Stein Couplings for Concentration of Measure

Jay Bartroff, Subhankar Ghosh, Larry Goldstein and Ümit Işlak
University of Southern California
[arXiv:0906.3886] [arXiv:1304.5001] [arXiv:1402.6769]

Borchard Symposium, June/July 2014


Stein Mart
GRAND OPENING

Concentration of Measure

Distributional tail bounds can be provided in cases where exact computation is intractable.

Concentration of measure results can provide exponentially decaying bounds with explicit constants.

Concentration of Measure

Distributional tail bounds can be provided in cases where exact computation is intractable.

Concentration of measure results can provide exponentially decaying bounds with explicit constants.

Bounded Difference Inequality

If $Y = f(X_1, \ldots, X_n)$ with $X_1, \ldots, X_n$ independent, and for every $i = 1, \ldots, n$ the differences of the function $f : \Omega^n \to \mathbb{R}$

$$\sup_{x_i, x_i'} |f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)|$$

are bounded by c_i, then

$$\mathbb{P}(|Y - \mathbb{E}[Y]| \geq t) \leq 2 \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right).$$

Self Bounding Functions

The function $f(\mathbf{x}), \mathbf{x} = (x_1, \ldots, x_n)$ is (a, b) self bounding if there exist functions $f_i(\mathbf{x}^i), \mathbf{x}^i = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ such that

$$\sum_{i=1}^{n} (f(\mathbf{x}) - f_i(\mathbf{x}^i)) \leq af(\mathbf{x}) + b$$

and

$$0 \leq f(\mathbf{x}) - f_i(\mathbf{x}^i) \leq 1 \quad \text{for all } \mathbf{x}.$$

Self Bounding Functions

For say, the upper tail, with $c = (3a-1)/6$, $Y = f(X_1, \ldots, X_n)$, with $X_1, \ldots, X_n$ independent, for all $t \geq 0$,

$$\mathbb{P}(Y - \mathbb{E}[Y] \geq t) \leq \exp\left(-\frac{t^2}{2(a\mathbb{E}[Y] + b + c_+ t)}\right).$$

Mean in the denominator can be very competitive with the factor $\sum_{i=1}^n c_i^2$ in the bounded difference inequality.

If $(a, b) = (1, 0)$, say, the denominator of the exponent is $2(\mathbb{E}[Y] + t/3)$, and as $t \to \infty$ rate is $\exp(-3t/2)$.

Self Bounding Functions

For say, the upper tail, with $c = (3a-1)/6$, $Y = f(X_1, \ldots, X_n)$, with $X_1, \ldots, X_n$ independent, for all $t \geq 0$,

$$\mathbb{P}(Y - \mathbb{E}[Y] \geq t) \leq \exp\left(-\frac{t^2}{2(a\mathbb{E}[Y] + b + c_+ t)}\right).$$

Mean in the denominator can be very competitive with the factor $\sum_{i=1}^n c_i^2$ in the bounded difference inequality.

If $(a, b) = (1, 0)$, say, the denominator of the exponent is $2(\mathbb{E}[Y] + t/3)$, and as $t \to \infty$ rate is $\exp(-3t/2)$.

Use of Stein's Method Couplings

- Stein's method developed for evaluating the quality of distributional approximations through the use of characterizing equations.
- Implementation of the method often involves coupling constructions, with the quality of the resulting bounds reflecting the closeness of the coupling.
- Such couplings can be thought of as a type of distributional perturbation that measures dependence.
- Concentration of measure should hold when perturbation is small.

Use of Stein's Method Couplings

- Stein's method developed for evaluating the quality of distributional approximations through the use of characterizing equations.
- Implementation of the method often involves coupling constructions, with the quality of the resulting bounds reflecting the closeness of the coupling.
- Such couplings can be thought of as a type of distributional perturbation that measures dependence.
- Concentration of measure should hold when perturbation is small.

Use of Stein's Method Couplings

- Stein's method developed for evaluating the quality of distributional approximations through the use of characterizing equations.
- Implementation of the method often involves coupling constructions, with the quality of the resulting bounds reflecting the closeness of the coupling.
- Such couplings can be thought of as a type of distributional perturbation that measures dependence.
- Concentration of measure should hold when perturbation is small.

Use of Stein's Method Couplings

- Stein's method developed for evaluating the quality of distributional approximations through the use of characterizing equations.
- Implementation of the method often involves coupling constructions, with the quality of the resulting bounds reflecting the closeness of the coupling.
- Such couplings can be thought of as a type of distributional perturbation that measures dependence.
- Concentration of measure should hold when perturbation is small.

