1. Packet Routing

Packets arriving at a switch are routed to either destination $A$ (with probability $p$) or destination $B$ (with probability $1 - p$). The destination of each packet is chosen independently of each other. In the time interval $[0, 1]$, the number of arriving packets is Poisson($\lambda$).

(a) Show that the number of packets routed to $A$ is Poisson distributed. With what parameter?

(b) Are the number of packets routed to $A$ and to $B$ independent?

**Solution:**

(a) Let $X, Y$ be random variables which are equal to the number of packets routed to the destinations $A, B$ respectively. Let $Z = X + Y$. We are given that $Z \sim$ Poisson($\lambda$). We prove that $X$ has the Poisson distribution with mean $p\lambda$.

$$
\mathbb{P}(X = x) = \sum_{z=x}^{\infty} \mathbb{P}(X = x, Z = z)
= \sum_{z=x}^{\infty} \mathbb{P}(Z = z)\mathbb{P}(X = x \mid Z = z)
= \sum_{z=x}^{\infty} e^{-\lambda} \frac{\lambda^z}{z!} \binom{z}{x} p^x (1 - p)^{z-x}
= e^{-\lambda} \sum_{z=x}^{\infty} \frac{\lambda^z}{z!} \frac{z!}{x!(z-x)!} p^x (1 - p)^{z-x}
= e^{-\lambda} (\lambda p)^x \sum_{z=x}^{\infty} \frac{\lambda(1 - p)}{(z-x)!} (1 - p)^{z-x}
= e^{-\lambda} (\lambda p)^x \frac{x!}{x!} e^{\lambda(1-p)}
= e^{-\lambda p} (\lambda p)^x.
$$
(b) We prove that $X$ and $Y$ are independent.

\[\mathbb{P}(X = x, Y = y) = \sum_{z=0}^{\infty} \mathbb{P}(X = x, Y = y, Z = z)\]

\[= \sum_{z=0}^{\infty} \mathbb{P}(X = x, Y = y \mid Z = z)\mathbb{P}(Z = z)\]

\[= \mathbb{P}(X = x, Y = y \mid Z = x + y)\mathbb{P}(Z = x + y)\]

\[= \frac{(x+y)!}{x!y!} \left( \frac{e^{-\lambda} \lambda^{x+y}}{(x+y)!} \right) p^x (1-p)^y\]

\[= \frac{e^{-\lambda p (\lambda p)^x}}{x!} \cdot \frac{e^{-\lambda(1-p)(\lambda(1-p))^{y}}}{y!}\]

\[= \mathbb{P}(X = x)\mathbb{P}(Y = y).\]

2. Compact Arrays

Consider an array of $n$ entries, where $n$ is a positive integer. Each entry is chosen uniformly randomly from \{0, \ldots, 9\}. We want to make the array more compact, by putting all of the non-zero entries together at the front of the array. As an example, suppose we have the array

\[\begin{bmatrix} 6, 4, 0, 0, 5, 3, 0, 5, 1, 3 \end{bmatrix}.\]

After making the array compact, it now looks like

\[\begin{bmatrix} 6, 4, 5, 3, 5, 1, 3, 0, 0, 0 \end{bmatrix}.\]

Let $i$ be a fixed positive integer in \{1, \ldots, n\}. Suppose that the $i$th entry of the array is non-zero (assume that the array is indexed starting from 1). Let $X$ be a random variable which is equal to the index that the $i$th entry has been moved after making the array compact. Calculate $\mathbb{E}[X]$ and $\text{var}(X)$.

**Solution:**

Let $X_j$ be the indicator that the $j$th entry of the original array is 0, for $j \in \{1, \ldots, i-1\}$. Then, the $i$th entry is moved backwards $\sum_{j=1}^{i-1} X_j$, positions, so

\[\mathbb{E}[X] = i - \sum_{j=1}^{i-1} \mathbb{E}[X_j] = i - \frac{i-1}{10} = \frac{9i+1}{10}.\]

The variance is also easy to compute, since the $X_j$ are independent. Then, $\text{var}(X_j) = (1/10)(9/10) = 9/100$, so

\[\text{var}(X) = \text{var} \left( i - \sum_{j=1}^{i-1} X_j \right) = \sum_{j=1}^{i-1} \text{var}(X_j) = \frac{9(i-1)}{100}.\]
3. Message Segmentation

The number of bytes $N$ in a message has a geometric distribution with parameter $p$. Suppose that the message is segmented into packets, with each packet containing $m$ bytes if possible, and any remaining bytes being put in the last packet. Let $Q$ denote the number of full packets in the message, and let $R$ denote the number of bytes left over.

