

Assignment Number: 1

Student Name: Gurpreet Singh

Roll Number: 150259

Date: February 14, 2018

We can have a naive sampling algorithm similar to the idea of Rejection sampling discussed in Lecture 3. This is based on the idea that we flip each coin twice, and invert the outcome of the second coin, *i.e.* if the second flip gives Heads, count it as Tails and vice versa.

Since there are K coins, there will be $2K$ coin flips in total. If the number of Heads is greater than K , then we say the outcome is Heads, if it is less than K , we predict Tails, otherwise, we reject. The algorithm is given more formally in [Algorithm 1](#)

Algorithm 1: SAMPLE (A Naive Approach)

Iterate until a sample is accepted.

1. For each $k = 1 \dots K$, flip the coin C_k twice and invert the second outcome.
2. Count the total number of heads, say n , and do the following
 - (a) If $n > K$, accept and report the outcome as Heads
 - (b) If $n < K$, accept and report the outcome as Tails
 - (c) If $n = K$, reject

Claim 1.1. [Algorithm 1](#) returns Heads or Tails, each with probability exactly $1/2$.

Proof. Since we have K coins, *i.e.* $C_1, C_2 \dots C_K$, and we are flipping each coin twice, we can alternately say that we have $2K$ coins, with the additional K coins being $C'_1, C'_2 \dots C'_K$. However, since we are inverting the outcome of the second flip, we can say that the bias of the coin C'_k is actually $1 - p_k$.

Therefore, effectively, we have $2K$ coins with biases $p_1, (1 - p_1) \dots p_K, (1 - p_K)$. The claim holds true because the resulting distribution for the number of heads obtained from flipping these $2K$ coins is symmetric. To show this, consider probability of getting the number of heads equal to r .

Note. I have denoted the number of heads as N_H and number of tails as N_T

$$\mathbb{P}[N_H = r] = \mathbb{P}[N_T = 2K - r]$$

The probability of getting $2K - r$ Tails is the same as getting the same number of heads if the outcome of each coin (consider C'_k coin as not inverted, but an entirely new coin) is inverted. This is equal to saying that the bias of each coin is now inverted. Let us denote these new coins as $\bar{C}_1, \bar{C}'_1 \dots \bar{C}_K, \bar{C}'_K$, and the probability of getting r heads as $\tilde{\mathbb{P}}[N_H = r]$.

As stated earlier, we can write

$$\begin{aligned} \mathbb{P}[N_H = r] &= \mathbb{P}[N_T = 2K - r] \\ &= \tilde{\mathbb{P}}[N_H = 2K - r] \\ &= \tilde{\mathbb{P}}[N_T = r] \end{aligned}$$

The bias of the coins $\bar{C}_1, \bar{C}'_1 \dots \bar{C}_K, \bar{C}'_K$ are simply the inverse of the bias of the original coins, *i.e.* $(1 - p_1), p_1 \dots (1 - p_K), p_K$, which is the same as that of the original coins, with only the sequence being

different. Hence there will be no effect on the number of Heads. Therefore, we can say $\mathbb{P}[N_H = r] = \tilde{\mathbb{P}}[N_H = r]$ or equivalently $\mathbb{P}[N_T = r] = \tilde{\mathbb{P}}[N_T = r]$. Hence,

$$\begin{aligned} \implies \mathbb{P}[N_H = r] &= \tilde{\mathbb{P}}[N_T = r] \\ &= \mathbb{P}[N_T = r] \end{aligned}$$

Therefore, the probability of getting r Heads is the same as the probability of getting $2K - r$ Heads. Hence, the probability $\mathbb{P}[N_H < K] = \mathbb{P}[N_H > K]$.

Using this, we can say that given the number of Heads will never be equal to K (since we are rejecting that case), the probability of getting the outcome Heads is the same as the probability of getting the outcome Tails, and therefore is equal to $1/2$. \square

Therefore I have stated that we have a Naive algorithm which will give the outcome Heads with probability $1/2$. However, this algorithm is not necessarily desired as we do not want to flip all the coins ($2K$ flips) in each iteration. Therefore, our aim is to optimize the number of coins we actually want to use.

Before that, let us see the optimal bias we require if there is only one coin. This is, in fact, very easy to see as the probability of getting one Heads and one Tails is if the both the flips give either Heads or Tails (since second flip's outcome will be reversed). This is equal to $p^2 + (1 - p)^2$. This is clearly minimized for $p = 1/2$.