Stein's Method and Concentration Inequalities

- Raič (2007) applies the Stein equation to obtain Cramér type moderate deviations relative to the normal for some graph related statistics.
- Chatterjee (2007) derives tail bounds for Hoeffding's combinatorial CLT and the net magnetization in the Curie-Weiss model from statistical physics based on Stein's exchangeable pair coupling.
- Goldstein and Ghosh (2011) show bounded size bias coupling implies concentration.
- Chen and Röellin (2010) consider general 'Stein couplings' of which the exchangeable pair and size bias (but not zero bias) are special cases; $E[Gf(W') - Gf(W)] = E[Wf(W)]$.
- Paulin, Mackey and Tropp (2012,2013) extend exchangeable pair method to random matrices.

Stein's Method and Concentration Inequalities

- Raič (2007) applies the Stein equation to obtain Cramér type moderate deviations relative to the normal for some graph related statistics.
- Chatterjee (2007) derives tail bounds for Hoeffding's combinatorial CLT and the net magnetization in the Curie-Weiss model from statistical physics based on Stein's exchangeable pair coupling.
- Goldstein and Ghosh (2011) show bounded size bias coupling implies concentration.
- Chen and Röellin (2010) consider general 'Stein couplings' of which the exchangeable pair and size bias (but not zero bias) are special cases; $E[Gf(W') - Gf(W)] = E[Wf(W)]$.
- Paulin, Mackey and Tropp (2012,2013) extend exchangeable pair method to random matrices.

Stein's Method and Concentration Inequalities

- Raič (2007) applies the Stein equation to obtain Cramér type moderate deviations relative to the normal for some graph related statistics.
- Chatterjee (2007) derives tail bounds for Hoeffding's combinatorial CLT and the net magnetization in the Curie-Weiss model from statistical physics based on Stein's exchangeable pair coupling.
- Goldstein and Ghosh (2011) show bounded size bias coupling implies concentration.
- Chen and Röellin (2010) consider general 'Stein couplings' of which the exchangeable pair and size bias (but not zero bias) are special cases; $E[Gf(W') - Gf(W)] = E[Wf(W)]$.
- Paulin, Mackey and Tropp (2012,2013) extend exchangeable pair method to random matrices.

Stein's Method and Concentration Inequalities

- Raič (2007) applies the Stein equation to obtain Cramér type moderate deviations relative to the normal for some graph related statistics.
- Chatterjee (2007) derives tail bounds for Hoeffding's combinatorial CLT and the net magnetization in the Curie-Weiss model from statistical physics based on Stein's exchangeable pair coupling.
- Goldstein and Ghosh (2011) show bounded size bias coupling implies concentration.
- Chen and Röellin (2010) consider general 'Stein couplings' of which the exchangeable pair and size bias (but not zero bias) are special cases; $E[Gf(W') - Gf(W)] = E[Wf(W)]$.
- Paulin, Mackey and Tropp (2012,2013) extend exchangeable pair method to random matrices.

Stein's Method and Concentration Inequalities

- Raič (2007) applies the Stein equation to obtain Cramér type moderate deviations relative to the normal for some graph related statistics.
- Chatterjee (2007) derives tail bounds for Hoeffding's combinatorial CLT and the net magnetization in the Curie-Weiss model from statistical physics based on Stein's exchangeable pair coupling.
- Goldstein and Ghosh (2011) show bounded size bias coupling implies concentration.
- Chen and Röellin (2010) consider general 'Stein couplings' of which the exchangeable pair and size bias (but not zero bias) are special cases; $E[Gf(W') - Gf(W)] = E[Wf(W)]$.
- Paulin, Mackey and Tropp (2012,2013) extend exchangeable pair method to random matrices.

Exchangeable Pair Couplings

Let $(\mathbf{X}, \mathbf{X}')$ be exchangeable,

$$F(X, X') = -F(X', X) \quad \text{and} \quad \mathbb{E}[F(X, X')|X] = f(X)$$

with

$$\frac{1}{2}\mathbb{E}[|(f(X) - f(X'))F(X, X')||X] \leq c.$$

Then $Y = f(\mathbf{X})$ satisfies

$$\mathbb{P}(|Y| \geq t) \leq 2\exp\left(-\frac{t^2}{2c}\right).$$

No independence assumption.

Exchangeable Pair Couplings

Let $(\mathbf{X}, \mathbf{X}')$ be exchangeable,

$$F(X, X') = -F(X', X) \quad \text{and} \quad \mathbb{E}[F(X, X')|X] = f(X)$$

with

$$\frac{1}{2}\mathbb{E}[|(f(X) - f(X'))F(X, X')||X] \leq c.$$

Then $Y = f(\mathbf{X})$ satisfies

$$\mathbb{P}(|Y| \geq t) \leq 2\exp\left(-\frac{t^2}{2c}\right).$$

No independence assumption.

