Randomization in Computing
CSE 6512
Lecture 4 - Notes
Sep 08 2011

By
Huda Alhazmi

MARKOV'S INEQUALITY

Let X be any non negative random variable with a mean μ, then probability $[X \geq a\mu] \leq \frac{1}{a}$; for any non negative real number a

Proof:
Let $f(x) = 1 \; if \; x \geq a\mu$
$\qquad\qquad = 0 \; otherwise$
Note: $f(x) \leq \frac{x}{a\mu}$.

Probability $[X \geq a\mu] = E[f(X)]$
$\Rightarrow prob.[X \geq a\mu] = E[f(X)] \leq \frac{E(X)}{a\mu} = \frac{1}{a}$.

CHEBYSHEV'S INEQUALITY

Let X be a random variable with a mean μ and std. deviation σ, then probability $[|x - \mu| \geq a\sigma] \leq \frac{1}{a^2}$; for any non negative real number *a.*

Proof:
Probability $[|x - \mu| \geq a\sigma] = prob.[(x - \mu)^2 \geq a^2\sigma^2]$
Using Markov's inequality, the right hand side $\leq \frac{E(x-\mu)^2}{a^2\sigma^2} = \frac{1}{a^2}$.

CHERNOFF BOUNDS

A Bernoulli trials has two outcomes *success* and *failure*, with prob.[success]=p and prob.[failure]=(1-p).
A binomial random variable with parameters n & p refers to the number of successes in n independent Bernoulli trails, the success prob. being p.
$Notation: X = B(n, p)$

Lemma 1
$\mu = np$
$If \; X = B(n,p) then$
$$prob.[X \geq (1 + \epsilon)np] \leq \exp\left(\frac{-\epsilon^2 np}{3}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots\ldots.(1)$$

$$prob.[X \leq (1-\epsilon)np] \leq \exp\left(\frac{-\epsilon^2 np}{2}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots\ldots.(2)$$

$$prob.[X \geq m] \leq \exp\left(\frac{np}{m}\right)^m e^{m-np} \ldots\ldots\ldots.\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(3)$$

$for\ any\ m > np\ and\ for\ any\ 0 < \epsilon < 1.$

Example:

Flip a 2-side coin 1000 times.
Let X= numbers of heads, $p = \frac{1}{2}$
$\Rightarrow \mu = 500\,, \sigma = \sqrt{np(1-p)} = \sqrt{250} = 15.811$

We would like to know the prob.$[X \geq 600]$.

Prob.$[X = i] = \binom{n}{i} p^i (1-p)^{n-i}$

- Using Markov's inequality, prob. $[X \geq 600] \leq \frac{1}{1.2} = \frac{5}{6} = 0.83 \ldots\ldots(4)$
- Using Chebyshev's inequality
  a≈6.32, the prob.$[X \geq 600] \leq \frac{1}{6.32^2} = 0.025 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..\ldots\ldots.(5)$
- Using Chernoff's bound
  $(1+\epsilon)np = 600$
  $\Rightarrow (1+\epsilon) = \frac{600}{500} = 1.2 \Rightarrow \epsilon = 0.2$
  Using ineq. (1) in Chernoff's bound:
  $Prob\ [X \geq 600] \leq \exp\left(\frac{-0.04\times 50}{3}\right) = \exp\left(\frac{-20}{3}\right) = 0.0013$

Fact 1:
Let X be a binomial with mean μ then,
$prob.[X \geq \mu + \sqrt{3 \propto \mu \log_e n}] \leq n^{-\propto}$
$\Rightarrow (1+\epsilon)\mu = \mu + \sqrt{3 \propto \mu \log_e n}$

$$\Rightarrow \epsilon = \sqrt{\frac{3 \propto \log_e n}{\mu}}$$

Using fact (1): $prob.[X \geq (1+\epsilon)\mu] \leq \exp\left[\frac{-3\propto \log_e n}{\mu} . \frac{\mu}{3}\right] = n^{-\propto}$

Fact 2:

Let X be a binomial with mean μ then $prob.\left[X \le \mu - \sqrt{2 \propto \mu \log_e n}\right] \le n^{-\propto}$.
This can be proven in a similar manner.

