Upper tail estimates: Arithmetic progressions and the missing log

Lutz Warnke

Georgia Tech

What is this talk about?

Motivation

Behaviour of a function of independent random variables $\xi_1, \ldots, \xi_N$:

$$X = F(\xi_1, \ldots, \xi_N)$$

- The random variable X often counts the number of certain objects

Tail estimates

Want *exponential bounds* for the lower/upper tail:

$$\mathbb{P}(X \leq (1-\varepsilon)\mathbb{E}X) \quad \text{and} \quad \mathbb{P}(X \geq (1+\varepsilon)\mathbb{E}X)$$

- Allow us to show that whp $X \approx \mathbb{E}X$
- Exponential decay useful in union bound arguments

Topic of his talk

Some *best possible* upper tail estimates (exponentially small)

Upper tail is more interesting

Lower tail: $\mathbb{P}(X \le (1-\varepsilon)\mathbb{E}X)$

Janson's + Suen's inequality give good upper bounds

- Janson's inequality often *best possible* (lower bounds of Janson–W.)

Upper tail: $\mathbb{P}(X \ge (1+\varepsilon)\mathbb{E}X)$

Best methods often leave logarithmic gap factors in the exponent, e.g.,

$$\exp\Big(-C\Psi\log(1/p)\Big) \le \mathbb{P}(X \ge 2\mathbb{E}X) \le \exp\Big(-c\Psi\Big),$$

- Moment based method of Janson–Oleszkiewicz–Ruciński
- Closing the gap is technical challenge ('infamous upper tail problem')

Upper tail is more interesting

Upper tail: $\mathbb{P}(X \geq (1+\varepsilon)\mathbb{E}X)$

Best methods often leave logarithmic gap factors in the exponent, e.g.,

$$\exp\Big(-C\Psi\log(1/p)\Big) \leq \mathbb{P}(X \geq 2\mathbb{E}X) \leq \exp\Big(-c\Psi\Big),$$

- Moment based method of Janson–Oleszkiewicz–Ruciński
- Closing the gap is technical challenge ('infamous upper tail problem')

Why should we care?

- Natural probability question in concentration of measure
- Requires deeper understanding of the problem (how $X \geq 2\mathbb{E}X$ arises)
- Extra $\log(1/p)$ might help in removing log-factors from other results
- Test / develop methods for proving concentration inequalities

Case study: Arithmetic progressions

$[n]_p$ = random subset: $j \in [n]$ included independently with probability p
X = number of k-term arithmetic progressions in $[n]_p$

Lower tail: exponential decay

Janson's inequality + Janson–W. result (lower bound) gives

$$\mathbb{P}(X \leq (1-\varepsilon)\mathbb{E}X) = \exp\Big(-\Theta(\varepsilon^2)\min\big\{\mathbb{E}X,\ \mathbb{E}|[n]_p|\big\}\Big)$$

Upper tail: logarithmic gap

Janson–Ruciński obtained via a moment-based method

$$\exp\Big(-C_\varepsilon\sqrt{\mathbb{E}X}\log(1/p)\Big) \leq \mathbb{P}(X \geq (1+\varepsilon)\mathbb{E}X) \leq \exp\Big(-c_\varepsilon\sqrt{\mathbb{E}X}\Big)$$

Case study: Arithmetic progressions

$[n]_p$ = random subset: $j \in [n]$ included independently with probability p
X = number of k-term arithmetic progressions in $[n]_p$

Upper tail: logarithmic gap

Janson–Ruciński obtained via a moment-based method

$$\exp\Big(-C_\varepsilon \sqrt{\mathbb{E}X}\log(1/p)\Big) \le \mathbb{P}(X \ge (1+\varepsilon)\mathbb{E}X) \le \exp\Big(-c_\varepsilon \sqrt{\mathbb{E}X}\Big)$$

Resolving the tail behavior of k-term APs (W. 2013+)

We establish the missing logarithm using new techniques:

$$\mathbb{P}(X \ge (1+\varepsilon)\mathbb{E}X) = \exp\Big(-\Theta(1)\min\big\{\mathbb{E}X,\ \sqrt{\mathbb{E}X}\log(1/p)\big\}\Big),$$

and can also recover the 'correct' dependence on ε

Why are lower and upper tails so different?

