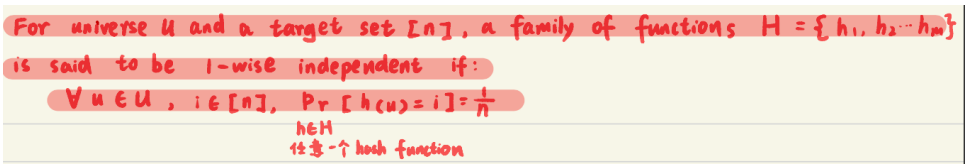
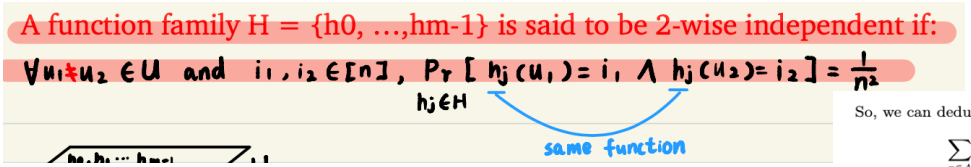
1. 1-wise independent



2. 2-wise independent



3. Random variable

Definition: A random variable is a function over a probability space.

Example: X is a random variable of the number of heads when we toss 2 coins:

$$X \in \{0,1,2\}, \quad \Pr(X=0)=\frac{1}{4}, \Pr(X=1)=\frac{1}{2}, \Pr(X=2)=\frac{1}{4}$$

4. Expectation

Definition: For a random variable X, the expectation of X:

$$\mathbb{E}[X] = \sum_x x \cdot \Pr[X=x]$$

Example: If we toss 3 coins and X represents the number of heads:

$$\mathbb{E}[X] = 0\times\frac{1}{8}+1\times\frac{3}{8}+2\times\frac{3}{8}+3\times\frac{1}{8}=\frac{3}{2}$$

5. Linearity of expectation (not require independence)

The **linearity of expectation** is a fundamental property in probability theory that states:

$$\mathbb{E}[X+Y] = \mathbb{E}[X]+\mathbb{E}[Y]$$

for any two random variables X and Y. This property holds even if the random variables are not independent. More generally, for any finite collection of random variables $X_1, X_2, \ldots, X_n$, the linearity of expectation can be expressed as:

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

6. Expected number of collisions (using 2-wise independence)

Suppose we have a set S of items $\{x_1, x_2, \ldots, x_s\}$. We want to compute $h(x_1), h(x_2), \ldots, h(x_s)$ and see how many collisions happen. (The number of collisions is a random variable; we want to understand an average expected number of collisions.)

Let us choose $h \sim H$ (where H is **2-wise independent**). The number of collisions:

$$X_c : \#\text{collisions} = \sum_{1\le i<j\le s} \mathbb{1}\cdot[h(x_i)=h(x_j)] \tag{2.13}$$

So, we can deduce:

$$\sum_{x\in A}[\mathbb{E}[\mathcal{E}_{x,h}^2]] = \sum_{x\in A}[\mathbb{E}[\mathcal{E}_{x,h}]] = |A|\cdot\mathbb{E}[\mathcal{E}_{x,h}] = p\cdot 2^k\cdot\frac{1}{2^k} = p$$

The second term is a bit tricky:

$$\sum_{x\in A}[\mathbb{E}[\mathcal{E}_{x,h}\mathcal{E}_{x',h}]] = \sum_{x\in A}\Pr_{h\in\mathcal{H}}[h(x)=h(x')=0] = \Pr_{h\in\mathcal{H}}[h(x)=0\wedge h(x')=0]$$

(we assume $x \neq x'$)

The result of this equation is calculated in the previous question. Therefore:

$$\sum_{x\in A}[\mathbb{E}[\mathcal{E}_{x,h}\mathcal{E}_{x',h}]] = \sum_{x\in A}[\Pr_{h\in\mathcal{H}}[h(x)=0]\cdot\Pr_{h\in\mathcal{H}}[h(x')=0]] = \sum_{x\in A}[\frac{1}{2^k}\cdot\frac{1}{2^k}] = \sum_{x\in A}[\frac{1}{2^{2k}}]$$

$$\sum_{x\neq x'}\mathbb{E}[\mathcal{E}_{x,h}\mathcal{E}_{x',h}] = 2\cdot\binom{|A|}{2}\cdot\frac{1}{2^{2k}} = \frac{|A|(|A|-1)}{2}\cdot\frac{1}{2^{2k}}$$

$$= 2\cdot\frac{p\cdot 2^k(p\cdot 2^k-1)}{2}\cdot\frac{1}{2^{2k}}$$

$$= 2\cdot\frac{p\cdot(p\cdot 2^k-1)}{2\cdot 2^k}$$

$$= 2\cdot\frac{p^2\cdot 2^k-p}{2\cdot 2^k}$$

$$= p^2-\frac{p}{2^k}$$

Thus, the final result is:

$$\mathbb{E}[\mathcal{E}^2] = p+p^2-\frac{p}{2^k}$$

c) **Sub-question 1:**

$$(\mathcal{E}-1)^2 = \mathcal{E}^2-2\mathcal{E}+1$$

$$\mathbb{E}[(\mathcal{E}-1)^2] = \mathbb{E}[(\mathcal{E}^2-2\cdot\mathcal{E}+1)]$$

$$= \mathbb{E}[\mathcal{E}^2]-2\cdot\mathbb{E}[\mathcal{E}]+1$$

Based on the previous question, we know $\mathbb{E}[\mathcal{E}^2] = p+p^2-\frac{p}{2^k}$, and $\mathbb{E}[\mathcal{E}] = p$. Thus:

$$\mathbb{E}[(\mathcal{E}-1)^2] = \mathbb{E}[\mathcal{E}^2]-2\cdot\mathbb{E}[\mathcal{E}]+1 = p+p^2-\frac{p}{2^k}-2\cdot p+1$$

We want to find $\mathbb{E}[X_c]$:

$$\mathbb{E}[X_c] = \mathbb{E}\left[\sum_{1\le i<j\le s} \mathbb{1}\cdot[h(x_i)=h(x_j)]\right]$$

By linearity of expectation:

$$= \sum_{1\le i<j\le s} \mathbb{E}\left[\mathbb{1}\cdot[h(x_i)=h(x_j)]\right] \tag{2.15}$$

$$= \sum_{1\le i<j\le s} \Pr[h(x_i)=h(x_j)] \tag{2.16}$$

As $h(x_i)$ and $h(x_j)$ have the same value and we have n choices for that value (marked as k):

$$= \sum_{1\le i<j\le s}\sum_{1\le k\le n} \Pr[h(x_i)=k \wedge h(x_j)=k] \tag{2.17}$$

Given H is 2-wise independent, $\Pr[h(x_i)=i_1 \wedge h(x_j)=i_2] = \frac{1}{n^2}$ for any $x_i \neq x_j \in U$ and $i_1, i_2 \in [n]$:

$$= \sum_{1\le i<j\le s}\sum_{1\le k\le n} \frac{1}{n^2} = \sum_{1\le i<j\le s} n\cdot\frac{1}{n^2} = \sum_{1\le i<j\le s}\frac{1}{n} \tag{2.18}$$

Thus,

$$\mathbb{E}[X_c] = \sum_{1\le i<j\le s}\frac{1}{n} = \binom{s}{2}\frac{1}{n} = \frac{s(s-1)}{2n} \asymp \frac{s^2}{2n} \tag{2.19}$$

As n increases, the average/expected number of collisions goes down.

Now we can choose the size of n such that $n = 10s^2$ and the expected number of collisions will be $\le \frac{1}{20}$.

7. Storing a function h

If we choose H to be "all function", then storing a single $h \in H$ could require: $|U| \log n$ bits

Expectation describes the general statistics of a random variable. It doesn't give us any insight into the actual values that the R.V takes. Markov's inequality will explore obtaining the probability of a R.V taking values greater than or equal to a particular value (t).

8. **Markov's inequaltiy**

Suppose Z is a non-negative random variable, then $\forall t \ge 1$, $\Pr[Z \ge t\cdot\mathbb{E}[Z]] \le \frac{1}{t}$

Meaning, for example, if the expected value of Z is 1, the probability $Z \ge 2$ ($E[Z]$ is 1 and t is 2 here) necessarily *must* be $\le \frac{1}{2}$.

Proof:

$$E[Z] = \sum_{x\ge 0} x \times Pr[Z=x] = \sum_{x\ge tE[Z]} x \times Pr[Z=x] + \sum_{0\le x<tE[Z]} x \times Pr[Z=x]$$

We know $\sum_{x\ge tE[Z]} x \times Pr[Z=x] \le E[Z]$ (as the sum is equal to $E[Z] - \sum_{x<tE[Z]} x \times Pr[Z=x]$, which is necessarily less than or equal to $E[Z]$).

Since the lower bound of the variable x in the sum $\sum_{x\ge tE[Z]} x \times Pr[Z=x]$ is $tE[Z]$, we can establish:

$$tE[Z] \sum_{x\ge tE[Z]} Pr[Z=x] \le \sum_{x\ge tE[Z]} x \times Pr[Z=x]$$

which implies:

$$tE[Z] \sum_{x\ge tE[Z]} Pr[Z=x] = tE[Z] \times Pr[Z \ge tE[Z]] \le \sum_{x\ge tE[Z]} x \times Pr[Z=x] \implies$$

$$tPr[Z \ge tE[Z]] \le \frac{\sum_{x\ge tE[Z]} x \times Pr[Z=x]}{E[Z]} \le 1 \implies [\textit{divide everything by } t]$$

$$Pr[Z \ge tE[Z]] \le \frac{1}{t}$$

9. Quantity of Collisions

3.2.2 Quantity of Collisions

It should be <= m^2/n.

