Concentration of Measures by Bounded Couplings

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

May 2013

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.

Azuma Hoeffding Inequality

If $Y_k, k = 0, 1 \ldots, n$ is a martingale satisfying $|Y_k - Y_{k-1}| \leq c_k$ for $k = 1, \ldots, n$ with constants $c_1, \ldots, c_n$, then

$$P\left(|Y_n - Y_0| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right)$$

Handles dependence, requires martingale, some boundedness.

Azuma Hoeffding Inequality

If $Y_k, k = 0, 1 \ldots, n$ is a martingale satisfying $|Y_k - Y_{k-1}| \leq c_k$ for $k = 1, \ldots, n$ with constants $c_1, \ldots, c_n$, then

$$P(|Y_n - Y_0| \geq t) \leq 2 \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right)$$

Handles dependence, requires martingale, some boundedness.

Longest Common Subsequence Problem

Let $L_{m,n}(X_1,\ldots,X_m,X_{m+1},\ldots,X_n)$ be the length of the longest common subsequence between two, say, i.i.d. sequences of length m and $n-m$ from some discrete alphabet.

Using $Y_k = E[L_{m,n}|X_1,\ldots,X_k]$ is a martingale satisfying $|Y_{k-1}-Y_k| \le 1$ one attains the two sided tail bound $2\exp\left(-t^2/2n\right)$.

Though the distribution of $L_{m,n}$ is intractable (even the constant $c = \lim_{m\to\infty} L_{m,m}/2m$ is famously unknown), much can be said about its tails.

Longest Common Subsequence Problem

Let $L_{m,n}(X_1,\ldots,X_m,X_{m+1},\ldots,X_n)$ be the length of the longest common subsequence between two, say, i.i.d. sequences of length m and $n-m$ from some discrete alphabet.

Using $Y_k = E[L_{m,n}|X_1,\ldots,X_k]$ is a martingale satisfying $|Y_{k-1}-Y_k| \leq 1$ one attains the two sided tail bound $2\exp\left(-t^2/2n\right)$.

Though the distribution of $L_{m,n}$ is intractable (even the constant $c = \lim_{m\to\infty} L_{m,m}/2m$ is famously unknown), much can be said about its tails.

Longest Common Subsequence Problem

Let $L_{m,n}(X_1, \ldots, X_m, X_{m+1}, \ldots, X_n)$ be the length of the longest common subsequence between two, say, i.i.d. sequences of length m and $n - m$ from some discrete alphabet.

Using $Y_k = E[L_{m,n}|X_1, \ldots, X_k]$ is a martingale satisfying $|Y_{k-1} - Y_k| \leq 1$ one attains the two sided tail bound $2\exp\left(-t^2/2n\right)$.

Though the distribution of $L_{m,n}$ is intractable (even the constant $c = \lim_{m\to\infty} L_{m,m}/2m$ is famously unknown), much can be said about its tails.

Talagrand Isoperimetric Inequality

Let $L(x_1, \ldots, x_n)$ be a real valued function for $x_i \in \mathbb{R}^d, i = 1, \ldots, n$ such that there exists weight functions $\alpha_i(x)$ such that

$$L(x_1, \ldots, x_n) \leq L(y_1, \ldots, y_n) + \sum_{i=1}^{n} \alpha_i(x) \mathbf{1}(x_i \neq y_i)$$

and $\sum_{i=1}^{n} \alpha_i(x)^2 \leq c$ for some constant c. Then for $X_1, \ldots, X_n$, i.i.d. $\mathcal{U}([0,1]^d)$,

$$P\left(|L(X_1, \ldots, X_n) - M_n| \geq t\right) \leq 4 \exp(-t^2/4c^2)$$

where M_n is the median of $L(X_1, \ldots, X_n)$.

Applications: Steiner Tree, Travelling Salesman Problem.

Need to construct weights $\alpha_i(x)$.

Talagrand Isoperimetric Inequality

Let $L(x_1, \ldots, x_n)$ be a real valued function for $x_i \in \mathbb{R}^d, i = 1, \ldots, n$ such that there exists weight functions $\alpha_i(x)$ such that

$$L(x_1, \ldots, x_n) \leq L(y_1, \ldots, y_n) + \sum_{i=1}^{n} \alpha_i(x) \mathbf{1}(x_i \neq y_i)$$

and $\sum_{i=1}^{n} \alpha_i(x)^2 \leq c$ for some constant c. Then for $X_1, \ldots, X_n$, i.i.d. $\mathcal{U}([0,1]^d)$,

$$P\left(|L(X_1, \ldots, X_n) - M_n| \geq t\right) \leq 4 \exp(-t^2/4c^2)$$

where M_n is the median of $L(X_1, \ldots, X_n)$.
Applications: Steiner Tree, Travelling Salesman Problem.
Need to construct weights $\alpha_i(x)$.

