PROJECT 3 - SOME INTERESTING DRV's

Name - Tamoghna Chattopadhyay
STUD ID - 8541324935

PROBLEM STATEMENTS

QUES 1 : Sum of Uniform RV's

Define : $N = \text{Min} \left\{ n : \sum_{i=1}^{n} U_i > 1 \right\}$

where $\{U_i\}$ are iid uniform (0,1) RVs
Find (by simullation) : $\hat{m} = E[N]$ an estimator for mean
Can you guess (or derive) the true value for $E[N]$?

QUES 2 : Minima of uniform RV's

Define : $N = \text{Min} \{ n : U_1 \leq U_2 \leq \cdots \leq U_{n-1} > U_n \}$
i.e., the n^{th} term is the first that is less than its predecessor, where $\{U_i\}$ are independent identically distributed (iid) Uniform (0,1) RV's.
Find (by simulation) : $\hat{m} = E[N]$ an estimator for mean
Can you guess (or derive) the true value for $E[N]$?

QUES 3 : Maxima of Uniform RV's

Consider the sequence of iid uniform RV's $\{U_i\}$. If $U_j > \max_{i=1, j-1} \{U_i\}$, we say U_j is a record.

Let X_i be an RV for the distance from the $i-1^{st}$ record to the i^{th} record. Clearly, $X_1 = 1$ always.

Distribution of records : Using simulation, obtain (and graph) a probability histogram for X_2 and X_3 and compute the sample means.
Can you find an analytical expression for $P(X_2 = k)$?
What does this say about $E[X_2]$?

QUES 1: Sum of Uniform Random Variables

1.1 Introduction

$$N = \min \{ n : \sum_{i=1}^{n} U_i > 1 \}$$

The given RV is uniformly distributed in the interval (0,1). Instead of 1, in the above equation, consider it to be generalized for the sum to be greater than x. Let $f(x) = E[N]$. A recursion is derived by conditioning on U_1.

$$f(x) = \int_0^1 (E[N] \mid U_1 = a)\, da$$

The integral can also be written as:

$$f(x) = \int_0^x (E[N] \mid U_1 = a)\, da + \int_x^1 (E[N] \mid U_1 = a)\, da$$

If the value of U_1 is greater than x, as in the second integral, the sum already has exceeded x, which implies that $E[N] = 1$

If $U_1 < x$, $(E[N] \mid U_1 = a) = 1 + f(x-a)$

Using the above equation, the integral can be written as,

$$f(x) = 1 + \int_0^x f(x-a)\, da$$

$$= 1 + \int_0^x f(a)\, da$$

And can be solved using a differential equation the solution of which is given by,

$$\frac{d}{dx} f(x) = f(x) \text{ for } 0 \le x < 1$$

$f(x) = ce^{x}$; $x \to 0$; $f(x) \to 0$.
and the conditions $= e^{x}$ and $x = 1$; $f(x) = e$

The distribution can also be determined by :

$$P_{x_1 + x_2 + \cdots x_n}(u) = \iiint \cdots \int \delta(x_1 + x_2 + \cdots x_n - u)\, dx_1 dx_2 \cdots dx_n$$

The probability that the sum of n RV's > 1 and $n-1 < 1$ is given as

$$P_n^{[1]} = \int_1^n P(x_1 + x_2 + \cdots + x_n) u \, du - \int_1^{n-1} P(x_1 + x_2 + \cdots + x_n - 1) u \, du$$

$$\therefore P_n^{[1]} = (1 - 1/n!) - (1 - 1/(n-1)!)$$

$$= 1/(n(n-2)!)$$

So,

$$E[N] = \sum_{n=1}^{\infty} n P_n$$

$$= \sum_{n=1}^{\infty} \frac{1}{(n-2)!}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!}$$

$$= e$$

$$\boxed{\therefore E[N] = e}$$

N = 100
n = 10
E =

1.2 SIMULATION METHODOLOGY

A sequence of N uniformly distributed RV's is generated using the rand function in MATLAB, where N is a very large value taken as input from user. The URV's are denoted by U_i. The U_i's are summed from 1 to the value i, and compared to 1. The value of i, is exactly the min value of n. The process is repeated for N runs to obtain better results or greater accuracy. To finally get the expected value, E[N], we consider the mean of the N iterations performed.

