Probabilistic Inequalities and Examples

Lecture 3

Aug 30, 2022

Outline

Probabilistic Inequalities

Markov's Inequality

Chebyshev's Inequality

Bernstein-Chernoff-Hoeffding bounds

Some examples

Motivation

- Random variable $Q = \#comparisons$ made by randomized **QuickSort** on an array of n elements.
- We proved that $\mathrm{E}[Q] \leq 2n \ln n$.
- But we want to know more because expectation is only one basic piece of information. For instance what is $\Pr[Q \geq 10n \ln n]$? What is $Var[Q]$?
- Of course we would like to know the full distribution of Q but it is not feasible in many cases because Q is the outcome of a non-trivial algorithm.
- Even when we know the full distribution we don't want complex formulas but nice simple closed forms that help us understand the behaviour of a random variable in intuitive ways.

Binomial distribution

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Let X be the random variable that counts the number of 1s.

Binomial distribution

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Let X be the random variable that counts the number of 1s.

X has the well known Binomial distribution with $p = 1/2$:
$\Pr[X = k] = \binom{n}{k} 1/2^n$.

$\mathrm{E}[X] = n/2$

$Var[X] = n/4$

Binomial distribution

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Let X be the random variable that counts the number of 1s.

X has the well known Binomial distribution with $p = 1/2$:
$\Pr[X = k] = \binom{n}{k} 1/2^n$.

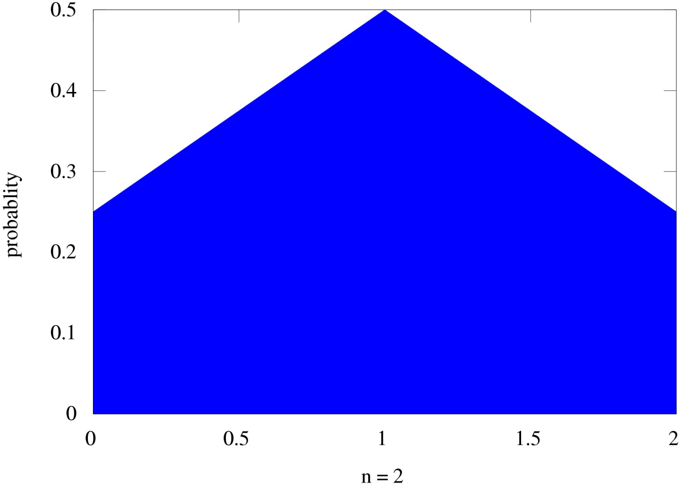
$E[X] = n/2$

$Var[X] = n/4$

Despite knowing the exact distribution it is hard to grasp how X behaves without some analysis of binomial coefficients etc. Let's plot.

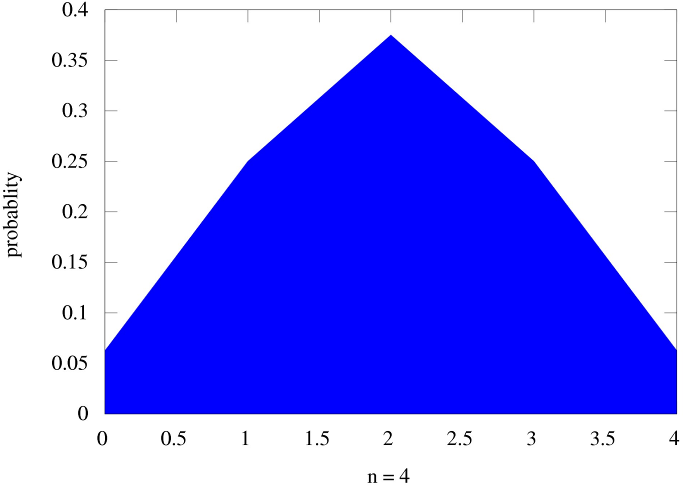
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



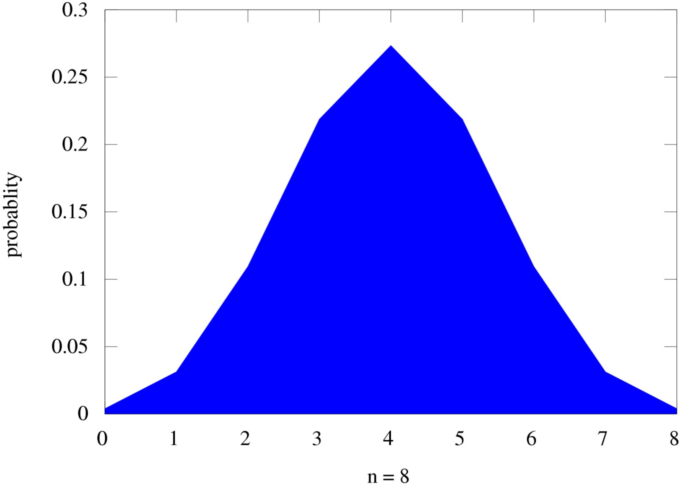
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



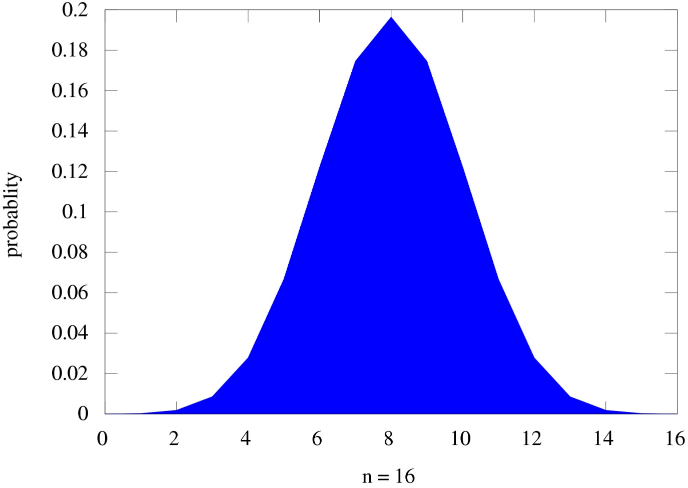
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



