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:

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

- Allow us to show that whp $X \approx E[X]$
- 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: \( \Pr(X \leq (1 - \varepsilon)EX) \)

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

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

Upper tail: \( \Pr(X \geq (1 + \varepsilon)EX) \)

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

\[
\exp\left(-C\Psi \log\left(\frac{1}{p}\right)\right) \leq \Pr(X \geq 2EX) \leq \exp\left(-c\Psi\right),
\]

- Moment based method of Janson–Oleszkiewicz–Ruciński
- Closing the gap is technical challenge (‘infamous upper tail problem’)
Upper tail: \( \mathbb{P}(X \geq (1 + \varepsilon) \mathbb{E}X) \)

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

\[
\exp\left(-C \Psi \log(1/p)\right) \leq \mathbb{P}(X \geq 2 \mathbb{E}X) \leq \exp\left(-c \Psi\right),
\]

- 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

\([n]_p = \text{random subset: } j \in [n] \text{ included independently with probability } p\)

\(X = \text{number of } k\text{-term arithmetic progressions in } [n]_p\)

### Lower tail: exponential decay

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

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

### Upper tail: logarithmic gap

Janson–Ruciński obtained via a moment-based method

\[ \exp\left(-C_\varepsilon \sqrt{\mathbb{E}X \log(1/p)}\right) \leq \mathbb{P}(X \geq (1 + \varepsilon)\mathbb{E}X) \leq \exp\left(-c_\varepsilon \sqrt{\mathbb{E}X}\right) \]
Case study: Arithmetic progressions

$[n]_p = \text{random subset: } j \in [n] \text{ included independently with probability } p$

$X = \text{number of } k\text{-term arithmetic progressions in } [n]_p$

Upper tail: logarithmic gap

Janson–Ruciński obtained via a moment-based method

$$\exp\left(-C\varepsilon \sqrt{\mathbb{E}X \log(1/p)}\right) \leq \mathbb{P}(X \geq (1 + \varepsilon)\mathbb{E}X) \leq \exp\left(-c\varepsilon \sqrt{\mathbb{E}X}\right)$$

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

We establish the missing logarithm using new techniques:

$$\mathbb{P}(X \geq (1 + \varepsilon)\mathbb{E}X) = \exp\left(-\Theta(1) \min\{\mathbb{E}X, \sqrt{\mathbb{E}X \log(1/p)}\}\right),$$

and can also recover the ‘correct’ dependence on $\varepsilon$
Lower and upper tails are quite different (for $k$-term APs)

Ignoring polylogarithmic factors:

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

\[- \log P(X \geq 2E X) \cong \sqrt{E X} \cong np^{k/2}\]

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’
Assume that basic application of Kim–Vu gives
\[
P(X \geq 2\mathbb{E}X) \leq \exp\left(-c(\mathbb{E}X)^{1/q}\right).
\]
Then under some additional ‘strictly-balanced-like condition’ we obtain
\[
P(X \geq 2\mathbb{E}X) \leq \exp\left(-c \min\{\mathbb{E}X, (\mathbb{E}X)^{1/q} \log(1/p)\}\right).
\]

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
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 ($x_1 + \cdots + x_r = y_1 + \cdots + y_s$)
Intuitive punchline of our results (W. 2015+)

Assume that basic application of Kim–Vu gives

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

Then under some additional ‘strictly-balanced-like condition’ we obtain

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

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

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

Then under some additional ‘strictly-balanced-like condition’ we obtain

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

Exponent resembles two different behaviours:

- Poisson behaviour: \( \exp(-c\mathbb{E}X) \)
- ‘Clustered behaviour’: \( \exp(-c(\mathbb{E}X)^{1/q} \log(1/p)) = p^{c(\mathbb{E}X)^{1/q}} \)
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

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

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

**High-level proof strategy**

1. Define good events $G_i$ which imply that $X = e(H_p)$ is small:
   
   \[ \text{all } G_i \text{ hold } \implies X < (1 + \varepsilon)\mathbb{E}X \]

2. Show that these ‘good’ events $G_i$ are very unlikely to fail:
   
   \[ \mathbb{P}(\text{some } G_i \text{ fails}) \leq \exp(-\cdots) \]

3. Via 1+2 we then have
   
   \[ \mathbb{P}(X \geq (1 + \varepsilon)\mathbb{E}X) \leq \mathbb{P}(\text{some } G_i \text{ fails}) \leq \exp(-\cdots) \]
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

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

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

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

- \textit{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
Informal summary

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

\[
\leq \exp\left(-c_\varepsilon \mu^{1/q}\right) \quad \longrightarrow \quad \leq \exp\left(-d_\varepsilon \min\{\mu, \mu^{1/q} \log(1/p)\}\right),
\]

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

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

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

Relative degree events (\( Q_j < R_j \))

- \( \mathcal{D}_j \triangleq \text{for all } \mathcal{F} \subseteq \mathcal{H}_p : \Delta_{j+1}(\mathcal{F}) \leq R_{j+1} \) implies \( \Delta_j(\mathcal{F}) \leq R_j \)
- \( \mathcal{D}_j^+ \triangleq \text{for all } \mathcal{F} \subseteq \mathcal{H}_p : \Delta_{j+1}(\mathcal{F}) \leq Q_{j+1} \) 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 \( e(\mathcal{H}_p \setminus \mathcal{J}) < \varepsilon \mathbb{E} X / 2 \)

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

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

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

Relative degree events (\( Q_j < R_j \))

\[
\begin{align*}
\mathcal{D}_j & \triangleq \text{for all } \mathcal{F} \subseteq \mathcal{H}_p: \Delta_{j+1}(\mathcal{F}) \leq R_{j+1} \text{ implies } \Delta_j(\mathcal{F}) \leq R_j \\
\mathcal{D}_j^+ & \triangleq \text{for all } \mathcal{F} \subseteq \mathcal{H}_p: \Delta_{j+1}(\mathcal{F}) \leq Q_{j+1} \text{ implies } \Delta_j(\mathcal{F}) \leq Q_j
\end{align*}
\]

Sparsification event (by deleting edges)

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

First good event revisited

\[
\mathcal{G} \triangleq \text{for all } \mathcal{F} \subseteq \mathcal{H}_p: \Delta_1(\mathcal{F}) \leq Q_1 \text{ implies } e(\mathcal{F}) \leq (1 + \varepsilon/2)\mathbb{E}X
\]
**New concentration inequality (simplified)**

\((\xi_i)_{i \in A}\): independent random variables

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

<table>
<thead>
<tr>
<th>Well-behaved variant of the sum (X := \sum_{\alpha \in \mathcal{I}} Y_\alpha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restriction to <em>subsum</em> where each (Y_\beta) depends on (\leq C) variables</td>
</tr>
</tbody>
</table>

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

- \(\alpha \cap \beta = \emptyset\) implies that \(Y_\alpha\) and \(Y_\beta\) are independent

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

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

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