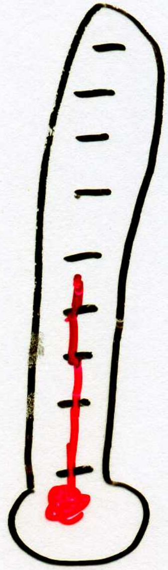
Constructive Proofs of Concentration Bounds

Russell Impagliazzo
UCSD & IAS

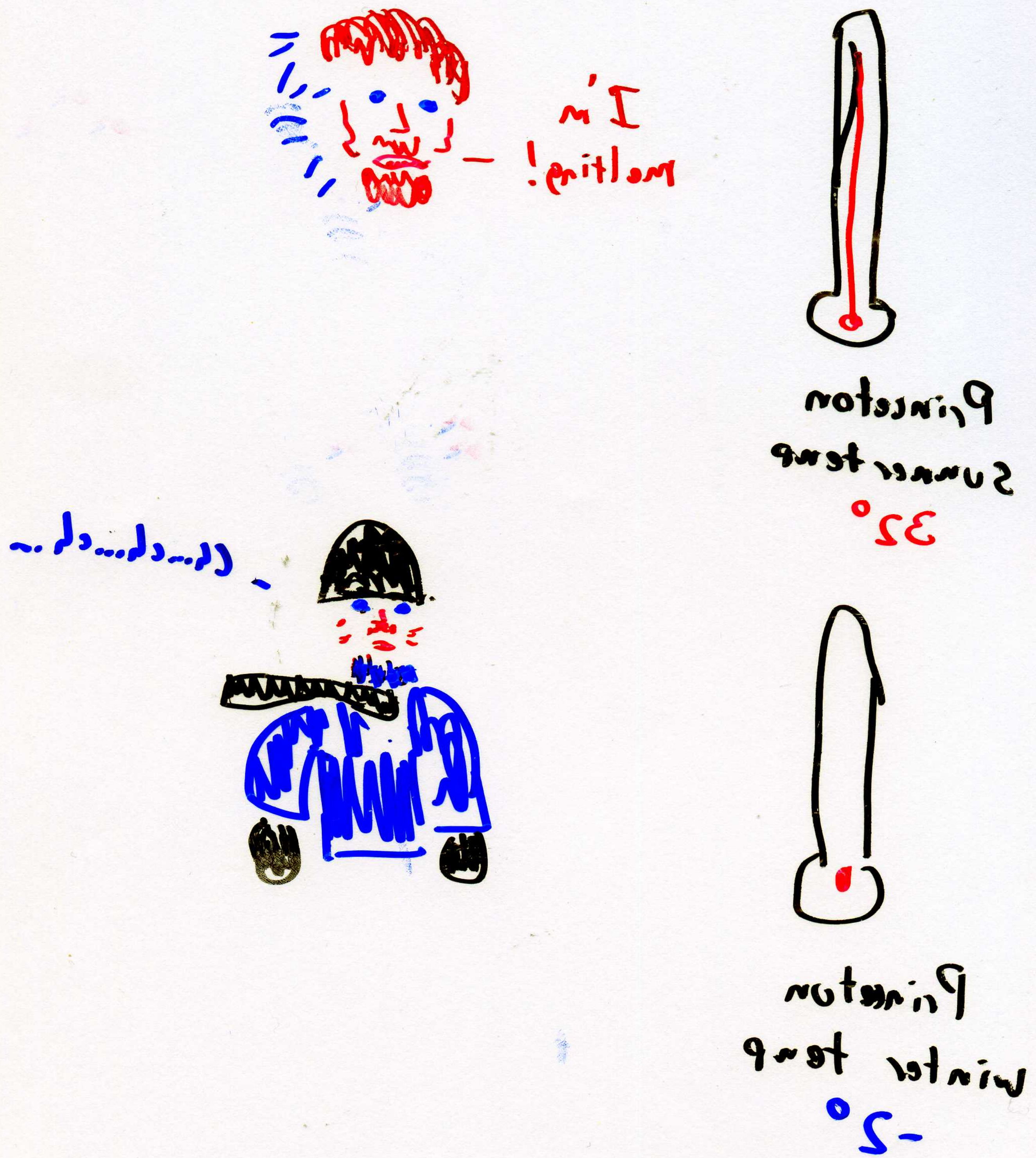
Valentine Kabanets
SFU

Averages can be misleading



Princeton, NJ
Average Temp
15°

San Diego, CA
Average Temp
18°


I'm melting!
Princeton summer temp
32°
Ch..ch..ch..
Princeton winter temp
-2°


At last, some beach weather!
San Diego summer temp
22°


Brr... Glad I wore a long-sleeve today!
San Diego winter temp
14°

Concentration bounds

When are random variables almost always close to their expected values?

Chernoff bound:
Sum of independent Boolean variables

PS
Sum of negatively correlated variables

Azuma's inequality
Martingale with bounded difference

Expander walk "Chernoff bound"
Number of times random walk visits subset of nodes

exidle

Security based on hard problems requires reliable intractability not occasional intractability

CAPTCHA: puzzle to distinguish humans from AI

20% success rate vs. Gmail CAPTCHA in 2008 (Wikipedia)

Hardness amplification

$f(x)$, occasionally hard $\Rightarrow$

$F(X)$, reliably hard

Direct product Y82

$$X = x_1, x_2, \dots x_K$$

$$F(X) = f^{(K)}(X) = f(x_1), f(x_2), \dots f(x_K)$$

XOR Y82

$$X = x_1, x_2, \dots x_K$$

$$F(X) = f^{\oplus K}(X) =$$

$$f(x_1) \oplus f(x_2) \oplus \dots \oplus f(x_K)$$

Applications

Cryptography Yao 82, Levin 87, BIN 97, CHS 05, DIJK 09, ...