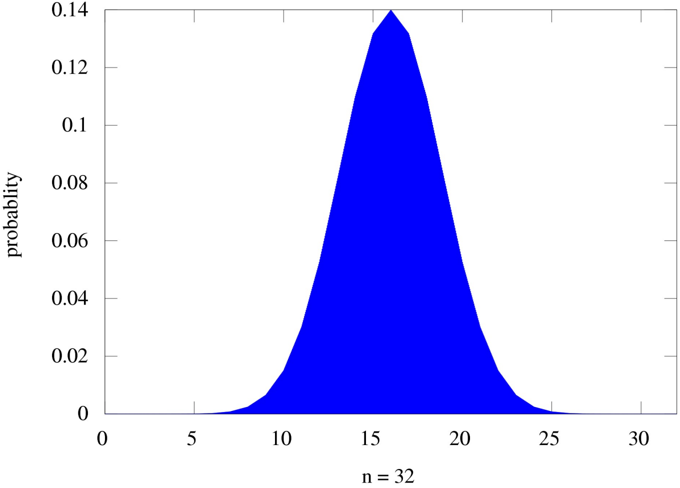
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



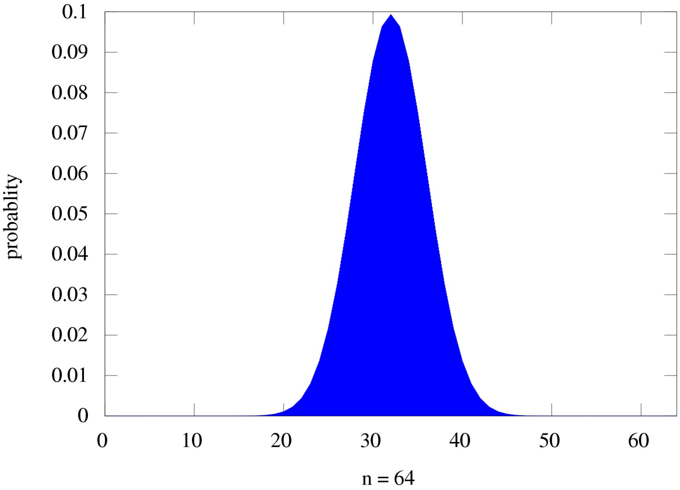
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



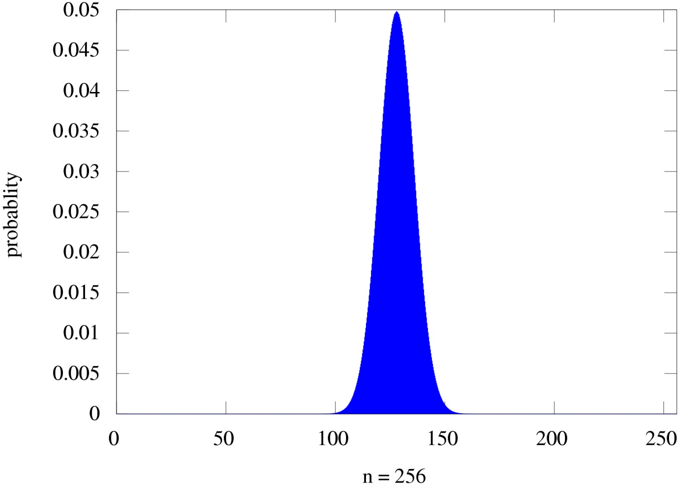
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



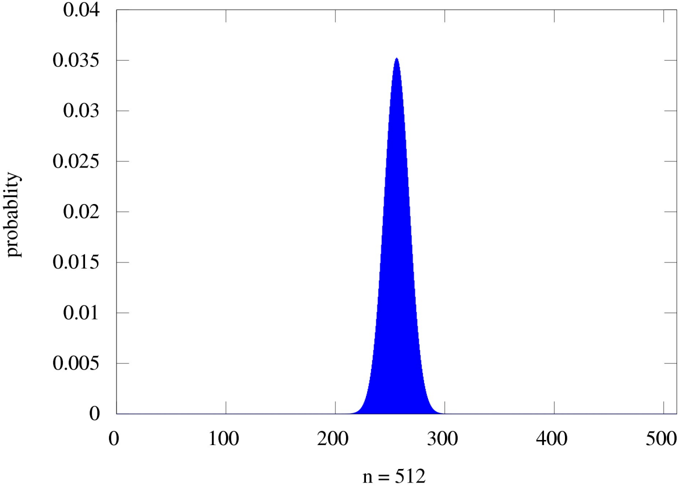
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



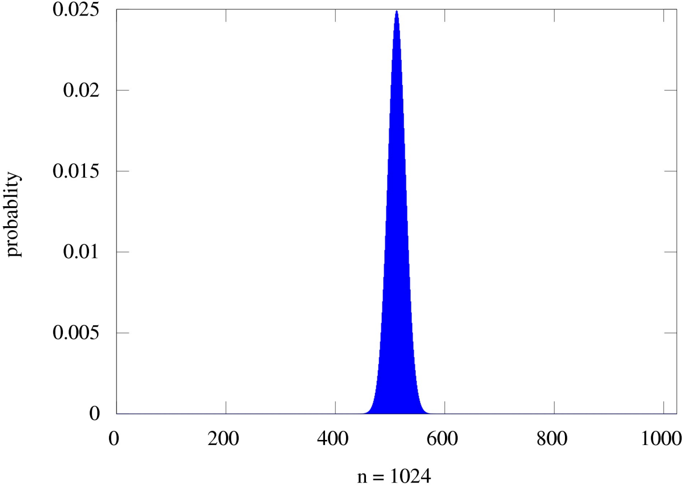
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



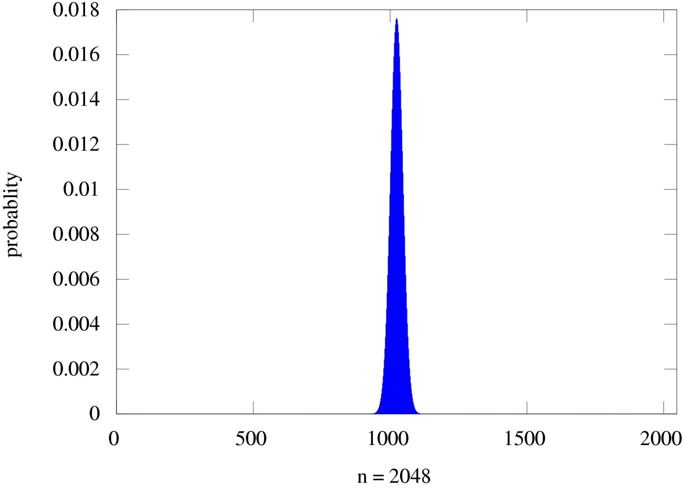
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



