(1-q)^4\right) + rp\left(1-(1-q)^3\right) + r^2p\left(1-(1-q)^2\right) + r^3p\left(1-(1-q)\right) + \underbrace{r^4 p}_{P(A_5)}$$

(c) <u>Solution 1</u>: Using series of increasing events.

Let $A^{(N)}$ denote that w_p occurs before w_p in the N first trials. Then, $A^{(N)}$ can be partitioned in the same way presented above, which gives

$$P\left(A^{(N)}\right) = \sum_{k=1}^{N} \underbrace{\left(1-(1-q)^{N-k}\right)}_{w_q \text{ occurs at least once in the following } N-n \text{ trials}} \; \underbrace{p}_{w_p \text{ occurs on the } k\text{-th trial}} \; \underbrace{r^{k-1}}_{w_r \text{ occurs } k-1 \text{ times first}}$$

Now call A the event where w_p occurs before w_q in an infinite number of trials. Then, the event A is the disjoint union of the events $A^{(N)}$, meaning $A^{(N)} \to A$ as $N \to \infty$. Then

$$P(A) = \lim_{N\to\infty} P\left(A^{(N)}\right) = \lim_{N\to\infty} \sum_{k=1}^{N} r^{k-1} p \left(1-(1-q)^{N-k}\right)$$

$$= p \lim_{N\to\infty} \sum_{k=1}^{N} r^{k-1} - \sum_{k=1}^{N} \underline{r^{k-1}}\,\underline{(1-q)^{N-k}}\,\underline{(1-q)}(1-q)^{-1}$$

$$= p \lim_{N\to\infty} \sum_{k=0}^{N} r^{k} - (1-q)^{N-1} \sum_{k=0}^{N-1} \left(\frac{r}{1-q}\right)^{N-1} \quad 1-q = p+r,\ \frac{r}{p+r} < 1$$

$$= p \lim_{N\to\infty} \frac{1-r^N}{1-r} - (1-q)^{N-1} \frac{1-\left(\frac{r}{1-q}\right)^{N}}{1-\frac{r}{1-q}}$$

because $r<1$ & $1-q<1$

$$= p \lim_{N\to\infty} \frac{1-\cancelto{0}{r^N}}{1-r} - \frac{\cancelto{0}{(1-q)^{N-1}} - \cancelto{0}{r^N}(1-q)^{-1}}{1-\frac{r}{1-q}} \qquad \text{taking the limit}$$

$$= p \cdot \frac{1}{1-r} = \frac{p}{1-r} = \frac{p}{p+q}$$

Solution 2: Partitioning the event.

Partition the event A into events A_n, where A_n is the event where w_r occurs in the first $n-1$ trials & w_p occurs in the last trial. As long as this happens, we are sure that w_q will happen eventually, because the probability of w_q NOT occurring in an infinite number of trials is zero: $(1-q)^N \to 0$ as $N \to \infty$.

Then, the A_n's partition A, and so

$$A = \bigsqcup_{n=1}^{\infty} A_n \quad \text{meaning} \quad P(A) = \sum_{n=1}^{\infty} P(A_n)$$

geometric series with $r<1$

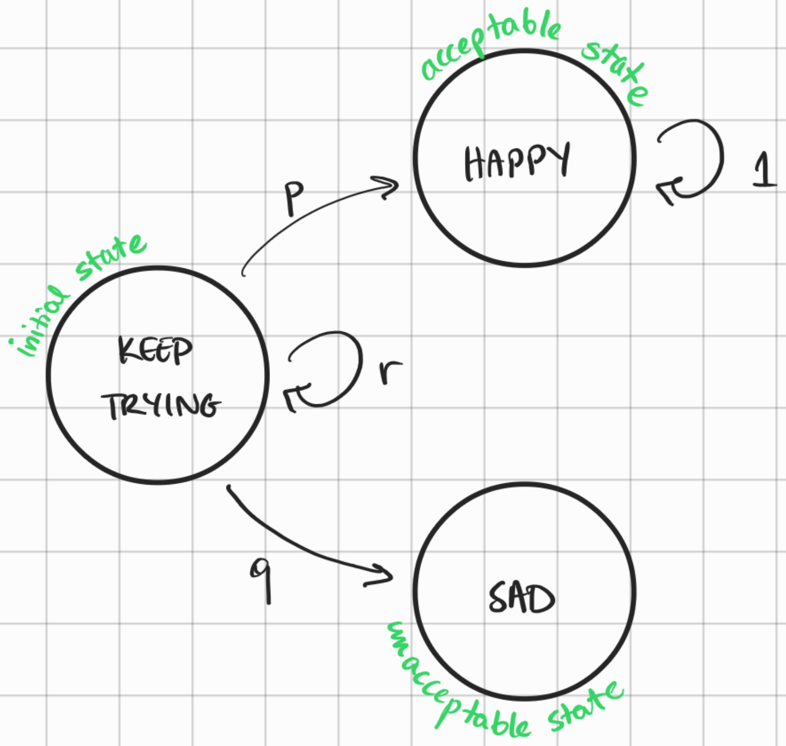
$$P(A_n) = r^{n-1} p \implies P(A) = \sum_{n=1}^{\infty} p r^{n-1} = p \sum_{n=0}^{\infty} r^n = p \cdot \frac{1}{1-r}$$