Curie Weiss Model

Consider a graph on n vertices V with symmetric neighborhoods N_j, and Hamiltonian

$$H_h(\sigma) = -\frac{1}{2n} \sum_{j \in V} \sum_{k \in N_j} \sigma_j \sigma_k - h \sum_{i \in V} \sigma_j,$$

and the measure on 'spins' $\sigma = (\sigma_i)_{i \in V}, \sigma_i \in \{-1, 1\}$

$$p_{\beta,h}(\sigma) = Z_{\beta,h}^{-1} e^{-\beta H_h(\sigma)}.$$

Interested in the average net magentization

$$m = \frac{1}{n} \sum_{i \in V} \sigma_i.$$

Curie Weiss Model

Consider a graph on n vertices V with symmetric neighborhoods N_j, and Hamiltonian

$$H_h(\sigma) = -\frac{1}{2n}\sum_{j\in V}\sum_{k\in N_j}\sigma_j\sigma_k - h\sum_{i\in V}\sigma_j,$$

and the measure on 'spins' $\sigma = (\sigma_i)_{i\in V}, \sigma_i \in \{-1, 1\}$

$$p_{\beta,h}(\sigma) = Z_{\beta,h}^{-1} e^{-\beta H_h(\sigma)}.$$

Interested in the average net magentization

$$m = \frac{1}{n}\sum_{i\in V}\sigma_i.$$

Consider the complete graph.

Curie Weiss Concentration

Choose vertex V uniformly and sample σ'_V from the conditional distribution of σ_V given $\sigma_j, j \notin N_V$. Yields an exchangeable pair allowing the result above to imply, taking $h = 0$ for simplicity,

$$\mathbb{P}\left(|m - \tanh(\beta m)| \geq \frac{\beta}{n} + t\right) \leq 2e^{-nt^2/(4+4\beta)}.$$

The magnetization m is concentrated about the roots of the equation

$$x = \tanh(\beta x).$$

Curie Weiss Concentration

Choose vertex V uniformly and sample σ'_V from the conditional distribution of σ_V given $\sigma_j, j \notin N_V$. Yields an exchangeable pair allowing the result above to imply, taking $h = 0$ for simplicity,

$$\mathbb{P}\left(|m - \tanh(\beta m)| \geq \frac{\beta}{n} + t\right) \leq 2e^{-nt^2/(4+4\beta)}.$$

The magnetization m is concentrated about the roots of the equation

$$x = \tanh(\beta x).$$

Size Bias Couplings

For a nonnegative random variable Y with finite nonzero mean μ, we say that Y^s has the Y-size bias distribution if

$$\mathbb{E}[Yg(Y)] = \mu \mathbb{E}[g(Y^s)] \quad \text{for all } g.$$

- Size biasing may appear, undesirably, in sampling.
- For sums of independent variables, size biasing a single summand size biases the sum.
- The closeness of a coupling of a sum Y to Y^s is a type of perturbation that measures the dependence in the summands of Y.
- If X is a non trivial indicator variable then $X^s = 1$.

Size Bias Couplings

For a nonnegative random variable Y with finite nonzero mean μ, we say that Y^s has the Y-size bias distribution if

$$\mathbb{E}[Yg(Y)] = \mu \mathbb{E}[g(Y^s)] \quad \text{for all } g.$$

- Size biasing may appear, undesirably, in sampling.
- For sums of independent variables, size biasing a single summand size biases the sum.
- The closeness of a coupling of a sum Y to Y^s is a type of perturbation that measures the dependence in the summands of Y.
- If X is a non trivial indicator variable then $X^s = 1$.

Size Bias Couplings

For a nonnegative random variable Y with finite nonzero mean μ, we say that Y^s has the Y-size bias distribution if

$$\mathbb{E}[Yg(Y)] = \mu\mathbb{E}[g(Y^s)] \quad \text{for all } g.$$

- Size biasing may appear, undesirably, in sampling.
- For sums of independent variables, size biasing a single summand size biases the sum.
- The closeness of a coupling of a sum Y to Y^s is a type of perturbation that measures the dependence in the summands of Y.
- If X is a non trivial indicator variable then $X^s = 1$.

