SEEM5020 2026 Spring: Homework 2 Solution

- **Q1. Chernoff Bound (Lower Tail)**

  Proof: For any $t > 0$, we have:

  $$\mathbb{P}[X \leq (1-\delta)\mu] = \mathbb{P}[e^{-tX} \geq e^{-t(1-\delta)\mu}].$$

  Applying Markov's Inequality:

  $$\mathbb{P}[e^{-tX} \geq e^{-t(1-\delta)\mu}] \leq \frac{\mathbb{E}[e^{-tX}]}{e^{-t(1-\delta)\mu}}.$$

  Since $X = \sum X_i$ are independent:

  $$\mathbb{E}[e^{-tX}] = \prod_{i=1}^{n} \mathbb{E}[e^{-tX_i}].$$

  For a Bernoulli variable X_i:

  $$\mathbb{E}[e^{-tX_i}] = p_i e^{-t} + (1-p_i) = 1 + p_i(e^{-t} - 1).$$

  Using the inequality $1 + x \leq e^x$:

  $$\mathbb{E}[e^{-tX_i}] \leq e^{p_i(e^{-t}-1)}.$$

  Thus,

  $$\mathbb{E}[e^{-tX}] \leq \prod_{i=1}^{n} e^{p_i(e^{-t}-1)} = e^{(e^{-t}-1)\sum p_i} = e^{\mu(e^{-t}-1)}.$$

  Substituting back into the Markov bound:

  $$\mathbb{P}[X \leq (1-\delta)\mu] \leq \frac{e^{\mu(e^{-t}-1)}}{e^{-t(1-\delta)\mu}} = e^{\mu(e^{-t}-1+t(1-\delta))}.$$

  To minimize the exponent, we set the derivative with respect to t to 0: $-e^{-t} + (1-\delta) = 0 \implies e^{-t} = 1-\delta \implies t = -\ln(1-\delta)$. Since $\delta \in (0,1)$, $t > 0$. Substituting $t = -\ln(1-\delta)$:

  $$\text{RHS} = e^{\mu((1-\delta)-1-(1-\delta)\ln(1-\delta))} = e^{\mu(-\delta-(1-\delta)\ln(1-\delta))} = \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}.$$

  To prove the second inequality, we take the logarithm of the term inside the power μ:

  $$-\delta - (1-\delta)\ln(1-\delta).$$

  Using the ineqaulity $(1-\delta)\ln(1-\delta) + \delta \geq \delta^2/2$, we have:

  $$-\delta - (1-\delta)\ln(1-\delta) \leq -\delta^2/2.$$

  Thus,

  $$\mathbb{P}[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}.$$

- **Q2. Hoeffding's Inequality (Additive Bound)**

  Proof: We apply the Chernoff bound technique. For any $t > 0$:

  $$\mathbb{P}[X \geq \mu + \lambda] = \mathbb{P}[X - \mu \geq \lambda] = \mathbb{P}[e^{t(X-\mu)} \geq e^{t\lambda}] \leq e^{-t\lambda}\mathbb{E}[e^{t(X-\mu)}].$$

Since $X = \sum X_i$ and $\mu = \sum p_i$, we can write $X - \mu = \sum_{i=1}^{n}(X_i - p_i)$. Due to independence:

$$\mathbb{E}[e^{t(X-\mu)}] = \mathbb{E}\left[\prod_{i=1}^{n} e^{t(X_i-p_i)}\right] = \prod_{i=1}^{n} \mathbb{E}[e^{t(X_i-p_i)}].$$

Let $Y_i = X_i - p_i$. Note that $\mathbb{E}[Y_i] = 0$. Since $X_i \in \{0, 1\}$, the range of Y_i is $[0 - p_i, 1 - p_i]$, so $b - a = (1 - p_i) - (-p_i) = 1$.