(a) Find the joint PMF of $Q$ and $R$. Pay attention on the support of the joint PMF.

(b) Find the marginal PMFs of $Q$ and $R$.

(c) Repeat part (b), given that we know that $N > m$.

Note: you can use the formulas

$$\sum_{k=0}^{n} a^k = \frac{1 - a^{n+1}}{1 - a}, \text{ for } a \neq 1$$

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1 - x}, \text{ for } |x| < 1$$

in order to simplify your answer.

Solution:

(a) Given any $N$ there is a unique way to write it as $N = Qm + R$, where $R \in \{0, 1, \ldots, m\}$. Therefore

$$\mathbb{P}(Q = q, R = r) = \mathbb{P}(N = qm + r)$$

$$= (1 - p)^{qm+r-1}p.$$ 

The range of values of $(Q, R)$ is $\{ (q, r) \mid q \geq 0, 0 \leq r < m \} - \{(0, 0)\}$. Note that since $\mathbb{P}(N = 0) = 0$ we cannot have both $Q = 0$ and $R = 0$ at the same time.

(b) The marginal PMF of $Q$ is

$$\mathbb{P}(Q = q) = \sum_{r=0}^{m-1} (1 - p)^{qm+r-1}p$$

$$= (1 - p)^{qm-1} \sum_{r=0}^{m-1} (1 - p)^r$$

$$= (1 - p)^{qm-1}(1 - (1 - p)^m), \text{ for } q = 1, 2, \ldots,$$

$$\mathbb{P}(Q = 0) = \sum_{r=1}^{m-1} (1 - p)^{r-1}p$$

$$= 1 - (1 - p)^{m-1}.$$ 

The marginal PMF of $P$ is

$$\mathbb{P}(R = r) = \sum_{q=0}^{\infty} (1 - p)^{qm+r-1}p$$
Almost fixed points of a permutation

We now need to calculate the sizes of the sets $A_i$ in three ways. Note that this calculation is true so long as you have the three choices, $X$. To find the expectation from above, we need to observe to write $ \sum_{i,j} E \left[ \mathbb{1}_{i,j} \right] = \mathbb{E} \left[ \sum_i \mathbb{1}_i \right] = \mathbb{E} \left[ \sum_i E \left[ \mathbb{1}_i \right] \right] = \sum_i P(i \text{ is an almost fixed point}) = n^2 \frac{2}{n} = 3$. (We get that $P(i \text{ is a fixed point})$ from counting the number of such permutations; there are 3 choices for what $i$ can map to, and once we pick it for $i$, we can permute the rest in $(n-1)!$ ways.) Note that this calculation is true so long as you have the three choices, and for $n=1$ or $n=2$, the expectations are 1 and 2 respectively.

Continuing from the sum interpretation above, we get that $\mathbb{E} X^2 = \mathbb{E} \left[ \sum_{i,j} \mathbb{1}_{i,j} \right] = \mathbb{E} \left[ \sum_i \mathbb{1}_i^2 + 2 \sum_{i,j} \mathbb{1}_i \mathbb{1}_j \right]$. We note that the first sum is equal to $\mathbb{E} X$ since $\mathbb{1}_i^2 = \mathbb{1}_i$. Now, we simply look at the probability that both $i$ and $j$ may be almost fixed; to do so, we find it convenient to divide the sum into three sets $A_1 = \{ \{i,j\} \mid j = i+1 \text{ (mod $n$), $i,j \in [n]$} \}$, $A_2 = \{ \{i,j\} \mid j = i+2 \text{ (mod $n$), $i,j \in [n]$} \}$, $A_3 = \{ \{i,j\} \mid \{i,j\} \not\in A_1, A_2, i,j \in [n] \}$

We now need to calculate the sizes of the sets $A_1, A_2$ and $A_3$. When $n$ is sufficiently large (for this problem, $n \geq 5$), we have that $|A_1| = |A_2| = n$ and $|A_3| = \binom{n}{2} - 2n$.