Hence, if we are choosing just one coin, we wish the bias of that coin to be close to $1/2$. In fact, this holds even if the number of coins is more than one, as given in the following claim.

Claim 1.2. The minimum probability of rejection will be obtained when $\forall k = 1 \dots K' \quad p_k \approx 1/2$ (K' is the number of coins we have chosen).

Proof. Let us denote the probability of obtaining N_H number of heads considering only the first k coins as $p(N_H, k)$. Therefore, if we take all K' coins, we can write the probability of getting N_H heads as follows

$$\begin{aligned} \mathbb{P}[N = N_H] &= p(N_H, K') \\ &= p_{K'} (1 - p_{K'}) (p(N_H - 2, K' - 1) + p(N_H, K' - 1)) \\ &\quad + (p_{K'}^2 + (1 - p_{K'})^2) p(N_H - 1, K' - 1) \end{aligned}$$

If we put $N_H = K'$, we get the rejection probability (as discussed before Algorithm 1). Since we want to minimize the probability of rejection, we can differentiate (partial) with respect to p_k and compute the optimal value of the biases. Consider the partial derivative w.r.t. to $p_{K'}$

$$\frac{\delta(\mathbb{P}[N = K'])}{\delta(p_{K'})} = (1 - 2p_{K'}) (p(N_H, K' - 1) - 2p(N_H - 1, K' - 1) + p(N_H - 2, K' - 1))$$

Hence, in order to minimize the gradient (get close to minima), we need to set $p_{K'} \approx 1/2$. Also, since all the coins are exchangeable, we can say that this will in fact hold for all the coins we have chosen (by replacing coin K' with some other coin k). \square

Hence, for all K' coins we choose, we wish the bias of each coin to be close to $1/2$. Since we might not actually have K' coins such that each chosen coin will have bias $1/2$, we need to optimize the value of K' as well.

This leads us to the problem of finding the correct trade off between a higher value of K' and the bias of each chosen coin. The intuitive sense is that if use more number of coins, the probability of rejecting one sample decreases, and hence we might converge faster. However, it comes with the added number of flips per sample (equal to $2K'$). We, in fact, can say that the optimal value of $K' = 1$, given we choose the coin with bias close to $1/2$ (this is possible since we have at least one coin with bias equal to $1/2$).

In order to show this, consider the minimum probability of convergence which is obtained when the bias of each coin is equal to $1/2$. Suppose if the bias of the coin which is closest to $1/2$ is p . Therefore, the probability of rejection if we choose K' coins is at least $p^{K'}$. Considering the case of only one coin, we choose the coin with the bias closest to $1/2$, and therefore the probability of rejection for $2K'$ flips (same as the case of one sampling iteration for K' coins) is equal to $p^{K'}$. This is the case of minimum probability of rejection in case of K' coins.

Therefore, it is optimal to have only one coin which has a bias closest to $1/2$, instead of having K' coins. Since we know that at least one of the coins has bias equal to $1/2$. Our goal is now to try and find out the coin with bias equal to $1/2$.

Taking our cue from the idea of point-wise convergence (and Alice and Bob's example, Lec 3), we can say that the most (asymptotically) optimal strategy to find this coin is to estimate the bias of each coin (by flipping them n times, and estimating bias as ratio of number of heads and number of flips, as discussed in class), and to consider the coin with the estimated bias closest to $1/2$.

This strategy works since if we flip a coin (say C_k) n times, we can be sure with probability greater than $1 - \delta$ that our estimation is at least ϵ -correct. More formally

$$\mathbb{P}[|\hat{p}_k - p_k| > \epsilon] \leq \delta = 2 \exp\left(-\frac{n\epsilon^2}{2}\right) \quad [\text{Chernoff's Bound}]$$

The probability that we have estimated all the biases of all the coins at least ϵ -correct, *i.e.* we have uniform convergence, we have

$$\mathbb{P}\left[\sup_k |\hat{p}_k - p_k| > \epsilon\right] \leq \delta = 2K \exp\left(-\frac{n\epsilon^2}{2}\right) \quad [\text{Uniform Convergence}]$$

Therefore, if we keep ϵ to a small value, with sufficient number of flips (n), we can ensure with a high probability that the biases we have estimated are approximately correct. Therefore, we can now simply choose the coin with the estimated bias closest to $1/2$. Therefore, we have our preprocessing algorithm, given in [Algorithm 2](#), and then the sample algorithm in [Algorithm 3](#).

