

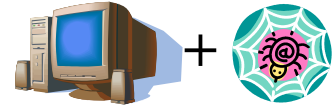


Avoiding Large Deviations (Chernoff Bound)



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 $E[\text{\#successes}]$ is large, then

$\Pr\{\text{never succeeding}\}$ is

“exponentially small”:

$$\Pr\{R = 0\} \leq e^{-\mu R}$$



Deviation from the Mean

This is a deviation from mean result:

$\Pr\{\text{observed value far from expected value}\}$ is SMALL



Deviation from the Mean

$$\Pr\{R = 0\} \leq e^{-\mu R}$$

R below μ by $\underbrace{\mu}_{\text{far}}$ (if μ_R large) $\underbrace{\quad}_{\text{small}}$

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



In Class Problem

Problem 1



Chernoff Bound

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

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

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

only when μ is large.



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

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

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)
- $c=(1+\epsilon)$: $\beta(c) = \Theta(\epsilon^2)$

4	9	13	7
12	10	6	
3	1	8	11
15	5	14	2

The Lottery

Example: **Pick 4**

Pick a lottery number

0000, 0001, ..., 9999

4	9	13	7
12	10	6	
3	1	8	11
15	5	14	2

The Lottery

1,000,000 people buy \$1 ticket

$\mu ::=$ Expected # winners =

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

4	9	13	7
12	10	6	
3	1	8	11
15	5	14	2

The Lottery

How much reserve \$\$
does lottery need?

Must be prepared for more
than expected # winners:
say a day with 1000 winners?

4	9	13	7
12	10	6	
3	1	8	11
15	5	14	2

Chernoff Bound for Lottery

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

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

Chance of even 173 extra winners
is *negligible*. Small reserve \$\$ OK.

4	9	13	7
12	10	6	
3	1	8	11
15	5	14	2

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.

4	9	13	7
12	10	6	
3	1	8	11
15	5	14	2

Akamai Server Network

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

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

4	8	13	7
12	10	5	
3	1	6	14
15	2	11	9

Designing One Server to Survive Overload

Probability rate fluctuates by 1%:

$$\begin{aligned} \Pr \{T \geq 1.01 M\} \\ &\leq e^{-(1.01 \log 1.01 + 1.01 - 1)M} \\ &\leq 2 \cdot 10^{-22} \quad (\text{very small})!!! \end{aligned}$$

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

4	8	13	7
12	10	5	
3	1	6	14
15	2	11	9

The Whole Server Network

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

Use Boole's inequality:

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

4	8	13	7
12	10	5	
3	1	6	14
15	2	11	9

Chernoff vs. Binomial Bounds

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

$$R = \sum T_i \text{ is Binomial.}$$

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

4	8	13	7
12	10	5	
3	1	6	14
15	2	11	9

In Class Problems

Problems 2,3