Now consider the RV Z representing number of hash collisions again. We know $E[Z] = \frac{m}{n^2}$, so e.g. if we have $n = 10m^2 \implies E[Z] = \frac{1}{20}$. Applying Markov's Inequality (with $t = 20$), we have:

$$Pr[Z \geq 1] = Pr[Z \geq 20E[Z]] \leq \frac{1}{20}$$

$$\implies Pr[Z = 0] = 1 - Pr[Z \geq 1] \implies 1 - \frac{1}{20} = \frac{19}{20}$$

Meaning, the probability of having no hash collisions when we have $n = 10m^2$ is $\frac{19}{20}$. In general, if we choose some $n = O(m^2)$, we can ensure with high probability that there will be no collisions ($Z = 0$).

M represents the number of elements being hashed. N represents the number of possible hash values.

10. Modulus arithmetic

Arithmetic mod p. Suppose p is a prime. 如果p不是prime，则公式5不成立

1. $(a + b) \bmod p = a \bmod p + b \bmod p$
2. $(a \cdot b) \bmod p = a \bmod p \cdot b \bmod p$ Only consider integers from 1 to p-1
3. $a \cdot (b + c) = (a \cdot b) \bmod p + (a \cdot c) \bmod p$
4. $\forall a, \exists! b, (a+b) \bmod p = 0$
5. $\forall a \neq 0, \exists! b, a \cdot b = 1 \rightarrow a \bmod p \cdot b \bmod p = 1$

存在唯一一个b $(a \cdot b) \bmod p = 1$

11. Construct a pairwise independent hash functions family

Suppose U is a universe. Choose a prime $|U| \leq p \leq 2|U|$. Then, we can construct a pairwise independent hash function family H such that

$$\forall h \in H,\ h : U \to [p] \text{ and } |H| = p^2$$

A random function family fulfilling these condition has size $n^{|U|} \asymp p^{\frac{p}{2}}$.

Each $h \in H$ is indexed by 2 elements $a, b \in [p]$ which define its behavior. In particular:

$$h_{a,b}(u) = (a \times u + b)(mod\ p)$$

We want to show H is a 2-wise ind. hash function family. We must evaluate the probability for some fixed choice of a, b (all calculations are mod p):

$$Pr_{h \sim H}[h(u_1) = i_1 \wedge h(u_2) = i_2] = Pr_{a,b \sim [p]}[a \times u_1 + b = i_1 \wedge a \times u_2 + b = i_2]$$

We now nee to find the number of pair of values of (a,b) that satisfy this condition. i.e. We have to find the solution to the 2 equations: $a \times u_1 + b = i_1$ and $a \times u_2 + b = i_2$.

By solving the equations, we have the fact that:

$$a(u_1 - u_2) = i_1 - i_2$$

As p is a prime, we know there exists some *unique* element $(u_1 - u_2)^{-1}$ which is the multiplicative inverse of $(u_1 - u_2)$ in the field of p. This means there is exactly one choice of a pair a, b in the above equation such that for the unique inverse element $(u_1 - u_2)^{-1}$, we have:

$$a(u_1 - u_2) = i_1 - i_2 \implies a(u_1 - u_2)(u_1 - u_2)^{-1} = (i_1 - i_2)(u_1 - u_2)^{-1} \implies$$

$$a = (i_1 - i_2)(u_1 - u_2)^{-1},\ b = i_1 - a \times u_1$$

Which is to say that there is a unique solution to the 2 equations.

Meaning the probability of

$$Pr_{a,b \sim [p]}[a \times u_1 + b = i_1 \wedge a \times u_2 + b = i_2] = \frac{1}{p} \times \frac{1}{p} = \frac{1}{p^2}$$

Thus we have constructed a family of functions H s.t. the necessary conditions for p.i. are met for all $h \in H$.

- Suppose the domain is $[p]$ (p is a prime number), then $\forall a, b \in [p]$, $H = \{h_{a,b} : h_{a,b}(u) = au + b\}$ is a family of pairwise independent hash functions $\iff Pr_{a,b \in [p]}[h_{a,b}(u) = i_1 \wedge h_{a,b}(u') = i_2] = 1/p^2$.
- If U is our universe, then we stipulate that $p \geq |U| \geq p/2$ (you need p to be larger than the size of universe but we don't want it to be too much bigger otherwise, that will cost in terms of memory). The fact that there always exists such a prime is a non-trivial fact from number theory (called Bertrand's postulate).
- If you have any pairwise independent hash function family $U \to [n]$, as long as $n \gg m^2$, we get no collisions.
- But we don't want to create a table of size of universe ← computationally and storage-wise expensive & defeating purpose of hashing.

This should be p>10m^2.

- As long as ~~$p^2 \geq 10m^2$~~ {where m is the worst case input}, with high probability, we can be sure there are no collisions.

12. Size of hash table

Size of the hash table $= \Theta(n) = \Theta(m^2)$, where m is the number of items.

Example:

$$|U| = 10^7, \quad m = 10^3 \Rightarrow \text{Size of hash table} = \Theta(10^6)$$

And memory to store the definition of each hash function, we need to store $a, b \in Z_p$.

$$\text{Number of bits to store } a \in Z_p = \log p$$

13. **Variance** of a random variable

Give a real valued random variable X, $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$
$$= \mathbb{E}[X^2 + (\mathbb{E}[X])^2 - 2X \cdot \mathbb{E}[X]]$$
$$= \mathbb{E}[X^2] + (\mathbb{E}[X])^2 - 2\mathbb{E}[X] \cdot \mathbb{E}[X]$$
$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0 \text{ because } \mathbb{E}[(X - \mathbb{E}[X])^2] \geq 0$$

Fact: $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$

Another definition of variance:

Suppose we make 2 independent draws from X.

$$\mathbb{E}_{x,y \in X}[(x - y)^2] = \mathbb{E}_{x,y \in X}[x^2 + y^2 - 2xy]$$
$$= \mathbb{E}_{x \in X}[x^2] + \mathbb{E}_{y \in X}[y^2] - 2\mathbb{E}_{x \in X}[x]\mathbb{E}_{y \in X}[y]$$
$$= 2\mathbb{E}_{x \in X}[x^2] - 2(\mathbb{E}_{x \in X}[x])^2 = 2 \cdot var(X)$$

problem: find z such that $\mathbb{E}_{x \in X}[(x - z)^2]$ is minimized.

$$\mathbb{E}_{x \in X}[x^2] + z^2 - 2z\,\mathbb{E}_{x \in X}[x^2]$$

This is minimized at $z = \mathbb{E}[X]$

14. **Chebyshev's inequality**

- **Sample Mean** $\bar{X}$: If you're estimating the population mean μ from a sample, the sample mean $\bar{X} = \frac{1}{n}\sum X_i$ is an unbiased estimator of μ because:

$$E[\bar{X}] = \mu$$

- **Sample Proportion** $\hat{p}$: If you are estimating the probability p of an event (like in the coin flip example you shared), the sample proportion $\hat{p} = \frac{\text{Number of successes}}{n}$ is an unbiased estimator of p, meaning:

$$E[\hat{p}] = p$$

where $\sigma^2 = \mathrm{Var}(X)$. σ is known as the standard deviation. Thus, for any random variable, the chance that you are t standard deviations away from the mean is at most $\frac{1}{t^2}$. The concentration of the random variable is around its mean.

Chebyshev's inequality: $\Pr[|X-E[X]|>t\cdot\sigma] \le \frac{1}{t^2}$ (σ: sigma)

For any random variable X, if $\mu = \mathbb{E}[X]$, $\sigma^2 = \mathrm{Var}(X)$, then $\forall t \ge 0$, $\Pr[|X-\mu| \ge t\cdot\sigma] \le \frac{1}{t^2}$. scale of concentration (偏离mean的值)

Sep 10 Lecture 5

For a real valued random variable X, $\mathrm{Var}(X) = \mathbb{E}_{x\in X}[(X-\mu)^2]$ when $\mu = \mathbb{E}[X]$.

$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

Chebyshev's inequality proof:

Let a random variable $y = |x - \mathbb{E}[x]|$ and x is also a random variable.

$y^2 = |x-\mathbb{E}[x]|^2 = (x-\mathbb{E}[x])^2 = x^2 + (\mathbb{E}[x])^2 - 2x\cdot\mathbb{E}[x]$

$\mathbb{E}[y^2] = \mathbb{E}[x^2] + (\mathbb{E}[x])^2 - 2(\mathbb{E}[x])^2$

$= \mathrm{Var}(x)$

$= \sigma^2$

$\Pr[|x-\mathbb{E}[x]| \ge t\cdot\sigma] = \Pr[(x-\mathbb{E}[x])^2 \ge t^2\cdot\sigma^2]$

$= \Pr[y^2 \ge t^2\cdot\sigma^2] \le \frac{\mathbb{E}[y^2]}{t^2\cdot\sigma^2} = \frac{\sigma^2}{t^2\cdot\sigma^2} = \frac{1}{t^2}$ (By Markov's)

Markov's: For any non-negative Z, $t \ge 0$, $\Pr[Z \ge t] \le \frac{\mathbb{E}[Z]}{t}$.

