Approximation & Randomization
Approximation and Randomization

- **Approximation**
  Return $\hat{f}(A)$ instead of $f(A)$ where
  
  $$|f(A) - \hat{f}(A)| \leq \epsilon f(A)$$

  is a $(1 + \epsilon)$-approximation of $f(A)$. 
Approximation and Randomization

- **Approximation**
  Return $\hat{f}(A)$ instead of $f(A)$ where
  \[
  \left| f(A) - \hat{f}(A) \right| \leq \epsilon f(A)
  \]
  is a $(1 + \epsilon)$-approximation of $f(A)$.

- **Randomization**
  Return $\hat{f}(A)$ instead of $f(A)$ where
  \[
  \Pr \left[ \left| f(A) - \hat{f}(A) \right| \leq \epsilon f(A) \right] \geq 1 - \delta
  \]
  is a $(1 + \epsilon, \delta)$-approximation of $f(A)$. 
Markov Inequality

Let $X \geq 0$ be a random variable. Then for all $a > 0$,

$$
\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.
$$
Markov and Chebyshev inequalities

- **Markov Inequality**
  Let $X \geq 0$ be a random variable. Then for all $a > 0$, $\Pr[X \geq a] \leq \frac{E[X]}{a}$.

- **Chebyshev’s Inequality**
  Let $X \geq 0$ be a random variable. Then for all $a > 0$, $\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}$. 
**Applications: the coupon collector problem**

- **Coupon collector**
  
  Suppose that each of box of cereal contains one of \( n \) different coupons. Once you obtain one of every type of coupon, you can send in for a prize.

  Assuming that the coupon in each box is chosen independently and uniformly at random from the \( n \) possibilities, how many boxes of cereal must you buy before you obtain at least one of every type pf coupon?
Coupons collector

Suppose that each box of cereal contains one of \( n \) different coupons. Once you obtain one of every type of coupon, you can send in for a prize.

Assuming that the coupon in each box is chosen independently and uniformly at random from the \( n \) possibilities, how many boxes of cereal must you buy before you obtain at least one of every type of coupon?

Analysis (on board)
Chernoff-Hoeffding bounds

Let $X_1, \ldots, X_k$ be independent random variables, let $\Delta_i = \max\{X_i\} - \min\{X_i\}$, let $X = \sum_{i=1}^{k} X_i$. For any $a > 0$,

$$\Pr[|X - E[X]| \geq a] \leq 2e^{-\frac{2a^2}{\sum_i \Delta_i^2}}.$$}

Often, $\Delta_1 = \Delta_2 = \cdots = \Delta_k = \Delta$, thus we can write

$$\Pr[|X - E[X]| \geq a] \leq 2e^{-\frac{2a^2}{k \cdot \Delta^2}}.$$
Applications: coin flips

- Coin flips
  
  We take $n$ independent fair coin flips. How many heads we will see?
Applications: coin flips

- **Coin flips**
  
  We take \( n \) independent fair coin flips. How many heads we will see?

- **Estimating a parameter**

  Suppose that we are interested in evaluating the probability (say, \( p \)) that a particular gene mutation occurs in the population.

  Given a DNA sample, a lab test can determine if it carries the mutation.

  Lab test is expensive so that we would like to obtain a relatively reliable estimate from a small \# of sample.

  Assume we have tested \( n \) samples and we noticed \( X = \bar{p}n \) of these samples have the mutation. What can we say about the unknown (true) value \( p \)?
§1.1 Distinct Elements

How many distinct elements? Approximation needed.