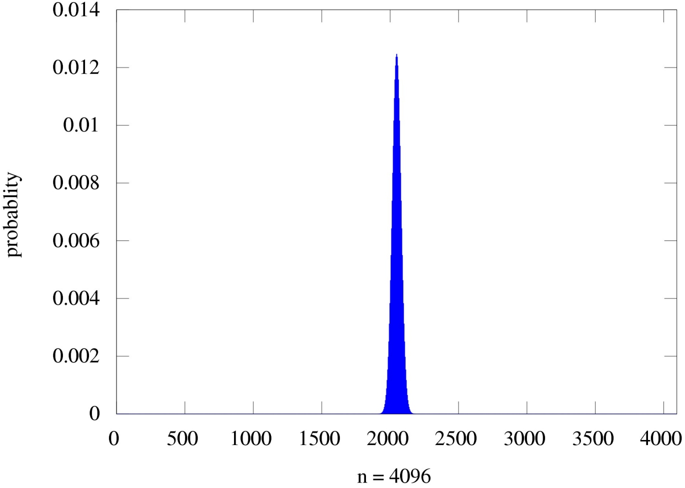
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k}1/2^n$.



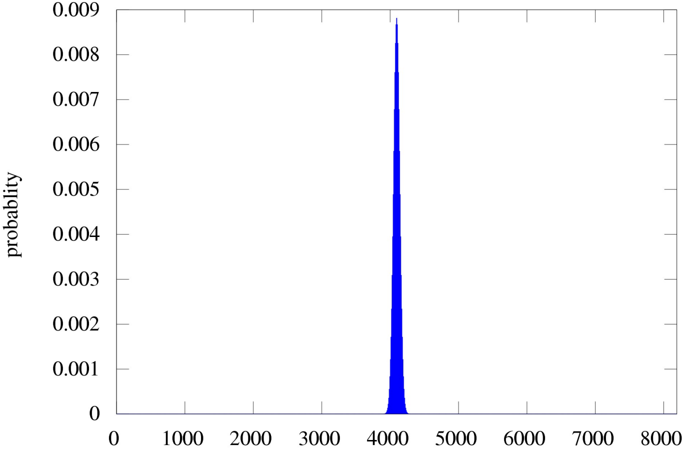
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



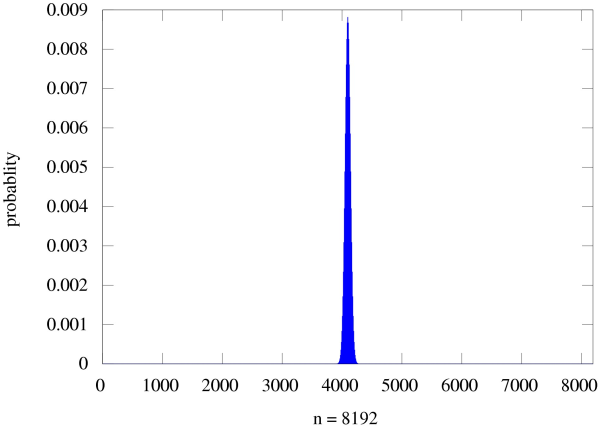
Massive randomness.. Is not that random.

Consider flipping a fair coin n times independently, head gives 1, tail gives zero. How many 1s? Binomial distribution: k w.p. $\binom{n}{k} 1/2^n$.



n = 8192

Massive randomness.. Is not that random.



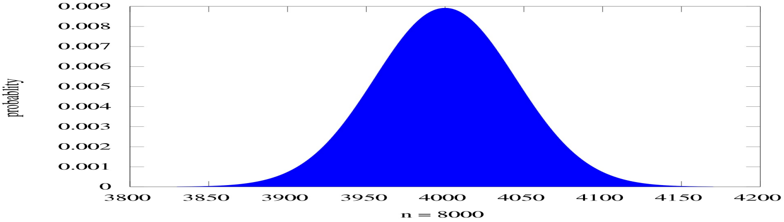
This is known as **concentration of measure**.
This is a related to the **law of large numbers** and *Chernoff bounds*

Side note...

Law of large numbers (weakest form)...

Informal statement of law of large numbers

For n large enough, the middle portion of the binomial distribution looks like (converges to) the normal/Gaussian distribution.



Part I

Inequalities

Randomized QuickSort

- Random variable $Q = \#comparisons$ made by randomized **QuickSort** on an array of n elements.
- We proved that $\mathrm{E}[Q] \leq 2n \ln n$.
- What is $\Pr[Q \geq 10n \ln n]$?

Question: Can we say anything interesting knowing just the expectation?

Markov's Inequality

Markov's inequality

Let X be a **non-negative** random variable over a probability space $(\Omega, \Pr)$ and let $\mu = \mathrm{E}[X]$. For any $t > 0$, $\Pr[X \geq t\mu] \leq 1/t$. Equivalently, for any $a > 0$, $\Pr[X \geq a] \leq \frac{\mu}{a}$.

Markov's Inequality

Markov's inequality

Let X be a **non-negative** random variable over a probability space $(\Omega, \Pr)$ and let $\mu = \mathrm{E}[X]$. For any $t > 0$, $\Pr[X \geq t\mu] \leq 1/t$. Equivalently, for any $a > 0$, $\Pr[X \geq a] \leq \frac{\mu}{a}$.

Meaningful only when $t > 1$. Example: $\Pr[X \geq 3\mu] \leq 1/3$.

Markov's Inequality

Markov's inequality

Let X be a **non-negative** random variable over a probability space $(\Omega, \Pr)$ and let $\mu = \mathrm{E}[X]$. For any $t > 0$, $\Pr[X \geq t\mu] \leq 1/t$. Equivalently, for any $a > 0$, $\Pr[X \geq a] \leq \frac{\mu}{a}$.

Meaningful only when $t > 1$. Example: $\Pr[X \geq 3\mu] \leq 1/3$. Proof?

Markov's Inequality

Markov's inequality

Let X be a **non-negative** random variable over a probability space $(\Omega, \Pr)$ and let $\mu = \mathrm{E}[X]$. For any $t > 0$, $\Pr[X \geq t\mu] \leq 1/t$. Equivalently, for any $a > 0$, $\Pr[X \geq a] \leq \frac{\mu}{a}$.

Meaningful only when $t > 1$. Example: $\Pr[X \geq 3\mu] \leq 1/3$. Proof? Simple averaging argument.
Split range of X into two disjoint intervals $I_1 = [0, t\mu)$ and $I_2 = [t\mu, \infty)$. This is because X is non-negative.

If $\Pr[X \in I_2] > 1/t$ then $\mathrm{E}[X] > (1/t)(t\mu) > \mu$ a contradiction!

