The Probabilistic Method

Week 6: Expectation, Variance, and Beyond

Joshua Brody

CS49/Math59

Fall 2015

Reading Quiz

What is the law of large numbers?

(A) **For any large number, there is always a larger number.**

(B) **As n gets large, the sample mean of n identically distributed random variables closely approximates the expected value with high probability.**

(C) **As n gets large, the expected value of the sum of n variables approaches the sum of the expected value of n variables.**

(D) **$\Pr[X \geq \alpha] \leq E[X]/\alpha$**

(E) **None of the above**

Reading Quiz

What is the law of large numbers?

(A) **For any large number, there is always a larger number.**

(B) **As n gets large, the sample mean of n identically distributed random variables closely approximates the expected value with high probability.**

(C) **As n gets large, the expected value of the sum of n variables approaches the sum of the expected value of n variables.**

(D) **$\Pr[X \geq \alpha] \leq E[X]/\alpha$**

(E) **None of the above**

Expectation and Variance

- $E[X] = \sum_w X(w)\, P(w)$
- $Var[X] := E[(X-E[X])^2]$

Markov's Inequality:

$$Pr[X \geq \alpha] \leq E[X]/\alpha$$

Clicker Question

There are **300k** workers in Delaware County.

- average income: **40k**
- variance: **100 million** **$(10k)^2$**

How many can make **$100k**?

(A) **at most 120k workers**

(B) **at most 88k workers**

(C) **at most 12k workers**

(D) **at most 8333 workers**

(E) **none of the above**

Clicker Question

There are **300k** workers in Delaware County.

- average income: **40k**
- variance: **100 million** **$(10k)^2$**

How many can make **$100k**?

(A) **at most 120k workers**

(B) **at most 88k workers**

(C) **at most 12k workers**

(D) **at most 8333 workers**

(E) **none of the above**

Markov/Chebyshev Inequalities

There are **300k** workers in Delaware County.

- average income: **40k**
- variance: **100 million (10k)^2**

How many can make **$100k**?

	max % of workers	max # of workers
Markov	**≤ 40%**	**≤ 120,000**
Markov + min wage	**≤ 29%**	**≤ 88,235**
Chebyshev	**≤ 3%**	**≤ 8,333**

Clicker Question

$X_1, \ldots, X_n$: fair coins, $X = \sum_i X_i$

What is $\Pr[X \geq n/2 + c\sqrt{n}]$?