Size Bias Couplings

For a nonnegative random variable Y with finite nonzero mean μ, we say that Y^s has the Y-size bias distribution if

$$\mathbb{E}[Yg(Y)] = \mu\mathbb{E}[g(Y^s)] \quad \text{for all } g.$$

- Size biasing may appear, undesirably, in sampling.
- For sums of independent variables, size biasing a single summand size biases the sum.
- The closeness of a coupling of a sum Y to Y^s is a type of perturbation that measures the dependence in the summands of Y.
- If X is a non trivial indicator variable then $X^s = 1$.

Bounded Size Bias Coupling implies Concentration

Let Y be a nonnegative random variable with finite positive mean μ. Suppose there exists a coupling of Y to a variable Y^s having the Y-size bias distribution that satisfies $Y^s \leq Y + c$ for some $c > 0$ with probability one. Then,

$$\max\left(\mathbf{1}_{t\geq 0}\mathbb{P}(Y-\mu \geq t), \mathbf{1}_{-\mu\leq t\leq 0}\mathbb{P}(Y-\mu \leq t)\right) \leq b(t;\mu,c)$$

where

$$b(t;\mu,c) = \left(\frac{\mu}{\mu+t}\right)^{(t+\mu)/c} e^{t/c}.$$

Ghosh and Goldstein (2011), Improvement by Arratia and Baxendale (2013)

Poisson behavior, rate $\exp(-t\log t)$ as $t\to\infty$.

Bounded Size Bias Coupling implies Concentration

Let Y be a nonnegative random variable with finite positive mean μ. Suppose there exists a coupling of Y to a variable Y^s having the Y-size bias distribution that satisfies $Y^s \le Y + c$ for some $c > 0$ with probability one. Then,

$$\max\left(\mathbf{1}_{t\ge 0}\mathbb{P}(Y-\mu \ge t), \mathbf{1}_{-\mu \le t \le 0}\mathbb{P}(Y-\mu \le t)\right) \le b(t;\mu,c)$$

where

$$b(t;\mu,c) = \left(\frac{\mu}{\mu+t}\right)^{(t+\mu)/c} e^{t/c}.$$

Ghosh and Goldstein (2011), Improvement by Arratia and Baxendale (2013)

Poisson behavior, rate $\exp(-t\log t)$ as $t \to \infty$.

Bounded Coupling Concentration Inequality

For the right tail, say, using that for $x \geq 0$ the function $h(x) = (1+x)\log(1+x) - x$ obeys the bound

$$h(x) \geq \frac{x^2}{2(1+x/3)},$$

we have

$$\mathbb{P}(Y - \mu \geq t) \leq \exp\left(-\frac{t^2}{2c(\mu + t/3)}\right).$$

Proof of Upper Tail Bound

For $\theta \geq 0$,

$$e^{\theta Y^s} = e^{\theta(Y + Y^s - Y)} \leq e^{c\theta} e^{\theta Y}. \tag{1}$$

With $m_{Y^s}(\theta) = \mathbb{E}e^{\theta Y^s}$, and similarly for $m_Y(\theta)$,

$$\mu m_{Y^s}(\theta) = \mu \mathbb{E} e^{\theta Y^s} = \mathbb{E}[Y e^{\theta Y}] = m_Y'(\theta)$$

so multiplying by μ in (1) and taking expectation yields

$$m_Y'(\theta) \leq \mu e^{c\theta} m_Y(\theta).$$

Integration yields

$$m_Y(\theta) \leq \exp\left(\frac{\mu}{c}\left(e^{c\theta} - 1\right)\right)$$

and the bound is obtained upon choosing $\theta = \log(t/\mu)/c$ in

$$\mathbb{P}(Y \geq t) = \mathbb{P}(e^{-\theta t} e^{\theta Y} \geq 1) \leq e^{-\theta t + \frac{\mu}{c}\left(e^{c\theta} - 1\right)}.$$

Size Biasing Sum of Exchangeable Indicators

Suppose X is a sum of nontrivial exchangeable indicator variables $X_1, \ldots, X_n$, and that for $i \in \{1, \ldots, n\}$ the variables $X_1^i, \ldots, X_n^i$ have joint distribution

$$\mathcal{L}(X_1^i, \ldots, X_n^i) = \mathcal{L}(X_1, \ldots, X_n | X_i = 1).$$

Then

$$X^i = \sum_{j=1}^{n} X_j^i$$

has the X-size bias distribution X^s, as does the mixture X^I when I is a random index with values in $\{1, \ldots, n\}$, independent of all other variables.

In more generality, pick index I with probability $P(I = i)$ proportional to EX_i.