Lower and upper tails are quite different (for k-term APs)

Ignoring polylogarithmic factors:

$$-\log \mathbb{P}(X \le 0.5\mathbb{E}X) \cong \min\{\mathbb{E}X,\ \mathbb{E}|[n]_p|\} \cong \min\{n^2p^k, np\}$$

$$-\log \mathbb{P}(X \ge 2\mathbb{E}X) \cong \sqrt{\mathbb{E}X} \cong np^{k/2}$$

One conceptual key difference

- Can create many APs by adding small interval $[m] = \{1, \ldots, m\}$
- Can *not* significantly *reduce* number of APs by removing few elements (extreme case: all/most numbers contained in only $O(1)$ APs)

Take-home message

- Lower tail mainly governed by '*global* behaviour'
- Upper tail mainly governed by '*local* behaviour'

Flavour of our results

Intuitive punchline of our results (W. 2015+)

Assume that basic application of Kim–Vu gives

$$\mathbb{P}(X \geq 2\mathbb{E}X) \leq \exp\left(-c(\mathbb{E}X)^{1/q}\right).$$

Then under some additional 'strictly-balanced-like condition' we obtain

$$\mathbb{P}(X \geq 2\mathbb{E}X) \leq \exp\left(-c \min\{\mathbb{E}X,\ (\mathbb{E}X)^{1/q} \log(1/p)\}\right).$$

Improvement conceptually important

- Exponential decay best possible for additive combinatorics examples
- The 'strictly-balanced' condition can *not* be dropped
- Proof develops new tools/ideas for obtaining extra logarithmic factor

Flavour of our results

Intuitive punchline of our results (W. 2015+)

Assume that basic application of Kim–Vu gives

$$\mathbb{P}(X \geq 2\mathbb{E}X) \leq \exp\left(-c(\mathbb{E}X)^{1/q}\right).$$

Then under some additional 'strictly-balanced-like condition' we obtain

$$\mathbb{P}(X \geq 2\mathbb{E}X) \leq \exp\left(-c \min\{\mathbb{E}X,\ (\mathbb{E}X)^{1/q} \log(1/p)\}\right).$$

Best possible for examples in additive combinatorics:

- k-term arithmetic progressions
- Schur triples ($x_1 + x_2 = x_3$)
- Additive quadruples ($x_1 + x_2 = y_1 + y_2$)
- (r, s)-sums ($x_1 + \cdots + x_r = y_1 + \cdots + y_s$)

Flavour of our results

Intuitive punchline of our results (W. 2015+)

Assume that basic application of Kim–Vu gives

$$\mathbb{P}(X \geq 2\mathbb{E}X) \leq \exp\left(-c(\mathbb{E}X)^{1/q}\right).$$

Then under some additional 'strictly-balanced-like condition' we obtain

$$\mathbb{P}(X \geq 2\mathbb{E}X) \leq \exp\left(-c\min\left\{\mathbb{E}X,\ (\mathbb{E}X)^{1/q}\log(1/p)\right\}\right).$$

The 'strictly-balanced' condition can not be dropped:

- There are families of examples where exponent is of order $(\mathbb{E}X)^{1/q}$, i.e., we do *not* have an extra logarithmic factor

Flavour of our results

Intuitive punchline of our results (W. 2015+)

Assume that basic application of Kim–Vu gives

$$\mathbb{P}(X \geq 2\mathbb{E}X) \leq \exp\left(-c(\mathbb{E}X)^{1/q}\right).$$

Then under some additional 'strictly-balanced-like condition' we obtain

$$\mathbb{P}(X \geq 2\mathbb{E}X) \leq \exp\left(-c \min\{\mathbb{E}X,\ (\mathbb{E}X)^{1/q}\log(1/p)\}\right).$$

Exponent resembles two different behaviours:

- Poisson behaviour: $\exp\bigl(-c\mathbb{E}X\bigr)$
- 'Clustered behaviour': $\exp\bigl(-c(\mathbb{E}X)^{1/q}\log(1/p)\bigr) = p^{c(\mathbb{E}X)^{1/q}}$

Proof setup

We take a *combinatorial* point of view to concentration (no induction)

Random induced subhypergraph

Given a k-uniform hypergraph $\mathcal{H}$ with vertex set $V = [n]$, let

$$\mathcal{H}_p = \mathcal{H}\big[V_p\big],$$

i.e., hypergraph induced by random subset $V_p := [n]_p$ of the vertices

Counting the number of edges

Many counting problems can be written as

$$X = e(\mathcal{H}_p)$$

Example: k-term arithmetic progressions

Edge set: k-element subsets of $[n]$ corresponding to arithm. progressions

Proof strategy (1/2)

Our approach relies on a blend of *combinatorial* + probabilistic arguments

High-level proof strategy

1. Define good events $\mathcal{G}_i$ which imply that $X = e(\mathcal{H}_p)$ is small:

$$\text{all } \mathcal{G}_i \text{ hold} \quad \Longrightarrow \quad X < (1+\varepsilon)\mathbb{E}X$$