\[
\mathbb{P}(R = 0) = \sum_{q=1}^{\infty} (1 - p)^{q} p
\]

\[
= p(1 - p)^{m-1} \sum_{q=0}^{\infty} ((1 - p)^{m})^q
\]

\[
= p(1 - p)^{m-1} \frac{1}{1 - (1 - p)^m}.
\]

(c) Due to the memoryless property of the geometric distribution, the PMF of $R$ will be exactly the same as in part (b), while the PMF of $Q$ is

\[
\mathbb{P}(Q = q) = (1 - p)^{(q-1)m-1} (1 - (1 - p)^m), \text{ for } q = 2, 3, \ldots,
\]

\[
\mathbb{P}(Q = 1) = 1 - (1 - p)^{m-1}.
\]

4. Almost fixed points of a permutation

Let $\Omega$ be the set of all permutations of the numbers $1, 2, \ldots, n$. Let an almost fixed point be defined as follows: If we put the numbers $i \in 1, 2, \ldots, n$ around a circle in clockwise order (such that 1 and $n$ are next to each other) and then assign another number $\omega(i) \in 1, 2, \ldots, n$ to it, if the number $\omega(i)$ is next to $i$ (or is equal to $i$), we will say that $i$ is almost a fixed point. So, for the permutation $\omega(1) = 5, \omega(2) = 3, \omega(3) = 1, \omega(4) = 4, \omega(5) = 2$, we have that 1, 2, and 4 are almost fixed points.

Now, let $X(\omega)$ denote the number of almost fixed points in $\omega \in \Omega$. Find $\mathbb{E}[X]$ and $\text{var}(X)$.

Solution:

To find the expectation from above, we need to observe to write $X$ as a sum of indicator random variables indicating whether $i$ is an almost fixed point or not, and doing so will give the answer as $3$ by $\mathbb{E}[X] = \mathbb{E} \left[ \sum_i \mathbb{1}_i \right] = \sum_i \mathbb{E} \left[ \mathbb{1}_i \right] = \sum_i P(i \text{ is an almost fixed point}) = \frac{n^2}{n} = 3$. (We get that $P(i \text{ is a fixed point})$ from counting the number of such permutations; there are $3$ choices for what $i$ can map to, and once we pick it for $i$, we can permute the rest in $(n-1)!$ ways.) Note that this calculation is true so long as you have the three choices, and for $n = 1$ or $n = 2$, the expectations are $1$ and $2$ respectively.
Thus, the second sum can be rewritten as: 
\[ 2 \sum_{i,j} 1_{i,j} = 2 \sum_{\{i,j\} \in A_1} 1_{i,j} + 2 \sum_{\{i,j\} \in A_2} 1_{i,j} + 2 \sum_{\{i,j\} \in A_3} 1_{i,j}. \]

Now, computing the probabilities gives us that
\[ E\left[ 2 \sum_{i,j} 1_{i,j}\right] = 2 \sum_{\{i,j\} \in A_1} \frac{7}{n(n-1)} + 2 \sum_{\{i,j\} \in A_2} \frac{8}{n(n-1)} + 2 \sum_{\{i,j\} \in A_3} \frac{9}{n(n-1)} = 2\left(\frac{15}{n-1} + \binom{n}{2} - 2n \frac{9}{n(n-1)}\right), \]
which again holds only for \( n \geq 5 \).

So the final answer is:
\[ \text{var}(X) = 3 + 2\left(\frac{15}{n-1} + \binom{n}{2} - 2n \frac{9}{n(n-1)}\right) - 9 \]
which now holds for \( n \geq 5 \).

We note that the variance for \( n = 1, 2, 3 \) is 0 since everything is an almost fixed point. For \( n = 4 \) we count the sizes of the sets as \( |A_1| = 4, |A_2| = 2, |A_3| = 0 \). Now, we note that the probability that both \( i \) and \( j \) are fixed points if they are in \( |A_2| \) will be \( \frac{7}{n(n-1)} \). So, simplifying what we had before, we now get 
\[ 3 + 2\left(\frac{7}{2}\right) - 9 = 1 \]
as our variance.

5. Introduction to Information Theory

Define the entropy of a discrete random variable \( X \) to be
\[ H(X) \triangleq -\sum_x p(x) \log p(x) = -E[\log p(X)], \]
where \( p(\cdot) \) is the PMF of \( X \). Here, the logarithm is taken with base 2, and entropy is measured in bits.

(a) Prove that \( H(X) \geq 0 \).