Lemma 2:

Let X be any sequence of n elements and let S be a random sample from X with $|S|$=s. Let
$Rank(w, S) = j, and\ rank(w, X) = r_j$.
$Then, prob. [\left|r_j - j\frac{n}{s}\right| > 4 \propto \frac{n}{\sqrt{S}}\sqrt{\log n}] \le n^{-\propto}$.

Proof:

Let Y be any subset of X with $|Y| = y$

The expected numbers of sample keys in $Y = B(y, \frac{s}{n})$
Using Fact 1, the actual number of sample keys in Y is no more than $\frac{ys}{n} + \sqrt{3 \propto . \frac{ys}{n}\log_e n}$ with prob. $\ge (1 - n^{-\propto})$.
Similarly, with the same probability, the number of sample keys is $\ge \frac{ys}{n} - \sqrt{3 \propto \frac{ys}{n}\log_e n}$.
Let Y be the smallest q elements of X. If the number of sample keys in $Y\ is \ge j$, then $r_j \le q$. Also, if Y has $\le j$ sample keys then $r_j \ge q$.

Let $q = j\frac{n}{s} + \frac{n}{s}\sqrt{4 \propto j \log_e n}$

Using Fact 2, the expected number of sample keys in this set $Y = j + \sqrt{4 \propto j \log_e n}$

Using Fact 1, the actual number of sample keys is $\ge j + \sqrt{4 \propto j \log_e n} - \sqrt{3 \propto (j + \sqrt{4 \propto j \log_e n}) \log_e n}$ with prob.$\ge ((1 - n^{-\propto})$. $RHS \ge j$

$\Rightarrow r_j \le j\frac{n}{s} + \frac{n}{s}\sqrt{4 \propto j \log_e n}$

$$\leq j\frac{n}{s} + \frac{n}{\sqrt{s}}\sqrt{4 \propto \log_e n}$$

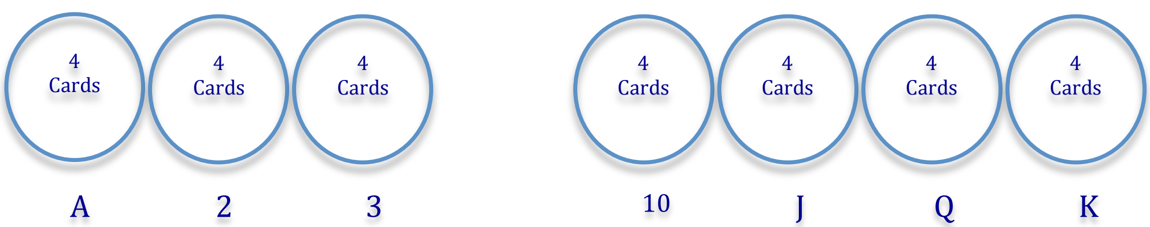
Similarly, we can show that $r_j \geq j\frac{n}{s} - \frac{n}{\sqrt{s}}\sqrt{4 \propto \log_e n}$.

PRINCIPLE OF DEFERRED DECISIONS

Clock Solitaire Game

A solitaire game is played with cards:

- Split a pack of cards into 13 groups of 4 cards and label the groups as {A,2,3,…..,10,J,Q,K}.



- Draw a card from K; if you see *i* go to group *i* and draw a card; if that is *j* draw a card from *j.*

- Repeat in the same manner until the game ends. The game ends when we try to draw from an empty group. If all the cards have been drawn when the game ends you win; if not you lose.

What is the probability you win?
This probability computation will be complex if we look at all possible distribution of cards into 13 groups. Instead notice that the game ends when we try to draw from the group K. This is because only this group has 3 cards to begin with (after the first card is drawn). If the 52nd card drawn is a K, then the game is won. The probability of this happening is $1/13$.

The principle of deferred decisions is to not assume that the entire set of random choices is made in advance. Rather, at each step of the process, we only fix the random choices that must be revealed to the process.