Markov's Inequality

Markov's inequality

Let X be a **non-negative** random variable over a probability space $(\Omega, \Pr)$ and let $\mu = \mathrm{E}[X]$. For any $t > 0$, $\Pr[X \geq t\mu] \leq 1/t$. Equivalently, for any $a > 0$, $\Pr[X \geq a] \leq \frac{\mu}{a}$.

Proof:

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{\omega \in \Omega} X(\omega) \Pr[\omega] \\
&= \sum_{\omega,\ 0 \leq X(\omega) < a} X(\omega) \Pr[\omega] + \sum_{\omega,\ X(\omega) \geq a} X(\omega) \Pr[\omega] \\
&\geq \sum_{\omega \in \Omega,\ X(\omega) \geq a} X(\omega) \Pr[\omega] \\
&\geq a \sum_{\omega \in \Omega,\ X(\omega) \geq a} \Pr[\omega] \\
&= a \Pr[X \geq a]
\end{aligned}
$$

Markov's Inequality

Markov's inequality

Let X be a **non-negative** random variable over a probability space $(\Omega, \Pr)$ and let $\mu = \mathrm{E}[X]$. For any $a > 0$, $\Pr[X \geq a] \leq \frac{\mu}{a}$. Equivalently, for any $t > 0$, $\Pr[X \geq t\mu] \leq 1/t$.

Proof:

$$\begin{aligned} \mathrm{E}[X] &= \int_0^\infty z f_X(z) dz \\ &\geq \int_a^\infty z f_X(z) dz \\ &\geq a \int_a^\infty f_X(z) dz \\ &= a \Pr[X \geq a] \end{aligned}$$

Randomized QuickSort

- Random variable $Q = \#comparisons$ made by randomized **QuickSort** on an array of n elements.
- We proved that $\mathrm{E}[Q] \leq 2n \ln n$.

Question: What is $\Pr[Q \geq 10n \ln n]$?

By Markov's inequality at most $1/5$.

Chebyshev's Inequality: Variance

Variance

Given a random variable X over probability space $(\Omega, \Pr)$, variance of X is the measure of how much does it deviate from its mean value. Formally, $Var(X) = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$

Derivation

Define $Y = (X - \mathrm{E}[X])^2 = X^2 - 2X\,\mathrm{E}[X] + \mathrm{E}[X]^2$.

$$\begin{aligned} Var(X) &= \mathrm{E}[Y] \\ &= \mathrm{E}[X^2] - 2\,\mathrm{E}[X]\,\mathrm{E}[X] + \mathrm{E}[X]^2 \\ &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 \end{aligned}$$

Tightness of Markov's Inequality

Exercise: Prove that Markov's inequality is tight.

More formally: for any given $t > 1$ describe a simple probability space and a non-negative random variable X with $\mu = \mathrm{E}[X]$ finite such that $\Pr[X \geq t\mu] = 1/t$.

Thus, improving on Markov's inequality requires additional knowledge/assumption on distribution of X.

Chebyshev's Inequality: Variance

Independence

Random variables X and Y are called mutually independent if

$$\forall x, y \in \mathbb{R}, \ \Pr[X = x \wedge Y = y] = \Pr[X = x] \Pr[Y = y]$$

Lemma

If X and Y are independent random variables then $Var(X + Y) = Var(X) + Var(Y)$.

Chebyshev's Inequality: Variance

Independence

Random variables X and Y are called mutually independent if
$$\forall x, y \in \mathbb{R},\ \Pr[X = x \wedge Y = y] = \Pr[X = x] \Pr[Y = y]$$

Lemma

If X and Y are independent random variables then $Var(X + Y) = Var(X) + Var(Y)$.

Lemma

If X and Y are mutually independent, then $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$.

Chebyshev's Inequality

Chebyshev's Inequality

If $Var X < \infty$, for any $a \geq 0$, $\Pr[|X - \mathrm{E}[X]| \geq a] \leq \frac{Var(X)}{a^2}$

Chebyshev's Inequality

Chebyshev's Inequality

If $\mathit{VarX} < \infty$, for any $a \geq 0$, $\Pr[|X - \mathrm{E}[X]| \geq a] \leq \frac{Var(X)}{a^2}$

Proof.

$Y = (X - \mathrm{E}[X])^2$ is a non-negative random variable. Apply Markov's Inequality to Y for a^2.

$$\begin{aligned} \Pr[Y \geq a^2] \leq {}^{\mathrm{E}[Y]}/_{a^2} &\Leftrightarrow \Pr[(X - \mathrm{E}[X])^2 \geq a^2] \leq {}^{Var(X)}/_{a^2} \\ &\Leftrightarrow \Pr[|X - \mathrm{E}[X]| \geq a] \leq {}^{Var(X)}/_{a^2} \end{aligned}$$

□

Chebyshev's Inequality

Chebyshev's Inequality

If $\mathit{Var}X < \infty$, for any $a \geq 0$, $\Pr[|X - \mathrm{E}[X]| \geq a] \leq \frac{\mathit{Var}(X)}{a^2}$

Proof.

$Y = (X - \mathrm{E}[X])^2$ is a non-negative random variable. Apply Markov's Inequality to Y for a^2.

$$\begin{aligned} \Pr[Y \geq a^2] \leq \mathrm{E}[Y]/a^2 &\Leftrightarrow \Pr[(X - \mathrm{E}[X])^2 \geq a^2] \leq \mathit{Var}(X)/a^2 \\ &\Leftrightarrow \Pr[|X - \mathrm{E}[X]| \geq a] \leq \mathit{Var}(X)/a^2 \end{aligned}$$

□

$\Pr[X \leq \mathrm{E}[X] - a] \leq \mathit{Var}(X)/a^2$ AND
$\Pr[X \geq \mathrm{E}[X] + a] \leq \mathit{Var}(X)/a^2$

Chebyshev's Inequality

Chebyshev's Inequality

Given $a \geq 0$, $\Pr[|X - \mathrm{E}[X]| \geq a] \leq \frac{Var(X)}{a^2}$ equivalently for any $t > 0$, $\Pr[|X - \mathrm{E}[X]| \geq t\sigma_X] \leq \frac{1}{t^2}$ where $\sigma_X = \sqrt{Var(X)}$ is the standard deviation of X.

Example: Random walk on the line

- Start at origin 0. At each step move left one unit with probability $1/2$ and move right with probability $1/2$.
- After $\boldsymbol{n}$ steps how far from the origin?