变式: suppose $t = t'\cdot\mathbb{E}[Z]$

$\Pr[Z \ge t'\cdot\mathbb{E}[Z]] \le \frac{\mathbb{E}[Z]}{t'\mathbb{E}[Z]} = \frac{1}{t'}$

Markov's inequality: Suppose Z is a non-negative random variable, then $\forall t \ge 1$, $\Pr[Z \ge t\cdot\mathbb{E}[Z]] \le \frac{1}{t}$

For any estimation task, we need to consider error and confidence. Let $\delta, \mathcal{E} > 0$ be confidence and error parameters.

Choose $M \ge \frac{1}{4\mathcal{E}^2\delta}$. Lower $\mathcal{E}$ and δ lead to higher M.

$$Pr[|\hat{p} - p| \ge \mathcal{E}] \le \frac{1}{4\mathcal{E}^2 M} \le \delta.$$

Summary: given unknown point, set $\mathcal{E}$, ie 0.01. Set confidence $\delta := 0.05$, meaning the answer will be terribly wrong with 5% chance. If you toss coin M times, such that $M \ge \frac{10000*20}{4} = 50000$ times, output $\hat{p}$ estimate will be within 0.01 of p with 95% probability.

15. Application of Chebyshev's inequality

Suppose we have a coin with unknown bias p. (Pr[H]=p, Pr[T]=1-p)
Strategy: to estimate p, toss the coin M times and let $\hat{p}$*M the number of heads. Output $\hat{p}$. $\hat{p}$ is a random variable. $\hat{p}$ 是根据实验结果的估计值.

Let Xi=1 if the ith coin toss is head
=0 otherwise
E[Xi] =p
Output $\hat{p}$=(X1+X2+…+Xμ)/M

$\mathbb{E}[\hat{p}] = \frac{1}{M}\cdot\sum_{i=1}^{M}\mathbb{E}[x_i] = \frac{1}{M}\cdot M\cdot p = p$ unbiased estimator

Compute $\mathrm{Var}(P)$:

Lemma: suppose x, y are random variables, then $\mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(y)$

proof: $Z = x+y$, $\mathrm{Var}(Z) = \mathbb{E}_{\substack{x\in x\\ y\in y}}[((x+y)-(\mathbb{E}[Z]))^2]$ only require they are pairwise independent

$= \mathbb{E}_{\substack{x\in x\\ y\in y}}[((x+y)-(\mathbb{E}[x]+\mathbb{E}[y]))^2]$

$= \mathbb{E}[((x-\mathbb{E}[x])+(y-\mathbb{E}[y]))^2]$

$= \mathbb{E}[(x-\mathbb{E}[x])^2] + \mathbb{E}[(y-\mathbb{E}[y])^2] + 2\mathbb{E}[(x-\mathbb{E}[x])\cdot(y-\mathbb{E}[y])]$

$= \mathrm{Var}(x) + \mathrm{Var}(y)$ (0 because x, y are random variables they are independent from each other)

Back to the coin example

$\mathrm{Var}(x_1+\cdots+x_M) = \sum_{i=1}^{M}\mathrm{Var}(x_i)$

Fact: $\mathrm{Var}(c\cdot x) = c^2\cdot\mathrm{Var}(x)$

$\Rightarrow \mathrm{Var}(\hat{p}) = \frac{1}{M^2}\cdot\sum_{i=1}^{M}\mathrm{Var}(x_i) = \frac{\mathrm{Var}(x_1)}{M}$ because all x are identically distributed

$\mathbb{E}[x_1] = p$, $\mathbb{E}[x_1^2] = p$, $\mathrm{Var}(x_1) = p - p^2 = p\cdot(1-p)$

$\mathrm{Var}(\hat{p}) = \frac{p\cdot(1-p)}{M}$ standard deviation $= \frac{1}{\sqrt{M}}\cdot\sqrt{p\cdot(1-p)}$

$\Pr[|\hat{p}-p| \ge \frac{t}{\sqrt{M}}\cdot\sqrt{p\cdot(1-p)}] \le \frac{1}{t^2}$ by Markov's

$\Pr[|\hat{p}-p| \ge \underset{\text{epsilon}}{\varepsilon}] \le \frac{p\cdot(1-p)}{\varepsilon^2 M} \le \frac{1}{4\varepsilon^2 M}$ (Let $t = \frac{\varepsilon\sqrt{M}}{\sqrt{p\cdot(1-p)}}$)

Choose M:

Let $\underset{\text{delta}}{\delta}, \varepsilon > 0$ be the confidence and error parameter.

choose $M \ge \frac{1}{4\varepsilon^2\delta} \Rightarrow \Pr[|\hat{p}-p| \ge \varepsilon] \le \frac{1}{4\varepsilon^2 M} \le \delta$

$\varepsilon = 0.01$ $\delta = 0.05$ $M = \frac{1}{4}\cdot 10000\cdot 20 = 50000$

(离真实p的误差 概率)

16. Implication of Chebyshev's

Theorem 6.1 *Chebyshev's Inequality: For any r.v. X,*

$$\Pr[|X-\mu| \ge t\sigma] \le \frac{1}{t^2}$$

where $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$.

Consequence: Given a coin with unknown bias p, we can estimate p with error $\pm\varepsilon$ and confidence $1-\delta$ with $\theta\left(\frac{1}{\delta\varepsilon^2}\right)$ independent trials.

To see this, consider an r.v. X normalized to be bounded in $[-1, 1]$ (i.e. average income, voting). If $X_1, \cdots, X_M$ are independent copies of X and $Y = (X_1+\cdots+X_M)/M$, we know $\mathbb{E}[Y] = \mathbb{E}[X]$ and $\mathrm{Var}(Y) = \frac{1}{M}\mathrm{Var}(X) \le \frac{2}{M}$. Applying Chebyshev's, we get

$$\Pr[|Y-\mathbb{E}[Y]| \ge t\sigma_y] \le \frac{1}{t^2}$$

Denoting $\mu = \mathbb{E}[X] = \mathbb{E}[Y]$ and substituting our upper bound on σ_y,

$$\Pr\left[|Y-\mu| \ge t\sqrt{\frac{2}{M}}\right] \le \frac{1}{t^2}$$

When $M = \frac{2}{\delta\varepsilon^2}$, we can set $t = \frac{1}{\sqrt{\delta}}$ to get

$$\therefore \Pr[|Y-\mu| \ge \varepsilon] \le \delta$$
$$\Rightarrow \Pr[|Y-\mu| \ge \varepsilon] \ge 1-\delta$$

which shows $M = \theta\left(\frac{2}{\delta\varepsilon^2}\right)$ suffices to get error ε with confidence $\ge 1-\delta$.

As it turns out, we *must* have M proportional to $\frac{1}{\varepsilon^2}$ to guarantee error ε with high probability. However, it turns out we can improve M's dependence on δ when X is bounded.

17. Chenoff-Hoeffding Bound

Chenoff-Hoeffding bound: Let $x_1 \cdots x_m$ be independent random variable supported in $[-1, 1]$. Let $y = \left(\frac{x_1 + \cdots x_m}{m}\right)$, $\mathbb{E}[y] = \left(\sum_{i=1}^{m} \mathbb{E}[x_i]\right)/m$

$\Pr[|y - \mathbb{E}[y]| \geq \varepsilon] \leq e^{\frac{-\varepsilon^2 m}{4}}$

$e^{-\frac{\varepsilon^2 m}{4}} \leq \delta \iff \frac{\varepsilon^2 m}{4} \geq \ln\left(\frac{1}{\delta}\right) \Rightarrow m \geq \frac{4}{\varepsilon^2} \cdot \ln\left(\frac{1}{\delta}\right)$

only require pairwise independent. Larger m required.

Chebyshev's : To get error $\pm\varepsilon$ with confidence $1-\delta$, $m \geq \frac{1}{\delta\varepsilon^2}$.

Chenoff : To get error $\pm\varepsilon$ with confidence $1-\delta$, $m \geq \frac{1}{\delta\varepsilon^2}$. $m \geq \frac{1}{\varepsilon^2} \cdot \ln\left(\frac{1}{\delta}\right)$.

require random independent. Smaller m required.

This is significantly better than our bound obtained by Chebyshev's. For example, if we wanted probability of failure of $\delta \leq \frac{1}{e^{100}}$, Chebyshev's inequality tells us we'd need $\geq e^{100}$ samples, while Chernoff-Hoeffding's bound suggests we only need ≥ 100.

Note that the requirements to use Chernoff-Hoeffding are tighter than those for Chebyshev's: only pairwise independence was needed for Chebyshev's, but *mutual* independence for all $X_1, \cdots, X_M$ is needed for the Chernoff-Hoeffding bound.

18. Counting distinct elements in a stream

Given N objects in stream, each of which is drawn from a universe of size M, how can we count the number of distinct elements?

Solution outline: We define a uniform continuous hashing function $h : U \to [0, 1]$. Whenever a new object x_i arrives, we calculate $h(x_i)$. We have a running tracker of the minimum such hash we've seen so far, which we can denote c_i at iteration i; if $h(x_i) < c_{i-1}$, we set $c_i = h(x_i)$. Otherwise, $c_i = c_{i-1}$. Finally, we output $1/c_n$.

We will design a randomized algorithm which will achieve a C-approximation.