Derandomization "Boolean functions hard ≈ 1/2 the time can replace random bits" Yao 82, Levin 87, BFNW, I 95, IW 97, ...

Average-case complexity "If some Dist NP problems are easy of the time, then they're easy almost all the time" O'D, Trev, IJK, IJKW

Less direct

Error correcting codes

ABNNR 92, Trev 03, I 02,
IJK 06

PCPs and hardness of approximation

Raz 95, Rao 08, H 07,
IJKW 09, DM 10

Lower bounds

Direct product
constructions
are
simple
intuitive
generic
flexible
useful

BUT DO THEY WORK?

NOT ALWAYS

and NOT ALWAYS

INTUITIVELY even

when they do

Uniform models:

$f^{(k)}$ can be computable with $1/n$ advantage, but f has constant hardness

Cryptographic protocols: BIN97, PW07

Parallel composition of protocols may not help soundness

PCPs Raz 98

Parallel repetition for constraint satisfaction problems may not improve soundness exponentially w/ intuitive constant imã

Problem: Humans are imperfect

CAPTTCHA's are ambiguous

Human success $\approx 80\% < 1$

As CAPTTCHA length $K \to \infty$

DP Theorem

Bot success $\to 0$

but human success $\to 0$ too!

Thresholded direct product [IJK, DIJK]

Use direct product construction.

$$X = x_1, x_2, \ldots x_k$$

$$f^{(k)}(x) = f(x_1), \ldots f(x_k)$$

But accept answers

$a_1 \ldots a_k$ if

large fraction correct

$$H(\vec{a}, f^{(k)}(x)) \leq \Theta k.$$

bot correctness < Θ < human correctness

$K \to \infty$

Human success $\to 1$

(Chernoff bounds)

Bot success $\to 0$?