Applying Hoeffding's Lemma to Y_i:

$$\mathbb{E}[e^{tY_i}] \leq e^{\frac{t^2(1)^2}{8}} = e^{\frac{t^2}{8}}.$$

Substituting this back into the product:

$$\mathbb{E}[e^{t(X-\mu)}] \leq \prod_{i=1}^{n} e^{\frac{t^2}{8}} = e^{\frac{nt^2}{8}}.$$

Now, substitute this into the initial Markov bound:

$$\mathbb{P}[X \geq \mu + \lambda] \leq e^{-t\lambda} \cdot e^{\frac{nt^2}{8}} = e^{\frac{nt^2}{8} - t\lambda}.$$

To find the tightest bound, we minimize the exponent $f(t) = \frac{n}{8}t^2 - \lambda t$ with respect to t.

$$f'(t) = \frac{n}{4}t - \lambda = 0 \implies t = \frac{4\lambda}{n}.$$

Substituting $t = \frac{4\lambda}{n}$ back into the exponent:

$$\frac{n}{8}\left(\frac{4\lambda}{n}\right)^2 - \lambda\left(\frac{4\lambda}{n}\right) = \frac{n}{8}\frac{16\lambda^2}{n^2} - \frac{4\lambda^2}{n} = \frac{2\lambda^2}{n} - \frac{4\lambda^2}{n} = -\frac{2\lambda^2}{n}.$$

Therefore:

$$\mathbb{P}[X \geq \mu + \lambda] \leq e^{-\frac{2\lambda^2}{n}}.$$

- **Q3. Weighted Reservoir Sampling (A-Res)**

Proof: The A-Res algorithm selects the k items with the largest values of $K_i = U_i^{1/w_i}$. Let us apply a strictly monotonic transformation to K_i. Since the transformation preserves order, the set of selected indices will remain the same.

Consider the natural logarithm function $\ln(x)$, which is strictly increasing for $x > 0$.

$$\ln(K_i) = \ln\left(U_i^{1/w_i}\right) = \frac{1}{w_i}\ln(U_i).$$

Selecting the largest K_i is equivalent to selecting the largest $\frac{\ln(U_i)}{w_i}$.

Now, consider multiplying by -1. This reverses the inequality.

$$-\ln(K_i) = -\frac{\ln(U_i)}{w_i} = K_i'.$$

Therefore, finding the indices i that maximize K_i is equivalent to finding the indices i that maximize $\ln(K_i)$, which is equivalent to finding the indices i that minimize $-\ln(K_i) = K_i'$.

Thus, maintaining the k elements with the smallest $K_i' = -\frac{\ln(U_i)}{w_i}$ is mathematically equivalent to the original algorithm.

- **Q4. Morris(a) Algorithm**

I. Unbiased Estimator Let X_t be the counter value after t updates. Let $Y_t = (1+a)^{X_t}$. The estimator is $\hat{N}_t = \frac{Y_t - 1}{a}$. We compute the conditional expectation of Y_{t+1} given X_t:

$$\begin{aligned}\mathbb{E}[Y_{t+1}|X_t] &= (1+a)^{X_t+1} \cdot \frac{1}{(1+a)^{X_t}} + (1+a)^{X_t} \cdot \left(1 - \frac{1}{(1+a)^{X_t}}\right) \\ &= (1+a) + (1+a)^{X_t} - 1 \\ &= Y_t + a.\end{aligned}$$

Taking the total expectation:

$$\mathbb{E}[Y_{t+1}] = \mathbb{E}[\mathbb{E}[Y_{t+1}|X_t]] = \mathbb{E}[Y_t] + a.$$

Since $X_0 = 0$, we have $\mathbb{E}[Y_0] = (1+a)^0 = 1$. By recurrence, $\mathbb{E}[Y_t] = 1 + ta$. Thus, for the final count N:

$$\mathbb{E}[\hat{N}_N] = \frac{\mathbb{E}[Y_N] - 1}{a} = \frac{(1+Na) - 1}{a} = N.$$

II. Variance Let $\Delta_t = \hat{N}_{t+1} - \hat{N}_t$ be the increment of the estimator at step $t+1$. If X increments (probability $1/Y_t$), $\Delta_t = \frac{(1+a)Y_t - 1}{a} - \frac{Y_t - 1}{a} = Y_t$. If X does not increment, $\Delta_t = 0$.

