Tuesday, October 01, 2013
11:29 AM

# Randomized Algorithms

Coupon collector problem: $n$ different coupons, sampled with replacement from uniform distribution.

Q: How many coupons before one of each kind shows up?

Define $X_i$ = # of rounds between $i$th and $(i+1)$st coupon

$$X_i \sim Geo\left(\frac{n-i}{n}\right) \equiv Geo\left(1-\frac{i}{n}\right)$$

$$E[X_i] = \frac{n}{n-i}.$$

$$\Rightarrow \mathbb{E}\left[\sum_{i=0}^{n-1} X_i\right] = \sum_{i=0}^{n-1}\mathbb{E}[X_i] = \sum_{i=0}^{n-1}\frac{n}{n-i} = n H_n = \Theta(n\log n).$$

Linearity of expectation: for any set of random variables $X_1, \ldots, X_n$ (not necessarily independent!!)

$$\mathbb{E}\left[\sum X_i\right] = \sum \mathbb{E}[X_i]$$

## Balls and bins

Suppose we throw $n$ balls uniformly at random into $n$ bins.

Let $X_i$: # of balls in the $i$th bin

$$\mathbb{E}[X_i] = 1 \quad \forall i$$ ← not very interesting

Let $X = \max_i X_i$

What is $\mathbb{E}[X]$?

Stirling's approximation

For any $i$, $\mathbb{P}[X_i \geq k] \leq \binom{n}{k}\frac{1}{n^k} \leq \left(\frac{ne}{k}\right)^k \frac{1}{n^k} = \left(\frac{e}{k}\right)^k$

$$\Pr\left[X_i \geq \frac{e\ln n}{\ln\ln n}\right] \leq \left(\ln\ln n\right)^{\frac{e\ln n}{\ln\ln n}} < \frac{1}{n^2}.$$

$$\Pr\left[X_i \geq \frac{e \ln n}{\ln\ln n}\right] \leq \left(\frac{\ln\ln n}{\ln n}\right)^{\frac{e \ln n}{\ln\ln n}} < \frac{1}{n^2}.$$

## Union Bound

$$\Pr\left[\bigcup_i E_i\right] \leq \sum_i \Pr[E_i]$$

Using union bound, $\underbrace{\Pr\left(X \geq \frac{e \ln n}{\ln\ln n}\right) < \frac{1}{n}}$.

whp, $X = O\left(\frac{\ln n}{\ln\ln n}\right)$ (also called a tail bound)

$$\mathbb{E}[X] \leq \left(1 - \frac{1}{n}\right)\frac{\ln n}{\ln\ln n} + \frac{1}{n}\cdot n = O\left(\frac{\ln n}{\ln\ln n}\right)$$

We used a tail bound to obtain a bound on the expectation. Often we want the reverse: use expectation to obtain a tail bound.

**Markov's inequality**: $\Pr[X \geq t] \leq \mathbb{E}[X]/t$ for $X \geq 0$.

Proof: $\mathbb{E}[X] \geq \Pr[X \geq t]\cdot t + \Pr[X < t]\cdot 0$

Often too weak by itself (e.g. in balls in bins, gives probability bound of $\ln\ln n / \ln n$ which is too weak for union bound) but useful for proving stronger inequalities.

## Chebyshev's Inequality

$$\mathbb{P}[|X - \mu| \geq t\sigma] \leq 1/t^2$$

(Note: $X$ need not be non-negative)

Proof: Using Markov,

$$\mathbb{P}[(X-\mu)^2 \geq t^2\sigma^2] \leq \frac{\mathbb{E}[(X-\mu)^2]}{t^2\sigma^2} = \frac{\sigma^2}{t^2\sigma^2} = \frac{1}{t^2}$$

# Chernoff Bounds

For independent $X_1, \dots, X_n$ where $X_i = 1$ w.p. $p_i$ and $0$ otherwise,

$$\mu = \mathbb{E}[X] = \sum_{i=1}^n p_i$$

$$\mathbb{P}\left[X = \sum_{i=1}^n X_i > (1+\varepsilon)\mu\right] < \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu}$$

$$\mathbb{P}\left[X < (1-\varepsilon)\mu\right] < \left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^{\mu}$$

Proof: let us prove only positive deviation.

$$\mathbb{P}\left[X > (1+\varepsilon)\mu\right] = \mathbb{P}\left[e^{tX} > \exp(t(1+\varepsilon)\mu)\right] \underset{\text{Markov}}{\leq} \frac{\mathbb{E}[e^{tX}]}{\exp(t(1+\varepsilon)\mu)}$$

$e^{tX}$: monotonically increasing function of $X$ for any $t > 0$

independence of $X_i$'s

$1+y \leq e^y$

$$= \frac{\prod_{i=1}^n \mathbb{E}[e^{tX_i}]}{\exp(t(1+\varepsilon)\mu)} = \frac{\prod_{i=1}^n (e^t \cdot p_i + (1-p_i))}{\exp(t(1+\varepsilon)\mu)} \leq \frac{\prod_{i=1}^n \exp(p_i(e^t - 1))}{\exp(t(1+\varepsilon)\mu)}$$

$$= \frac{\exp((e^t-1)\mu)}{\exp(t(1+\varepsilon)\mu)} \underset{t = \ln(1+\varepsilon)}{=} \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{\mu}$$

## Simpler (more useful bounds):

$$\varepsilon < 1: \begin{cases} \Pr[X > (1+\varepsilon)\mu] \leq \exp(-\varepsilon^2\mu/3) \\ \Pr[X < (1-\varepsilon)\mu] \leq \exp(-\varepsilon^2\mu/2) \end{cases}$$

# Generic Counting via Sampling

Suppose you have a universe $U$ and subset $S \subseteq U$.
Your goal is to estimate $|S|$, given a uniform sampling procedure for $U$.

Algorithm: $X_i = \begin{cases} 1 & \text{if sample in } S \\ 0 & \text{o.w.} \end{cases}$

Output: $\left( \dfrac{\sum_{i=1}^{N} X_i}{N} \right) |U| = \tilde{N}_S$ ← unbiased estimator

$\mathbb{E}[\tilde{N}_S] = |S|$

Q: How many samples ($N$) do you need?

Suppose $|S|/|U| = \theta$

$\Pr\left[ \left| \tilde{N}_S - |S| \right| > \varepsilon |S| \right] = \Pr\left[ \left| \sum_{i=1}^{N} X_i - \theta N \right| > \varepsilon \theta N \right]$

$< e^{-\frac{\varepsilon^2 \theta N}{3}}$, by Chernoff

For $O(1)$ error probability, $N \geq \dfrac{3}{\varepsilon^2 \theta} = \dfrac{3|U|}{\varepsilon^2 |S|}$

↑ large if $|S|$ is small compared to $|U|$

For $O(1/n^d)$ error probability, repeat $d \ln n$ times (BOOSTING).