(What we need to show)

IJK '08 Thresholded direct product theorem

Unger '09

Strong XOR Theorem $\rightarrow$

Thresholded DP

Abstract probabilistic view

$$z_i = \begin{cases} 1 & \text{if adv. succeeds on challenge } i \\ 0 & \text{o.w.} \end{cases}$$

δ failure chance

$\theta < \delta$ threshold

Strong XOR $\forall S \subseteq \{1, \dots k\}$

$$\text{Prob}\left[\bigoplus_{i \in S} z_i = 1\right] \le \frac{1}{2} + (1-2\delta)^{|S|}$$

$\rightarrow$ Strong TDP

$$\text{Prob}\left[\#\{z_i = 0\} \ge \theta k\right] \le e^{-(\delta-\theta)^2 k/2}$$

Here

Factor Unger's theorem

XOR → DP → Thresholded DP

$E[\bigwedge_{i \in S} Z_i] \lesssim (1-\delta)^{|S|} \rightarrow$

$\text{Prob}[\sum Z_i \geq (1-\Theta)K] \leq e^{-(\Theta-\delta)^2 K/2}$

Like Chernoff bound and generalization (Panconesi, Srinivasan)

But actually gives "new" proofs of concentration bounds

Constructive: From failure of concentration, find failure of independence

New results

Effective version

If adversary successfully attacks thresholded DP, we find S with attack on DP for S effectively

Limits on basic approach

Reductions in abstract setting cannot have ideal preservation of advantage

Circumventing limits via conditioning

Basic Chernoff bound

[Bernstein]

Z_i, $i = 1 \ldots K$, iid Boolean variables,

$$\text{Prob}[Z_i = 1] = \mu$$

Let $\gamma > \mu$, $D(\mu \| \gamma) =$

$$\mu \log \frac{\gamma}{\mu} + (1-\mu) \log\left(\frac{1-\gamma}{1-\mu}\right)$$

$$\geq (\mu - \gamma)^2 / 2$$

Let $P_\gamma = \text{Prob}\left[\sum Z_i \geq \gamma K\right]$

$$P_\gamma \leq e^{-K D(\mu \| \gamma)}$$

(Think $\mu = 1 - \delta$, $\gamma = 1 - \theta$)

Proof

$$\forall S \subseteq \{1 .. k\}$$

$$\text{Prob}\left[\bigwedge_{i \in S} Z_i\right] \le u^{|S|}$$

(Only place we use independence, [PS])

For q to be determined,

pick S by putting each $i \in S$

ind. w/ prob. q

What is $\text{Prob}\left[\bigwedge_{i \in S} Z_i\right]$?

Pick S, then Z:

$$\boxed{u}\ \boxed{u}\ \underset{}{\overset{\notin S}{\boxed{1}}}\ \boxed{u}\ \overset{\notin S}{\boxed{1}}\ \boxed{u}\ \boxed{u}\ \boxed{1}\quad u^{|S|}$$

Pick Z, then S

$$\overset{\notin S}{\boxed{0}}\ \overset{\notin S}{\boxed{0}}\ \boxed{1}\ \boxed{1}\ \boxed{1} \ldots \boxed{1} \ldots \overset{\notin S}{\boxed{0}} : (1-q)^{\#0\text{'s}}$$

$$\mathrm{Prob}\left[\bigwedge_{i\in S} Z_i\right] \le$$

$$\sum_{t=0}^{K} \mathrm{Prob}\left[|S|=t\right]\cdot u^t$$

$$= \sum_{t=0}^{K} \binom{K}{t} q^t (1-q)^{K-t} u^t =$$

$$(qu+(1-q))^K =$$

$$(1-q(1-u))^K$$

No i has both
$i \in S$ (prob q)
and $Z_i = 0$ (prob $(1-u)$)

On the other hand,

with prob p_γ,

$\sum z_i > \gamma K$.