Algorithm 2: PRE-PROCESS

For all $k = 1 \dots K$

1. Flip the coin C_k N times, given

$$N = -\frac{2}{\epsilon^2} \log\left(\frac{2k}{\delta}\right)$$

where δ is the confidence parameter.

2. Estimate the bias \hat{p}_k as

$$\hat{p}_k = \frac{N_H}{N}$$

Choose the coin C'_k such that

$$C'_k = \arg \min_{C_k} |\hat{p}_k - 0.5|$$

Return C'_k

We can, now, also formalise the bound on the number of flips we require to probably find a coin with the bias closest to $1/2$. Since we know that there is at least one coin with bias $1/2$. Therefore, if we find the bias estimates ϵ -correctly, we can say that there will be at least one coin such that the estimate of the bias of that coin will be in the range $(0.5 - \epsilon, 0.5 + \epsilon)$.

Let us say the coin satisfying the above property and having the estimated bias closest to 0.5 is C_r . Then, from the arguments of uniform convergence, we know $p_r \in (\hat{p}_r - \epsilon, \hat{p}_r + \epsilon)$ with probability at least $(1 - \delta)$. Hence, we can write

$$\mathbb{P}[|\hat{p}_r - 0.5| > 2\epsilon] \leq 2K \exp\left(-\frac{n\epsilon^2}{2}\right)$$

Hence, with high probability, we can say that the coin we have predicted to be the fairest coin has a bias close to 0.5. We can therefore use this coin with a high confidence that for the sampling algorithm (Rejection Sampling) in order to create a fair coin.

Algorithm 3: SAMPLE

Let C_r be the coin from the pre-processing step. Until we have an accepted sample, iterate

1. Flip the coin C_r two times, and return as follows

Flip 1	Flip 2	Outcome
H	H	H
H	T	\perp
T	H	\perp
T	T	T

2. If Outcome is H or T, accept and return, otherwise reject and continue

In case Alice knew that this is in fact the fair coin, she could directly output the outcome of the coin, however, this is not the case, and hence she needs to sample using the above algorithm only.

We can also say that the probability of getting a valid sample in n iterations of the SAMPLE algorithm is equal to $(p_r^2 + (1 - p_r)^2)^{n-1}$, where p_r is the bias of the coin we use to do the sampling.

Note. The only case where SAMPLE will fail is when the bias of the coin we have taken is equal to 0 or 1. However, the probability for this to occur will be very small, given ϵ is small, and n (the number of preprocessing flips) is large.

As stated earlier, we can now also formally prove that the probability of getting Heads (or Tails) is exactly 1/2. This is same as the getting a heads for any number of flips, *i.e.*

$$\begin{aligned} \mathbb{P}[\text{SAMPLE} = H] &= \sum_{n=1}^{\infty} (p_r^2 + (1 - p_r)^2)^{n-1} p_r(1 - p_r) \\ &= \frac{p_r(1 - p_r)}{1 - (p_r^2 + (1 - p_r)^2)} \\ &= \frac{1}{2} \end{aligned}$$

If we were to find the number of tosses (m) required to say with at least $1 - \Delta$ confidence, the SAMPLE will end, we can say that

$$\begin{aligned} 1 - \Delta &\leq \sum_{n=1}^m (p_r^2 + (1 - p_r)^2)^{n-1} 2p_r(1 - p_r) \\ &= 1 - (p_r^2 + (1 - p_r)^2)^m \end{aligned}$$

For ease of notation, say $P_r = p_r^2 + (1 - p_r)^2$, then

$$\implies m \geq \frac{\log(\Delta)}{\log(P_r)}$$

$$\implies m_H = \left\lceil \frac{\log(\Delta)}{\log(P_r)} \right\rceil$$

Hence, if we want the SAMPLE algorithm to end with confidence $1 - \Delta$, we need to iterate at least m_H times.