2. Show that these 'good' events $\mathcal{G}_i$ are *very* unlikely to fail:

$$\mathbb{P}(\text{some } \mathcal{G}_i \text{ fails}) \le \exp\bigl(- \cdots\bigr)$$

3. Via 1+2 we then have

$$\mathbb{P}(X \ge (1+\varepsilon)\mathbb{E}X) \le \mathbb{P}(\text{some } \mathcal{G}_i \text{ fails}) \le \exp\bigl(- \cdots\bigr)$$

Proof strategy (2/2)

One exemplary 'good event' (proof uses several)

For ALL $\mathcal{F} \subseteq \mathcal{H}_p$ with small max-degree we have $e(\mathcal{F}) < (1+\varepsilon/2)\mathbb{E}X$

- In words: ALL subhypergraphs with small max-degree have few edges

Sparsification idea (simplified)

1. Use combinatorial arguments to gradually decrease the max-degree

$$\mathcal{H}_p = \mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \cdots \supseteq \mathcal{F}_{q-1} \supseteq \mathcal{F}_q$$

2. 'Good events' then ensure that the number of edges satisfies

$$X = e(\mathcal{H}_p) = \underbrace{e(\mathcal{F}_q)}_{<(1+\varepsilon/2)\mathbb{E}X} + \underbrace{\sum_{1 \le i < q} e(\mathcal{F}_i \setminus \mathcal{F}_{i+1})}_{\le \varepsilon \mathbb{E}X/2} < (1+\varepsilon)\mathbb{E}X$$

A SURPRISING INEQUALITY

One exemplary 'good event' (proof uses several)

For all $\mathcal{F} \subseteq \mathcal{H}_p$ with small max-degree we have $e(\mathcal{F}) < (1+\varepsilon/2)\mathbb{E}X$

- Statement for all subhypergraphs might seem too ambitious, but

A useful insight (W. 2013+)

We get Chernoff-like tail estimate for

$$\mathbb{P}(\text{there is } \mathcal{F} \subseteq \mathcal{H}_p \text{ with } \Delta_1(\mathcal{F}) \leq C \text{ and } e(\mathcal{F}) \geq \mu + t)$$

WITHOUT taking a union bound over all subhypergraphs $\mathcal{F} \subseteq \mathcal{H}_p$

- Estimates for all $\mathcal{F} \subseteq \mathcal{H}_p$ enable additional combinatorial arguments

A SURPRISING+USEFUL INEQUALITY

One exemplary 'good event' (proof uses several)

For all $\mathcal{F} \subseteq \mathcal{H}_p$ with small max-degree we have $e(\mathcal{F}) < (1 + \varepsilon/2)\mathbb{E}X$

- Statement for all subhypergraphs might seem too ambitious, but

Chernoff-like estimate for all subhypergraphs (W. 2013+)

If $\mathcal{H}$ is a k-uniform with $\mu = \mathbb{E}e(\mathcal{H}_p)$, then for $C, t > 0$ we have

$$\mathbb{P}(\text{there is } \mathcal{F} \subseteq \mathcal{H}_p \text{ with } \Delta_1(\mathcal{F}) \leq C \text{ and } e(\mathcal{F}) \geq \mu + t)$$
$$\leq \exp\left(-\frac{\varphi(t/\mu)\mu}{kC}\right) \leq \exp\left(-\frac{t^2}{2kC(\mu + t/3)}\right),$$

where $\varphi(x) = (1 + x)\log(1 + x) - x$

- *NO* union bound over all subhypergraphs $\mathcal{F} \subseteq \mathcal{H}_p$ needed
- Estimates for all $\mathcal{F} \subseteq \mathcal{H}_p$ enable additional combinatorial arguments

Summary

Informal summary

Can often improve estimates for $\mathbb{P}(X \geq (1+\varepsilon)\mathbb{E}X)$ by logarithmic factor:

$$\leq \exp\Big(-c_\varepsilon\, \mu^{1/q}\Big) \quad \longrightarrow \quad \leq \exp\Big(-d_\varepsilon \min\big\{\mu,\ \mu^{1/q}\log(1/p)\big\}\Big),$$

where $\mu = \mathbb{E}X$ and p is as in random subset $[n]_p$ or random graph $G_{n,p}$

Remarks

- Sharp for several additive combinatorics examples (incl. arithm. progr.)
- More combinatorial approach + new tail inequalities
- Estimates for all $\mathcal{F} \subseteq \mathcal{H}_p$ enable additional combinatorial arguments