Talagrand Isoperimetric Inequality

Let $L(x_1, \ldots, x_n)$ be a real valued function for $x_i \in \mathbb{R}^d, i = 1, \ldots, n$ such that there exists weight functions $\alpha_i(x)$ such that

$$L(x_1, \ldots, x_n) \leq L(y_1, \ldots, y_n) + \sum_{i=1}^{n} \alpha_i(x) \mathbf{1}(x_i \neq y_i)$$

and $\sum_{i=1}^{n} \alpha_i(x)^2 \leq c$ for some constant c. Then for $X_1, \ldots, X_n$, i.i.d. $\mathcal{U}([0,1]^d)$,

$$P\left(|L(X_1, \ldots, X_n) - M_n| \geq t\right) \leq 4 \exp(-t^2/4c^2)$$

where M_n is the median of $L(X_1, \ldots, X_n)$.
Applications: Steiner Tree, Travelling Salesman Problem.
Need to construct weights $\alpha_i(x)$.

Use of Stein's Method Couplings

- Stein's method developed for distributional approximation (normal, Poisson) through use of characterizing equation.
- 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 results should hold under similar sets of favorable conditions.

Use of Stein's Method Couplings

- Stein's method developed for distributional approximation (normal, Poisson) through use of characterizing equation.
- 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 results should hold under similar sets of favorable conditions.

Use of Stein's Method Couplings

- Stein's method developed for distributional approximation (normal, Poisson) through use of characterizing equation.
- 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 results should hold under similar sets of favorable conditions.

Use of Stein's Method Couplings

- Stein's method developed for distributional approximation (normal, Poisson) through use of characterizing equation.
- 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 results should hold under similar sets of favorable conditions.

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.
- 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.
- 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.
- Paulin, Mackey and Tropp (2012,2013) Extend exchangeable pair method to random matrices.

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

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

- 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

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

- 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

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

- 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

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

- 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 Coupling implies Concentration Inequality

Let Y be a nonnegative random variable with mean and variance μ and σ^2 respectively, both finite and positive. Suppose there exists a coupling of Y to a variable Y^s having the Y-size bias distribution that satisfies $|Y^s - Y| \le C$ for some $C > 0$ with probability one. Let $A = C\mu/\sigma^2$ and $B = C/2\sigma$.

a) If $Y^s \ge Y$ with probability one, then

$$P\left(\frac{Y-\mu}{\sigma} \le -t\right) \le \exp\left(-\frac{t^2}{2A}\right) \quad \text{for all } t > 0.$$

b) If the moment generating function of Y is finite at $2/C$, then

$$P\left(\frac{Y-\mu}{\sigma} \ge t\right) \le \exp\left(-\frac{t^2}{2(A+Bt)}\right) \quad \text{for all } t > 0.$$

Bounded Coupling implies Concentration Inequality

Let Y be a nonnegative random variable with mean and variance μ and σ^2 respectively, both finite and positive. Suppose there exists a coupling of Y to a variable Y^s having the Y-size bias distribution that satisfies $|Y^s - Y| \le C$ for some $C > 0$ with probability one. Let $A = C\mu/\sigma^2$ and $B = C/2\sigma$.

a) If $Y^s \ge Y$ with probability one, then

$$P\left(\frac{Y-\mu}{\sigma} \le -t\right) \le \exp\left(-\frac{t^2}{2A}\right) \quad \text{for all } t > 0.$$

b) If the moment generating function of Y is finite at $2/C$, then

$$P\left(\frac{Y-\mu}{\sigma} \ge t\right) \le \exp\left(-\frac{t^2}{2(A+Bt)}\right) \quad \text{for all } t > 0.$$

Outline of Proof

By the convexity of the exponential function, for all $x \neq y$,

$$\frac{e^y - e^x}{y - x} = \int_0^1 e^{ty+(1-t)x} dt \leq \int_0^1 (te^y + (1-t)e^x) dt = \frac{e^y + e^x}{2}.$$

Hence, when $|Y^s - Y| \leq C$, we obtain

$$Ee^{\theta Y^s} - Ee^{\theta Y} \leq \frac{C\theta}{2}\left(Ee^{\theta Y^s} + Ee^{\theta Y}\right).$$