C-approximation: It outputs $\hat{M}$ such that the number of distinct item (M_d) satisfies:

$$\frac{1}{C} \leq \frac{\hat{M}}{M_d} \leq C$$

Push C $\to 1 + \varepsilon$

For example, we can have an initial goal of achieving $C = 5$, then with more ideas, we make $C \to 1 + \varepsilon$

7.2 Initial algorithm

Consider a random function $h : [N] \to [0, 1]$.

(Aside: suppose you have a random function $h_q : [N] \to [q] \iff h_q : [N] \to \{0, \frac{1}{q}, \frac{2}{q}, ..., \frac{q-1}{q}\}$.

1. Initialize $r = 2$
2. For every item x_i, compute $h(x_i)$, $r \leftarrow min\{r, h(x_i)\}$
3. Output $\frac{1}{r}$

Claim 1: If there are M_d distinct numbers, then $r \in min\{y_1, y_2, ..., y_{M_d}\}$ where each y_i is uniformly and randomly chosen from $[0, 1]$.

Claim 2:

$$\mathbb{E}[r] = \frac{1}{M_d + 1}$$

7.2.1 Calculating $\mathbb{E}[r^2]$

By definition, $\mathbb{E}[r^2] = \int_0^1 y^2 \cdot \Pr(y)\, dy$, that is

$$\begin{aligned}
\mathbb{E}[r^2] &= \int_0^1 y^2 M_d (1-y)^{M_d - 1}\, dy \\
&= \int_0^1 (1-z)^2 M_d z^{M_d - 1}\, dz \\
&= \int_0^1 (1 - 2z + z^2) M_d z^{M_d - 1}\, dz \\
&= \int_0^1 M_d z^{M_d - 1}\, dz - 2\int_0^1 M_d z^{M_d}\, dz + \int_0^1 M_d z^{M_d + 1}\, dz \\
&= 1 - \frac{2M_d}{M_d + 1} + \frac{M_d}{M_d + 2}
\end{aligned}$$

With this, we can compute

$$\begin{aligned}
Var(r) &= \mathbb{E}[r^2] - (\mathbb{E}[r])^2 \\
&= 1 - \frac{2M_d}{M_d + 1} + \frac{M_d}{M_d + 2} - \frac{1}{(M_d + 1)^2} \\
&\leq \left(\frac{1}{M_d + 1}\right)^2 \\
\sigma_r &\leq \frac{1}{M_d + 1}
\end{aligned}$$

7.3 Modified Algorithm

1. Choose k hash functions $h_1, h_2, ..., h_k$.
2. Initialize $r_1, r_2, ..., r_k \leftarrow 2$.
3. For each x_i, computed $h_j(x_i)$; $\forall 1 \leq j \leq k$, $r_j \leftarrow min[r_j, h_j(x_i)]$
4. $\hat{r} = \frac{r1+r2+...+rk}{k}$, Output $\frac{1}{\hat{r}}$

7.3.1 Calculating $\mathbb{E}[\hat{r}]$

By linearnality of expectations, $\mathbb{E}[\hat{r}] = \frac{1}{k}\sum_{i=1 \text{ to } k} \mathbb{E}[\hat{r}_i]$. Since $\mathbb{E}[\hat{r}_i] = \frac{1}{M_d+1}$, $\mathbb{E}[\hat{r}] = \frac{1}{k} * \frac{k}{M_d+1} = \frac{1}{M_d+1}$.

$$\begin{aligned}
\mathbb{E}[\hat{r}^2] &= \frac{1}{k^2}\sum_{i=1 \text{ to } k} \mathbb{E}[\hat{r}_i{}^2] \\
&= \frac{1}{k^2} * k * \left(1 - \frac{2M_d}{M_d + 1} + \frac{M_d}{M_d + 2}\right) \\
&= \frac{1}{k}\left(1 - \frac{2M_d}{M_d + 1} + \frac{M_d}{M_d + 2}\right)
\end{aligned}$$

7.3.2 Calculating Var($\hat{r}$)

$$\begin{aligned}
Var(\hat{r}) &= \frac{1}{k^2}\sum_{i=1}^{k} Var(r_i) \\
&= \frac{1}{k} \cdot \left(1 - \frac{2M_d}{M_d + 1} + \frac{M_d}{M_d + 2} - \frac{1}{(M_d + 1)^2}\right) \\
&\leq \frac{1}{k}\left(\frac{1}{M_d + 1}\right)^2 \\
\sigma_r &\leq \frac{1}{\sqrt{k}} \frac{1}{M_d + 1}
\end{aligned}$$

Choose $k = \frac{1}{\epsilon^2}$ and apply Chebychev inequality with $t = \frac{1}{\sqrt{\epsilon}}$, we have

$$\Pr[|\hat{r} - \frac{1}{M_d + 1}| \geq \frac{\sqrt{\epsilon}}{M_d + 1}] \leq \epsilon$$

Which implies that $\hat{r} \in \left(\frac{1-\sqrt{\epsilon}}{M_d+1}, \frac{1+\sqrt{\epsilon}}{M_d+1}\right)$ with probability at least $1 - \epsilon$

19. Algorithm to count the number of distinct items in a stream & minimizing memory

Naive solution: $M_d \cdot \log N$

Randomized solution: Choose a random hash function $h_1 \cdots h_k \in [N] \to [0, 1]$

$r^{(1)} \leftarrow \min_{1 \leq i \leq m} h_1(x_i)$ ..... $r^{(k)} \leftarrow \min_{1 \leq i \leq m} h_k(x_i)$ $\quad k = \frac{1}{\varepsilon^2}$

$r \leftarrow (r^{(1)} + \cdots + r^{(k)})/k$: outputs $\frac{1}{r}$

With probability $1 - \varepsilon$, $\frac{1}{r}$ is within $1 \pm \varepsilon$ of M_d

Memory: k · (storing hash functions + register)

enough to store $\Theta(\log N)$ bits to error $\frac{1}{100N}$

$r \cdot \hat{r}$ $\quad r \geq \frac{1}{n} \cdot |\hat{r} - r| \leq \frac{1}{100n}$

$0.99 \leq \left(\frac{1}{\hat{r}}\right)/\left(\frac{1}{r}\right) \leq 1.01$

Instead of using a completely independent hash function, let's sample $h_1 \cdots h_k \in$ pairwise independent hash function family.

Suppose $a \neq b \in U$. $h_1(a), h_1(b) \in [0, 1]$. $h_1(a)$ & $h_1(b)$ are independent of each other.

$r^{(1)} \leftarrow \min_{1 \leq i \leq m} h_1(x^{(i)})$

Claim: With probability 60%, $\frac{M_d}{5} \leq \frac{1}{r^{(1)}} \leq 5 \cdot M_d$ Good answer

proof:

$r^{(1)} \leftarrow \min_{1 \leq i \leq m} h(x^{(i)})$

$\Pr\left[h(x^{(1)}) \leq \frac{1}{5M_d}\right] = \frac{1}{5M_d}$

HW 1

1. [15 points] Let $\mathcal{H}$ be a family of pairwise independent hash functions mapping $\{0,1\}^n$ to $\{0,1\}^k$. Let $A \subseteq \{0,1\}^n$ such that $|A| = p \cdot 2^k$. Suppose $1/4 \le p \le 1/2$. Let $\mathbf{0} \in \{0,1\}^k$ (this is the string with all k zeros).

1. For any $x \in \{0,1\}^n$, what is $\Pr_{h\in\mathcal{H}}[h(x) = \mathbf{0}]$? For $x \neq x' \in \{0,1\}^n$, what is $\Pr_{h\in\mathcal{H}}[h(x) = h(x') = \mathbf{0}]$? Justify.
2. Define the random variable $\mathcal{E}_{x,h}$ as 1 if $h(x) = \mathbf{0}$ and 0 otherwise. Define $\mathcal{E} = \sum_{x\in A} \mathcal{E}_{x,h}$. What is $\mathbf{E}[\mathcal{E}]$ and $\mathbf{E}[\mathcal{E}^2]$? Give your answer in terms of p and k (Hint: it's nothing but linearity of expectation).
3. Compute $\mathbf{E}[(\mathcal{E}-1)^2]$ (in terms of p and k) and use this to show that $\Pr[\mathcal{E} = 1] \ge 1/16$.

Then, it's easy to deduce that the probability that both $h(x)$ and $h(x')$ equal the 0 string is the product of these two individual probabilities. Therefore:

$$\Pr_{h\in\mathcal{H}}[h(x) = h(x') = 0] = \Pr_{h\in\mathcal{H}}[h(x) = 0 \wedge h(x') = 0]$$

$$= \Pr_{h\in\mathcal{H}}[h(x) = 0] \cdot \Pr_{h\in\mathcal{H}}[h(x') = 0] = \frac{1}{2^k} \cdot \frac{1}{2^k} = \frac{1}{2^{2k}}$$

(b) **Sub-question 1:**

Based on the previous question and the definition of this random variable $\mathcal{E}_{x,h}$, we know $\Pr[h(x) = 0] = \frac{1}{2^k}$, so it's easy to compute $\mathbb{E}[\mathcal{E}_{x,h}]$:

$$\mathbb{E}[\mathcal{E}_{x,h}] = \Pr[h(x) = 0] = \frac{1}{2^k}$$

We also know:

$$\mathcal{E} = \sum_{x\in A} \mathcal{E}_{x,h}$$

So,

$$\mathbb{E}[\mathcal{E}] = \mathbb{E}[\sum_{x\in A} \mathcal{E}_{x,h}] = \sum_{x\in A} \mathbb{E}[\mathcal{E}_{x,h}] = |A| \cdot \mathbb{E}[\mathcal{E}_{x,h}]$$

$$|A| \cdot \mathbb{E}[\mathcal{E}_{x,h}] = p \cdot 2^k \cdot \frac{1}{2^k} = p$$

Thus, $\mathbb{E}[\mathcal{E}]$ is equal to p.

