Central moment inequalities using Stein's method

Nathan Ross (University of Melbourne)

Joint Work with: Andrew Barbour (Zürich)
Yuting Wen

Introduction

Concentration inequalities bound the deviation of a variable from its mean:

$$\mathbb{P}\big(|W - \mathbb{E}W| \geq t\big) \leq \cdots$$

Useful in many applications: statistics (p-values), theoretical computer science, discrete math, statistical physics,... (LLNs, union bounds)

Introduction

Cramér-Chernoff method: For $\theta > 0$,

$$\mathbb{P}\big(W - \mathbb{E}W \geq t\big) \leq e^{-\theta t}\,\mathbb{E}e^{\theta(W-\mathbb{E}W)},$$

bound the mgf on the RHS and optimize over θ.

Main approach to bounding the mgf in applications is

1. Martingale difference (Hoeffding,...,McDiarmid,...)
   - Application: sums of independent random variables
2. Entropy method (...,Boucheron-Lugosi-Massart,...)
   - Application: "low influence" functions of independent rvs
3. **Stein's method** (Chatterjee, Goldstein,...)
   - Application: variables amenable to Stein's method couplings

Introduction

Stein's method approach to concentration relies on couplings:

- local dependence,
- exchangeable pairs,
- size biasing,
- other ad-hoc.

Requires extra assumptions, such as boundedness.

Introduction

Related approach, when mgf is unavailable: for any $q > 0$,

$$\mathbb{P}\big(|W - \mathbb{E}W| \geq t\big) \leq t^{-q}\mathbb{E}\left[|W - \mathbb{E}W|^q\right],$$

and bound the central moment on the RHS.

Main approach to bounding the central moment in applications is

1. Martingale difference (Rosenthal,...)
2. Entropy method (...,Boucheron-Lugosi-Massart,...)
3. **Exchangeable pairs** (Chatterjee,...)

Note: Can still get exponential tail bounds if good central moment bounds for large q.

This talk

Use **Stein couplings** to bound the central moments.

Main strengths are

- weaker assumptions on the couplings than for bounding mgf,
- applications with dependence that aren't amenable to other approaches.

Outline of the talk:

1. Stein couplings,
2. General result,
3. Applications:
   - local statistics of sparse Erdős-Rényi graphs,
   - neighbourhood statistics of germ grain models.

Stein Couplings. Chen and Röllin (2010)

The ordered collection of random variables (W, W', G) with $\mathbb{E}W =: \mu$ are said to form a **Stein coupling** if

$$\mathbb{E}\big[G\big(f(W') - f(W)\big)\big] = \mathbb{E}\big[(W-\mu)f(W)\big],$$

for all functions f such that the expectations exist.

Examples

Stein Coupling:

$$\mathbb{E}\big[G\big(f(W') - f(W)\big)\big] = \mathbb{E}\big[(W-\mu)f(W)\big]$$

Exchangeable pairs:

(W, W') is exchangeable, and for some $a > 0$,

$$\mathbb{E}\big[(W' - W)|W\big] = -a(W-\mu).$$

Set $G = (W' - W)/(2a)$.

Examples

Stein Coupling:

$$\mathbb{E}\big[G\big(f(W') - f(W)\big)\big] = \mathbb{E}\big[(W-\mu)f(W)\big]$$

Size biasing:

W' has the size-bias distribution of $W \geq 0$:

$$\mathbb{E}f(W') = \frac{\mathbb{E}[Wf(W)]}{\mu}.$$

Set $G = \mu$.

Examples

Stein Coupling:

$$\mathbb{E}\big[G\big(f(W') - f(W)\big)\big] = \mathbb{E}\big[(W-\mu)f(W)\big]$$

Local dependence:

- $W = \sum_{i=1}^{n} X_i$, with $\mathbb{E}X_i = 0$.
- For $1 \le i \le n$, there are $\mathcal{N}_i \subseteq \{1, \dots, n\}$, such that X_i is independent of $W_i := W - \sum_{j \in \mathcal{N}_i} X_j$.

Set $W' := W_I$, where I is a uniform index from $\{1, \dots, n\}$.

Set $G = -nX_I$.

Stein Couplings

These **Stein couplings** are widely used for both distributional approximation and concentration inequalities.

It is known how to construct them for many applications.

There are other ad-hoc constructions in the literature.

General result

NOTATION: $\|X\|_q = \left(\mathbb{E}[X^q]\right)^{1/q}$

Theorem: Let (W, W', G) be a **Stein coupling** and suppose that

$$\boxed{\|G\|_{2k} \leq A, \quad \text{and} \quad \|W' - W\|_{2k} \leq B.}$$

Then

$$\|W - \mu\|_{2k} \leq \sqrt{(2k-1)AB}\, \exp\left\{\sqrt{\frac{B(2k-1)}{A}}\right\}.$$

General Result

From previous slide:

$$\|W - \mu\|_{2k} \leq \sqrt{(2k-1)AB}\exp\left\{\sqrt{\frac{B(2k-1)}{A}}\right\}.$$

In many applications, $\sigma^2 := \mathrm{Var}(W) \to \infty$ and

- A is of order σ^2,
- B is of constant order.

Then if $k \ll \sigma^2$, we have

$$\|(W-\mu)/\sigma\|_{2k} \leq c\sqrt{2k-1}.$$

Order in k is tight for sums of random variables, or the normal distribution.