With $m(\theta) = Ee^{\theta Y}$, the size bias relation yields

$$m'(\theta) = E[Ye^{\theta Y}] = \mu E[e^{\theta Y^s}].$$

Hence $m(\theta)$ satisfies the differential inequality

$$m'(\theta) \leq \mu \left(\frac{1 + C\theta/2}{1 - C\theta/2}\right) m(\theta) \quad \text{for all } 0 < \theta < 2/C.$$

Outline of Proof

By the convexity of the exponential function, for all $x \neq y$,

$$\frac{e^y - e^x}{y - x} = \int_0^1 e^{ty+(1-t)x} dt \leq \int_0^1 (te^y + (1-t)e^x) dt = \frac{e^y + e^x}{2}.$$

Hence, when $|Y^s - Y| \leq C$, we obtain

$$Ee^{\theta Y^s} - Ee^{\theta Y} \leq \frac{C\theta}{2}\left(Ee^{\theta Y^s} + Ee^{\theta Y}\right).$$

With $m(\theta) = Ee^{\theta Y}$, the size bias relation yields

$$m'(\theta) = E[Ye^{\theta Y}] = \mu E[e^{\theta Y^s}].$$

Hence $m(\theta)$ satisfies the differential inequality

$$m'(\theta) \leq \mu \left(\frac{1 + C\theta/2}{1 - C\theta/2}\right) m(\theta) \quad \text{for all } 0 < \theta < 2/C.$$

Outline of Proof

By the convexity of the exponential function, for all $x \neq y$,

$$\frac{e^y - e^x}{y - x} = \int_0^1 e^{ty+(1-t)x} dt \leq \int_0^1 (te^y + (1-t)e^x) dt = \frac{e^y + e^x}{2}.$$

Hence, when $|Y^s - Y| \leq C$, we obtain

$$Ee^{\theta Y^s} - Ee^{\theta Y} \leq \frac{C\theta}{2}\left(Ee^{\theta Y^s} + Ee^{\theta Y}\right).$$

With $m(\theta) = Ee^{\theta Y}$, the size bias relation yields

$$m'(\theta) = E[Ye^{\theta Y}] = \mu E[e^{\theta Y^s}].$$

Hence $m(\theta)$ satisfies the differential inequality

$$m'(\theta) \leq \mu \left(\frac{1 + C\theta/2}{1 - C\theta/2}\right) m(\theta) \quad \text{for all } 0 < \theta < 2/C.$$

Outline of Proof

By the convexity of the exponential function, for all $x \neq y$,

$$\frac{e^y - e^x}{y - x} = \int_0^1 e^{ty+(1-t)x} dt \leq \int_0^1 (te^y + (1-t)e^x) dt = \frac{e^y + e^x}{2}.$$

Hence, when $|Y^s - Y| \leq C$, we obtain

$$Ee^{\theta Y^s} - Ee^{\theta Y} \leq \frac{C\theta}{2} \left(Ee^{\theta Y^s} + Ee^{\theta Y} \right).$$

With $m(\theta) = Ee^{\theta Y}$, the size bias relation yields

$$m'(\theta) = E[Ye^{\theta Y}] = \mu E[e^{\theta Y^s}].$$

Hence $m(\theta)$ satisfies the differential inequality

$$m'(\theta) \leq \mu \left(\frac{1 + C\theta/2}{1 - C\theta/2} \right) m(\theta) \quad \text{for all } 0 < \theta < 2/C.$$

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.

Size Bias Exchangeable Indicators

For a given function g

$$E[Xg(X)] = \sum_{j=1}^{n} E[X_j g(X)] = \sum_{j=1}^{n} P[X_j = 1] E[g(X)|X_j = 1].$$

By exchangeability, $E[g(X)|X_j = 1] = E[g(X)|X_i = 1]$ for all $j = 1, \ldots, n$, so

$$E[Xg(X)] = \left(\sum_{j=1}^{n} P[X_j = 1] \right) E[g(X)|X_i = 1] = E[X]E[g(X^i)],$$

hence $X^i = X^s$.