1.3 RESULTS AND OBSERVATIONS

The results of the simulations are shown below. It is noticed that, with increasing the value of N, the expected value which is observed by the simulation is closer to the expected value (e) determined analytically. The graph also shows a comparison between the expected values for increasing number of iterations.

Thus, by simply adding random variables between 0 and 1 till a threshold, we have estimated a fundamental constant in Maths.

N = 1000
n = 100
E = 2.6930

N = 10000
n = 100
E = 2.7255

N = 100000
n = 100
E = 2.7176

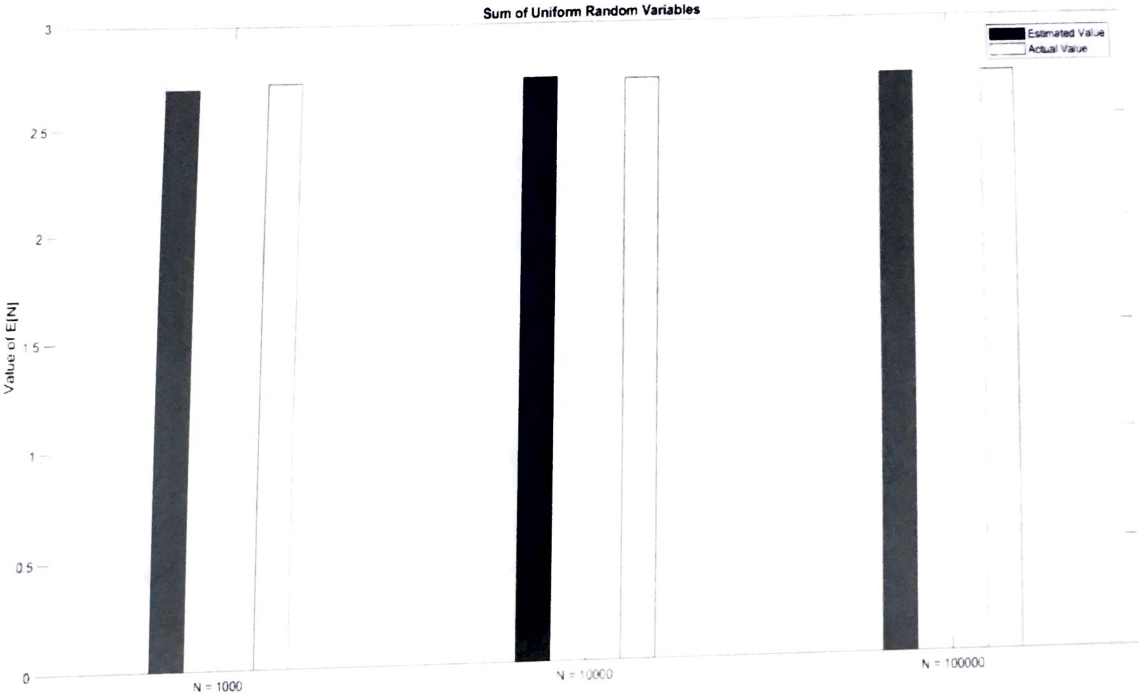


Figure: Expected Value vs e

1.4 MATLAB CODE

```
% Sum of Uniform Random Varibles

N = 1000;  % Large Sequence
n = 100;
K = zeros(1,N);

for x = 1:N

    % Generate uniformly distribted random numbers
    u = rand(1,n);
    i = 1;

    % Check whether the sum of Ui and Ui+1 <=1
    while sum(u(1:i))<=1
        i = i+1;
    end
```

```
    % Obtaining the minimum of the sum
    K(x) = i;
end

E1 = mean(K)

N = 10000;   % Large Sequence
n = 100;
K = zeros(1,N);

for x = 1:N

    % Generate uniformly distribted random numbers
    u = rand(1,n);
    i = 1;

    % Check ehether the sum of Ui and Ui+1 <=1
    while sum(u(1:i))<=1
        i = i+1;
    end

    % Obtaining the minimum of the sum
    K(x) = i;
end

E2 = mean(K)

N = 100000;   % Large Sequence
n = 100;
K = zeros(1,N);

for x = 1:N

    % Generate uniformly distribted random numbers
    u = rand(1,n);
    i = 1;

    % Check ehether the sum of Ui and Ui+1 <=1
    while sum(u(1:i))<=1
        i = i+1;
    end

    % Obtaining the minimum of the sum
    K(x) = i;
end

E3 = mean(K)

e = [E1 exp(1); E2 exp(1); E3 exp(1)]
bar(e,0.33)
title('Sum of Uniform Random Variables');
ylabel('Value of E[N]');
legend('Estimated Value','Actual Value');
set(gca, 'XTick', 1:3, 'XTickLabel', {'N = 1000','N = 10000', 'N = 100000'})
```