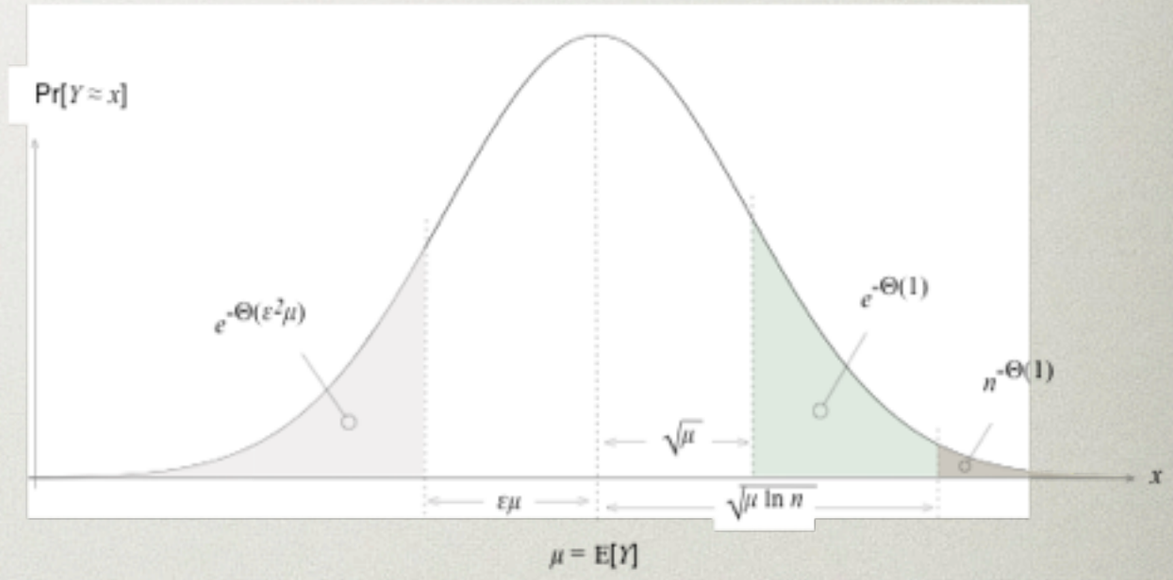
(A) at most $\exp(-cn)$

(B) at most $\exp(-c^2 n)$

(C) at most $\exp(-c^2)$

(D) at most $\exp(-c^2 n^2)$

(E) none of the above



Clicker Question

$X_1, \ldots, X_n$: fair coins, $X = \sum_i X_i$

What is $\Pr[X \geq n/2 + c\sqrt{n}]$?