Example: Random walk on the line

- Start at origin 0. At each step move left one unit with probability $1/2$ and move right with probability $1/2$.
- After n steps how far from the origin?

At time i let X_i be -1 if move to left and 1 if move to right.
Y_n position at time n
$Y_n = \sum_{i=1}^{n} X_i$

Example: Random walk on the line

- Start at origin 0. At each step move left one unit with probability $1/2$ and move right with probability $1/2$.
- After n steps how far from the origin?

At time i let X_i be -1 if move to left and 1 if move to right.
Y_n position at time n
$Y_n = \sum_{i=1}^{n} X_i$

$E[Y_n] = 0$ and $Var(Y_n) = \sum_{i=1}^{n} Var(X_i) = n$

Example: Random walk on the line

- Start at origin 0. At each step move left one unit with probability $1/2$ and move right with probability $1/2$.
- After n steps how far from the origin?

At time i let X_i be -1 if move to left and 1 if move to right.
Y_n position at time n
$Y_n = \sum_{i=1}^{n} X_i$

$\mathrm{E}[Y_n] = 0$ and $Var(Y_n) = \sum_{i=1}^{n} Var(X_i) = n$

By Chebyshev: $\Pr\left[|Y_n| \geq t\sqrt{n}\right] \leq 1/t^2$

Chernoff Bound: Motivation

In many applications we are interested in X which is sum of *independent* and *bounded* random variables.

$X = \sum_{i=1}^{k} X_i$ where $X_i \in [0, 1]$ or $[-1, 1]$ (normalizing)

Chebyshev not strong enough. For random walk on line one can prove

$$\Pr\left[|Y_n| \geq t\sqrt{n}\right] \leq 2exp(-t^2/2)$$

Chernoff Bound: Non-negative case

Lemma

Let $X_1, \ldots, X_k$ *be* k *independent binary random variables such that, for each* $i \in [k]$, $\mathrm{E}[X_i] = \Pr[X_i = 1] = p_i$*. Let* $X = \sum_{i=1}^{k} X_i$*. Then* $\mathrm{E}[X] = \sum_i p_i$*.*

Chernoff Bound: Non-negative case

Lemma

Let $X_1, \ldots, X_k$ be k independent binary random variables such that, for each $i \in [k]$, $\mathrm{E}[X_i] = \Pr[X_i = 1] = p_i$. Let $X = \sum_{i=1}^{k} X_i$. Then $\mathrm{E}[X] = \sum_i p_i$.

- *Upper tail bound: For any $\mu \geq \mathrm{E}[X]$ and any $\delta > 0$,*

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

- *Lower tail bound: For any $0 < \mu < \mathrm{E}[X]$ and any $0 < \delta < 1$,*

$$\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$

Chernoff Bound: Non-negative case, simplifying

When $0 < \delta < 1$ an important regime of interest we can simplify.

Lemma

Let $X_1, \ldots, X_k$ be k independent random variables such that, for each $i \in [1, k]$, X_i equals 1 with probability p_i, and 0 with probability $(1 - p_i)$. Let $X = \sum_{i=1}^{k} X_i$ and $\mu = \mathrm{E}[X] = \sum_i p_i$. For any $0 < \delta < 1$, it holds that:

- $\Pr[X \geq (1+\delta)\mu] \leq e^{\frac{-\delta^2 \mu}{3}}$
- $\Pr[X \leq (1-\delta)\mu] \leq e^{\frac{-\delta^2 \mu}{2}}$
- *Hence by union bound:* $\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{\frac{-\delta^2 \mu}{3}}$

Chernoff Bound: Non-negative case

Important: non-negative case bound depends only on μ, not on k.

Regimes of interest for δ for upper tail.

- $0 \leq \delta < 1$: $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\delta^2}{3}\cdot\mu}$
- $\delta \geq 1$: $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\delta}{3}\cdot\mu}$
  (useful when δ is close to a small constant)
- $\delta \geq 1$: $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{(1+\delta)\ln(1+\delta)}{4}\cdot\mu}$.
  (useful when δ is large)

Chernoff Bound: general

Lemma

Let $X_1, \ldots, X_k$ be k independent random variables such that, for each $i \in [1, k]$, $X_i \in [-1, 1]$.

Chernoff Bound: general

Lemma

Let $X_1, \ldots, X_k$ be k independent random variables such that, for each $i \in [1, k]$, $X_i \in [-1, 1]$. Let $X = \sum_{i=1}^{k} X_i$. For any $a > 0$,

$$\Pr[|X - \mathrm{E}[X]\,| \geq a] \leq 2exp(\frac{-a^2}{2n}).$$

When variables are not positive the bound depends on n while in the non-negative case there is no dependence on n (dimension-free)

Chernoff Bound: general

Lemma

Let $X_1, \ldots, X_k$ be k independent random variables such that, for each $i \in [1, k]$, $X_i \in [-1, 1]$. Let $X = \sum_{i=1}^{k} X_i$. For any $a > 0$,

$$\Pr[|X - \mathrm{E}[X]\,| \geq a] \leq 2exp(\frac{-a^2}{2n}).$$

When variables are not positive the bound depends on n while in the non-negative case there is no dependence on n (dimension-free)

Applying to random walk:

$$\Pr\left[|Y_n| \geq t\sqrt{n}\right] \leq 2exp(-t^2/2).$$

Extensions and variations

Hoeffding extension: Theorems hold as long as X_i is bounded — variables do not have to be $\{0, 1\}$.

- For non-negative $X_i \in [0, 1]$
- For general $X_i \in [-1, 1]$

Averaging version: Bound $X = \frac{1}{k}(\sum_{i=1}^{k} X_i)$ instead of the sum. Use variable $Y = kX$ and bound on Y.

Scaling variables: If X_i is in $[0, B]$ use $Y_i = X_i/B$.

Shifting variables: If $X_i \in [a_i, b_i]$ where $b_i - a_i$ is small consider $Y_i = X_i - a_i$.

Many variations and generalization. See pointers on course webpage.

Part II

Balls and Bins

Balls and Bins

- m balls and n bins
- Each ball thrown independently and uniformly in a bin
- Want to understand properties of bin loads
- Fundamental problem with many applications

Balls and Bins

- m balls and n bins
- Each ball thrown independently and uniformly in a bin
- Want to understand properties of bin loads
- Fundamental problem with many applications

- Z_{ij} indicator for ball i falling into bin j
- $X_j = \sum_{i=1}^{m} Z_{ij}$ is number of balls in bin j
- $\sum_{j=1}^{n} Z_{ij} = 1$ deterministically
- $\mathrm{E}[Z_{ij}] = 1/n$ for all i, j, and hence $\mathrm{E}[X_j] = m/n$ for each bin j

Maximum load

Question: Suppose we throw n balls into n bins. What is the expectation of the *maximum* load?