QUES 2: <u>Minima of Uniform Random Variables</u>

2.1 <u>INTRODUCTION</u>

$$N = \min \{ n : U_1 \leq U_2 \leq \dots \leq U_{n-1} > U_n \}$$

A sequence, U_i is generated over N repetitions and compared with the next value in the sequence, i.e., U_{i+1}, until $U_{i+1} > U_i$. This process is repeated over N runs too and every time the criterion is satisfied, $i = i+1$ is recorded. Taking the mean of these values gives us the expected value of the expression.

The evaluation of the mean of this term is similar to Ques (1). So it can easily be said that analytically, the expected value $E[N] = e$.

Consider the probability distribution function for the given expression from $U_1 > U_2$ to $U_1 \leq U_2 \leq \dots \leq U_{n-1} > U_n$. Now, checking every element in the distribution, we have,

$$P_{U1 > U2} = \frac{1}{2!}$$

$$P_{U1 <= U2 > U3} = \frac{1}{3!} * 2$$

$$P_{U_1 \leq U_2 \leq U_3 > U_4} = \frac{1}{4!} \times 3$$

Generalizing,

$$P_{U1 \leq U2 \leq \dots \leq U_{n-1} > U_n} = \frac{1}{n!} * (n-1)$$

So,

$$E[N] = \sum_{n=1}^{\infty} n \, P_{U_1 \le U_2 \le \ldots \le U_{n-1} > U_n}$$

$$= \sum_{n=1}^{\infty} \frac{n}{n(n-2)!}$$

$$= \sum_{n=1}^{\infty} \frac{1}{(n-2)!} = \sum_{n=0}^{\infty} \frac{1}{n!} = e$$

2.2 SIMULATION METHODOLOGY

A sequence of N uniformly distributed RV's is generated using the rand function in MATLAB. N is a very large value. The URV's are denoted by U_i. The U_i's are compared $U_i < U_{i+1}$. The value of i+1 is recorded. The process is repeated for N runs to obtain better results or greater accuracy. To finally get the expected value, E[N], we consider the mean of the N iterations performed.

2.3 RESULTS AND OBSERVATIONS

The results of the simulation are shown below. It is noticed that, with increasing the value of N, the expected value which is observed by simulation is closer to the expected value (e) determined analytically. The graph also shows a comparison between the expected values for increasing number of its iterations.