Size Biasing Sum of Exchangeable Indicators

Suppose X is a sum of nontrivial exchangeable indicator variables $X_1, \ldots, X_n$, and that for $i \in \{1, \ldots, n\}$ the variables $X_1^i, \ldots, X_n^i$ have joint distribution

$$\mathcal{L}(X_1^i, \ldots, X_n^i) = \mathcal{L}(X_1, \ldots, X_n | X_i = 1).$$

Then

$$X^i = \sum_{j=1}^{n} X_j^i$$

has the X-size bias distribution X^s, as does the mixture X^I when I is a random index with values in $\{1, \ldots, n\}$, independent of all other variables.

In more generality, pick index I with probability $P(I = i)$ proportional to EX_i.

Applications

1. The number of local maxima of a random function on a graph
2. The number of vertices in an Erdős-Rényi graph exceeding pre-set thresholds
3. The d-way covered volume of a collection of m balls placed uniformly over a volume m subset of $\mathbb{R}^p$

Applications

1. The number of local maxima of a random function on a graph
2. The number of vertices in an Erdős-Rényi graph exceeding pre-set thresholds
3. The d-way covered volume of a collection of m balls placed uniformly over a volume m subset of $\mathbb{R}^p$

Applications

1. The number of local maxima of a random function on a graph
2. The number of vertices in an Erdős-Rényi graph exceeding pre-set thresholds
3. The d-way covered volume of a collection of m balls placed uniformly over a volume m subset of $\mathbb{R}^p$

Local Maxima on Graphs

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a given graph, and for every $v \in \mathcal{V}$ let $\mathcal{V}_v \subset \mathcal{V}$ be the neighbors of v, with $v \in \mathcal{V}$. Let $\{C_g, g \in \mathcal{V}\}$ be a collection of independent and identically distributed continuous random variables, and let X_v be the indicator that vertex v corresponds to a local maximum value with respect to the neighborhood $\mathcal{V}_v$, that is

$$X_v(C_w, w \in \mathcal{V}_v) = \prod_{w \in \mathcal{V}_v \setminus \{v\}} 1(C_v > C_w), \quad v \in \mathcal{V}.$$

The sum

$$Y = \sum_{v \in \mathcal{V}} X_v$$

is the number of local maxima on $\mathcal{G}$.

Local Maxima on Graphs

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a given graph, and for every $v \in \mathcal{V}$ let $\mathcal{V}_v \subset \mathcal{V}$ be the neighbors of v, with $v \in \mathcal{V}$. Let $\{C_g, g \in \mathcal{V}\}$ be a collection of independent and identically distributed continuous random variables, and let X_v be the indicator that vertex v corresponds to a local maximum value with respect to the neighborhood $\mathcal{V}_v$, that is

$$X_v(C_w, w \in \mathcal{V}_v) = \prod_{w \in \mathcal{V}_v \setminus \{v\}} 1(C_v > C_w), \quad v \in \mathcal{V}.$$

The sum

$$Y = \sum_{v \in \mathcal{V}} X_v$$

is the number of local maxima on $\mathcal{G}$.

Size Biasing $\{X_v, v \in \mathcal{V}\}$

If $X_v = 1$, that is, if v is already a local maxima, let $\mathbf{X}^v = \mathbf{X}$. Otherwise, interchange the value C_v at v with the value C_w at the vertex w that achieves the maximum C_u for $u \in \mathcal{V}_v$, and let $\mathbf{X}^v$ be the indicators of local maxima on this new configuration. Then Y^s, the number of local maxima on $\mathbf{X}^I$, where I is chosen proportional to $\mathbb{E}X_v$, has the Y-size bias distribution.

We have

$$Y^s \le Y + c \quad \text{where} \quad c = \max_{v \in \mathcal{V}} \max_{w \in \mathcal{V}_v} |\mathcal{V}_w|.$$

Size Biasing $\{X_v, v \in \mathcal{V}\}$

If $X_v = 1$, that is, if v is already a local maxima, let $\mathbf{X}^v = \mathbf{X}$. Otherwise, interchange the value C_v at v with the value C_w at the vertex w that achieves the maximum C_u for $u \in \mathcal{V}_v$, and let $\mathbf{X}^v$ be the indicators of local maxima on this new configuration. Then Y^s, the number of local maxima on $\mathbf{X}^I$, where I is chosen proportional to $\mathbb{E}X_v$, has the Y-size bias distribution.