Assignment Number: 1

Student Name: Gurpreet Singh

Roll Number: 150259

Date: February 14, 2018

Suppose the set of random variables over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is denoted by \mathcal{V} . We need to prove that \mathcal{V} is a vector space (over \mathbb{R}). In order to prove this, we need the addition and scalar multiplication operators to be well defined in the vector space.

$$\text{Vector Addition:} \quad \mathbf{v}_1 + \mathbf{v}_2 \in \mathcal{V} \quad (1)$$

$$\text{Scalar Multiplication:} \quad c \cdot \mathbf{v}_1 \in \mathcal{V} \quad (2)$$

Note. \mathbf{v}_i denote elements of the vector space we are concerned with

\mathcal{V} is said to be a vector space if it follows all the following properties

$$\mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3) = (\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 \quad (3)$$

$$\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1 \quad (4)$$

$$\mathbf{0} \in \mathcal{V}, \text{ s.t. } \mathbf{v} + \mathbf{0} = \mathbf{v} \quad (5)$$

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0} \quad (6)$$

$$c \cdot (\mathbf{v}_1 + \mathbf{v}_2) = c \cdot \mathbf{v}_1 + c \cdot \mathbf{v}_2 \quad (7)$$

$$(c + d) \mathbf{v} = c \cdot \mathbf{v} + d \cdot \mathbf{v} \quad (8)$$

$$c(d\mathbf{v}) = (cd) \mathbf{v} \quad (9)$$

$$\mathbf{1} \in \mathcal{V}, \text{ s.t. } \mathbf{1}(\mathbf{v}) = \mathbf{v} \quad (10)$$

We first show the properties that for a space \mathcal{V} over random variables defined over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the properties (1) and (2) hold true. Consider two random variables $X_1, X_2 \in \mathcal{V}$, then, we need to show

$$X = X_1 + X_2 \in \mathcal{V}$$

$$X = c \cdot X_1 \in \mathcal{V}$$

where $c \in \mathbb{R}$ is a constant. In order to show that these quantities are in the vector space, *i.e.* X is indeed a random variable over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we need to show that the events defined by $\{w \in \Omega : X(w) \leq t\} \in \mathcal{F}$.

Consider the case of multiplication first. Since X_1 is a random variable, we know that

$$\begin{aligned} \{w \in \Omega : X_1(w) \leq t\} &\in \mathcal{F} \\ \implies \{w \in \Omega : c \cdot X_1(w) \leq c \cdot t\} &\in \mathcal{F} \end{aligned}$$

Since $t' = c \cdot t \in \mathbb{R}$ (properties of the real space), we can write this as

$$\begin{aligned} \{w \in \Omega : c \cdot X_1(w) \leq t'\} &\in \mathcal{F} \\ \implies \{w \in \Omega : X(w) \leq t'\} &\in \mathcal{F} \end{aligned}$$

Hence scalar multiplication holds for the space \mathcal{V} . For the case of addition, I have referred to the scribes for [EE510, Dr. Krishna Jagannathan](#).

Firstly, when we are adding two random variables, say $X_1, X_2 \in \mathcal{V}$, since both X_1 and X_2 take finite values, their sum (addition operation), if measurable will also take finite values. Therefore, we only need to prove that $X = X_1 + X_2$ is measurable at all $w \in \Omega$ and is defined over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. An alternate approach to this, as stated earlier is to prove that for all $t \in \mathbb{R}$, the set $\{w \in \Omega : X(w) \leq t\} \in \mathcal{F}$. We

can write

$$\begin{aligned}\{w \in \Omega : X(w) \leq t\} &= \{w \in \Omega : X_1(w) + X_2(w) \leq t\} \\ &= \bigcup_{r \in \mathbb{R}} \{w \in \Omega : X_1(w) \leq r \text{ and } X_2(w) \leq t - r\}\end{aligned}$$

However, this is an \mathbb{R} uncountably infinite set, and hence, we cannot take the union over this. However, from the realization that rationals are dense in the real space, *i.e.* for all $r \in \mathbb{R}$, we can find a sequence of rationals $q_1, q_2 \dots \in \mathbb{Q}$, such that $\lim_{n \rightarrow \infty} q_n = r$. Hence, we can transform the union as follows

$$\{w \in \Omega : X(w) \leq t\} = \bigcup_{q \in \mathbb{Q}} (\{w \in \Omega : X_1(w) \leq q\} \cap \{w \in \Omega : X_2(w) \leq t - q\})$$

Since \mathbb{Q} is countably infinite (from arguments from Abstract Algebra), we can say that the above union exists. Also, since \mathcal{F} is closed under both intersection and union, and for all $r \in \mathbb{R}$, both the sets $\{w \in \Omega : X_1(w) \leq r\}$ and $\{w \in \Omega : X_2(w) \leq r\}$ are in \mathcal{F} , the intersection and union over such sets will result in a set which also belongs in \mathcal{F} .

Hence, we can claim X is measurable and therefore a bounded (finitely) random variable defined over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and therefore by definition $X \in \mathcal{V}$. This proves property of Vector Addition (1).