$$= \frac{p}{1-r} = \frac{p}{p+q}$$

Solution 3: Noticing the recursion

Let's say we started performing the infinite number of trials. Call this state as KEEP TRYING. If w_r occurs when we are in KEEP TRYING, we remain there. If w_q occurs when we KEEP TRYING, that experiment is not in our desired event set, so that run is rejected and we become SAD (another state). No matter what happens when we are SAD, we remain there. If, finally, w_p occurs, we become HAPPY (another state). We are HAPPY because as stated previously, after we observe w_p, we will definitely observe w_q eventually.

Below is a diagram depicting this story. The transition probabilities are indicated on the arrows.



We want to find the probability of transitioning from KEEP TRYING to HAPPY, call this probability $P(KT \to H)$. At each trial we remain at KEEP TRYING with probability r, after which the desired $KT \to H$ probability is again $P(KT \to H)$. There is also p probability of actually making the $KT \to H$ transition. So, we have

$$P(KT \to H) = P(KT \to H) \cdot r + p$$

Solving for $P(KT \to H)$, we obtain

$$P(KT \to H) = \frac{p}{1-r} = \frac{p}{p+q}$$

EE230 HW1 MATLAB Question: Parameter Estimation

Name, Surname

ID No.

Table of Contents

Task: Estimating the probability of success from Bernoulli trials. 1
  Part I. Sampling 1
  Part II. Estimation 2
  Part III. Evaluation 2
Appendix A. Function Declarations 5
  A.1. Estimator 5

Task: Estimating the probability of success from Bernoulli trials.

The Bernoulli random variable is a mathematical abstraction for a possibly unfair coin toss, expressed as follows:

$$X \in \{0, 1\} \qquad \mathrm{P}(X = 1) = p,\ \mathrm{P}(X = 0) = 1 - p,\ \ p \in [0, 1]$$

The $X = 1$ case is called a *success*, and consequently $X = 0$ is called as *failure*. In this work, you will be trying to estimate the probability of success p of a Bernoulli random variable from `N` independent & identically distributed (iid.) samples.

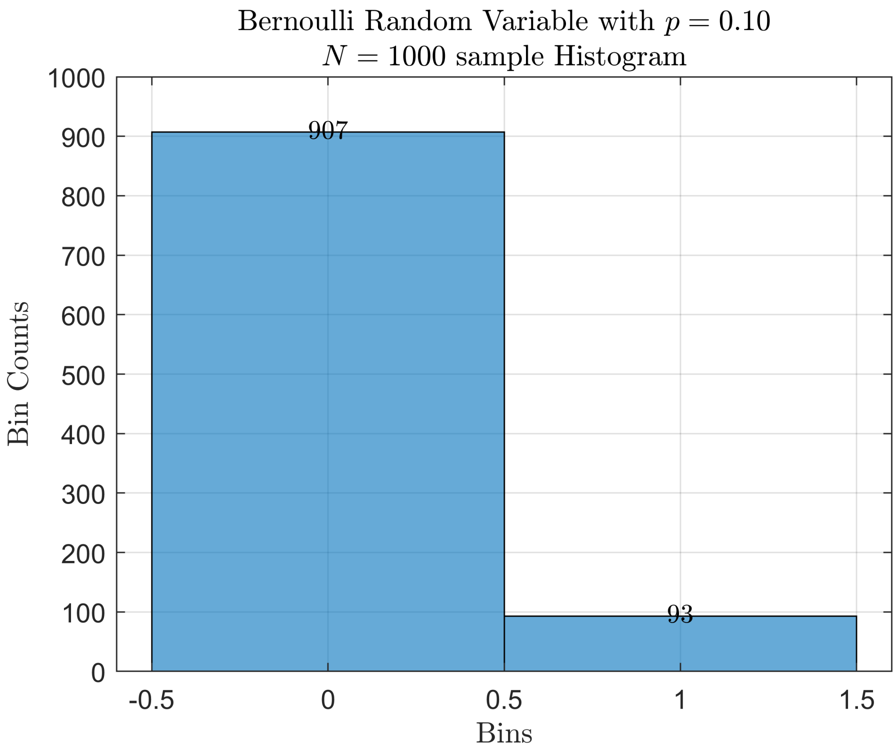
Part I. Sampling

I first start by sampling from a Bernoulli random variable, using the `randsample()` function.

```
N = 1000;
p = 0.1;

samples = randsample([0 1], N, true, [1-p p]);
```

