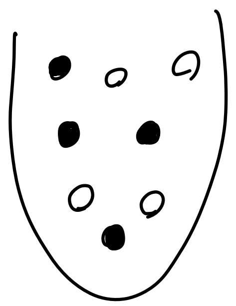
Applications of Chernoff bounds

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

U



n balls

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

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

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

randomized approximated → $O(\log n)$

↳ and that's why we can do exit polls!

We'll output a quantity β such that

$$\Pr\left(\frac{|\beta-\alpha|}{\alpha} > \varepsilon\right) \text{ is very low } \left(\text{e.g. } < \frac{1}{n^2}\right)$$

ε: relative error; $\Pr$: confidence threshold

APPROXIMATE-α $(U, \varepsilon, \alpha_{min})$

```
n = |U|
K = f(n, ε, α_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  ↪ β
```

Complexity: $O(K)$

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

k indicator random variables

$X_i = 1$ if the i-th extracted ball is white

$$\Pr(X_i = 1) = \alpha$$

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

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

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

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

we'll apply this Chernoff bound:

$$\Pr\left(|X-\mu| > \varepsilon\mu\right) < 2e^{-\frac{\mu\varepsilon^2}{2}} \qquad 0<\varepsilon\leq 1$$

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

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

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

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

$$\Downarrow$$

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

Load balancing

n servers

n jobs/requests 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

Simple algorithm: assign each request to a server chosen uniformly at random

General model: "balls-and-bins"

Theorem: (famous result) if n requests are assigned uniformly at random to n servers, then with probability $\geq 1 - \frac{1}{n}$ every server has

$$\leq \frac{3 \ln n}{\ln \ln n}$$

requests, assuming sufficiently high n.

Proof:

Consider a fixed server:

$X_i = 1$ if i-th request is assigned to that server

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

X_i's are independent

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

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

we'll apply

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

$$\| $$

$$\frac{3\ln n}{\ln\ln n} \Rightarrow \delta = \frac{3\ln n}{\ln\ln n} - 1$$

$$\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \overset{?}{\leq} \frac{1}{n^2}$$

$$\Updownarrow \quad \text{take logs of both sides}$$

$$\delta - (1+\delta)\ln(1+\delta) \leq -2\ln n$$

$$\Updownarrow$$

$$\frac{3\ln n}{\ln\ln n} - 1 - \frac{3\ln n}{\ln\ln n}\ln\left(\frac{3\ln n}{\ln\ln n}\right) \leq -2\ln n$$

$$\Updownarrow$$

$$\frac{3\ln n}{\ln\ln n} - 1 - \frac{3\ln n}{\ln\ln n}\left(\ln 3 + \ln\ln n - \ln\ln\ln n\right) \leq -2\ln n$$

$$\Updownarrow$$

$$\frac{3}{\ln\ln n} - \frac{1}{\ln n} - \frac{3}{\ln\ln n}\left(\ln 3 + \ln\ln n - \ln\ln\ln n\right) \leq -2$$

$$\Updownarrow$$

$$\frac{3}{\ln\ln n} - \frac{1}{\ln n} - \frac{3\ln 3}{\ln\ln n} - 3 + \frac{3\ln\ln\ln n}{\ln\ln n} \leq -2$$

$$\Updownarrow \quad n \text{ sufficiently high}$$

$$o(1) + o(1) + o(1) - 3 + o(1) \leq -2 \quad \checkmark$$

Now let's apply the union bound to see that the same is true for every server simultaneously:

E_i = the i-th server gets more than $\frac{3 \ln n}{\ln \ln n}$ requests

$$\Pr\left(\exists \text{ server that gets more than } \frac{3 \ln n}{\ln \ln n} \text{ requests}\right)$$

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

$$< n \frac{1}{n^2} = \frac{1}{n}$$

In other words, the prob. that no server gets more than $\frac{3 \ln n}{\ln \ln n}$ requests is $\geq 1 - \frac{1}{n}$

(can be shown to be tight: some server gets $\Omega\left(\frac{\log n}{\log \log n}\right)$ req.)

(improved alg.: choose 2 servers at random and assign the request to the least loaded $\longrightarrow$ max load drops to $O(\log \log n)$!)