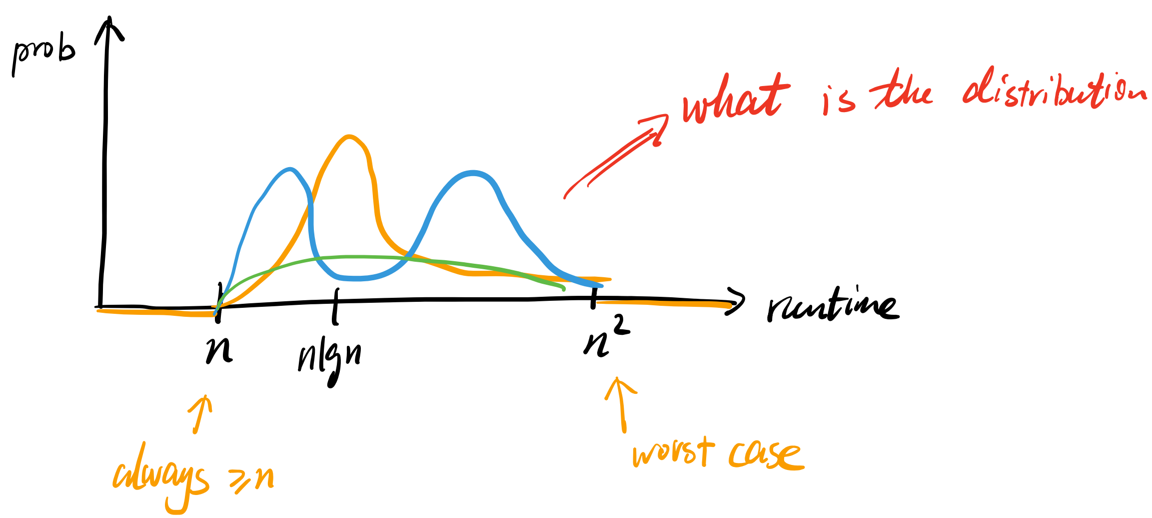
Randomized Algo : Distribution

Quick Sort

$T(a)$ = runtime of quick sort of sequence a

$$T(n) = \max_{|a|=n} \mathbb{E}[T(a)] \leq O(n \lg n)$$

Distribution:



Q: $\Pr[T(a) \geq \frac{n^2}{10}] \approx \frac{1}{2}$?

$\leq \frac{1}{\sqrt{n}}$?

$\leq \frac{1}{2}\sqrt{n}$?

def
$\boxed{\text{a randomized algo runs in time } T(n) \text{ w/ probability } 1-\delta(n)}$

For all inputs x of size n, then algo runs in time $T(n)$ and produces correct answer w/ probability $p = 1-\delta(n)$

rmk: quick sort will always eventually be correct. ← randomness in runtime.

Correctness is random, but runtime is not

Q $\Pr\{X = \mathbb{E}[X]\} \approx 1$? Want to argue: expected runtime is what we expect.

lemma: **Markov's Inequality**

$$X \geq 0, \quad a \geq 0, \quad \Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$$

Pf: $\mathbb{E}[X] = \underbrace{\mathbb{E}[X \mid X \geq a]}_{\geq a} \cdot \underbrace{\Pr[X \geq a]}_{\text{want}} + \underbrace{\mathbb{E}[X \mid X < a]}_{\geq 0} \cdot \underbrace{\Pr[X < a]}_{\geq 0}$

$$\geq a \Pr[X \geq a] \Rightarrow \text{done.}$$

cor $X \geq 0,\ k \geq 1,\ \Pr[X \geq k \cdot \mathbb{E}[X]] \leq \frac{1}{k}$

cor $\Pr[T(a) \geq \underbrace{k \cdot \Theta(n \lg n)}_{\frac{n^2}{10}}] \leq \frac{1}{k}$

$\Rightarrow \leq O\left(\frac{\lg n}{n}\right)$

Q do better than quick sort ?