Since we have proved properties (1) and (2), we can move on to show the properties from (3) - (10). Before that, consider property (5). We can easily find a random variable (say X_0), such that the property holds true.

Consider $X_0 : \Omega \rightarrow \mathbb{R}$ such that for all $w \in \Omega$, $X_0(w) = 0$. It is easy to see that this satisfies the property (5). Consider $X = X_1 + X_0$. From property (1), we have that $X \in \mathcal{V}$. We can therefore write

$$\begin{aligned}\forall t \in \mathbb{R}, \{w \in \Omega : X(w) \leq t\} &= \{w \in \Omega : (X_0 + X_1)(w) \leq t\} \\ &= \{w \in \Omega : X_1(w) \leq t\}\end{aligned}$$

Therefore, we can say $X \equiv X_1$. Hence $X_0 \equiv 0$ is the “zero vector” for the vector space \mathcal{V} . Also, the same variable is the “zero vector” for the real space. This suggests that the vector space is indeed a linear subspace of the real space. Hence, all the properties (3) - (10) follow from the properties of the real space ()

Therefore, the vector space \mathcal{V} is in fact a vector space. Hence, the set of random variables over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ form a vector space (over \mathbb{R}).

Assignment Number: 1

Student Name: Gurpreet Singh

Roll Number: 150259

Date: February 14, 2018

For an event $A \in \mathcal{F}$, the σ -sub algebra generated by that event would simply be the set $\sigma(A) = \{\phi, A, A^C, \Omega\}$, since this satisfies all the properties of a σ -sub algebra, *i.e.* closed under complement, and contains both the null event and the sure event.

Therefore, we need to prove $A_1, A_2 \in \mathcal{F}$ are independent iff $\sigma(A_1)$ and $\sigma(A_2)$ are independent. Let us consider the forward implication, *i.e.* $A_1 \perp A_2 \implies \sigma(A_1) \perp \sigma(A_2)$.

Claim 1.3. For any two events $A_1, A_2 \in \mathcal{F}$, if $A_1 \perp A_2 \implies \sigma(A_1)\sigma(A_2)$.

Proof. In order to prove two sigma algebras are independent, we need to prove for any events $E_1 \in \sigma(A_1)$ and $E_2 \in \sigma(A_2)$, $E_1 \perp E_2$.

Let us prove an identity of complement and independent events first. Consider two events $A_1, A_2 \in \mathcal{F}$ such that $A_1 \perp A_2$, then

$$\begin{aligned} A_1 \wedge A_2^C &= A_1 \wedge (\Omega \setminus A_2) \\ &= A_1 \setminus (A_1 \wedge A_2) \\ \implies \mathbb{P}[A_1 \wedge A_2^C] &= \mathbb{P}[A_1 \setminus (A_1 \wedge A_2)] \end{aligned}$$

Since $A_1 \wedge A_2 \subseteq A_1$, therefore

$$\begin{aligned} \mathbb{P}[A_1 \wedge A_2^C] &= \mathbb{P}[A_1 \setminus (A_1 \wedge A_2)] \\ &= \mathbb{P}[A_1] - \mathbb{P}[(A_1 \wedge A_2)] \\ &= \mathbb{P}[A_1](1 - \mathbb{P}[A_2]) \quad [\text{as } A_1 \perp A_2] \end{aligned}$$

Hence

$$\mathbb{P}[A_1 \wedge A_2^C] = \mathbb{P}[A_1] \cdot \mathbb{P}[A_2^C] \quad (11)$$

Therefore, $A_1 \perp A_2 \implies A_1 \perp A_2^C$. From the same steps, we can get $A_1^C \perp A_2$. Further applying (11) to this, we get $A_1^C \perp A_2^C$.

Since we have $\sigma(A) = \{\phi, A, A^C, \Omega\}$, and we already know $A_1 \perp A_2$, therefore, from the properties mentioned above, we have $A_1^C \perp A_2^C$, $A_1^C \perp A_2$ and $A_1 \perp A_2^C$.

Also, for any event A , $\Omega \cap A = A \implies \mathbb{P}[\Omega \cap A] = 1 \cdot \mathbb{P}[A] = \mathbb{P}[\Omega] \cdot \mathbb{P}[A]$.

Similarly $\phi \cap A = \phi \implies \mathbb{P}[\phi \cap A] = 0 \cdot \mathbb{P}[A] = \mathbb{P}[\phi] \cdot \mathbb{P}[A]$.