(b) Entropy is often described as the average information content of a random variable. If \( H(X) = 0 \), then no new information is given by observing \( X \). On the other hand, if \( H(X) = m \), then observing the value of \( X \) gives you \( m \) bits of information on average.

Let \( X \) be a Bernoulli random variable with \( \mathbb{P}(X = 1) = p \). Would you expect \( H(X) \) to be greater when \( p = 1/2 \) or when \( p = 1/3 \)? Calculate \( H(X) \) in both of these cases and verify your answer.

(c) We now consider a binary erasure channel (BEC).

![Binary Erasure Channel Diagram](image)

Figure 1: The channel model for the BEC showing a mapping from channel input \( X \) to channel output \( Y \). The probability of erasure is \( p_e \).

The input \( X \) is a Bernoulli random variable with \( \mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 1/2 \). Each time that we use the channel the input \( X \) will either get erased with probability \( p_e \), or it will get transmitted correctly with
probability $1 - p_e$. Using the character "?" to denote erasures, the output $Y$ of the channel can be written as

$$Y = \begin{cases} X, & \text{with probability } 1 - p_e \\ ?, & \text{with probability } p_e. \end{cases}$$

Compute $H(Y)$.

(d) We defined the entropy of a single random variable as a measure of the uncertainty inherent in the distribution of the random variable. We now extend this definition for a pair of random variables $(X, Y)$, but there is nothing really new in this definition because the pair $(X, Y)$ can be considered to be a single vector-valued random variable. Define the joint entropy of a pair of discrete random variables $(X, Y)$ to be

$$H(X, Y) \triangleq -E[\log p(X, Y)],$$

where $p(\cdot, \cdot)$ is the joint PMF and the expectation is also taken over the joint distribution of $X$ and $Y$.

Compute $H(X, Y)$, for the BEC.

Solution:

(a) This follows since $\log p(x) \leq 0$ for $p(x) \leq 1$.

(b) As an extreme example, when $p = 1$, you already know that $X$ will be 1, so observing $X$ gives you no new information. Therefore, we expect that the entropy will be greatest when $p = 1/2$.

The entropy of a Bernoulli random variable with bias $p$ can be written as

$$H(X) = -p \log p - (1 - p) \log(1 - p).$$

When $p = 1/2$,

$$H(X) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = \frac{1}{2} + \frac{1}{2} = 1 \text{ bit}.$$  

When $p = 1/3$,

$$H(X) = -\frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3} \approx 0.918 \text{ bits}.$$  

(c) The random variable $Y$ takes on three values: 0, 1, and ?. The marginal PMF of $Y$ can be written as

$$Y = \begin{cases} 0, & \text{with probability } \frac{1 - p_e}{2} \\ 1, & \text{with probability } \frac{1 - p_e}{2} \\ ?, & \text{with probability } p_e. \end{cases}$$

Therefore the entropy of $Y$ is

$$H(Y) = -p_e \log p_e - (1 - p_e) \log \frac{1 - p_e}{2}$$

$$= 1 - p_e - p_e \log p_e - (1 - p_e) \log(1 - p_e).$$
(d) The joint PMF of \((X,Y)\) can be written as

\[
(X,Y) = \begin{cases} 
(0,0), & \text{with probability } \frac{1-p_e}{2} \\
(0,?), & \text{with probability } \frac{p_e}{2} \\
(1,1), & \text{with probability } \frac{1-p_e}{2} \\
(1,?), & \text{with probability } \frac{p_e}{2}.
\end{cases}
\]

Therefore the entropy of the pair \((X,Y)\) is

\[
H(X,Y) = -p_e \log \frac{p_e}{2} - (1-p_e) \log \frac{1-p_e}{2} = 1 - p_e \log p_e - (1-p_e) \log(1-p_e).
\]

6. Soliton Distribution

This question pertains to the **fountain codes** introduced in the lab.

Say that you wish to send \(n\) chunks of a message, \(X_1, \ldots, X_n\), across a channel, but alas the channel is a **packet erasure channel**; each of the packets you send is erased with probability \(p_e > 0\). Instead of sending the \(n\) chunks directly through the channel, we will instead send \(n\) packets through the channel, call them \(Y_1, \ldots, Y_n\). How do we choose the packets \(Y_1, \ldots, Y_n\)? Let \(p(\cdot)\) be a probability distribution on \(\{1, \ldots, n\}\); this represents the **degree distribution** of the packets.