Size Bias Exchangeable Indicators

Now mixing over an independent random index I, we have

$$\begin{aligned} Eg(X^I) &= \sum_{i=1}^n E[g(X^I), I = i] = \sum_{i=1}^n E[g(X^I)|I = i]P(I = i) \\ &= \sum_{i=1}^n Eg(X^i)P(I = i) = \sum_{i=1}^n Eg(X^s)P(I = i) \\ &= Eg(X^s)\sum_{i=1}^n P(I = i) = Eg(X^s). \end{aligned}$$

Size Bias Sum of Nonnegative Variables

For X_i a non-trivial indicator, recall $X_i^s = 1$. For nonnegative random variables $X_1, \ldots, X_n$ with finite mean and

$$X = \sum_{i=1}^{n} X_i,$$

construct X_i^s,

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

and select i independently with probability $P(I = i) = EX_i / EX$. Then

$$X^s = \sum_{j=1}^{n} X_j^I.$$

Applications

1. The number of local maxima of a random function on a graph
2. The number of urns containing exactly one ball in the uniform multinomial urn occupancy model
3. The number of lightbulbs switched on at the terminal time in the lightbulb process of Rao, Rao and Zhang
4. The number of relatively ordered subsequences of a random permutation
5. Sliding window statistics such as the number of m-runs in a sequence of independent coin tosses
6. The volume covered by the union of n balls placed uniformly over a volume n subset of $\mathbb{R}^d$

Applications

1. The number of local maxima of a random function on a graph
2. The number of urns containing exactly one ball in the uniform multinomial urn occupancy model
3. The number of lightbulbs switched on at the terminal time in the lightbulb process of Rao, Rao and Zhang
4. The number of relatively ordered subsequences of a random permutation
5. Sliding window statistics such as the number of m-runs in a sequence of independent coin tosses
6. The volume covered by the union of n balls placed uniformly over a volume n subset of $\mathbb{R}^d$

Applications

1. The number of local maxima of a random function on a graph
2. The number of urns containing exactly one ball in the uniform multinomial urn occupancy model
3. The number of lightbulbs switched on at the terminal time in the lightbulb process of Rao, Rao and Zhang
4. The number of relatively ordered subsequences of a random permutation
5. Sliding window statistics such as the number of m-runs in a sequence of independent coin tosses
6. The volume covered by the union of n balls placed uniformly over a volume n subset of $\mathbb{R}^d$

Applications

1. The number of local maxima of a random function on a graph
2. The number of urns containing exactly one ball in the uniform multinomial urn occupancy model
3. The number of lightbulbs switched on at the terminal time in the lightbulb process of Rao, Rao and Zhang
4. The number of relatively ordered subsequences of a random permutation
5. Sliding window statistics such as the number of m-runs in a sequence of independent coin tosses
6. The volume covered by the union of n balls placed uniformly over a volume n subset of $\mathbb{R}^d$

Applications

1. The number of local maxima of a random function on a graph
2. The number of urns containing exactly one ball in the uniform multinomial urn occupancy model
3. The number of lightbulbs switched on at the terminal time in the lightbulb process of Rao, Rao and Zhang
4. The number of relatively ordered subsequences of a random permutation
5. Sliding window statistics such as the number of m-runs in a sequence of independent coin tosses
6. The volume covered by the union of n balls placed uniformly over a volume n subset of $\mathbb{R}^d$

Applications

1. The number of local maxima of a random function on a graph
2. The number of urns containing exactly one ball in the uniform multinomial urn occupancy model
3. The number of lightbulbs switched on at the terminal time in the lightbulb process of Rao, Rao and Zhang
4. The number of relatively ordered subsequences of a random permutation
5. Sliding window statistics such as the number of m-runs in a sequence of independent coin tosses
6. The volume covered by the union of n balls placed uniformly over a volume n subset of $\mathbb{R}^d$

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

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

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

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

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 EX_v, has the Y-size bias distribution.

When $I = v$, the values X_u for $u \in \mathcal{V}_v$, and for $u \in \mathcal{V}_w$ may change, and we have

$$|Y^s - Y| \leq |\mathcal{V}_v(2)|$$

where $\mathcal{V}_v(2)$ are the neighbors, and the neighbors of neighbors of v.

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 EX_v, has the Y-size bias distribution.

When $I = v$, the values X_u for $u \in \mathcal{V}_v$, and for $u \in \mathcal{V}_w$ may change, and we have

$$|Y^s - Y| \leq |\mathcal{V}_v(2)|$$

where $\mathcal{V}_v(2)$ are the neighbors, and the neighbors of neighbors of v.