Sub-question 2:

$$\mathbb{E}[\mathcal{E}^2] = \mathbb{E}[(\sum_{x\in A} \mathcal{E}_{x,h})^2]$$

There would be two scenarios of $(\sum_{x\in A} \mathcal{E}_{x,h})^2$:

The first one is: $\sum_{x\in A} \mathcal{E}_{x,h}^2$, while the second one is $\sum_{x\neq x'} \mathcal{E}_{x,h} \cdot \mathcal{E}_{x',h}$.

Thus, we can rewrite the equation to the following form:

$$\mathbb{E}[(\sum_{x\in A} \mathcal{E}_{x,h})^2] = \mathbb{E}[\sum_{x\in A} \mathcal{E}_{x,h}^2] + \mathbb{E}[\sum_{x\neq x'} \mathcal{E}_{x,h}\mathcal{E}_{x',h}]$$

Simplify it a bit:

$$= \sum_{x\in A} [\mathbb{E}[\mathcal{E}_{x,h}^2] + \sum_{x\in A} [\mathbb{E}[\mathcal{E}_{x,h}\mathcal{E}_{x',h}]$$

Because $\mathcal{E}_{x,h}$ is either 0 or 1, $(\mathcal{E}_{x,h})^2$ is either 0 or 1. This implies $\mathcal{E}_{x,h}^2 = \mathcal{E}_{x,h}$.

2. [10 points] Consider a biased coin such that the probability of Heads is 1/3 and the probability of Tails is 2/3. We toss the biased coin n times, where each toss is independent of the others. A run is a maximal contiguous sequence of either Heads or Tails. For example, the sequence HTTHTHHH has 5 runs. The longest run in this sequence has length 3. Show that with probability at least $1 - 1/n$, a sequence of n independent tosses will contain a run of length at least $\log_2 n - 2\log_2 \log_2 n$.

Hint: For a suitable choice of t, split the your n coin tosses into n/t blocks of size t each. Now, what is the probability that any of them is all Tails? Next, note the probability of all tails in disjoint blocks is independent.

Solution: The probability of a block of size t is all T, assuming the odds of T is 1/2, is $(1/2)^t$. Hence, the probability any n/t block is not all T is

$$\left(1 - \frac{1}{2^t}\right)^{n/t} = \left(1 - \frac{2^{\log_2(\log_2 n)^2}}{n}\right)^{\frac{n}{\log_2 n - 2\log_2\log_2 n}} \quad (1)$$

$$= \left(1 - \frac{(\log_2 n)^2}{n}\right)^{\frac{n}{\log_2 n - 2\log_2\log_2 n}} \quad (2)$$

$$= \left(1 - \frac{(\log_2 n)^2}{n}\right)^{\frac{n}{(\log_2 n)^2} \cdot \frac{(\log_2 n)^2}{n} \cdot \frac{n}{\log_2 n - 2\log_2\log_2 n}} \quad (3)$$

$$\le \exp(-\frac{(\log_2 n)^2}{\log_2 n - 2\log_2\log_2 n}) \quad (4)$$

$$\le \exp(-\log_2 n) \le 1/n, \quad (5)$$

taking t to be $\log_2 n - 2\log_2\log_2 n$, where the inequality on the fourth line is by the fact that $(1 - \frac{1}{x})^x \le e^{-1}$ for all $x \ge 1$.

1. (a) **Sub-question 1:**

Because $\mathcal{H}$ is a 2-wise independent hash function family, every hash value in $\{0,1\}^k$ should be equally likely for any input x from $\{0,1\}^n$.

We also know $\mathbf{0} \in \{0,1\}^k$, so we deduce this string 0 is merely one of all the possible outputs of the hash function.

Because $\mathcal{H}$ is a 2-wise independent hash function family, it's also 1-wise independent. Based on the definition of 1-wise independent hash function, for every $x \in \{0,1\}^n$ and for every $y \in \{0,1\}^k$, each possible output value $y \in \{0,1\}^k$ is equally likely for any given input x:

$$\Pr_{h\in\mathcal{H}}[h(x) = y] = \frac{1}{\textit{the number of all the possible outputs of the hash function}}$$

We know this pairwise independent hash function family maps $\{0,1\}^n$ to $\{0,1\}^k$, so the size of inputs is 2^n, and the size of outputs is 2^k. Thus:

$$\Pr_{h\in\mathcal{H}}[h(x) = 0] = \frac{1}{2^k}$$

Sub-question 2:

Since $\mathcal{H}$ is a pairwise independent hash function family, the values of $h(x)$ and $h(x')$ are independent (we assume $x \neq x'$).

We have already known: $\Pr_{h\in\mathcal{H}}[h(x) = 0] = \frac{1}{2^k}$, so we can deduce the probability of $h(x) = 0$ and $h(x') = 0$ is $\frac{1}{2^k}$ and $\frac{1}{2^k}$, respectively, as the following equations show:

$$\Pr_{h\in\mathcal{H}}[h(x) = 0] = \frac{1}{2^k}$$

Similarly,

$$\Pr_{h\in\mathcal{H}}[h(x') = 0] = \frac{1}{2^k}$$

$$= p^2 - \frac{p}{2^k} - p + 1$$

Sub-question 2:

Since k should be much larger than p, we can deduce that:

$$\mathbb{E}[(\mathcal{E}-1)^2] = p^2 - \frac{p}{2^k} - p + 1 \le p^2 - p + 1$$

Based on Markov's inequality, we know that for any non-negative random variable $\mathcal{X}$ and a such that $a > 0$,

$$\Pr(\mathcal{X} \ge a) \le \frac{\mathbb{E}[\mathcal{X}]}{a}$$

By plugging in $\mathcal{X} = (\mathcal{E}-1)^2$, we can get the following equation:

$$\Pr[(\mathcal{E}-1)^2 \ge a] \le \frac{\mathbb{E}[(\mathcal{E}-1)^2]}{a}$$

Let a $= 1$, we can rewrite the equation to:

$$\Pr[(\mathcal{E}-1)^2 \ge 1] \le \mathbb{E}[(\mathcal{E}-1)^2]$$

Since $\mathbb{E}[(\mathcal{E}-1)^2] \le p^2 - p + 1$, we know:

$$\Pr[(\mathcal{E}-1)^2 \ge 1] \le p^2 - p + 1$$

$(\mathcal{E}-1)^2 \ge 1$ is only possible when $\mathcal{E} \neq 1$, so we can deduce:

$$\Pr[(\mathcal{E}-1)^2 \ge 1] = \Pr[(\mathcal{E} \neq 1)] \le p^2 - p + 1$$

Since we know $\frac{1}{4} \le p \le \frac{1}{2}$, we can plug in $p = \frac{1}{4}$ and get this equation:

$$\Pr[(\mathcal{E} \neq 1)] \le p^2 - p + 1 = \frac{13}{16}$$

Thus,

$$\Pr[(\mathcal{E} = 1)] = 1 - \Pr[(\mathcal{E} \neq 1)] \ge 1 - p^2 - p + 1 = 1 - \frac{13}{16} = \frac{3}{16}$$

$$\Pr[(\mathcal{E} = 1)] \ge \frac{3}{16} \ge \frac{1}{16}$$

Therefore,

$$\Pr[(\mathcal{E} = 1)] \ge \frac{1}{16}$$

3. [15 points] For any permutation $\sigma : [n] \to [n]$, we say $1 \le i_1 < i_2 < \ldots < i_k \le n$ is an increasing subsequence if $\sigma(i_1) < \sigma(i_2) \ldots < \sigma(i_k)$. For any permutation σ, let $LIS(\sigma)$ denote the length of the maximum increasing subsequence in σ.

Suppose we choose a random permutation $\sigma : [n] \to [n]$. Prove that with probability 9/10, $LIS(\sigma) = O(\sqrt{n})$.