(i) For \(i = 1, \ldots, n\), pick \(D_i\) (a random variable) according to the distribution \(p(\cdot)\). Then, choose \(D_i\) random chunks among \(X_1, \ldots, X_n\), and “assign” \(Y_i\) to the \(D_i\) chosen chunks.

(ii) For \(i = 1, \ldots, n\), let \(Y_i\) be the XOR of all of the chunks assigned for \(Y_i\) (the number of chunks assigned for \(Y_i\) is called the **degree** of \(Y_i\)).

(iii) Send each \(Y_i\) across the channel, along with metadata which describes which chunks were assigned to \(Y_i\).

The receiver on the other side of the channel receives the packets \(Y_1, \ldots, Y_n\) (for simplicity, assume that no packets are erased by the channel; in this problem, we are just trying to understand what we should do in the ideal situation of **no channel noise**), and decoding proceeds as follows:

(i) If there is a received packet \(Y\) with only one assigned chunk \(X_j\), then set \(X_j = Y\). Then, “peel off” \(X_j\): for all packets \(Y_i\) that \(X_j\) is assigned to, replace \(Y_i\) with \(Y_i \text{ XOR } X_j\). Remove \(Y\) and \(X_j\) (notice that this may create new degree-one packets, which allows decoding to continue).

(ii) Repeat the above step until all chunks have been decoded, or there are no remaining degree-one packets (in which case we declare failure).

In the lab, you will play around with the algorithm and watch it in action. Here, our goal is to work out what a good degree distribution \(p(\cdot)\) is.
Intuitively, a good degree distribution needs to occasionally have prolific packets that have high degree; this is to ensure that all packets are connected to at least one chunk. However, we need “most” of the packets to have low degree to make decoding easier. Ideally, we would like to choose \( p(\cdot) \) such that at each step of the algorithm, there is exactly one degree-one packet.

(a) Suppose that when \( k \) chunks have been recovered (\( k = 0, 1, \ldots, N - 1 \)), then the expected number of packets of degree \( d \) (for \( d > 1 \)) is \( f_k(d) \). Assuming we are in the ideal situation where there is exactly one degree-one packet for any \( k \): What is the probability that a given degree \( d \) packet is connected to the chunk we are about to peel off? Based on that, what is the expected number of packets of degree \( d \) whose degrees are reduced by one after the \( (k + 1) \)st chunk is peeled off?

(b) We want \( f_k(1) = 1 \) for all \( k = 0, 1, \ldots, n - 1 \). Show that in order for this to hold, then for all \( d = 2, \ldots, n \) we have \( f_k(d) = (n - k)/[d(d - 1)] \). From this, deduce what \( p(d) \) must be, for \( d = 1, \ldots, n \). (This is called the ideal soliton distribution.)

[Hint: You should get two different recursion equations since the only degree 1 node at peeling \( k + 1 \) is going to come from the peeling of degree 2 nodes at peeling \( k \), however, for other higher degree \( d \) nodes, there will be some probability that some degree \( d \) ones will remain from the previous iteration and some probability that they will come from \( d + 1 \) one that will be peeled off]

(c) Find the expectation of the distribution \( p(\cdot) \).

In practice, the ideal soliton distribution does not perform very well because it is not enough to design the distribution to work well in expectation.

**Solution:**

(a) Of the \( f_k(d) \) packets with degree \( d \), each packet has probability \( d/(n - k) \) (since there are \( n - k \) packets remaining) of being connected with the message packet which is peeled off at iteration \( k + 1 \). Thus, by linearity, the answer is \( f_k(d)d/(n - k) \).

(b) We certainly need \( f_0(1) = 1 \) and \( 1 = f_1(1) = f_0(2) \cdot 2/n \), so \( f_0(2) = n/2 \). For \( k = 0, 1, \ldots, n - 1 \), we have \( 1 = f_{k+1}(1) = f_k(2) \cdot 2/(n - k) \), so \( f_k(2) = (n - k)/2 \).