Maximum load

Question: Suppose we throw n balls into n bins. What is the expectation of the *maximum* load?

Theorem

Let $Y = \max_{j=1}^{n} X_j$ *be the maximum load. Then* $\Pr[Y > 10 \ln n / \ln \ln n] < 1/n^2$ *(high probability) and hence* $E[Y] = O(\ln n / \ln \ln n)$.

One can also show that $E[Y] = \Theta(\ln n / \ln \ln n)$.

Maximum load

Question: Suppose we throw n balls into n bins. What is the expectation of the *maximum* load?

Theorem

Let $Y = \max_{j=1}^{n} X_j$ *be the maximum load. Then* $\Pr[Y > 10 \ln n / \ln \ln n] < 1/n^2$ *(high probability) and hence* $\mathrm{E}[Y] = O(\ln n / \ln \ln n)$.

One can also show that $\mathrm{E}[Y] = \Theta(\ln n / \ln \ln n)$.
Proof technique: combine Chernoff bound and union bound which is powerful and general template

Maximum load

Focus on bin 1 without loss of generality since bins are symmetric. Simplifying notation $X = \sum_i Z_i$ where X is load of bin 1 and Z_i is indicator of ball i falling in bin.

- Want to know $\Pr[X \geq 12 \ln n / \ln \ln n]$

Maximum load

Focus on bin 1 without loss of generality since bins are symmetric. Simplifying notation $X = \sum_i Z_i$ where X is load of bin 1 and Z_i is indicator of ball i falling in bin.

- Want to know $\Pr[X \geq 12 \ln n / \ln \ln n]$
- $\mu = \mathrm{E}[X] = 1$
- $(1 + \delta) = 12 \ln n / \ln \ln n$. We are in large δ setting
- Apply the Chernoff upper tail bound (with simplification) :

$$\Pr[X \geq (1 + \delta)\mu] \leq e^{-\frac{(1+\delta)\ln(1+\delta)}{4} \cdot \mu}$$

Maximum load

Focus on bin 1 without loss of generality since bins are symmetric. Simplifying notation $X = \sum_i Z_i$ where X is load of bin 1 and Z_i is indicator of ball i falling in bin.

- Want to know $\Pr[X \geq 12 \ln n / \ln \ln n]$
- $\mu = \mathrm{E}[X] = 1$
- $(1 + \delta) = 12 \ln n / \ln \ln n$. We are in large δ setting
- Apply the Chernoff upper tail bound (with simplification) :

$$\Pr[X \geq (1 + \delta)\mu] \leq e^{-\frac{(1+\delta)\ln(1+\delta)}{4} \cdot \mu}$$

- Calculate/simplify and see that $\Pr[X \geq 12 \ln n / \ln \ln n] \leq 1/n^3$

Maximum load

- For each bin j, $\Pr[X_j \geq 12 \ln n / \ln \ln n] \leq 1/n^3$
- Let A_j be event that $X_j \geq 12 \ln n / \ln \ln n$
- By union bound

$$\Pr[\cup_j A_j] \leq \sum_j \Pr[A_j] \leq n \cdot 1/n^3 \leq 1/n^2.$$

- Hence, with probability at least $(1 - 1/n^2)$ **no bin** has load more than $12 \ln n / \ln \ln n$.

Maximum load

- For each bin j, $\Pr[X_j \geq 12 \ln n / \ln \ln n] \leq 1/n^3$
- Let A_j be event that $X_j \geq 12 \ln n / \ln \ln n$
- By union bound

$$\Pr[\cup_j A_j] \leq \sum_j \Pr[A_j] \leq n \cdot 1/n^3 \leq 1/n^2.$$

- Hence, with probability at least $(1 - 1/n^2)$ **no bin** has load more than $12 \ln n / \ln \ln n$.
- Let $Y = \max_j X_j$. $Y \leq n$. Hence

$$E[Y] \leq (1 - 1/n^2)(12 \ln n / \ln \ln n) + (1/n^2)n.$$

From a ball's perspective

Consider a ball i. How many other balls fall into the same bin as i?

From a ball's perspective

Consider a ball i. How many other balls fall into the same bin as i?

- Ball i is thrown first wlog. And lands in some bin j.
- Then the other $n - 1$ balls are thrown.
- Now bin j is fixed. Hence expected load on bin j is $(1 - 1/n)$.
- What is variance? What is a high probability bound?

Part III

Approximate Median

Approximate median

- **Input:** n distinct numbers $a_1, a_2, \ldots, a_n$ and $0 < \epsilon < 1/2$
- **Output:** A number x from input such that $(1 - \epsilon)n/2 \leq rank(x) \leq (1 + \epsilon)n/2$

Approximate median

- **Input:** n distinct numbers $a_1, a_2, \ldots, a_n$ and $0 < \epsilon < 1/2$
- **Output:** A number x from input such that $(1 - \epsilon)n/2 \leq rank(x) \leq (1 + \epsilon)n/2$

Algorithm:

- Sample with replacement k numbers from $a_1, a_2, \ldots, a_n$
- Output median of the sampled numbers

Approximate median

- **Input:** n distinct numbers $a_1, a_2, \ldots, a_n$ and $0 < \epsilon < 1/2$
- **Output:** A number x from input such that
  $(1 - \epsilon)n/2 \leq rank(x) \leq (1 + \epsilon)n/2$

Algorithm:

- Sample with replacement k numbers from $a_1, a_2, \ldots, a_n$
- Output median of the sampled numbers

Theorem

For any $0 < \epsilon < 1/2$ and $0 < \delta < 1$, if $k = \Omega(\frac{1}{\epsilon^2}\log(1/\delta))$, the algorithm outputs an ϵ-approximate median with probability at least $(1 - \delta)$.

Approximate median

- Let S be random sample chosen by algorithm
- Imagine sorting the numbers
- Split numbers into L (left), M (middle), and R (right)
- $M = \{y \mid (1-\epsilon)n/2 \leq rank(y) \leq (1+\epsilon)n/2\}$
- Algorithm makes a mistake only if $|S \cap L| \geq k/2$ or $|S \cap R| \geq k/2$. Otherwise it will output a number from M.