To prove this, two estimates would be helpful: for any $1 \le k \le n$

$$\binom{n}{k} \le (en/k)^k; \quad k! \ge (k/e)^k.$$

3 Question 3

From a set of n elements, we select k elements $1, i_1, i_2, \ldots, i_k$. The probability that its permutation $\sigma(i_1) < \sigma(i_2) < \cdots < \sigma(i_k)$ is:

$$\Pr = \frac{C(n,k)}{P(n,k)} = \frac{1}{k!}$$

Let X be a random variable, denoting the number of increasing sub-sequences of length k in a random permutation $\sigma : [n] \to [n]$. Then we can deduce that the expected value of X is:

$$\mathbb{E}(X) = \sum x \cdot \Pr(X = x) = \binom{n}{k} \frac{1}{k!} \leq (\frac{e^2 n}{k^2})^k$$

Let $k = C\sqrt{n}$, then the expectation of X under this k is:

$$\mathbb{E}(X_k) \leq (\frac{e}{c})^{2C \cdot \sqrt{n}}$$

Using Markov's inequality, we can get the following equation:

$$\Pr[X \geq t \cdot \mathbb{E}(X)] \leq \frac{1}{t}$$

Thus:

$$\Pr[X_k \geq t \cdot (\frac{e}{c})^{2C \cdot \sqrt{n}}] \leq \frac{1}{t}$$

Set $t = (\frac{e}{C})^{-2C \cdot \sqrt{n}}$. Then:

$$\Pr[X_k \geq 1] \leq \frac{1}{t}$$

This is equivalent to say that with probability less and equal to $\frac{1}{t}$, there is at least one increasing sub-sequence of length k in permutation. i.e.

$$\Pr[\text{LIS}(\sigma) \geq k] \leq \frac{1}{t}$$

To prove $\exists C' \in \mathbb{R}, \Pr(\text{LIS}(\sigma) < C'\sqrt{n}) > \frac{9}{10}$, is equivalent to prove that:

$$\Pr[\text{LIS}(\sigma) \geq C'\sqrt{n}] \leq \frac{1}{10}$$

Set $c = 2e$, then $t = (\frac{1}{2})^{-4e\sqrt{n}}$. Based on this, we can deduce:

$$\Pr[\text{LIS}(\sigma) \geq 2e\sqrt{n}] \leq (\frac{1}{2})^{4e\sqrt{n}}$$

Since $n \geq 1$, we can conclude:

$$\Pr[\text{LIS}(\sigma) \geq 2e\sqrt{n}] \leq (\frac{1}{2})^{4e\sqrt{n}} \leq (\frac{1}{2})^{4e} \leq (\frac{1}{2})^4 = \frac{1}{16} \leq \frac{1}{10}$$

Thus,

$$\exists C' = 2e \in \mathbb{R}, \Pr(\text{LIS}(\sigma) < C'\sqrt{n}) > \frac{9}{10}$$

Thus,

$$\text{LIS}(\sigma) = O(\sqrt{n})$$

Because we know $\text{Var}[f(Z)] \leq \mathbb{E}[(f(Z) - f(\mathbb{E}[Z]))^2]$, the equation becomes:

$$\text{Var}[f(Z)] \leq \mathbb{E}[(f(Z) - f(\mathbb{E}[Z]))^2] \leq \mathbb{E}[(Z - \mathbb{E}[Z])^2] = \text{Var}[Z]$$

Thus, when f is a 1-Lipschitz function.

$$\text{Var}[f(Z)] \leq \text{Var}[Z]$$

2. Suppose W is a random variable. Based on the definition of variance, we can write the variance of W as:

$$\text{Var}(W) = E[W^2] - (E[W])^2$$

Based on the question, we let $W = X - \beta Y$, where β is a constant.

$$W^2 = (X - \beta Y)^2 = X^2 - 2\beta XY + \beta^2 Y^2$$

Taking the expected values of both sides, we get:

$$E[W^2] = E[X^2] - 2\beta E[XY] + \beta^2 E[Y^2]$$

Because variance of any random variable should be greater or equal to 0, we know $\text{Var}(W) \geq 0$. Thus, $E[W^2] - (E[W])^2 \geq 0$

Let us assume $E[W] = 0$ to simplify this question a bit. Then:

$$E[W^2] \geq 0$$

$$E[X^2] - 2\beta E[XY] + \beta^2 E[Y^2] \geq 0$$

Then we want to find a value of β such that $E[W^2]$ can be as small as possible (which is 0 in this case).

Taking the derivative with respect to β:

$$\frac{d}{d\beta}\left(E[X^2] - 2\beta \cdot E[XY] + \beta^2 \cdot E[Y^2]\right) = -2 \cdot E[XY] + 2\beta \cdot E[Y^2] = 0$$

$$\beta = \frac{E[XY]}{E[Y^2]}$$

Plugging in $\beta = \frac{E[XY]}{E[Y^2]}$ to the original inequality $E[X^2] - 2\beta E[XY] + \beta^2 E[Y^2] \geq 0$, we get:

$$E[X^2] - 2\left(\frac{E[XY]}{E[Y^2]}\right)E[XY] + \left(\frac{E[XY]}{E[Y^2]}\right)^2 E[Y^2] \geq 0$$

$$E[X^2] - \frac{E[XY]^2}{E[Y^2]} \geq 0$$

$$E[XY]^2 \leq E[X^2]E[Y^2]$$

Taking the square root of both sides

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}$$

Thus,

$$E[|X \cdot Y|] \leq \sqrt{E[X^2]E[Y^2]}$$

1. [15 points] The goal of this question is to prove two basic probabilistic inequalities:
(i) [7 points] We call a function $f : \mathbb{R} \to \mathbb{R}$ to be 1-Lipschitz if for all $x, y \in \mathbb{R}$, $|f(x) - f(y)| \leq |x - y|$. Let $\mathbf{Z}$ be a real-valued random variable. Prove that $\mathsf{Var}[f(\mathbf{Z})] \leq \mathsf{Var}(\mathbf{Z})$.

Hint: Think about all the characterizations of variance we've learnt in the class. With the right one, this is a one line proof.

(ii) [8 points] Prove that for any two real-valued random variables $\mathbf{X}$ and $\mathbf{Y}$ (possibly correlated), $\mathbf{E}[|\mathbf{X} \cdot \mathbf{Y}|] \leq \sqrt{\mathbf{E}[\mathbf{X}^2]\mathbf{E}[\mathbf{Y}^2]}$.

Hint: To prove (ii), consider the random variable $\mathbf{W} = \mathbf{X} - \beta\mathbf{Y}$ and use the fact that $\mathsf{Var}(\mathbf{W}) \geq 0$ for every choice of β. Now put in a suitable choice of β.

1. We know that for any random variable X:

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

We can deduce that $\mathbb{E}[X]$ is the minimum mean squared deviation from any constant. Then, suppose c is a constant.

$$\text{Var}[X] = \min_c \mathbb{E}[(X - c)^2]$$

This implies that for any constant c,

$$\text{Var}[X] \leq \mathbb{E}[(X - c)^2]$$

Applying this equation to $X = f(Z)$, we can get:

$$\text{Var}[f(Z)] \leq \mathbb{E}[(f(Z) - c)^2]$$

If $c = f(\mathbb{E}[Z])$, then the equation becomes:

$$\text{Var}[f(Z)] \leq \mathbb{E}[(f(Z) - f(\mathbb{E}[Z]))^2]$$

Because f is 1-Lipschitz, for any $x, y \in \mathbb{R}$, we have:

$$|f(x) - f(y)| \leq |x - y|$$

If $x = Z$ and $y = \mathbb{E}[Z]$, then we can get:

$$|f(Z) - f(\mathbb{E}[Z])| \leq |Z - \mathbb{E}[Z]|$$

Squaring both sides and taking the expected values of both sides:

$$(f(Z) - f(\mathbb{E}[Z]))^2 \leq (Z - \mathbb{E}[Z])^2.$$

$$\mathbb{E}\left[(f(Z) - f(\mathbb{E}[Z]))^2\right] \leq \mathbb{E}[(Z - \mathbb{E}[Z])^2] = \text{Var}[Z]$$

2. [15 points] For any undirected graph $G = (V, E)$ and a subset $S \subseteq V$, we define $\mathsf{cut}(S) = \{(u, v) : u \in S, v \notin S\}$. In other words, $\mathsf{cut}(S)$ is the set of edges with one end in S and the other in $\overline{S}$. Given a graph $G = (V, E)$, finding the cut S which maximizes $|\mathsf{cut}(S)|$ is NP-hard. So, we often consider approximation algorithms for this problem.
(i) Consider choosing a subset $\mathbf{S} \subseteq V$ uniformly at random (i.e., include each vertex in $\mathbf{S}$ with probability 1/2). What is $\mathbf{E}[|\mathsf{cut}(\mathbf{S})|]$ and $\mathsf{Var}(\mathsf{cut}(\mathbf{S}))$?

(ii) Prove that for such a random subset $\mathsf{S} \subseteq V$, $\Pr[|\mathsf{cut}(\mathbf{S})| \geq |E|/2] \geq c/|E|$, for a positive constant $c > 0$.