We have

$$Y^s \leq Y + c \quad \text{where} \quad c = \max_{v \in \mathcal{V}} \max_{w \in \mathcal{V}_v} |\mathcal{V}_w|.$$

Self Bounding and Configuration Functions

The collection of sets $\Pi_k \subset \Omega^k, k = 0, \ldots, n$ is hereditary if $(x_1, \ldots, x_k) \in \Pi_k$ implies $(x_{i_1}, \ldots, x_{i_j}) \in \Pi_j$ for any $1 \le i_1 < \ldots < i_{i_j} \le k$. Let $f : \Omega^n \to \mathbb{R}$ be the function that assigns to $\mathbf{x} \in \Omega^n$ the size k of the largest subsequence of $\mathbf{x}$ that lies in Π_k. With $f_i(\mathbf{x})$ the function f evaluated on $\mathbf{x}$ after removing its i^{th} coordinate, we have

$$0 \le f(\mathbf{x}) - f_i(\mathbf{x}) \le 1 \quad \text{and} \quad \sum_{i=1}^{n} (f(\mathbf{x}) - f_i(\mathbf{x})) \le f(\mathbf{x})$$

as removing a single coordinate from x reduces f by at most one, and there at most $f = k$ 'important' coordinates. Hence, configuration functions are $(a, b) = (1, 0)$ self bounding.

Self Bounding Functions

The number of local maxima is a configuration function, with $(x_{i_1}, \ldots, x_{i_j}) \in \Pi_j$ when the vertices indexed by $i_1, \ldots, i_j$ are local maxima; hence the number of local maxima Y is a self bounding function. Hence, Y satisfies the concentration bound

$$\mathbb{P}(Y - \mathbb{E}[Y] \geq t) \leq \exp\left(-\frac{t^2}{2(\mathbb{E}[Y] + t/3)}\right).$$

Size bias bound is of Poisson type with tail rate $\exp(-t \log t)$.

Self Bounding Functions

The number of local maxima is a configuration function, with $(x_{i_1}, \ldots, x_{i_j}) \in \Pi_j$ when the vertices indexed by $i_1, \ldots, i_j$ are local maxima; hence the number of local maxima Y is a self bounding function. Hence, Y satisfies the concentration bound

$$\mathbb{P}(Y - \mathbb{E}[Y] \geq t) \leq \exp\left(-\frac{t^2}{2(\mathbb{E}[Y] + t/3)}\right).$$

Size bias bound is of Poisson type with tail rate $\exp(-t \log t)$.

Multinomial Occupancy Models

Let M_α be the degree of vertex $\alpha \in [m]$ in an Erdős-Rényi random graph. Then

$$Y_{\text{ge}} = \sum_{\alpha \in [m]} \mathbf{1}(M_\alpha \geq d_\alpha)$$

obeys the concentration bound $b(t; \mu, c)$ with $c = \sup_{\alpha \in [m]} d_\alpha + 1$.

Unbounded couplings can more easily be constructed than bounded ones, for instance, by giving the chosen vertex α the number of edges from the conditional distribution given $M_\alpha \geq d_\alpha$. A coupling bounded by $\sup_{\alpha \in [m]} d_\alpha$ may be constructed by adding edges, or not, sequentially, to the chosen vertex, with probabilities depending on its degree. Degree distributions are log concave.

Multinomial Occupancy Models

Let M_α be the degree of vertex $\alpha \in [m]$ in an Erdős-Rényi random graph. Then

$$Y_{\text{ge}} = \sum_{\alpha \in [m]} \mathbf{1}(M_\alpha \geq d_\alpha)$$

obeys the concentration bound $b(t; \mu, c)$ with $c = \sup_{\alpha \in [m]} d_\alpha + 1$.

Unbounded couplings can more easily be constructed than bounded ones, for instance, by giving the chosen vertex α the number of edges from the conditional distribution given $M_\alpha \geq d_\alpha$. A coupling bounded by $\sup_{\alpha \in [m]} d_\alpha$ may be constructed by adding edges, or not, sequentially, to the chosen vertex, with probabilities depending on its degree. Degree distributions are log concave.

Multinomial Occupancy Models

Let M_α be the degree of vertex $\alpha \in [m]$ in an Erdős-Rényi random graph. Then

$$Y_{\text{ge}} = \sum_{\alpha \in [m]} \mathbf{1}(M_\alpha \geq d_\alpha)$$

obeys the concentration bound $b(t; \mu, c)$ with $c = \sup_{\alpha \in [m]} d_\alpha + 1$.

Unbounded couplings can more easily be constructed than bounded ones, for instance, by giving the chosen vertex α the number of edges from the conditional distribution given $M_\alpha \geq d_\alpha$. A coupling bounded by $\sup_{\alpha \in [m]} d_\alpha$ may be constructed by adding edges, or not, sequentially, to the chosen vertex, with probabilities depending on its degree. Degree distributions are log concave.