Approximate median

- Let S be random sample chosen by algorithm
- Imagine sorting the numbers
- Split numbers into L (left), M (middle), and R (right)
- $M = \{y \mid (1-\epsilon)n/2 \leq rank(y) \leq (1+\epsilon)n/2\}$
- Algorithm makes a mistake only if $|S \cap L| \geq k/2$ or $|S \cap R| \geq k/2$. Otherwise it will output a number from M.

Lemma

$\Pr[|S \cap L| \geq k/2] \leq \delta/2$ *if* $k \geq \frac{10}{\epsilon^2}\log(1/\delta)$.

Analysis

- Let $Y = |S \cap L|$? What is $\mathrm{E}[Y]$?
- $Y = \sum_{i=1}^{k} X_i$ where X_i is indicator of sample i falling in L. Hence $\mathrm{E}[Y] = k(1-\epsilon)/2$
- Use Chernoff bound: $\Pr[Y \geq k/2] \leq \delta/2$ if $k \geq \frac{10}{\epsilon^2}\log(1/\delta)$.

Analysis continued

- $\Pr[|S \cap L| \geq k/2] \leq \delta/2$ if $k \geq \frac{10}{\epsilon^2}\log(1/\delta)$.
- By symmetry: $\Pr[|S \cap R| \geq k/2] \leq \delta/2$ if $k \geq \frac{10}{\epsilon^2}\log(1/\delta)$.
- By union bound at most δ probability that $|S \cap L| \geq k/2$ or $|S \cap R| \geq k/2$.
- Hence with $(1-\delta)$ probability median of S is an ϵ-approximate median

Part IV

Randomized QuickSort (Contd.)

Randomized QuickSort: Recall

Input: Array A of n numbers. **Output:** Numbers in sorted order.

Randomized QuickSort

1. Pick a pivot element *uniformly at random* from A.
2. Split array into 3 subarrays: those smaller than pivot, those larger than pivot, and the pivot itself.
3. Recursively sort the subarrays, and concatenate them.

Randomized QuickSort: Recall

Input: Array A of n numbers. **Output:** Numbers in sorted order.

Randomized QuickSort

1. Pick a pivot element *uniformly at random* from A.
2. Split array into 3 subarrays: those smaller than pivot, those larger than pivot, and the pivot itself.
3. Recursively sort the subarrays, and concatenate them.

Note: On *every* input randomized **QuickSort** takes $O(n \log n)$ time in expectation. On *every* input it may take $\Omega(n^2)$ time with some small probability.

Randomized QuickSort: Recall

Input: Array A of n numbers. **Output:** Numbers in sorted order.

Randomized QuickSort

1. Pick a pivot element *uniformly at random* from A.
2. Split array into 3 subarrays: those smaller than pivot, those larger than pivot, and the pivot itself.
3. Recursively sort the subarrays, and concatenate them.

Note: On *every* input randomized **QuickSort** takes $O(n \log n)$ time in expectation. On *every* input it may take $\Omega(n^2)$ time with some small probability.
Question: With what probability it takes $O(n \log n)$ time?

Randomized QuickSort: High Probability Analysis

Informal Statement

Random variable $Q(A) = \#$ comparisons done by the algorithm.

We will show that $\Pr[Q(A) \leq 32n \ln n] \geq 1 - 1/n^3$.

Randomized QuickSort: High Probability Analysis

Informal Statement

Random variable $Q(A) = \#$ comparisons done by the algorithm.

We will show that $\Pr[Q(A) \leq 32n \ln n] \geq 1 - {}^{1}/_{n^3}$.

If $n = 100$ then this gives $\Pr[Q(A) \leq 32n \ln n] \geq 0.99999$.

Randomized QuickSort: High Probability Analysis

Informal Statement

Random variable $Q(A) = \#$ comparisons done by the algorithm.

We will show that $\Pr[Q(A) \leq 32n \ln n] \geq 1 - 1/n^3$.

Outline of the proof

- If depth of recursion is k then $Q(A) \leq kn$.
- Prove that depth of recursion $\leq 32 \ln n$ with high probability. Which will imply the result.

Randomized QuickSort: High Probability Analysis

Informal Statement

Random variable $Q(A) = \#$ comparisons done by the algorithm.

We will show that $\Pr[Q(A) \leq 32n \ln n] \geq 1 - 1/n^3$.

Outline of the proof

- If depth of recursion is k then $Q(A) \leq kn$.
- Prove that depth of recursion $\leq 32 \ln n$ with high probability. Which will imply the result.
  1. Focus on a fixed element. Prove that it "participates" in $> 32 \ln n$ levels with probability at most $1/n^4$.
  2. By union bound, any of the n elements participates in $> 32 \ln n$ levels with probability at most

Randomized QuickSort: High Probability Analysis

Informal Statement

Random variable $Q(A) = \#$ comparisons done by the algorithm.

We will show that $\Pr[Q(A) \leq 32n \ln n] \geq 1 - 1/n^3$.

Outline of the proof

- If depth of recursion is k then $Q(A) \leq kn$.
- Prove that depth of recursion $\leq 32 \ln n$ with high probability. Which will imply the result.
  1. Focus on a fixed element. Prove that it "participates" in $> 32 \ln n$ levels with probability at most $1/n^4$.
  2. By union bound, any of the n elements participates in $> 32 \ln n$ levels with probability at most $1/n^3$.

Randomized QuickSort: High Probability Analysis

Informal Statement

Random variable $Q(A) = \#$ comparisons done by the algorithm.

We will show that $\Pr[Q(A) \leq 32n \ln n] \geq 1 - 1/n^3$.

Outline of the proof

- If depth of recursion is k then $Q(A) \leq kn$.
- Prove that depth of recursion $\leq 32 \ln n$ with high probability. Which will imply the result.
  1. Focus on a fixed element. Prove that it "participates" in $> 32 \ln n$ levels with probability at most $1/n^4$.
  2. By union bound, any of the n elements participates in $> 32 \ln n$ levels with probability at most $1/n^3$.
  3. Therefore, all elements participate in $\leq 32 \ln n$ w.p. $(1 - 1/n^3)$.

Useful lemma

Lemma

Consider $h = 32 \ln n$ *for* n *sufficiently large integer. Consider* h *independent unbiased coin tosses* $X_1, X_2, \ldots, X_h$ *and let* A *be the event that there are less than* $4 \ln n$ *heads. Then* $\Pr[A] \leq 1/n^4$.

Useful lemma

Lemma

Consider $h = 32 \ln n$ *for* n *sufficiently large integer. Consider* h *independent unbiased coin tosses* $X_1, X_2, \ldots, X_h$ *and let* A *be the event that there are less than* $4 \ln n$ *heads. Then* $\Pr[A] \leq 1/n^4$.

