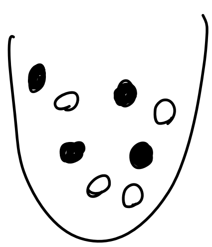
Applications of Chernoff bounds

<u>Exit polls</u> : approximate the % of voters that in an election voted for one of the available options, without counting all votes

U  n balls

Goal: approximate the true value of white balls $\alpha \cdot n$

Assumption: we know there are $\alpha_{min} \cdot n$ white balls

$$\text{Determine } \alpha \xrightarrow{\text{exact}} \Omega(n)$$

$$\xrightarrow[\text{approximated}]{\text{randomized}} O(\log n)$$

↳ and that's why we can do exit polls

We'll output a quantity β such that

$$\Pr\left(\frac{|\beta-\alpha|}{\alpha} > \varepsilon\right) \text{ is very low}$$

$\left(\text{e.g.} < \frac{1}{n^2}\right)$

relative error — confidence threshold

APPROXIMATE_α $(U, \varepsilon, \alpha_{min})$

$n = |U|$

$K = f(n, \varepsilon, \alpha_{min})$ \\ n° of extractions, to be determined in the analysis

$X = 0$

repeat K times

  p = RANDOM(U)

  if color(p) = white then X++

return X/K

$\hookrightarrow \beta$

Complexity: $O(K)$

What's the value of K that guarantees the high probability?

k indicator random variables

$X_i = 1$ is the extracted ball is white

$Pr(X_i = 1) = \alpha$

$X = \sum_{i=1}^{k} X_i$ n° of extracted white balls

$\mu = E[X] = E[\sum_{i=1}^{k} X_i] = \sum_{i=1}^{k} E[X_i] = k\alpha$

Event: "$\frac{|\beta - \alpha|}{\alpha} > \varepsilon$" $=$ "$\frac{|\frac{X}{k} - \alpha|}{\alpha} > \varepsilon$" $=$

$= \text{"}\underbrace{\frac{|X - \alpha k|}{\alpha k} > \varepsilon}\text{"}$

we'll use this Chernoff bound:

$$Pr(\overbrace{|X - \mu| > \varepsilon\mu}) < 2\underbrace{e^{-\frac{\mu\varepsilon^2}{2}}}_{\text{want: } \sim \frac{1}{n^2}} \qquad 0 < \varepsilon \leq 1$$

problem: α is unknown $\Rightarrow$ use α_{min}: $\alpha_{min} \leq \alpha$

$$2e^{-\frac{k\alpha\varepsilon^2}{2}}$$

$$\wedge\!\!/$$

$$2e^{-\frac{k\alpha_{min}\varepsilon^2}{2}} \longrightarrow \frac{2}{n^2}$$

$$-\frac{k\alpha_{min}\varepsilon^2}{2} = -\ln n^2 \longrightarrow e^{-\ln n^2} = \frac{1}{n^2}$$

$$\Rightarrow k = \frac{2\ln n^2}{\alpha_{min}\varepsilon^2} = O\left(\frac{\log n}{\varepsilon^2}\right)$$

Load balancing

n servers

n jobs, that arrive one by one

- distributed : no central control
- limited information : don't know the servers' loads

Goal : minimize max load over the n servers

Algorithm: assign each job to a server chosen uniformly at random

General model: "balls-and-bins"

Consider a fixed server:

$X_i = 1$ if i-th job gets assigned to that server

$Pr(X_i = 1) = \frac{1}{n}$

X_i's are independent

$X = \sum_{i=1}^{n} X_i$ = load of that server

$\mu = E[X] = \sum_{i=1}^{n} E[X_i] = n \cdot \frac{1}{n} = 1$

now: study X in high probability

we'll use:

$$Pr\left(X > (1+\delta)\mu\right) < \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$$

call

$1+\delta = c$

$$\Pr(X > c) < \frac{e^{c-1}}{c^c} < \left(\frac{e}{c}\right)^c$$

want: $\sim \frac{1}{n^2}$

$c = e\,f(n)$

$$\rightarrow \quad \left(\frac{e}{c}\right)^c = \left(\frac{1}{f(n)}\right)^{e f(n)} < \left(\frac{1}{f(n)}\right)^{2 f(n)}$$

it holds that $f(n)^{f(n)} = n \iff f(n) = \Theta\left(\frac{\log n}{\log\log n}\right)$

then

$$\Pr\left(X > \Theta\left(\frac{\log n}{\log\log n}\right)\right) < f(n)^{-2 f(n)}$$

$$= \left(f(n)^{f(n)}\right)^{-2} = n^{-2} = \frac{1}{n^2}$$

Now let's apply the union bound:

E_i = the i-th server gets more than $\Theta\left(\frac{\log n}{\log\log n}\right)$ jobs

$$\Pr\left(\exists \text{ server that gets more than } \Theta\left(\frac{\log n}{\log\log n}\right) \text{ jobs}\right)$$

$$= \Pr\left(\bigcup_{i=1}^{n} E_i\right) \underbrace{\leq}_{\text{union bound}} \sum_{i=1}^{n} \Pr(E_i) = n\frac{1}{n^2} = \frac{1}{n}$$

In other words, the prob. that no server gets more than $\Theta\left(\frac{\log n}{\log\log n}\right)$ jobs is $> 1 - \frac{1}{n}$