Multinomial Occupancy Models

Similar remarks apply to

$$Y_{\mathrm{ne}} = \sum_{\alpha \in [m]} \mathbf{1}(M_\alpha \neq d_\alpha).$$

For some models, not here but e.g. multinomial urn occupancy, the indicators of Y_{ge} are negatively associated, though not for Y_{ne}.

The d-way covered volume on m balls in $\mathbb{R}^p$

Let $X_1, \ldots, X_m$ be the uniform and independent over the torus $C_n = [0, n^{1/p})^p$, and unit balls $B_1, \ldots B_m$ placed at these centers. Then deviations of t or more from the mean by

$$V_k = \text{Vol}\left(\bigcup_{\substack{r \subset [m] \\ |r| \geq d}} \bigcap_{\alpha \in r} B_\alpha \right)$$

are bounded by $b(t; \mu, c)$ with $c = d\pi_p$.

Zero Bias Coupling

For the mean zero, variance σ^2 random variable, we say Y^* has the Y-zero bias distribution when

$$\mathbb{E}[Yf(Y)] = \sigma^2 \mathbb{E}[f'(Y^*)] \quad \text{for all smooth } f.$$

Restatement of Stein's lemma: Y is normal if and only if $Y^* =_d Y$.

If Y and Y^* can be coupled on the same space such that $|Y^* - Y| \le c$ a.s., then under a mild MGF assumption

$$\mathbb{P}(Y \ge t) \le \exp\left(-\frac{t^2}{2(\sigma^2 + ct)}\right),$$

and with $4\sigma^2 + Ct$ in the denominator under similar conditions.

Zero Bias Coupling

For the mean zero, variance σ^2 random variable, we say Y^* has the Y-zero bias distribution when

$$\mathbb{E}[Yf(Y)] = \sigma^2 \mathbb{E}[f'(Y^*)] \quad \text{for all smooth } f.$$

Restatement of Stein's lemma: Y is normal if and only if $Y^* =_d Y$.

If Y and Y^* can be coupled on the same space such that $|Y^* - Y| \le c$ a.s., then under a mild MGF assumption

$$\mathbb{P}(Y \ge t) \le \exp\left(-\frac{t^2}{2(\sigma^2 + ct)}\right),$$

and with $4\sigma^2 + Ct$ in the denominator under similar conditions.

Zero Bias Coupling

For the mean zero, variance σ^2 random variable, we say Y^* has the Y-zero bias distribution when

$$\mathbb{E}[Yf(Y)] = \sigma^2 \mathbb{E}[f'(Y^*)] \quad \text{for all smooth } f.$$

Restatement of Stein's lemma: Y is normal if and only if $Y^* =_d Y$.

If Y and Y^* can be coupled on the same space such that $|Y^* - Y| \le c$ a.s., then under a mild MGF assumption

$$\mathbb{P}(Y \ge t) \le \exp\left(-\frac{t^2}{2(\sigma^2 + ct)}\right),$$

and with $4\sigma^2 + Ct$ in the denominator under similar conditions.

Combinatorial CLT

Zero bias coupling can produce bounds for Hoeffdings statistic

$$Y = \sum_{i=1}^{n} a_{i\pi(i)}$$

when π is chosen uniformly over the symmetric group $\mathcal{S}_n$, and when its distribution is constant over cycle type.

Permutations π chosen uniformly from involutions, $\pi^2 = \text{id}$, without fixed points; arises in matched pairs experiments.

Combinatorial CLT

Zero bias coupling can produce bounds for Hoeffdings statistic

$$Y = \sum_{i=1}^{n} a_{i\pi(i)}$$

when π is chosen uniformly over the symmetric group $\mathcal{S}_n$, and when its distribution is constant over cycle type.

Permutations π chosen uniformly from involutions, $\pi^2 = \text{id}$, without fixed points; arises in matched pairs experiments.

Combinatorial CLT, Exchangeable Pair Coupling

Under the assumption that $0 \le a_{ij} \le 1$, using the exchangeable pair Chatterjee produces the bound

$$\mathbb{P}(|Y - \mu_A| \ge t) \le 2\exp\left(-\frac{t^2}{4\mu_A + 2t}\right),$$

while under this condition the zero bias bound gives

$$\mathbb{P}(|Y - \mu_A| \ge t) \le 2\exp\left(-\frac{t^2}{2\sigma_A^2 + 16t}\right),$$

which is smaller whenever $t \le (2\mu_A - \sigma_A^2)/7$, holding asymptotically everywhere if a_{ij} are i.i.d., say, as then $\mathbb{E}\sigma_A^2 < \mathbb{E}\mu_A$.

Matrix Concentration Inequalities

Application in high dimensional statistics, variable selection, matrix completion problem.