I proceed by visualizing my samples as a histogram. As the distribution I have is discrete, I expect the relative bin counts in the histogram to reflect the probabilities of success and of failure, provided that I have enough samples.

```
figure

h = histogram(samples);
title({['Bernoulli Random Variable with $p = ' num2str(p, '%.2f') '$'], ...
    ['$N = ' num2str(N) '$ sample Histogram']})
ylim([0 N])
ylabel('Bin Counts')
xlabel('Bins')

% Some optional decoration: Write the count above the bars
text(0, h.BinCounts(1), num2str(h.BinCounts(1)), ...
    HorizontalAlignment="center")
text(1, h.BinCounts(2), num2str(h.BinCounts(2)), ...
    HorizontalAlignment="center")
```



Part II. Estimation

Now that I have some samples from the distribution, I can code the intuitive estimator for the probabilit of success, by the provided formula.

```
p_est = sum(samples) / N
```

```
p_est = 0.1000
```

To be able to call this estimator more easily, I turn it into a function, which you can find in Appendix A.1. I call it again below to make sure it gives the same result as above.

```
p_est = estimator(samples)
```

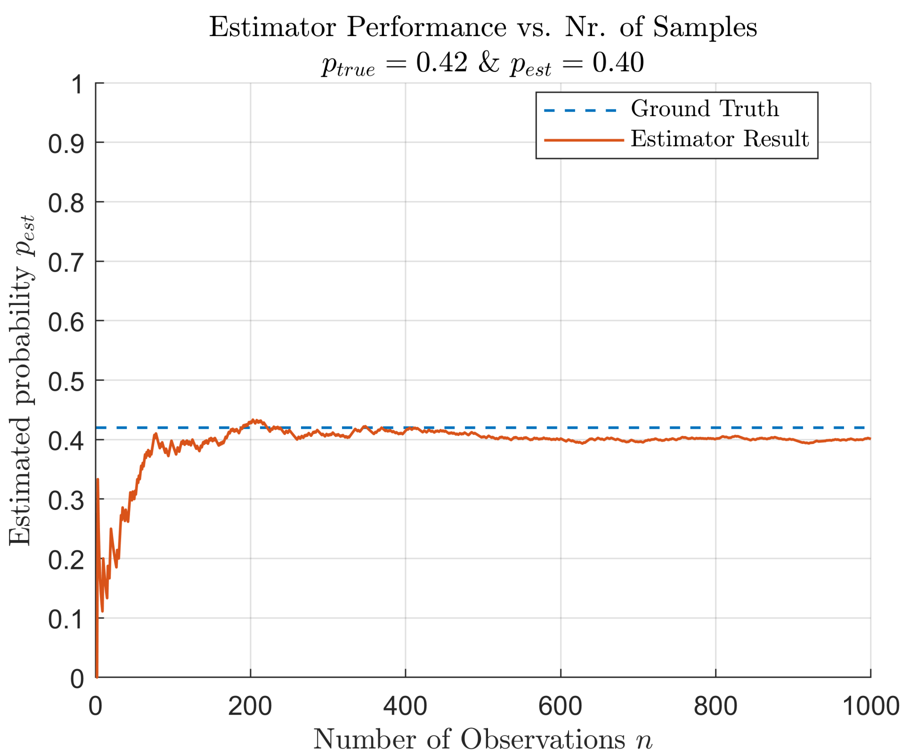
```
p_est = 0.1000
```

Part III. Evaluation

In this section I will be evaluating the performance of the estimator provided in the previous section. To do so, I generate a sufficiently large data set of $N = 1000$ samples, and I run my estimator in a for loop with the first n samples as n varies from 1 to N. This allows me to observe the estimator behaviour as a function of used samples, which I plot in the following figure.

```
p = 0.42;
N = 1000;

data = randsample([0 1], N, true, [1-p p]);

p_est_ = zeros(size(data));
```

```
for n = 1:N

    p_est_(n) = estimator(data(1:n));

end

figure
hold on
plot(1:N, p*ones(size(data)), '--', LineWidth=1)
plot(1:N, p_est_, LineWidth=1)
legend('Ground Truth', 'Estimator Result', ...
    Location='best', Interpreter='latex')
ylim([0 1])
title({'Estimator Performance vs. Nr. of Samples', ...
    ['$p_{true} = ' num2str(p, '%.2f') ...
    '$ \& $p_{est} = ' num2str(p_est_(end), '%.2f') '$']})
ylabel('Estimated probability $p_{est}$'); xlabel('Number of Observations $n$')
```