We compute the first two moments of Δ_t given Y_t:

$$\mathbb{E}[\Delta_t|Y_t] = Y_t \cdot \frac{1}{Y_t} = 1.$$

$$\mathbb{E}[\Delta_t^2|Y_t] = Y_t^2 \cdot \frac{1}{Y_t} = Y_t.$$

Now consider the second moment of $\hat{N}_{t+1}$:

$$\hat{N}_{t+1}^2 = (\hat{N}_t + \Delta_t)^2 = \hat{N}_t^2 + 2\hat{N}_t\Delta_t + \Delta_t^2.$$

Taking expectation conditioned on Y_t (note that $\hat{N}_t$ is determined by Y_t):

$$\mathbb{E}[\hat{N}_{t+1}^2|Y_t] = \hat{N}_t^2 + 2\hat{N}_t\mathbb{E}[\Delta_t|Y_t] + \mathbb{E}[\Delta_t^2|Y_t] = \hat{N}_t^2 + 2\hat{N}_t + Y_t.$$

Taking the total expectation:

$$\mathbb{E}[\hat{N}_{t+1}^2] = \mathbb{E}[\hat{N}_t^2] + 2\mathbb{E}[\hat{N}_t] + \mathbb{E}[Y_t].$$

Substituting $\mathbb{E}[\hat{N}_t] = t$ and $\mathbb{E}[Y_t] = at + 1$ (derived in Q4.I):

$$\mathbb{E}[\hat{N}_{t+1}^2] = \mathbb{E}[\hat{N}_t^2] + 2t + (at+1) = \mathbb{E}[\hat{N}_t^2] + (a+2)t + 1.$$

Since $\hat{N}_0 = 0$, $\mathbb{E}[\hat{N}_0^2] = 0$:

$$\mathbb{E}[\hat{N}_N^2] = \sum_{t=0}^{N-1}((a+2)t+1) = (a+2)\frac{N(N-1)}{2} + N.$$

Finally, the variance is:

$$\begin{aligned}\mathrm{Var}[\hat{N}_N] &= \mathbb{E}[\hat{N}_N^2] - (\mathbb{E}[\hat{N}_N])^2 \\ &= (a+2)\frac{N(N-1)}{2} + N - N^2 \\ &= \frac{aN(N-1)}{2} + N(N-1) + N - N^2 \\ &= \frac{aN(N-1)}{2} + (N^2 - N) + N - N^2 \\ &= \frac{aN(N-1)}{2}.\end{aligned}$$

III. Space Complexity Consider the case when $X \geq k$ where $k = \log_{1+a}(a(1+\epsilon)N+1)$. Using the conclusion from Q4.I and Q4.II and the fact that $a = 2\epsilon^2\delta$, we have:

$$\mathbb{P}[X \geq k] = \mathbb{P}[(1+a)^X \geq a(1+\epsilon)N+1] = \mathbb{P}[\hat{N} \geq (1+\epsilon)N] \leq \frac{aN(N-1)}{2\epsilon^2N^2} \leq \delta.$$

Thus, with probability at least $1-\delta$, the counter X is less than k. As $a \in (0,1]$, we have:

$$k = \frac{\ln(a(1+\epsilon)N+1)}{\ln(1+a)} \leq \frac{2}{a}\ln(a(1+\epsilon)N+1) = O(\frac{1}{a}\ln N).$$

The number of bits is:

$$\#\text{Bits} = O(\log_2 k) = O\left(\log\left(\frac{1}{a}\ln N\right)\right) = O(\log(1/a) + \log\log N).$$

Substituting $a = 2\epsilon^2\delta$, we have $\log(1/a) = O(\log(1/\epsilon) + \log(1/\delta))$ and thus:

$$\#\text{Bits} = O(\log\log N + \log(1/\delta) + \log(1/\epsilon)).$$

Remark: The number k might look like magic number to you, indeed you could derive this number on the draft by reversing the derivation steps in the proof.

Indeed, this upper bound is not tight, https://arxiv.org/pdf/2010.02116 shows $O(\log\log N + \log\log(1/\delta) + \log(1/\epsilon))$ bits suffices and is optimal.