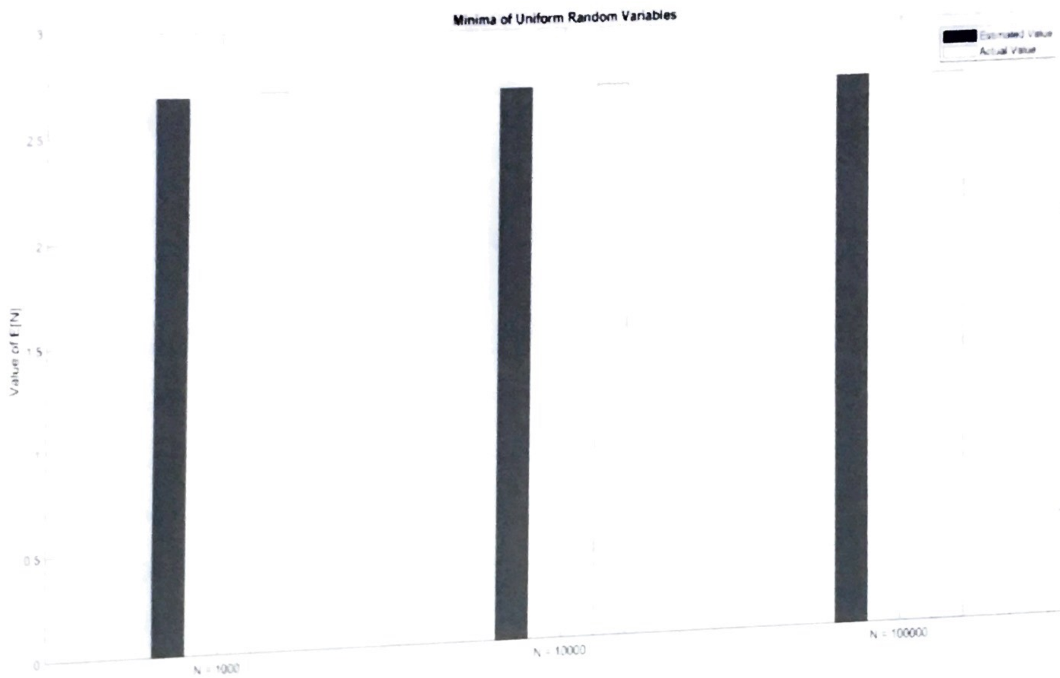


Figure: Expected Value vs e

N = 1000
n = 100
E = 2.7110

N = 10000
n = 100
E = 2.7109

N = 100000
n = 100
E = 2.7181

2.4 MATLAB CODE

```
% Minima of Uniform Random Varibles

N = 1000;   % Large Sequence
n = 100;
K = zeros(1,N);

for x = 1:N
    % Generate uniformly distribted random numbers
    u = rand(1,n);
    i = 1;

    % Check whether Ui<Ui+1
    while u(i)<u(i+1)
        i = i+1;
    end

    % Obtaining the minimum of the sum
```

```
    K(x) = i+1;
end

E1 = mean(K);

% Minima of Uniform Random Varibles

N = 10000;  % Large Sequence
n = 100;
K = zeros(1,N);

for x = 1:N

    % Generate uniformly distribted random numbers
    u = rand(1,n);
    i = 1;

    % Check whether Ui<Ui+1
    while u(i)<u(i+1)
        i = i+1;
    end

    % Obtaining the minimum of the sum
    K(x) = i+1;
end

E2 = mean(K);

% Minima of Uniform Random Varibles

N = 100000;  % Large Sequence
n = 100;
K = zeros(1,N);

for x = 1:N

    % Generate uniformly distribted random numbers
    u = rand(1,n);
    i = 1;

    % Check whether Ui<Ui+1
    while u(i)<u(i+1)
        i = i+1;
    end

    % Obtaining the minimum of the sum
    K(x) = i+1;
end

E3 = mean(K);

e = [E1 exp(1); E2 exp(1); E3 exp(1)]
bar(e,0.33)
title('Minima of Uniform Random Variables');
ylabel('Value of E[N]');
legend('Estimated Value','Actual Value');
set(gca, 'XTick', 1:3, 'XTickLabel', {'N = 1000','N = 10000', 'N = 100000'})
```

QUES 3: Maxima of Uniform Random Variables

3.1 INTRODUCTION

The process of finding records can be simulated for a large number of samples of uniform random variables. The records are then stored as U_y. The random numbers generated are then compared and the maximum is identified. The numbers (in U_y) are then compared with the maximum, and if they are more than the maximum, the records are updated, if not the same procedure is repeated.

We know that in the example considered, $X_1 = 1$ and $X_2 = 1$. The probability densities of $X_1, X_2, \cdots, X_n$ can be determined and they are given by:

$$P(X_2 = 1) = \frac{1}{2} \qquad P(X_2 = 3) = \frac{1}{12}$$

$$P(X_2 = 2) = \frac{1}{6} \qquad P(X_2 = 4) = \frac{1}{20}$$

and so on.

This can be generalized to

$$P(X_2 = k) = \frac{1}{k(k+1)}$$

The expected value E, can be given as

$$E[X] = \sum_{k=1}^{\infty} k\, P(X_2 = k)$$

$$= \sum_{k=1}^{\infty} \frac{1}{k+1} = \infty$$

Thus, the expected value cannot be determined because it is divergent.

3.2 SIMULATION METHODOLOGY

The trials are conducted for a fairly large sequence N and equivalent uniformly distributed random numbers are generated. The random numbers are then compared to the next number in the sequence.

Initially, the variable max is set to 0. Using max < u(y), the maximum of the sequence is identified, and assigned as a value to the variable max. This process is repeated for n trials and the records are updated for each new value of max.