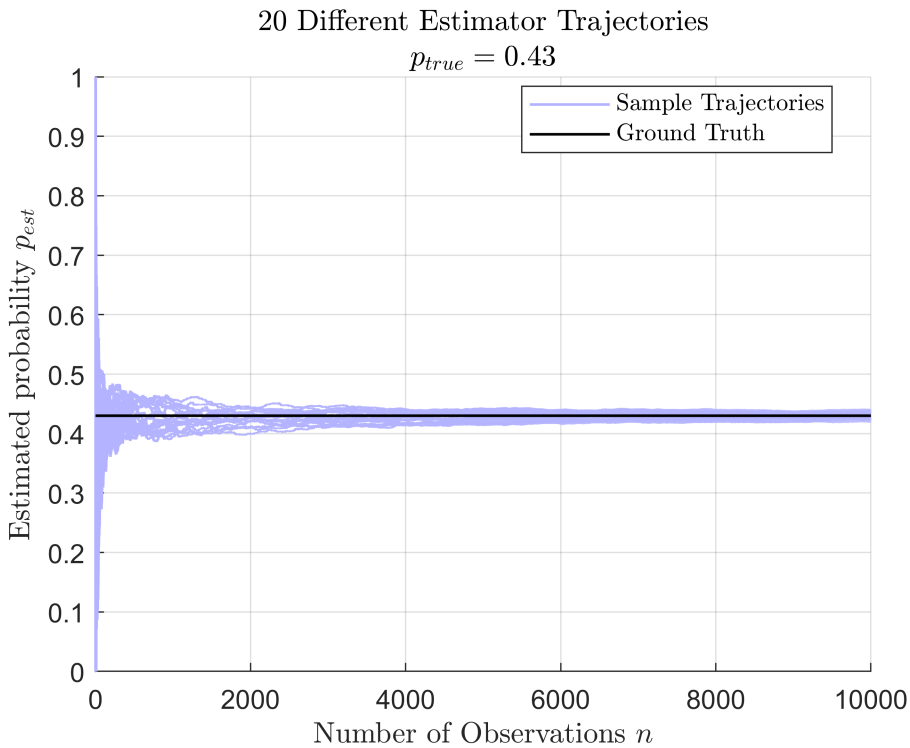


It is clear that the estimator eventually converges to the true value of p, although it may fluctuate and make faulty estimation when the observed number of samples is low. It is also visible that the trajectory is much more volatile/varied for low number of observations, which decreases as more observations are processed.

Different runs of this estimator with different data produces different estimator trajectories as a function of observed data, although all of these trajectories seem to converge to the true value eventually. Taking more and more samples should make the estimate more and more accurate.

Let's now visualize multiple estimator trajectories as they vary over different data sets.

```
figure
hold on
ylim([0 1])

K = 20; % Number of trajectories to be plotted

labels = cell(K,1); % To set the legend more easily

p = 0.43; % Keep these declarations out of the loop for efficiency
N = 10000;

for k = 1:K % Loop over trajectories

    labels{k} = '';

    data = randsample([0 1], N, true, [1-p p]);

    p_est_ = zeros(size(data));

    for n = 1:N % Loop over the n initial samples

        p_est_(n) = sum(data(1:n)) / n;

    end

    plot(1:N, p_est_, 'b', Color=[0.7 0.7 1], LineWidth=1);

end

labels{1} = 'Sample Trajectories';
labels{K+1} = 'Ground Truth';

plot(1:N, p*ones(size(data)), 'k', LineWidth=1)

title({[num2str(K) ' Different Estimator Trajectories'], ...
    ['$p_{true} = ' num2str(p, '%.2f') '$']})
legend(labels, ...
    Interpreter='latex', Location='best')
ylabel('Estimated probability $p_{est}$'); xlabel('Number of Observations $n$')
```



Appendix A. Function Declarations

A.1. Estimator

```
function [p_est] = estimator(samples)

    p_est = sum(samples) / length(samples);

end
```