Apply Chernoff bound (lower tail).

Useful lemma

Lemma

Consider $h = 32 \ln n$ *for* n *sufficiently large integer. Consider* h *independent unbiased coin tosses* $X_1, X_2, \ldots, X_h$ *and let* A *be the event that there are less than* $4 \ln n$ *heads. Then* $\Pr[A] \le 1/n^4$.

Apply Chernoff bound (lower tail).

- $X_i = 1$ if i is head, 0 otherwise. Let $Y = \sum_{i=1}^{h} X_i$ is number of heads.
- $\mu = \mathrm{E}[Y] = h/2 = 16 \ln n$.
- $\Pr[A] = \Pr[Y < 4 \ln n] = \Pr[Y < \mu/4]$.
- By Chernoff bound: $\Pr[Y \le (1-\delta)\mu] \le \exp(-\delta^2 \mu/2)$. Using $\delta = 3/4$ we have $\Pr[A] \le \exp(-4.5 \ln n) \le 1/n^{4.5}$.

Randomized QuickSort: High Probability Analysis

- Fix an element $s \in A$. We will track it at each level.
- Let S_i be the partition containing s at i^{th} level.
- $S_1 = A$ and $S_k = \{s\}$ where k is the last level for s (note k is a random variable). Define $S_\ell = \{s\}$ for all $k \leq \ell \leq n$ for technical convenience

Randomized QuickSort: High Probability Analysis

- Fix an element $s \in A$. We will track it at each level.
- Let S_i be the partition containing s at i^{th} level.
- $S_1 = A$ and $S_k = \{s\}$ where k is the last level for s (note k is a random variable). Define $S_\ell = \{s\}$ for all $k \leq \ell \leq n$ for technical convenience
- We call s lucky in i^{th} iteration, if balanced split: $|S_{i+1}| \leq (3/4)|S_i|$ and $|S_i \setminus S_{i+1}| \leq (3/4)|S_i|$.

Randomized QuickSort: High Probability Analysis

- Fix an element $s \in A$. We will track it at each level.
- Let S_i be the partition containing s at i^{th} level.
- $S_1 = A$ and $S_k = \{s\}$ where k is the last level for s (note k is a random variable). Define $S_\ell = \{s\}$ for all $k \leq \ell \leq n$ for technical convenience
- We call s lucky in i^{th} iteration, if balanced split: $|S_{i+1}| \leq (3/4)|S_i|$ and $|S_i \setminus S_{i+1}| \leq (3/4)|S_i|$.
- If $\rho =$#lucky rounds in first h rounds, then $|S_h| \leq (3/4)^\rho n$.

Randomized QuickSort: High Probability Analysis

- Fix an element $s \in A$. We will track it at each level.
- Let S_i be the partition containing s at i^{th} level.
- $S_1 = A$ and $S_k = \{s\}$ where k is the last level for s (note k is a random variable). Define $S_\ell = \{s\}$ for all $k \leq \ell \leq n$ for technical convenience
- We call s lucky in i^{th} iteration, if balanced split: $|S_{i+1}| \leq (3/4)|S_i|$ and $|S_i \setminus S_{i+1}| \leq (3/4)|S_i|$.
- If ρ =#lucky rounds in first h rounds, then $|S_h| \leq (3/4)^\rho n$.
- If $h \geq \rho = 4 \ln n$ then $S_h \leq 1$ implies s done.

Randomized QuickSort: High Probability Analysis

- Fix an element $s \in A$. We will track it at each level.
- Let S_i be the partition containing s at i^{th} level.
- $S_1 = A$ and $S_k = \{s\}$ where k is the last level for s (note k is a random variable). Define $S_\ell = \{s\}$ for all $k \leq \ell \leq n$ for technical convenience
- We call s lucky in i^{th} iteration, if balanced split: $|S_{i+1}| \leq (3/4)|S_i|$ and $|S_i \setminus S_{i+1}| \leq (3/4)|S_i|$.
- If ρ =#lucky rounds in first h rounds, then $|S_h| \leq (3/4)^\rho n$.
- If $h \geq \rho = 4 \ln n$ then $S_h \leq 1$ implies s done.

Lemma

Fix $h = 32 \ln n$. $|S_h| > 1$ only if less then $4 \ln n$ lucky rounds for s in the first h rounds.

How may rounds before $4 \ln n$ lucky rounds?

- Fix element s and $h = 32 \ln n$.
- $X_i = 1$ if s is lucky in iteration i

How may rounds before $4 \ln n$ lucky rounds?

- Fix element s and $h = 32 \ln n$.
- $X_i = 1$ if s is lucky in iteration i
- **Observation:** $X_1, \ldots, X_h$ are independent variables.
- $\Pr[X_i = 1] = \frac{1}{2}$ **Why?**

How may rounds before $4 \ln n$ lucky rounds?

- Fix element s and $h = 32 \ln n$.
- $X_i = 1$ if s is lucky in iteration i
- **Observation:** $X_1, \ldots, X_h$ are independent variables.
- $\Pr[X_i = 1] = \frac{1}{2}$ **Why?**
- Thus s not done after h iterations only if less than $4 \ln n$ lucky rounds in h rounds. Use Lemma to see probability less than $1/n^4$.

Randomized QuickSort w.h.p. Analysis

- n input elements. Probability that depth of recursion in **QuickSort** $> 32 \ln n$ is at most $\frac{1}{n^4} * n = \frac{1}{n^3}$.

Randomized QuickSort w.h.p. Analysis

- n input elements. Probability that depth of recursion in **QuickSort** $> 32 \ln n$ is at most $\frac{1}{n^4} * n = \frac{1}{n^3}$.

Theorem

With high probability (i.e., $1 - \frac{1}{n^3}$) the depth of the recursion of **QuickSort** *is $\leq 32 \ln n$. Due to n comparisons in each level, with high probability, the running time of* **QuickSort** *is $O(n \ln n)$.*

Randomized QuickSort w.h.p. Analysis

- n input elements. Probability that depth of recursion in **QuickSort** $> 32 \ln n$ is at most $\frac{1}{n^4} * n = \frac{1}{n^3}$.

Theorem

With high probability (i.e., $1 - \frac{1}{n^3}$) the depth of the recursion of **QuickSort** *is $\leq 32 \ln n$. Due to n comparisons in each level, with high probability, the running time of* **QuickSort** *is $O(n \ln n)$.*