Hence, we have proved that for all pairs over the σ -sub algebras of A_1 and A_2 and hence the σ -sub algebras themselves are independent. \square

Claim 1.4. For any two events $A_1, A_2 \in \mathcal{F}$, if $\sigma(A_1)\sigma(A_2) \implies A_1 \perp A_2$.

Proof. For the backward implication, we directly have that all pairs of the σ -sub algebras are independent of each other. Since A_1, A_2 is one such pair, we directly have their independence. Therefore $A_1 \perp A_2$. \square

From Claims 1.3 and 1.4, we can say for any two events $A_1, A_2 \in \mathcal{F}$, $A_1 \perp A_2 \iff \sigma(A_1) \perp \sigma(A_2)$.

Assignment Number: 1

Student Name: Gurpreet Singh

Roll Number: 150259

Date: February 14, 2018

Part 1

If we sample S without replacement, the events user i is sampled and user j is sampled won't be independent. More formally, consider E_i to be the event that user i is sampled out of the n sampled users. Then the probability

$$\mathbb{P}[E_i] = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

Now consider the probability of both E_i and E_j being sampled together.

$$\mathbb{P}[E_i \wedge E_j] = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}$$

Therefore, $\mathbb{P}[E_i \wedge E_j] \neq \mathbb{P}[E_i] \cdot \mathbb{P}[E_j]$, and hence the events E_i and E_j are not independent. This suggests that we cannot sample without replacement, even though there is a chance that the number of distinct users we get if we sample with replacement is lesser than n which in fact wastes our samples.

Part 2

As given in the question, the random variables we are considering, *i.e.* $c_{I_i I_j}$ and $c_{I_j I_k}$ are independent since we have a common user among them. This suggests that we cannot directly apply any tail bound we know of.

Therefore, we need $\hat{\mu}_S$ to be a sum of random variables such that they are independent. Hence, we tweak $\hat{\mu}_S$, in such a way, that for each user, there is only one term containing the dependence of that user.

In order to present this more formally, we can imagine the user-user matrix to be a perfect graph, with the number of messages between each user pair to be the weight of the edge between them. Therefore, we need to sample the edges such that there are no two edges with a common vertex.

Note. The user-user matrix is only between users who have been sampled in the first step

Considering only edges that do not share a vertex, we are ensuring that no edge weight is dependent on any other edge weight.

Suppose using the edge sampling method described above, we have $\lfloor \frac{k}{2} \rfloor$ (where k is the number of distinct users sampled in Step 1) edges (or edge weights). Let us represent these edge weights as $\bar{c}_1 \dots \bar{c}_{\lfloor k/2 \rfloor}$. Then we can write the new term for $\hat{\mu}_S$ as

$$\hat{\mu}_S = \frac{1}{\lfloor k/2 \rfloor} \sum_{i=1}^{\lfloor k/2 \rfloor} \bar{c}_i$$

Part 3

We can show that with high probability, $\hat{\mu}_S \approx \mu$. For this, we consider Hoeffding's Inequality.

Note. For the further portion of the solution, let $k' = \lfloor k/2 \rfloor$

Since we have k' independent random variables $\bar{c}_1 \dots \bar{c}_{k'}$, where each is considered to be identical (as a-priori, we do not distinguish between the number of messages, and since the users are sampled randomly, we do not know anything about the characteristics of the user, and hence about the number of messages he might send). Also, since each $\bar{c}_i \in [0, K]$, we can apply Hoeffding's Inequality on $\hat{\mu}_S$.

Therefore, we get

$$\mathbb{P} \left[\left| \frac{1}{k'} \sum_i^{k'} \bar{c}_i - \mathbb{E}[\bar{c}] \right| > \epsilon \right] \leq 2 \exp \left(-\frac{2k'\epsilon^2}{K^2} \right)$$

Since we have already stated that all the random variables \bar{c}_i are independent and identical, we can say that the expectation of each of these terms, and hence $\mathbb{E}[\bar{c}] = \mu$. Hence, we can write

$$\mathbb{P} [|\hat{\mu}_S - \mu| > \epsilon] \leq 2 \exp \left(-\frac{2k'\epsilon^2}{K^2} \right) \leq 2 \exp \left(-\frac{(k-1)\epsilon^2}{K^2} \right) = \delta$$

Hence, we can say with probability at least $1 - \delta$, where δ is as given above, that the estimated $\hat{\mu}_S$ is ϵ -close to the actual mean μ .