Conditional prob $\bigwedge_{i \in S} z_i = 1$

is prob $i \notin S$ $\forall i$ with

$z_i = 0 \geq (1-q)^{K-\gamma K}$

$\therefore$ $p_\gamma (1-q)^{K(1-\gamma)} \leq \text{Prob} \bigwedge_{i \in S} z_i \leq$

$(1-q(1-u))^K$

$$p_\gamma \leq \left(\frac{1-q(1-u)}{(1-q)^{1-\gamma}} \right)^K$$

So $p_\gamma \le h(q)^K$

$$h(q) = \frac{1-(1-\mu)q}{(1-q)^{1-\gamma}}$$

$h(q)$ is minimized at

$$q^* = \frac{\gamma-\mu}{\gamma(1-\mu)}$$

$$h(q^*) = \left(\frac{\mu}{\gamma}\right)^{\gamma}\left(\frac{1-\mu}{1-\gamma}\right)^{1-\gamma}$$

$$= e^{-D(\gamma\|\mu)}$$

What if not identical?

$$P[\,Z_i = 1\,] = \mu_i,$$

$$\mu = \frac{1}{K} \Sigma \mu_i$$

Only change

$$\text{Prob}\left[\bigwedge_{i \in S} Z_i\right] =$$

$$\sum_S q^{|S|}(1-q)^{K-|S|} \prod_{i \in S} \mu_i =$$

$$\prod_{i=1}^{K} \underbrace{(q\mu_i + 1 - q)}_{\text{maximized when all } \mu_i = \mu} \leq (q\mu + 1 - q)^K$$

What if not Boolean?

$Z_i \in [0,1]$, $\mu_i = E[Z_i]$

After picking Z_i, let

$Y_i = 1$ w/ prob Z_i.
0 o.w.

If $\sum Z_i \geq \gamma K$, constant conditional prob. that

$$\sum Y_i \geq \lfloor \gamma K \rfloor$$

(Siegal)

$\forall S$ $\quad \text{Prob} \bigwedge_{i \in S} Y_i = E\left(\prod_{i \in S} Z_i\right) = \prod_{i \in S} \mu_i$

Apply bound to Y_i, same order of bound for Z_i.

What if not independent?

Difference Martingale

$D_1, \ldots D_K$, $D_i \in [-1/2, \frac{1}{2}]$

$E[D_i \mid D_1 = d_1, \ldots D_{i-1} = d_{i-1}] = 0$

Azuma's Inequality [Bernstein]

$\text{Prob}(|\Sigma D_i| \geq \gamma K) = O(e^{-\gamma^2 K/2})$

Proof: Let $Z_i = \frac{1}{2} + D_i$.

$E[\prod_{i \in S} Z_i] = (\frac{1}{2})^{|S|}$

Apply prev. bound w/ $\mu = 1/2$,

$\gamma' = \frac{1}{2} + \gamma$

Expander walks

Let G be an expander

w/ second largest eigenvalue

λ, let $W \subseteq V(G)$,

$|W| = \mu |V(G)|$,

$\mu > 6\lambda$.

Let $v_1, \ldots v_k$ be a k-step

random walk in G.

AKS 87, AFWZ95

$$\text{Prob}[\forall i,\ v_i \in W] \leq (\mu + 2\lambda)^k$$

$$\forall S \subseteq \{1 \ldots k\} \quad \text{Prob}[\forall i \in S,\ v_i \in W] \leq (\mu + 2\lambda)^{|S|}$$

Corollary

$$\text{Prob}[\#v_i \in W \geq (\mu + \epsilon)k] \leq e^{-\epsilon^2 (1-\lambda) k / (2 \ln 4/\epsilon)}$$

Gil98 $\leq e^{-\epsilon^2 (1-\lambda) k / 4}$

Example

Pick $A \subseteq \{1 \dots N\}$,

$|A| = K$

For $W \subseteq \{1 \dots N\}$,

$|W| = \mu N$, bound

$\text{Prob}\left[|A \cap W| \geq \gamma K\right]$

$Z_i = 1$ if i'th element of $A \in W$

Not independent, but

$$\text{Prob}\left[\bigwedge_{i \in S} Z_i = 1\right] = \frac{|W|}{N} \cdot \frac{|W|-1}{N-1} \cdots$$

$$\frac{|W| - |S| + 1}{N - |S| + 1} \leq \mu^{|S|}$$

$\therefore$ Same bound as if independent.

