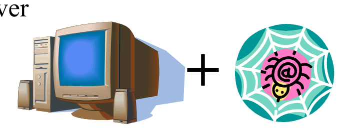
Mathematics for Computer Science

MIT 6.042J/18.062J

Avoiding Large Deviations (Chernoff Bound)

Design for Reliability

WEB Server



- On average gets μ queries per day
- How much access capacity should I build in?

Bernoulli Sums

Focus on random vars, R, that are sums of independent 0-1 variables:

$$R = \sum_{i=1}^{n} \underbrace{T_i}_{\text{Bernoulli variable}}$$

Probability of No Success

$T_i = 1$ means "success" on the i^{th} try.

$[R = 0]$ is the event that we never succeed.

Probability of No Success

Fundamental fact:

If E[#successes] is large, then Pr{never succeeding} is "*exponentially small*":

$$\Pr\{R = 0\} \le e^{-\mu_R}$$

Deviation from the Mean

This is a deviation from mean result:

Pr{observed value *far from* expected value} is *SMALL*

Deviation from the Mean

$$\Pr\{\underbrace{R=0}_{R \text{ below } \mu \text{ by } \underbrace{\mu}_{\text{far}}}\} \le \underbrace{e^{-\mu_R}}_{\text{small (if } \mu_R \text{ large)}}$$

- only need μ_R
- don't need n
- don't need $\Pr\{T_i = 1\}$

In Class Problem

Problem 1

Chernoff Bound

$$\Pr\{R \text{ above } \underbrace{c\mu}_{\text{far}}\} \le \underbrace{e^{-(c \log c - c + 1)\mu}}_{\text{small}}$$

(if $c \ge 1+\varepsilon$ and μ large)

Chernoff Bound

Still

- only need μ_R
- don't need n
- don't need $\Pr\{T_i = 1\}$

But …

Chernoff Bound

Let $\beta(c) ::= c \log c - c + 1$

$$\Pr\{R \ge c\mu\} \le \underbrace{e^{-\beta(c)\mu}}_{\text{small}}$$

only when μ is large.

Chernoff: $\Pr\{R \ge c\mu\} \le e^{-\beta(c)\mu}$

$$\beta(c) ::= c \log c - c + 1$$

Dependence on c?

c=1:	$\beta(c) = 0$	(useless)
c large:	$\beta(c) \approx c \log c$	LARGE
c=e:	$\beta(c) = 1$	(Problem 1)
c=(1+ε):	$\beta(c) = \Theta(\varepsilon^2)$	

The Lottery

Example: Pick 4

Pick a lottery number

0000, 0001, …., 9999

The Lottery

1,000,000 people buy $1 ticket

μ ::= Expected # winners =

$$\frac{1{,}000{,}000}{10{,}000} = 100$$

The Lottery

How much reserve $$

does lottery need?

Must be prepared for more

than expected # winners:

say a day with 1000 winners?

Chernoff Bound for Lottery

Let $c = e$, so $\beta(c) = 1$:

$$\Pr\{R \geq \underbrace{e \cdot \mu}_{273}\} \leq e^{-1 \cdot \mu} = \underbrace{e^{-100}}_{\text{Don't worry!}}$$

Chance of even 173 extra winners

is *negligible*. Small reserve $$ OK.

Large Deviation

System design must handle *rare* overloads to be reliable. That's why Chernoff more important in systems than "classical" results like the Central Limit Theorem.

Akamai Server Network

- Total Load $T = T_1 + T_2 + T_3 + \cdots + T_n$
- $T_i = 1$ if ith query routed to this server
- $T_i = 0$ if not
- Server handles average 1M calls/day:

$$E[T] = 1{,}000{,}000$$

Designing One Server to Survive Overload

Probability rate fluctuates by 1%:

$$\Pr\{T \geq 1.01\,M\}$$
$$\leq e^{-(1.01 \log 1.01 + 1.01 - 1)M}$$
$$\leq 2 \cdot 10^{-22} \quad (\textit{very small})!!!$$

1% excess capacity more than enough to make overload very unlikely.

The Whole Server Network

Akamai has a 1000 servers, and all get same average load per day.

Use Boole's inequality:

Pr {*any* server overloads}

$$\leq 1000 \Pr\{\text{this server overloads}\}$$
$$\leq 2 \cdot 10^{-19} \quad (\textit{still} \text{ very small})!!!$$

Chernoff vs. Binomial Bounds

If $\Pr\{T_i = 1\}$ same for all i, then

$R = \sum T_i$ is Binomial.

Even here, Chernoff bound is nearly as good as tight bounds based on Stirling.

In Class Problems

Problems 2,3