Note: Need k of order $\log(\sigma)$ for exponential tail bounds.

Example: Local Dependence

Local dependence:

- $W = \sum_{i=1}^{n} X_i$, with $\mathbb{E} X_i = 0$.
- For $1 \leq i \leq n$, there are $\mathcal{N}_i \subseteq \{1, \ldots, n\}$, such that X_i is <u>independent</u> of $W_i := W - \sum_{j \in \mathcal{N}_i} X_j$.

Set $W' := W_I$, where I is a uniform index from $\{1, \ldots, n\}$.

Set $G = -nX_I$.

Example: Local Dependence

Here $W' - W = -\sum_{j \in \mathcal{N}_I} X_j$ and $G = -nX_I$.

Recall assumption from General Result:

$$\boxed{\|G\|_{2k} \le A, \quad \text{and} \quad \|W' - W\|_{2k} \le B.}$$

Thus we can apply General Result with

$$A = n\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left|X_i\right|^{2k}\right)^{1/(2k)} \qquad B = \left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\Big|\sum_{j \in \mathcal{N}_i} X_j\Big|^{2k}\right)^{1/(2k)}.$$

In given application may be able to compute directly.

Example: Local Dependence

In general, we can set

$$d := \max_i |\mathcal{N}_i|, \quad m := \max_i ||X_i||_{2k},$$

and then take

$$A = nm, \quad B = dm,$$

implying

$$\|W\|_{2k} \leq \sqrt{n} m \sqrt{d(2k-1)} \exp\left\{\sqrt{\frac{d(2k-1)}{n}}\right\}.$$

Independent case: $d = 1$ and for $k \ll n$, RHS is of order

$$\sqrt{n} m \sqrt{2k-1}.$$

Application: n-hood statistics of sparse Erdős-Rényi

Setup:

- $\mathcal{G}_n$ is an Erdős-Rényi random graph on n vertices with edge probability λ/n,
- $N_r(i, \mathcal{G}_n)$ is the r-neighborhood of vertex i in $\mathcal{G}_n$,
- U is a function on vertex-labeled graphs of at most n vertices.

We look at

$$W := \sum_{i=1}^{n} U\big(N_r(i, \mathcal{G}_n)\big).$$

Examples of U:

- an indicator in terms of the degree of vertex i ($r = 1$),
- the number of triangles containing vertex i ($r = 2$).

Result

Recall

$$W := \sum_{i=1}^{n} U\big(N_r(i, \mathcal{G}_n)\big).$$

<u>Assume</u> U satisfies for some $c, \beta \geq 0$:

$$|U(G)| \leq c\,\big(\text{number of vertices of } G\big)^{\beta}.$$

Examples of U:

- an <u>indicator</u> in terms of the degree of vertex i: $\underline{c = 1, \beta = 0}$,
- the number of triangles containing vertex i: $\underline{c = 1/2, \beta = 2}$.

Then there is an explicit constant $C(\beta, r)$ such that

$$n^{-1/2}\,||W - \mathbb{E}W||_q \leq c\,C(\beta, r)\,\max\{\lambda, q(1+\beta)\}^{(1+2\beta)r+1/2}.$$

Application: n-hood statistics of germ grain models
Joint work with Peter Braunsteins (U Melb)

Setup:

- $\mathcal{V}_n$ is collection of $\underline{n \text{ points}}$ uniformly distributed on $[0, n^{1/2})^2$,
- $B_r(i, \mathcal{V}_n)$ is the radius-r n-hood around point i in $\mathcal{V}_n$,
- U is a function on point-labeled configurations.

We look at

$$W := \sum_{i=1}^{n} U\big(B_r(i, \mathcal{V}_n)\big).$$

Example of U:

- an <u>indicator</u> in terms of the number of points of $B_r(i, \mathcal{V}_n)$.

Result

Recall

$$W := \sum_{i=1}^{n} U\big(B_r(i, \mathcal{V}_n)\big).$$

Assume U satisfies for some $c, \beta \geq 0$:

$$|U(V)| \leq c\,\big(\text{number of points in } V\big)^{\beta}.$$

Example of U:

- an indicator in terms of the number of points of $B_r(i, \mathcal{V}_n)$: $c = 1, \beta = 0$.

Then there is an explicit constant $C(\beta)$ such that

$$n^{-1/2}\,||W - \mathbb{E}W||_q \leq c\,C(\beta)[\max\{9\pi r^2, q(1+\beta)\} + 1]^{\beta+3/2}.$$

Proof ideas

General Result: If $f(w) = (w - \mu)^{2k-1}$, then

$$\mathbb{E}[(W-\mu)^{2k}] = \mathbb{E}\big[(W-\mu)f(W)\big] = \mathbb{E}\big[G\big(f(W') - f(W)\big)\big].$$

Now use Binomial Theorem and manipulate...

Application: Use General Result.

- **Stein couplings** are similar to that for local dependence, but
- "dependency neighborhoods" are of random size, depending on the n-hoods of a randomly chosen vertex.

Extensions in the paper:

- allow for approximate **Stein coupling** (remainder term),
- sharper results assuming more about the couplings,
- detailed discussion using moment bounds for concentration.

Reference:

Central moment inequalities using Stein's method.

A.D. Barbour, N. Ross, Y. Wen

https://arxiv.org/abs/1802.10225

Thank You!