Matrix Concentration Inequalities

Paulin, Mackay and Tropp, Stein pair Kernel coupling. Take (Z, Z') exchangeable and $X \in \mathbb{H}^{d \times d}$ such that

$$X = \phi(Z) \quad \text{and} \quad X' = \phi(Z'),$$

and anti-symmetric function Kernel function K such that

$$E[K(Z, Z')|Z] = X.$$

With

$$V_X = \frac{1}{2}E[(X - X')^2|Z] \quad \text{and} \quad V^K = \frac{1}{2}E[K(Z, Z')^2|Z]$$

if there exist s, c, v such that

$$V_X \preceq s^{-1}(cX + vI) \quad \text{and} \quad V^K \preceq s(cX + vI),$$

then one has bounds, such as,

$$P(\lambda_{\max}(X) \geq t) \leq d \exp\left(\frac{-t^2}{2v + 2ct}\right)$$

Matrix Concentration by Size Bias

For X a non-negative random variable with finite mean, we say X^s has the X-size bias distribution when

$$E[Xf(X)] = E[X]E[f(X^s)]$$

Matrix Concentration by Size Bias

For X a positive definite random matrix with finite mean, we say X^s has the X-size bias distribution when

$$\operatorname{tr}(E[Xf(X)]) = \operatorname{tr}(E[X]E[f(X^s)]).$$

For a product $X = \gamma A$ with γ a non-negative, scalar random variable and A a fixed positive definite matrix, $X^s = \gamma^s A$.

Matrix Concentration by Size Bias

For X a positive definite random matrix with finite mean, we say X^s has the X-size bias distribution when

$$\operatorname{tr}\left(E[Xf(X)]\right) = \operatorname{tr}\left(E[X]E[f(X^s)]\right).$$

For a product $X = \gamma A$ with γ a non-negative, scalar random variable and A a fixed positive definite matrix, $X^s = \gamma^s A$.

Size Bias Matrix Concentration

If $X = \sum_{k=1}^n Y_k$, with $Y_1, \ldots, Y_n$ independent, then

$$\mathrm{tr}E[Xf(X)] = \sum_{k=1}^n \mathrm{tr}E[Y_k f(X)] = \sum_{k=1}^n \mathrm{tr}\left(E[Y_k]E[f(X^{(k)})]\right)$$

Size Bias Matrix Concentration

If $X = \sum_{k=1}^n Y_k$, with $Y_1, \ldots, Y_n$ independent, then

$$\mathrm{tr}E[Xf(X)] = \sum_{k=1}^n \mathrm{tr}E[Y_k f(X)] = \sum_{k=1}^n \mathrm{tr}\left(E[Y_k]E[f(X^{(k)})]\right)$$

May bound by

$$\sum_{k=1}^n \lambda_{\max}(E[Y_k])\mathrm{tr}E[f(X^{(k)})],$$

but doing so will produce a constant in the bound of value

$$\sum_{k=1}^n \lambda_{\max}(EY_k) \quad \text{rather than} \quad \lambda_{\max}(EX).$$

Summary

Concentration of measure results can provide exponential tail bounds on complicated distributions.

Most concentration of measure results require independence.

Size bias and zero bias couplings, or perturbations, measure departures from independence. Bounded couplings imply concentration of measure (and central limit behavior.)

Unbounded couplings can also be handled under special conditions – e.g., the number of isolated vertices in the Erdös-Rényi random graph (Ghosh, Goldstein and Raič).

Summary

Concentration of measure results can provide exponential tail bounds on complicated distributions.

Most concentration of measure results require independence.

Size bias and zero bias couplings, or perturbations, measure departures from independence. Bounded couplings imply concentration of measure (and central limit behavior.)

Unbounded couplings can also be handled under special conditions – e.g., the number of isolated vertices in the Erdös-Rényi random graph (Ghosh, Goldstein and Raič).

Summary

Concentration of measure results can provide exponential tail bounds on complicated distributions.

Most concentration of measure results require independence.

Size bias and zero bias couplings, or perturbations, measure departures from independence. Bounded couplings imply concentration of measure (and central limit behavior.)

Unbounded couplings can also be handled under special conditions – e.g., the number of isolated vertices in the Erdös-Rényi random graph (Ghosh, Goldstein and Raič).

Summary

Concentration of measure results can provide exponential tail bounds on complicated distributions.

Most concentration of measure results require independence.

Size bias and zero bias couplings, or perturbations, measure departures from independence. Bounded couplings imply concentration of measure (and central limit behavior.)

Unbounded couplings can also be handled under special conditions – e.g., the number of isolated vertices in the Erdös-Rényi random graph (Ghosh, Goldstein and Raič).