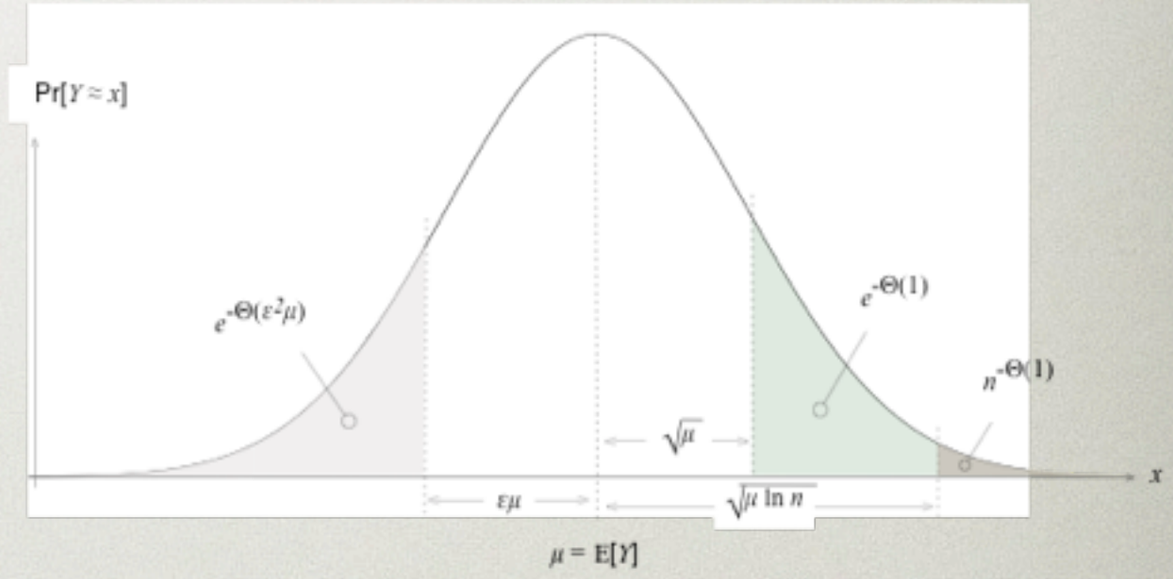
(A) at most $\exp(-cn)$

(B) at most $\exp(-c^2 n)$

(C) at most $\exp(-c^2)$

(D) at most $\exp(-c^2 n^2)$

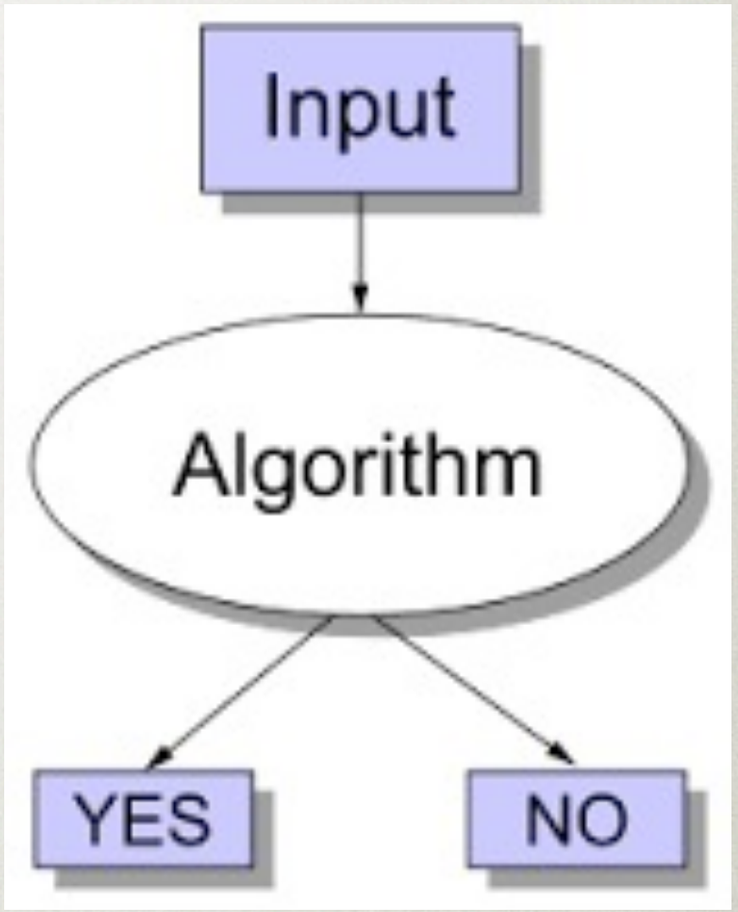
(E) none of the above



Error Reduction in Randomized Algorithms

Given randomized algorithm **A:**

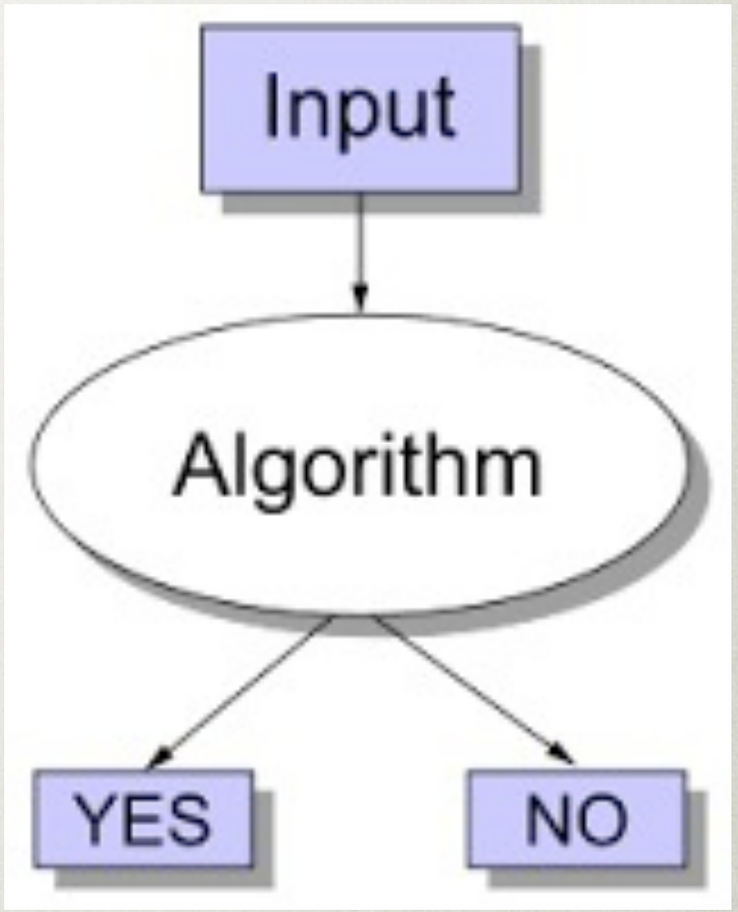
- answers **YES** or **NO**
  - *is input prime number?*
  - *does graph have a large clique?*
  - *is this a picture of a cat?*
- runs in **T** steps
- answers correctly with probability **2/3**



Error Reduction in Randomized Algorithms

Given randomized algorithm **A:**

- answers **YES** or **NO**
  - *is input prime number?*
  - *does graph have a large clique?*
  - *is this a picture of a cat?*
- runs in **T** steps
- answers correctly with probability **2/3**



Problem: Give efficient algorithm **A'** that answers correctly with probability **> 99%**.

The Probabilistic Method