Solution:

(i) For every edge $e \in E$, define indicator variable $\mathbf{I}_e$ to be 1 if $e \in cut(\mathbf{S})$ and 0 otherwise. By constructing $\mathbf{S}$, $\mathbf{I}_e = 1$ when exactly one of the endpoints is included, which happens with probability 1/2. By linearity of expectation, we have

$$\mathbf{E}[|cut(\mathbf{S})|] = \mathbf{E}[\sum_{e \in E} \mathbf{I}_e] = \sum_{e \in E} \mathbf{E}[\mathbf{I}_e] = \frac{|E|}{2}$$

$$\begin{aligned}
Var[|cut(\mathbf{S})|] =& Var[\sum_{e \in E} \mathbf{I}_e] \\
=& \mathbf{E}[(\sum_{e \in E} \mathbf{I}_e)^2] - \frac{|E|^2}{4} \\
=& \sum_{e \in E} \mathbf{E}[\mathbf{I}_e] + \sum_{e \in E} \sum_{e' \in E, e' \neq e} \mathbf{E}[\mathbf{I}_e \mathbf{I}_{e'}] - \frac{|E|^2}{4} \\
=& \frac{|E|}{2} + \sum_{e,e' \text{ disjoint}} \mathbf{E}[\mathbf{I}_e \mathbf{I}_{e'}] + \sum_{e,e' \text{ share an endpoint}} \mathbf{E}[\mathbf{I}_e \mathbf{I}_{e'}] - \frac{|E|^2}{4} \\
=& \frac{|E|}{2} + \frac{(\#\text{ of disjoint pairs of edges})}{4} \\
&+ \frac{(\#\text{ of adjacent pairs of edges})}{4} - \frac{|E|^2}{4} \\
=& \frac{|E|}{2} + \frac{|E|^2 - |E|}{4} - \frac{|E|^2}{4} \quad \text{(ii)} \\
=& \frac{|E|}{4}.
\end{aligned}$$

$$\begin{aligned}
\frac{|E|}{2} =& \mathbf{E}[|cut(\mathbf{S})|] && (5) \\
\leq& P(|cut(\mathbf{S})| < \frac{|E|}{2}) \cdot (\frac{|E|}{2} - \frac{1}{2}) + P(|cut(\mathbf{S})| \geq \frac{|E|}{2}) \cdot |E| && (6) \\
=& (1 - P(|cut(\mathbf{S})| \geq \frac{|E|}{2})) \cdot (\frac{|E|}{2} - \frac{1}{2}) + P(|cut(\mathbf{S})| \geq \frac{|E|}{2}) \cdot |E| && (7) \\
=& \frac{|E|}{2} - \frac{1}{2} + P(|cut(\mathbf{S})| \geq \frac{|E|}{2}) \cdot (\frac{1}{2} + \frac{|E|}{2}) && (8)
\end{aligned}$$

(Line (6) is in fact implicitly using the law of total expectation, where the two events are $|cut(\mathbf{S})| < |E|/2$ and $|cut(\mathbf{S})| \geq |E|/2$. When the first event happens it must be the case that $|cut(\mathbf{S})| \leq |E|/2 - 1/2$, where the fraction accounts for the possibility that $|E|$ is odd.)

Thus, we have

$$\begin{aligned}
&P(|cut(\mathbf{S})| \geq \frac{|E|}{2}) \cdot (\frac{1}{2} + \frac{|E|}{2}) \geq \frac{1}{2} \\
\Rightarrow& P(|cut(\mathbf{S})| \geq \frac{|E|}{2}) \geq \frac{1}{1 + |E|} \geq \frac{1}{2|E|}.
\end{aligned}$$

3. [15 points] Let us assign M items to N bins by allocating each item to a bin uniformly and independently at random. Let $\mathbf{C}$ denote the random variable which counts collisions – i.e., each pair (i, j) which gets mapped to the same bin, you add 1 to $\mathbf{C}$.

(i) [5 points] What is $\mathbf{E}[\mathbf{C}]$? Use this to show that if $M \leq 0.01\sqrt{N}$, then every item gets its own bin with probability at least 9/10.

(ii) [10 points] Compute $\mathbf{E}[\mathbf{C}^2]$. Use this to show that if $M \geq 0.01\sqrt{N}$, with some constant probability $\gamma > 0$, there is a collision, i.e., two items get assigned to the same bin.

1. If we assign M items to N bins uniformly and independently, we can easily get the following conclusion:

   Suppose there are two items x and y ∈ M. The possibility that x would collide with y is $1/N$.

   If we select two items from M, then the number of possible combinations is:

$$\binom{M}{2} = \frac{M(M-1)}{2}$$

   Thus,

$$E[C] = \frac{1}{N} \cdot \frac{M(M-1)}{2} = \frac{M(M-1)}{2N}$$

   When $M \leq 0.01\sqrt{N}$, we can rewrite $E[C]$ as follows:

$$E[C] \leq \frac{(0.01\sqrt{N})(0.01\sqrt{N}-1)}{2N} \approx \frac{0.0001N}{2N} = \frac{0.0001}{2}$$

$$E[C] \leq \frac{0.0001}{2} = 0.00005$$

   Therefore, the probability that every item gets its own bin is at least (using Markov's inequality):

$$P(C=0) = 1 - P(C \geq 1) \geq 1 - 0.00005 = 0.99995 \geq \frac{9}{10}$$

   Thus, when $M \leq 0.01\sqrt{N}$, every item gets its own bin with probability at least $\frac{9}{10}$.

Taking $M \leq 0.01\sqrt{N}$, $\mathbf{E}[\mathbf{C}] \leq \frac{M^2-M}{20000M^2} \leq 1/20000$. Therefore, $P(M \geq 1) \leq 1/20000$ by Markov's inequality, so each item gets a unique bin with probability at least 19999/20000.

2. To compute $E[C^2]$, we start by expressing C in terms of indicator random variables. Let X_{ij} be the indicator that two items i and j are assigned to the same bin. Then,

$$C = \sum_{1 \leq i < j \leq M} X_{ij}.$$

$$E[C] = \sum_{i<j} E[X_{ij}] = \binom{M}{2} \cdot \frac{1}{N} = \frac{M(M-1)}{2N}$$

   Now, we need to compute $E[C^2]$. Let X_{kl} be another indicator that two items k and l are assigned to the same bin:

$$C^2 = (\sum_{i<j} X_{ij})^2 = \sum_{i<j} X_{ij}^2 + \sum_{i<j,\, k<l} X_{ij}X_{kl}$$

   We know: $1 \leq i < j \leq M$ and $1 \leq k < l \leq M$.
   We know $X_{ij}^2 = X_{ij}$ since it's a indicator variable (it can either be 1 or 0). Then, we get the following equation:

$$E[C^2] = \sum_{i<j} E[X_{ij}] + \sum_{i<j,\, k<l} E[X_{ij}X_{kl}] = E[C] + \sum_{i<j,\, k<l} E[X_{ij}X_{kl}]$$

$$E[C^2] = \frac{M(M-1)}{2N} + \sum_{i<j,\, k<l} E[X_{ij}X_{kl}]$$

   Then, we need to consider computing $E[X_{ij}X_{kl}]$. We will discuss it under three different scenarios.
   (1)Assuming $(i, j) \neq (k, l)$: In other words, i, j, k, l are four different items. Because the two event are independent, we have:

$$E[X_{ij}X_{kl}] = E[X_{ij}] \cdot E[X_{kl}] = (\frac{1}{N})^2 = \frac{1}{N^2}$$

   The number of possible terms under this scenario is $\binom{M}{2}\binom{M-2}{2}$ as we want to select 4 different points from M.
   Thus,

$$\begin{aligned}\sum_{i<j,\, k<l} E[X_{ij}X_{kl}] &= \frac{M(M-1)}{2} \cdot \frac{(M-2)(M-3)}{2} \cdot \frac{1}{N^2} \\ &= \frac{M(M-1)(M-2)(M-3)}{4N^2}\end{aligned}$$

   (2)Assuming there is one overlapping item:

   When there is only one overlapping item, the probability that both pairs collide is the probability that the three items are all in the same bin (suppose i is the overlapping item):

$$E[X_{ij}X_{il}] = E[X_{ij}] \cdot E[X_{il}] = \frac{1}{N^2}$$

   The number of possible terms is $M \cdot \binom{M-1}{2} \cdot 2$ as we want to find one overlapping item from all items first and then find the other two items from the remaining $M-1$ items.
   Thus,

$$\begin{aligned}\sum_{i<j,\, k<l} E[X_{ij}X_{il}] &= M \cdot \binom{M-1}{2} \cdot 2 \cdot \frac{1}{N^2} \\ &= \frac{M(M-1)(M-2)}{N^2}\end{aligned}$$

   (3)Assuming (i, j) is the same with (k, l):
   This situation is the same with $\sum_{i<j} E[X_{ij}]$, so we won't consider it here.

   Thus,

$$\begin{aligned}E[C^2] &= E[C] + \frac{M(M-1)(M-2)(M-3)}{4N^2} + \frac{M(M-1)(M-2)}{N^2} \\ &\approx \frac{M^2}{2N} + \frac{M^4}{4N^2} + \frac{M^3}{N^2}\end{aligned}$$

   Computing $(E[C])^2$:

$$\begin{aligned}(E[C])^2 &\approx (\frac{M^2}{2N})^2 \\ &\approx \frac{M^4}{4N^2}\end{aligned}$$

   Calculating $\frac{(E[C])^2}{E[C^2]}$:

$$\frac{(E[C])^2}{E[C^2]} = \frac{(E[C])^2}{\frac{M^2}{2N} + \frac{M^4}{4N^2} + \frac{M^3}{N^2}} = \frac{\frac{M^4}{4N^2}}{\frac{M^2}{2N} + \frac{M^4}{4N^2} + \frac{M^3}{N^2}}$$

   The Paley-Zygmund inequality states for a non-negative random variable Z, $0 < \alpha < 1$:

$$Pr(Z > \alpha E[Z]) \geq (1-\alpha)^2 \frac{(E[Z])^2}{E[Z^2]}$$

Let $\alpha = \frac{1}{2}$, then:

$$\begin{aligned}Pr(C > \frac{E[C]}{2}) &\geq \left(1 - \frac{1}{2}\right)^2 \frac{(E[C])^2}{E[C^2]} \\ &\geq (\frac{1}{2})^2 \times \frac{(E[C])^2}{E[C^2]} \\ &\geq \frac{1}{4} \times \frac{(E[C])^2}{E[C^2]} \\ &\geq \frac{1}{4} \times \frac{\frac{M^4}{4N^2}}{\frac{M^2}{2N} + \frac{M^4}{4N^2} + \frac{M^3}{N^2}}\end{aligned}$$