Algo: quicksort − restart

- while
  - run $b =$ quicksort(a) for $2 \cdot \boxed{\Theta(n \lg n)}$ steps. $\rightarrow$ expected runtime.
  - if quicksort finishes, return b.

exponentially better

prop quicksort − restart is always correct.

prop quicksort − restart runs in $\Theta(k n \lg n)$ steps w/ probability $1 - \frac{1}{2^k}$

Pf: Let L be # loops. $\Rightarrow$ algo runtime $= L \cdot 2\,\Theta(n \lg n)$

Indicator RV:

$$X_i = \begin{cases} 1 & \text{if quicksort (QS) terminates in } 2\,\Theta(n \lg n) \\ 0 & \text{o.w.} \end{cases}$$

claim: $\Pr\{X_i = 0\} = \Pr\{\text{QS runtime} \geq 2 \cdot \mathbb{E}[\text{QS runtime}]\}$

$\leq \frac{1}{2}$ by Markov's Inequality.

Claim:

$$\Pr\{L > k\} = \Pr\{\underbrace{X_1 = 0 \wedge X_2 = 0 \wedge \cdots \wedge X_k = 0}_{\text{independent}}\} = \prod_{i=1}^{k} \Pr[X_i = 0] \leq \frac{1}{2^k}$$

Cor QS – restart runs in $\boxed{O(n\lg^2 n)}$ time w/ probability $1-\frac{1}{n}$

$O(c\cdot n\lg^2 n)$ time ~ $1-\frac{1}{n^c}$

$\geq \frac{n^2}{10}$ ~ $\leq \boxed{\frac{1}{2^{\Theta(n/\lg n)}}}$

Q Avoid penalty turning expected runtime to w/ high probability runtime.

Fact: Original QS in $\boxed{O(n\lg n)}$ w/ p $1-\frac{1}{n^c}$ $c=\Theta(1)$

Better

in $\geq \frac{n^2}{10}$ w/ p $\leq \boxed{\frac{1}{2^{\Theta(n)}}}$

Now, some probability tools for analysis.

Q Better tail bounds? What is expectations?

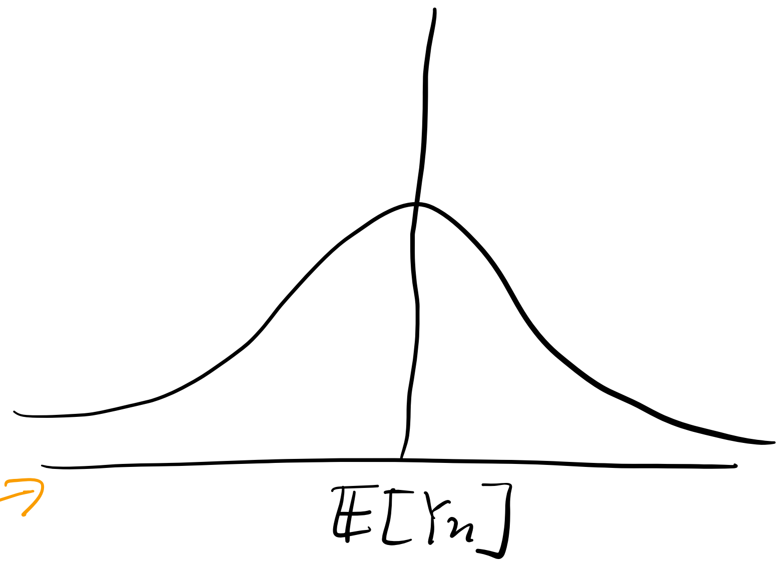
Central Limit Theorem

$X_1,\dots,X_i,\dots,$ identical and independent RV.

$\mathbb{E}[X_i^2] < \infty$ for all i.

Then $\underbrace{Y_n = \frac{X_1+\cdots+X_n}{\sqrt{n}}}$

$n\to\infty$, converges to Gaussian distribution.



Q How quickly this convergance happen ?