$K(x)$ = record (2) - record(1) gives the expected value of X_2
Similarly,
$K(x)$ = record (3) - record (2) gives the expected value of X_3

Finally, a probability histogram is calculated.

3.3 RESULTS AND OBSERVATION

The results obtained as a graph and the values of the probabilities are as shown below.

N = 100000

n = 100

a/sum(a)

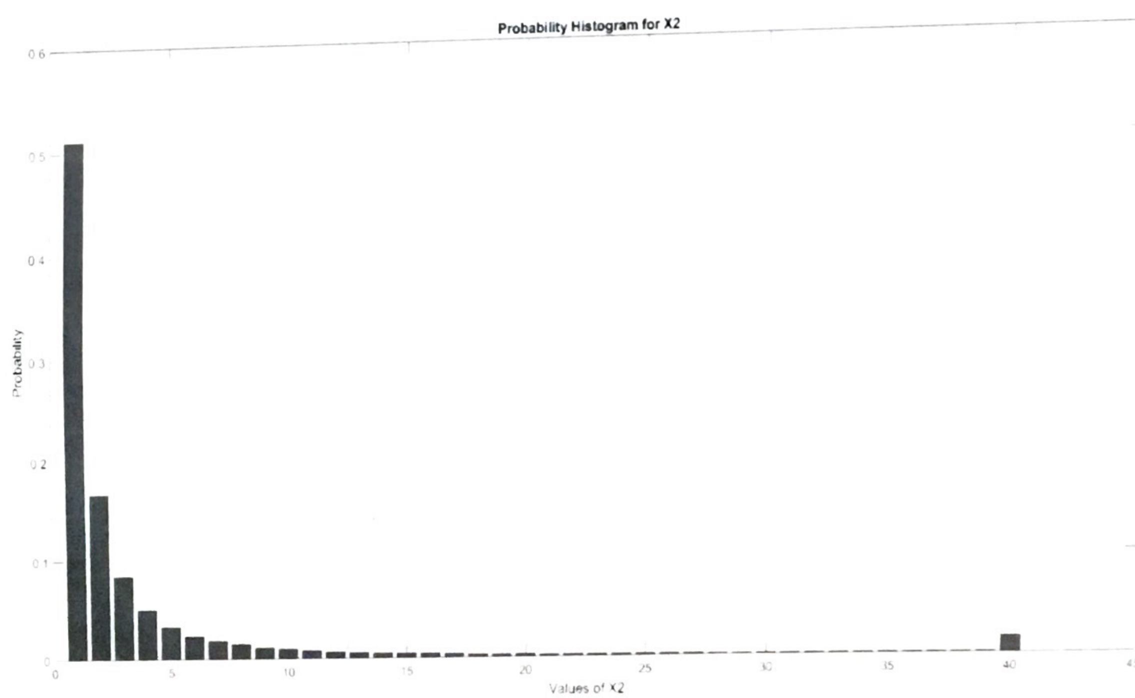
ans =

0.5084	0.1688	0.0841	0.0492
0.0328	0.0252	0.0174	0.0143
0.0109	0.0089	0.0075	0.0062
0.0053	0.0047	0.0044	0.0036
0.0030	0.0030	0.0028	0.0024
0.0020	0.0021	0.0017	0.0018
0.0014	0.0015	0.0013	0.0013
0.0010	0.0011	0.0009	0.0009
0.0010	0.0008	0.0008	0.0007
0.0006	0.0006	0.0006	0.0148

Value for mean of K and K(x)

mean(K)
ans = 4.1665

mean(K(x))
ans = 1



N = 100000
n = 100

a/sum(a)