We know C can only be non-negative integer, so $Pr(C \geq 1) = Pr(C > \frac{1}{2})$.
Since $M \geq 0.01\sqrt{N}$, we can set $M = \sqrt{2N}$. We know N is a positive integer:

$$Pr(C > \frac{E[C]}{2}) = Pr(C > \frac{\frac{M^2}{2N}}{2}) = Pr(C > \frac{M^2}{4N}) = Pr(C > \frac{1}{2})$$

$$\begin{aligned}Pr(C > \frac{1}{2}) &\geq \frac{1}{4} \times \frac{\frac{M^4}{4N^2}}{\frac{M^2}{2N} + \frac{M^4}{4N^2} + \frac{M^3}{N^2}} \\ &\geq \frac{4N^2}{\frac{4N^2}{\frac{2N}{2N} + \frac{4N^2}{4N^2} + \frac{2N \cdot \sqrt{2N}}{N^2}}} \cdot \frac{1}{4} \\ &\geq \frac{1}{1 + 1 + \frac{2N \cdot \sqrt{2N}}{N^2}} \cdot \frac{1}{4} \\ &\geq \frac{1}{2 + \frac{2N \cdot \sqrt{2N}}{N^2}} \cdot \frac{1}{4} \\ &\geq \frac{1}{8 + \frac{8N \cdot \sqrt{2N}}{N^2}}\end{aligned}$$

Since N is a positive integer, $\frac{1}{8 + \frac{8N \cdot \sqrt{2N}}{N^2}}$ is a positive constant.
Thus,

$$Pr(C \geq 1) = Pr(C > \frac{1}{2}) \geq \frac{1}{8 + \frac{8N \cdot \sqrt{2N}}{N^2}} \geq 0$$

Therefore, if $M \geq 0.01N$, with some constant probability $\gamma > 0$, there is a collision.

Practice problems

1) Let G be a random graph on n vertices obtained by joining any two vertices independently with probability $1/2$. Calculate the expected number of k-cliques as a function of n and k.

Now, prove that for $k_0 = 2\log_2 n + 2$, with probability $1 - n^{-\Theta(1)}$, $G(n, 1/2)$ does not have a clique of size more than k_0.

Solution:

In any fixed set of k vertices, the number of edges is $\binom{k}{2}$. Thus, for any fixed set of k vertices (say A), $Pr[A$ is a clique in $G] = 1/2^{\binom{k}{2}}$. Thus, for any set of vertices A, if $\mathcal{E}_A$ is a random variable which is 1 iff there is a clique on A, then $E[\mathcal{E}_A] = 1/2^{\binom{k}{2}}$. Thus, if $\mathcal{E}$ denotes the random variable corresponding to the number of k-cliques, then $\mathcal{E} = \sum_{A:|A|=k} \mathcal{E}_A$. By linearity of expectation, $E[\mathcal{E}] = \binom{n}{k} \cdot 2^{-\binom{k}{2}}$.

Now, by the approximation we learnt in class,

$E[\mathcal{E}] \leq 2^{-k(k-1)/2} \cdot (ne/k)^k = (2^{-(k-1)/2} \cdot ne/k)^k$. If you put $k_0 = 2\log_2 n + 2$, then the right hand side becomes at most $n^{-\Theta(1)}$. The bound now follows by Markov.

2) Let $p(x_1, \ldots, x_n) = \sum_{S \subseteq [n]} a_S \prod_{i \in S} x_i$ where $a_S \in \mathbb{R}$. Further, suppose $\sum_{S \subseteq [n]} a_S^2 = 1$.

Suppose $\mathbf{z}_1, \ldots, \mathbf{z}_n$ be independent unbiased ± 1 random variables. Prove that with $\Pr[|p(\mathbf{z}_1, \ldots, \mathbf{z}_n)| \geq c] \leq 1/c^2$.

Solution:

Observe that

$p(z_1, \ldots, z_n)^2 = \sum_{S,T} a_S a_T \prod_{i \in S} z_i \prod_{j \in T} z_j$.

Thus, by linearity of expectation, we have $E[p(z_1, \ldots, z_n)^2] = \sum_{S,T} a_S a_T E[\prod_{i \in S} z_i \prod_{j \in T} z_j]$. Now, note that whenever $S \neq T$, the term will vanish because if $S \neq T$, then one of the z_i's will appear as an odd power. Thus,

$E[p(z_1, \ldots, z_n)^2] = \sum_S a_S^2 = 1$.

Now, the claim follows by Markov.

3) Suppose Alice and Bob are two parties who both hold a vector each (say $a \in \{0,1\}^n$ and $b \in \{0,1\}^n$). Alice just has a and Bob just has b. They would like to verify whether $a = b$ while minimizing communication between them.

Of course, Alice can send her vector to Bob but that would mean $O(n)$ bits of communication. To obtain a better protocol, suppose a third party Eve samples a uniformly random string $r \in \{0,1\}^n$ and sends it to Alice and Bob. Now, Alice computes $\sum_{i=1}^n a_i r_i$ and Bob computes $\sum_{i=1}^n b_i r_i$. Then, Alice sends the parity of $\sum_{i=1}^n a_i r_i$ to Bob (i.e., if $\sum_{i=1}^n a_i r_i$ is even and 0 if it is odd). Bob also checks the parity of $\sum_{i=1}^n b_i r_i$.

(a) If $a = b$, then what is the probability that the parity of $\sum_{i=1}^n a_i r_i$ equals to parity of $\sum_{i=1}^n b_i r_i$?

(b) If $a \neq b$, then what is the probability that the parity of $\sum_{i=1}^n a_i r_i$ equals to parity of $\sum_{i=1}^n b_i r_i$?

(c) If instead of one uniformly random string r, suppose Eve sent them k uniformly random strings $r^{(1)}$, . . ., $r^{(k)}$ -- what will the probabilities in the first two parts be now?

Solution:

(a) If $a = b$, $\sum_{i=1}^n a_i r_i = \sum_{i=1}^n b_i r_i$ so the event happens with probability 1.

(b) (All calculations here are modulo 2 / in $\mathbb{F}_2$.) Since $a \neq b$, without loss of generality we assume $a_1 \neq b_1$. Then $P(a_1 r_1 \neq b_1 r_1) = 1/2$. Let $P(\sum_{i=2}^n a_i r_i = 0, \sum_{i=2}^n a_i r_i = 0) = \alpha, P(\sum_{i=2}^n a_i r_i = 0, \sum_{i=2}^n a_i r_i = 1) = \beta, P(\sum_{i=2}^n a_i r_i = 1, \sum_{i=2}^n a_i r_i = 0) = \theta$, then $P(\sum_{i=2}^n a_i r_i = 1, \sum_{i=2}^n a_i r_i = 1) = 1 - \alpha - \beta - theta$. It's straightforward to check that $P(\sum_{i=1}^n a_i r_i = \sum_{i=1}^n b_i r_i) = P(\sum_{i=2}^n a_i r_i = 0, \sum_{i=2}^n a_i r_i = 0) * P(a_1 r_1 = b_1 r_1) + P(\sum_{i=2}^n a_i r_i = 1, \sum_{i=2}^n a_i r_i = 1) * P(a_1 r_1 = b_1 r_1) + P(\sum_{i=2}^n a_i r_i = 0, \sum_{i=2}^n a_i r_i = 1) * P(a_1 r_1 \neq b_1 r_1) + P(\sum_{i=2}^n a_i r_i = 1, \sum_{i=2}^n a_i r_i = 0) * P(a_1 r_1 \neq b_1 r_1) = 1/2$.

(c) For part (a), each k iteration Alice and Bob will have the same bit so the probability of the event is 1. For the second part, it amounts to repeating the process in (b) k times where each time is independent of the other. This is a binomial distribution, and the probability that each k iteration provides the same bit is $1/2^k$.

4) Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent random variables where each $\mathbf{X}_i$ is an independent ± 1 unbiased random variable. Let $\mathbf{Z} = \sum_i \mathbf{X}_i$. Prove that there is some constant $c > 0$ such that $\Pr[|\mathbf{Z}| \geq 0.01\sqrt{n}] \geq c$.

Solution:

Since $Var[\mathbf{Z}] = \sum_i Var[\mathbf{X}_i] = n$, $E[\mathbf{Z}^2] = n$.

Notice that $E[\mathbf{Z}^4] = \sum_i E[\mathbf{X}_i^4] + 4\sum_i \sum_{j \neq i} E[\mathbf{X}_i^3 \mathbf{X}_j] + 3\sum_i \sum_{j \neq i} E[\mathbf{X}_i^2 \mathbf{X}_j^2] + 6\sum_i \sum_{j \neq i} \sum_{k \neq i, k \neq j} E[\mathbf{X}_i^2 \mathbf{X}_j \mathbf{X}_k] + \sum_i \sum_{j \neq i} \sum_{k \neq i, k \neq j} \sum_{l \neq i, l \neq j, l \neq k} E[\mathbf{X}_i \mathbf{X}_j \mathbf{X}_k \mathbf{X}_l] = n + 3n(n-1) = 3n^2 - 2n$.

Then, by Paley-Zygmund inequality, $\Pr(|\mathbf{Z}| > 0.01\sqrt{n}) = \Pr(\mathbf{Z}^2 > 0.0001n) = 0.9999^2 \frac{n^2}{3n^2 - 2n} > 0.9999^2 \frac{1}{3}$.