Proceed by induction. Suppose that for all \( d \leq d' \), where \( d' = 2, \ldots, n - 1 \), we know that \( f_k(d) = (n - k)/[d(d - 1)] \). Then, for \( k = 0, 1, \ldots, n - d - 1 \),

\[
\frac{n - k - 1}{d(d - 1)} = f_{k+1}(d) = f_k(d + 1)\frac{d + 1}{n - k} + f_k(d)\left(1 - \frac{d}{n - k}\right)
\]

\[
= f_k(d + 1)\frac{d + 1}{n - k} + \frac{n - k}{d(d - 1)}\left(1 - \frac{d}{n - k}\right)
\]

so \( f_k(d + 1) = (n - k)/[d(d + 1)] \).
Note that \( f_0(d) \), the expected number of degree-\( d \) received packets at the beginning of the algorithm, is exactly \( np(d) \), so:

\[
p(d) = \begin{cases} 
  \frac{1}{n}, & d = 1 \\
  \frac{1}{d(d-1)}, & d = 2, \ldots, n
\end{cases}
\]

(c) The expectation is

\[
\sum_{d=1}^{n} dp(d) = \frac{1}{n} + \sum_{d=2}^{n} \frac{1}{d(d-1)} = \frac{1}{n} + \sum_{d=2}^{n} \frac{1}{d-1} = \sum_{d=1}^{n} \frac{1}{d} \approx \ln n.
\]

7. [Bonus] Connected Random Graph

The bonus question is just for fun. You are not required to submit the bonus question, but do give it a try and write down your progress.

We start with the empty graph on \( n \) vertices, and iteratively we keep on adding undirected edges \( \{u,v\} \) uniformly at random from the edges that are not so far present in the graph, until the graph is connected. Let \( X \) be a random variable which is equal to the total number of edges of the graph. Show that \( \mathbb{E}[X] = O(n \log n) \).

**Hint:** consider the random variable \( X_k \) which is equal to the number of edges added while there are \( k \) connected components, until there are \( k-1 \) connected components. Don’t try to calculate \( \mathbb{E}[X_k] \), an upper bound is enough.

**Solution:**

The hint suggests that we should follow the approach used in the coupon collecting problem. Indeed, observe that we can write \( X = \sum_{k=2}^{n} X_k \). Suppose that \( p_k \) is the probability that we add an edge which brings us to \( k-1 \) connected components at the time when we first have \( k \) connected components, and let \( Y_k \) be a geometric random variable with probability of success \( p_k \). Note that as we continue to add edges, the probability of producing \( k-1 \) connected components from \( k \) connected components will increase, starting from \( p_k \). Then, \( \mathbb{E}[X_k] \leq \mathbb{E}[Y_k] \).

In order to bound \( p_k \), assume that there are \( k \) connected components so far and \( u \) is one endpoint of the edge that we are currently adding. Then there are at least \( k-1 \) other vertices to which we can connect \( u \) and reduce the number

1Intuitively, \( Y_k \) is “larger” than \( X_k \), although it is difficult to explain precisely what this means in the context of randomness. We say that \( Y_k \) **stochastically dominates** \( X_k \) if \( \mathbb{P}(Y_k \geq x) \geq \mathbb{P}(X_k \geq x) \) for each \( x \). Here, \( Y_k \) does indeed stochastically dominate \( X_k \), and this fact implies \( \mathbb{E}[Y_k] \geq \mathbb{E}[X_k] \). To explain why, we use a coupling argument: suppose that each time we add an edge, we flip a coin of probability \( p_k \); if the coin comes up heads, then we add an edge to the graph that connects two connected components. If the coin comes up tails, then we still have a chance of adding an edge that connects two connected components (this is because the probability of connecting two connected components is \( \geq p_k \) and can be strictly larger); however, it is clear that in this case, the number of flips until we have \( k-1 \) connected components (starting with \( k \) connected components) is at most the number of flips until we see heads, i.e., \( X_k \leq Y_k \). Thus, \( \mathbb{E}[X_k] \leq \mathbb{E}[Y_k] \).
of components. In total there are \( n - 1 \) other vertices to which we can connect \( u \) so the probability that this edge reduces the number of components is

\[
p_k \geq \frac{k - 1}{n - 1}.
\]

Putting it all together

\[
\mathbb{E}[X] = \sum_{k=2}^{n} \mathbb{E}[X_k] \leq \sum_{k=2}^{n} \mathbb{E}[Y_k] = \sum_{k=2}^{n} \frac{1}{p_k} \leq \sum_{k=2}^{n} \frac{n - 1}{k - 1} = (n - 1)H_{n-1}
\]

\[
= O(n \log n)
\]