DP → TDP

Attack on TDP → Attack on DP

a_{TDP} : Given K challenges,
solve $\geq k-\Theta K$ with prob >

$p > e^{-H(1-\Theta \| 1-\delta)K}$

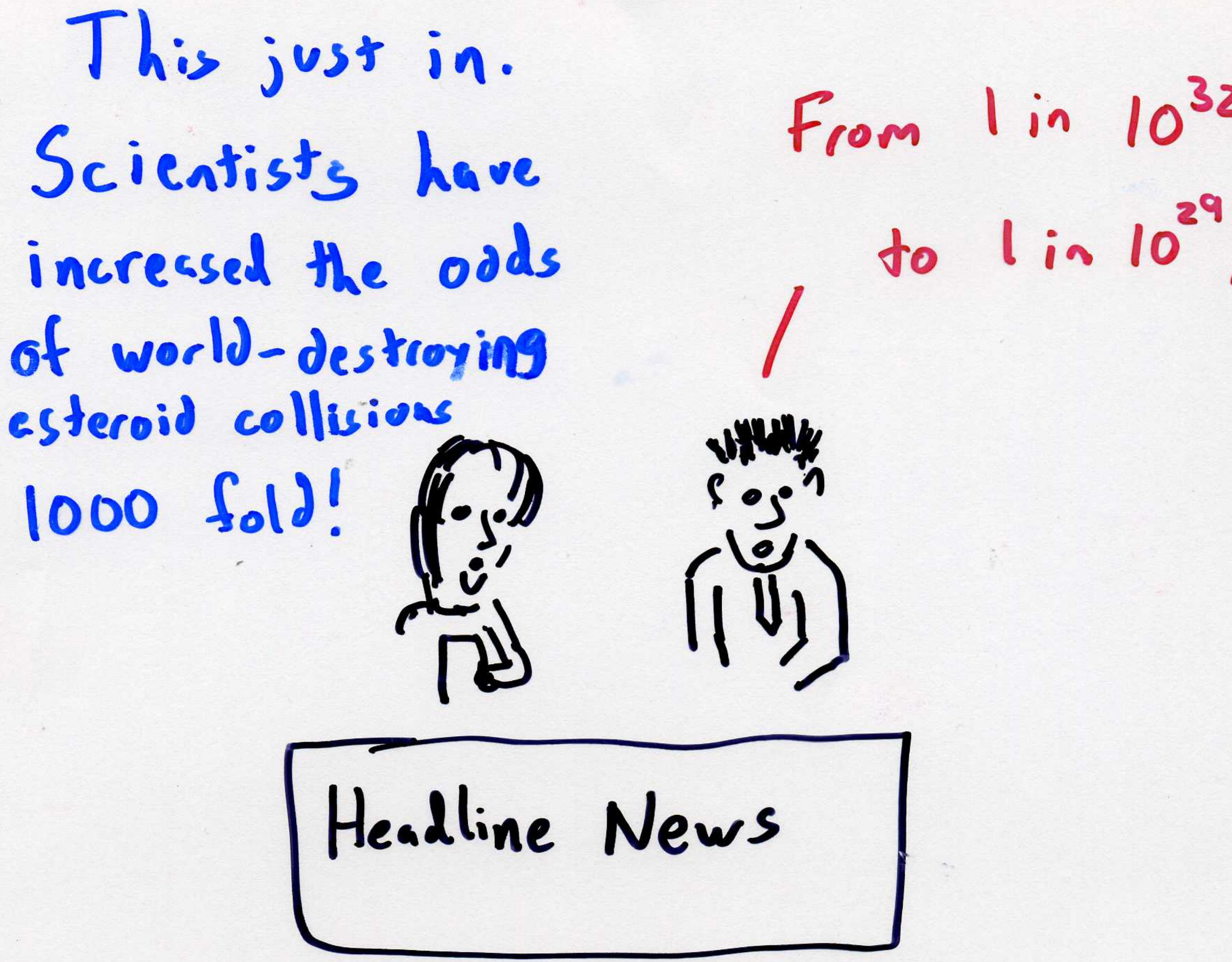
a_{DP} : Given K' challenges

Pick $S \subseteq \{1, \dots, K\}$, $|S| = K'$
uniformly.