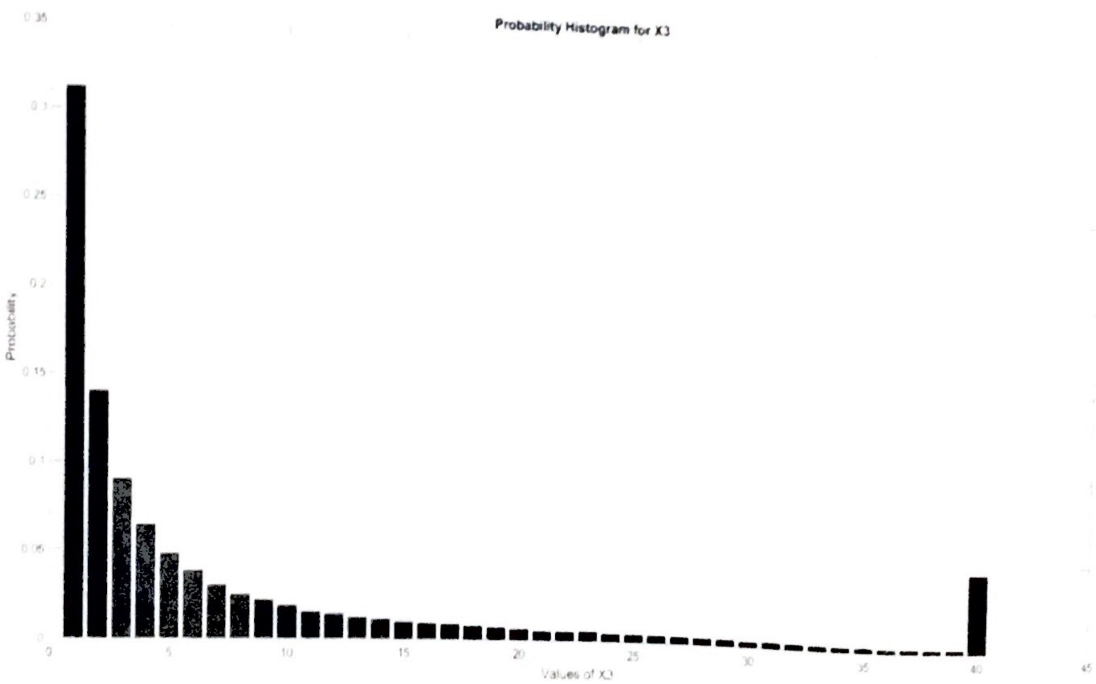
ans =

0.3106	0.1387	0.0898	0.0646
0.0488	0.0377	0.0312	0.0255
0.0211	0.0188	0.0162	0.0139
0.0117	0.0104	0.0100	0.0090
0.0078	0.0073	0.0066	0.0060
0.0058	0.0053	0.0053	0.0046
0.0043	0.0042	0.0035	0.0036
0.0029	0.0030	0.0030	0.0027
0.0027	0.0027	0.0024	0.0024
0.0019	0.0024	0.0020	0.0492

Value for mean of K and K(x)

mean(K)
ans = 7.4793

mean(K(x))
ans = 6



3.4 MATLAB CODE

```
% Maxima of Uniform Random Varibles

% Large Sequence
N = input('Enter N: ')
n = input('Enter n: ')
K1 = zeros(1,N);
K2 = zeros(1,N);

% Generating Uniformly distributed random numbers
for x=1:N

    u= rand(1,n);

    % Recording maximum value
    record=zeros(1,20);
    max=0;
    j=1;

    for y=1:n

        % Recording maximum value
        record(1)=1;
        if max < u(y)
            max = u(y);
            record(j)=y; j=j+1;
        end
    end

    % Values for X2 and X3
    K1(x)=record(2)-record(1);
    K2(x)=record(3)-record(2);
end

% mean(K(x))
t=1:1:40;
[a,b] = hist(K1,t)
bar(b, a/sum(a))
title('Probability Histogram for X2')
ylabel('Probability')
xlabel(' Values of X2')

figure;
[a,b] = hist(K2,t)
bar(b, a/sum(a))
title('Probability Histogram for X3')
ylabel('Probability')
xlabel(' Values of X3')
```

The histogram for the distribution of X_2 is interesting and looks like a geometric distribution, which can be further verified by checking the values of a/sum(a).

We can find that

$$P(X_2 = 1) \approx \frac{1}{2} = \frac{1}{1 \times 2}$$

$$P(X_2 = 2) \approx \frac{1}{6} = \frac{1}{2 \times 3}$$

$$P(X_2 = 3) \approx \frac{1}{12} = \frac{1}{3 \times 4}$$

Thus, generalizing, $P(X_2 = k) \approx \dfrac{1}{k(k+1)}$

And $$E[X_2] = \sum_{k=1}^{\infty} k\,(P(X_2 = k))$$

$$= \sum_{k=1}^{\infty} \frac{1}{k+1}$$

$$= \infty$$