Thm [Cheroff bound]

Given $X_1, \dots, X_n \in \{0, 1\}$ independent, $\mathbb{E}[X_i] = p_i$

define $X = X_1 + \dots + X_n$, $\delta \geq 0$,

then $\Pr\{X \geq (1+\delta)\,\mathbb{E}[X]\} \leq \left(\dfrac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}[X]}$

Also $\Pr\{|X - \mathbb{E}[X]| \geq \delta \cdot n\} \leq 2e^{-\delta^2 n/4}$

probability X deviates $\mathbb{E}[X]$ goes to zero exponentialy fast

EX] (polling) $S \subseteq \Omega$, what is $\dfrac{|S|}{|\Omega|}$

votes for A

Algo: - pick $\sigma_1, \dots, \sigma_t \in \Omega$ randomly

- return $\sum \dfrac{\mathbb{1}[\sigma_i \in S]}{t}$

What is t ?

Complexity: $O(t)$

Correctness:

Let $X_i = \mathbb{1}[\sigma_i \in S]$, then $\mathbb{E}[X_i] = \frac{|S|}{|\Omega|} = \mu$

$X = X_1 + \cdots + X_n$, $\boxed{\mathbb{E}[X] = t \cdot \mu}$ $\Rightarrow$ goal

Let $Y = \frac{X}{t}$, $\mathbb{E}[Y] = \mu$

What is $\Pr\{|Y - \mu| \geq \varepsilon\}$?

$$= \Pr\left\{\left|\frac{X}{t} - \frac{\mathbb{E}[X]}{t}\right| \geq \varepsilon\right\} = \Pr\left\{\frac{1}{t}|X - \mathbb{E}[X]| \geq \varepsilon\right\} \leq 2e^{-\varepsilon^2 t/4}$$

Proof of Chernoff Bound

idea: Markov's inequality on transformed variable.

pick $t > 0$ as parameter $\Rightarrow$ the map $x \longmapsto e^{tx}$ strictly increasing.

$$\Pr\{X \geq (1+\delta) \underbrace{\mathbb{E}[X]}_{= \mu}\} = \Pr\{e^{tX} \geq e^{t(1+\delta)\mu}\}$$

$$\leq \mathbb{E}[e^{tX}] / e^{t(1+\delta)\mu}$$ by Markov's inequality.

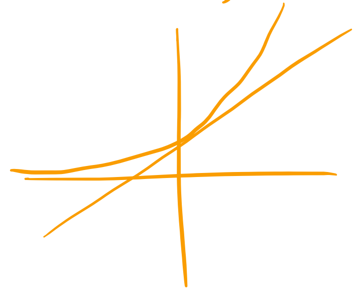
Meaning: $\mathbb{E}[e^{tX}] = \mathbb{E}[e^{t(X_1 + \cdots + X_n)}] = \mathbb{E}[e^{tX_1} \cdot e^{tX_2} \cdot \cdots \cdot e^{tX_n}]$

$$\xrightarrow{\text{independence}} \mathbb{E}[e^{tX_1}] \cdot \mathbb{E}[e^{tX_2}] \cdot \cdots \cdot \mathbb{E}[e^{tX_n}]$$

$$\mathbb{E}[e^{tX_i}] = e^t \Pr[X_i = 1] + e^0 \Pr[X_i = 0]$$

$$= e^t p_i + (1-p_i) = 1 + p_i(e^t - 1) \le e^{p_i(e^t-1)}$$

$(1+z \le e^z \quad \forall z)$



$$\mathbb{E}[e^{tX_1}] \cdot \mathbb{E}[e^{tX_2}] \cdot \cdots \cdot \mathbb{E}[e^{tX_n}] \le \prod e^{p_i(e^t-1)}$$

$$= e^{(e^t-1)\sum p_i} = e^{\mu(e^t-1)}$$

$$\Rightarrow \quad \mathbb{E}[e^{tX}] / e^{t(1+\delta)\mu} \le \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}}$$

choose $t = \ln(1+\delta)$ $\Rightarrow$
$$= \frac{e^{\mu(e^{\ln(1+\delta)}-1)}}{e^{(\ln(1+\delta))(1+\delta)\mu}}$$

$$= \frac{e^{\mu(1+\delta-1)}}{(1+\delta)^{(1+\delta)\mu}} = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$