Place real challenges in S.
Simulate challenges for $\bar{S}$
Return answers for S.

$\exists K'$, success prob $\geq (1-\delta)^{K'}$

Problem: could still be negligible


This just in. Scientists have increased the odds of world-destroying asteroid collisions 1000 fold!
From 1 in 10^{32}
to 1 in 10^{29}
Headline News

Theorem

$\forall p > 2e^{-H(\mu \| \delta)}$

$\exists q$ so that

$E[\alpha_s] \geq \Omega\left(p^{4/\delta(\mu-\delta)}\right)$

Corollary: $\forall$ constants

$0 < \delta < \mu < 1$, $\frac{1}{}$

breaking TDP with poly.

prob $\Rightarrow \frac{1}{poly}$ advantage for DP

breaking algorithm

Let probability of DP success for set S

be $(1-\delta)^{|S|} + \alpha_S =$

$$\mu^{|S|} + \alpha_S$$

Then using prev. analysis + additivity of expectations

$$(1-q(1-\mu))^K + \underset{S}{E}\,\alpha_S \geq p\,(1-q)^{(1-\gamma)K}$$

$$\underset{S}{E}\,\alpha_S \geq p \quad (1-q(1-\mu))^K \,(p - h(q))^K$$

What if $\gamma - \delta = \frac{1}{\log k}$?

Not polynomial.

Unfortunately, no "oblivious" reduction preserves advantage polynomially in this case.

$\forall \gamma, \delta, p \quad \exists$ distr. $Z_1, \dots Z_K$

so that $\quad \alpha_S = p^{\Omega\left(\frac{1}{\gamma - \delta}\right)}$

$\forall S \subseteq \{1, \dots K\}$

$Z_1, \dots Z_K = \begin{cases} \text{all } 0\text{'s} & \text{prob } 1-p \\ \text{random } \gamma K \ 1\text{'s} & \text{o.w.} \end{cases}$

Conditioning reductions

Another "witness of non-independence"

$S \subsetneq \{1, \dots k\}$,

Event : depends on Z_i, $i \in S$

$j \notin S$

$\text{Prob}[Z_j = 1 \mid \text{Event}] \overset{?}{=} \mu + \alpha$

Visibility: α large

Prob [Event] large

Theorem

$$\exists p_0 = O(e^{-\Omega((\gamma-\mu)^2 k)})$$

$\forall p > p_0$ $\text{Prob}[\Sigma Z_i \geq \gamma k] \geq p \rightarrow$

$\exists S, j \notin S, \gamma'$

$$\text{Prob}[\sum_{i \in S} Z_i \geq \gamma' |S|] = \Omega(p)$$

$$\text{Prob}[Z_j \mid \sum_{i \in S} Z_i \geq \gamma' |S|] \geq \mu + \Omega(\gamma - \mu)$$

Proof

Pick S at random

$j \in \bar{S}$ at random.

$$\gamma' = \frac{\gamma + \mu}{2}$$

If $\sum z_i \geq \gamma k$,

whp $\sum_{i \in S} z_i \geq \gamma' \frac{k}{2}$

$$\sum_{i \notin S} z_i \geq \gamma' \frac{k}{2}$$

Very small prob. that

$\sum_{i \in S} z_i \geq \gamma' \frac{k}{2}$ but

$$\sum_{i \notin S} z_i < \left(\frac{\gamma' + \mu}{2}\right) \frac{k}{2}$$

Gives "generic" proof

of TDP but:

a) must be able to
evaluate success
in simulations

b) full proof requires
sampling lemma technique
from [Rez, IJK]

Ends up looking similar to
known proofs, applying
to similar applications