Open problem

Obtain 'missing log' for subgraph counts in $G_{n,p}$ (only special cases known)

Relative Estimates: more good events

$$\Delta_j(\mathcal{H}) = \max_{S \subseteq V(\mathcal{H}) : |S|=j} \left|\left\{ f \in \mathcal{H} : \; S \subseteq f \right\}\right|$$

$$= \text{upper bound for \# edges containing any } j \text{ vertices of } \mathcal{H}$$

Relative degree events ($Q_j < R_j$)

$\mathcal{D}_j \triangleq$ for all $\mathcal{F} \subseteq \mathcal{H}_p$: $\Delta_{j+1}(\mathcal{F}) \le R_{j+1}$ implies $\Delta_j(\mathcal{F}) \le R_j$

$\mathcal{D}_j^+ \triangleq$ for all $\mathcal{F} \subseteq \mathcal{H}_p$: $\Delta_{j+1}(\mathcal{F}) \le Q_{j+1}$ implies $\Delta_j(\mathcal{F}) \le Q_j$

Sparsification event (by deleting edges)

$\mathcal{E} \triangleq$ $\Delta_1(\mathcal{H}_p) \le R_1$ implies existence of subhypergraph $\mathcal{J} \subseteq \mathcal{F}$ with $\Delta_{k-1}(\mathcal{J}) \le Q_{k-1}$ and $e(\mathcal{H}_p \setminus \mathcal{J}) < \varepsilon \mathbb{E} X/2$

Remarks

- Sparsification in spirit of Rödl–Ruciński 'deletion lemma', which focuses mainly on (i) the removal of *vertices* and (ii) global object counts
- New approach: combinatorics + BK-inequality ('disjoint occurrence')

Relative Estimates: more good events

$$\Delta_j(\mathcal{H}) = \max_{S \subseteq V(\mathcal{H}) : |S| = j} \left|\left\{ f \in \mathcal{H} : S \subseteq f \right\}\right|$$

$$= \text{upper bound for } \# \text{ edges containing any } j \text{ vertices of } \mathcal{H}$$

Relative degree events ($Q_j < R_j$)

$\mathcal{D}_j \triangleq$ for all $\mathcal{F} \subseteq \mathcal{H}_p$: $\Delta_{j+1}(\mathcal{F}) \leq R_{j+1}$ implies $\Delta_j(\mathcal{F}) \leq R_j$

$\mathcal{D}_j^+ \triangleq$ for all $\mathcal{F} \subseteq \mathcal{H}_p$: $\Delta_{j+1}(\mathcal{F}) \leq Q_{j+1}$ implies $\Delta_j(\mathcal{F}) \leq Q_j$

Sparsification event (by deleting edges)

$\mathcal{E} \triangleq$ $\Delta_1(\mathcal{H}_p) \leq R_1$ implies existence of subhypergraph $\mathcal{J} \subseteq \mathcal{F}$ with $\Delta_{k-1}(\mathcal{J}) \leq Q_{k-1}$ and $e(\mathcal{H}_p \setminus \mathcal{J}) < \varepsilon \mathbb{E}X/2$

First good event revisited

$\mathcal{G} \triangleq$ for all $\mathcal{F} \subseteq \mathcal{H}_p$: $\Delta_1(\mathcal{F}) \leq Q_1$ implies $e(\mathcal{F}) \leq (1 + \varepsilon/2)\mathbb{E}X$

New concentration inequality (simplified)

$(\xi_i)_{i \in \mathcal{A}}$: independent random variables
$(Y_\alpha)_{\alpha \in \mathcal{I}}$: indicator random variables with $Y_\alpha = F(\xi_i : i \in \alpha) \in \{0, 1\}$

Well-behaved variant of the sum $X := \sum_{\alpha \in \mathcal{I}} Y_\alpha$

Restriction to *subsum* where each Y_β depends on $\leq C$ variables

$$X_C := \max_{\mathcal{J} \subseteq \mathcal{I}} \left\{ \sum_{\alpha \in \mathcal{J}} Y_\alpha : \max_{\beta \in \mathcal{J}} \sum_{\alpha \in \mathcal{J} : \alpha \cap \beta \neq \emptyset} Y_\alpha \leq C \right\}$$

- $\alpha \cap \beta = \emptyset$ implies that Y_α and Y_β are independent

Chernoff-type upper tail estimate, simplified (W. 2013+)

If $\mu = \mathbb{E}X$, then for all $C, t > 0$ we have

$$\mathbb{P}(X_C \geq \mu + t) \leq \cdots \leq \exp\left(-\frac{t^2}{2C(\mu + t/3)}\right)$$