Example: Local Maxima, $\mathbb{Z}^p$ mod n

For $p \in \{1, 2, \ldots\}$ and $n \geq 5$ let $\mathcal{V} = \{1, \ldots, n\}^p$ modulo n in $\mathbb{Z}^p$ and set $\mathcal{E} = \{\{v, w\} : \sum_{i=1}^{p} |v_i - w_i| = 1\}$. Then

$$|Y^s - Y| \leq 2p^2 + 2p + 1,$$

and Y has mean and variance, respectively,

$$\mu = \frac{n}{2p+1} \quad \text{and} \quad \sigma^2 = n\left(\frac{4p^2 - p - 1}{(2p+1)^2(4p+1)}\right).$$

Right tail concentration inequality holds with

$$A = \frac{(2p+1)(4p+1)(2p^2+2p+1)}{4p^2 - p - 1} \quad \text{and} \quad B = \frac{2p^2+2p+1}{2\sigma}.$$

Number of Non-Isolated Balls under Uniform Allocation

- Say n balls are thrown independently into one of m equally likely urns. For $d \in \{0, 1, \ldots\}$ consider the number of urns containing d balls; $d = 0$ is a particularly well studied special case. The case $d = 1$ corresponds to the number of isolated balls. We equivalently study the number Y of non-isolated balls.
- Easy to construct an unbounded size bias coupling – import or export balls from a uniformly chosen urn so that it has the desired occupancy.
- A construction of Penrose and Goldstein yields a coupling of Y to Y^s satisfying $|Y^s - Y| \leq 2$.

Number of Non-Isolated Balls under Uniform Allocation

- Say n balls are thrown independently into one of m equally likely urns. For $d \in \{0, 1, \ldots\}$ consider the number of urns containing d balls; $d = 0$ is a particularly well studied special case. The case $d = 1$ corresponds to the number of isolated balls. We equivalently study the number Y of non-isolated balls.
- Easy to construct an unbounded size bias coupling – import or export balls from a uniformly chosen urn so that it has the desired occupancy.
- A construction of Penrose and Goldstein yields a coupling of Y to Y^s satisfying $|Y^s - Y| \leq 2$.

Number of Non-Isolated Balls under Uniform Allocation

- Say n balls are thrown independently into one of m equally likely urns. For $d \in \{0, 1, \ldots\}$ consider the number of urns containing d balls; $d = 0$ is a particularly well studied special case. The case $d = 1$ corresponds to the number of isolated balls. We equivalently study the number Y of non-isolated balls.
- Easy to construct an unbounded size bias coupling – import or export balls from a uniformly chosen urn so that it has the desired occupancy.
- A construction of Penrose and Goldstein yields a coupling of Y to Y^s satisfying $|Y^s - Y| \leq 2$.

Non-Isolated Balls, Coupling

Let X_i be the location of ball $i = 1, \ldots, n$. Select a balls $I \neq J$, uniformly from $\{1, 2, \ldots, n\}$, and independently of $X_1, \ldots, X_n$. With M_i the number of balls in the urn containing ball i, and $N \sim \text{Bin}(1/m, n-1)$, import ball J into the urn containing ball I with probability π_{M_I}, where

$$\pi_k = \begin{cases} \frac{P(N>k|N>0)-P(N>k)}{P(N=k)(1-k/(n-1))} & \text{if } 0 \leq k \leq n-2 \\ 0 & \text{if } k = n-1. \end{cases}$$

We have $|Y^s - Y| \leq 2$, as at most the occupancy of two urns can affected by the movement of a single ball. Can check also that $\pi_0 = 1$, so if ball I is isolated we always move ball J to urn X_I.

Non-Isolated Balls, Coupling

Let X_i be the location of ball $i = 1, \ldots, n$. Select a balls $I \neq J$, uniformly from $\{1, 2, \ldots, n\}$, and independently of $X_1, \ldots, X_n$. With M_i the number of balls in the urn containing ball i, and $N \sim \text{Bin}(1/m, n-1)$, import ball J into the urn containing ball I with probability π_{M_I}, where

$$\pi_k = \begin{cases} \frac{P(N>k|N>0) - P(N>k)}{P(N=k)(1-k/(n-1))} & \text{if } 0 \leq k \leq n-2 \\ 0 & \text{if } k = n-1. \end{cases}$$

We have $|Y^s - Y| \leq 2$, as at most the occupancy of two urns can affected by the movement of a single ball. Can check also that $\pi_0 = 1$, so if ball I is isolated we always move ball J to urn X_I.

Non-Isolated Balls, Concentration

For positive functions f and h depending on n write $f \asymp h$ when $\lim_{n\to\infty} f/h = 1$.

If m and n both go to infinity such that $n/m \to \alpha \in (0,\infty)$, then with $g(\alpha)^2 = e^{-\alpha} - e^{-2\alpha}(\alpha^2 - \alpha + 1) > 0$, the mean and variance of Y satisfy

$$\mu \asymp n(1 - e^{-\alpha}) \quad \text{and } \sigma^2 \asymp ng(\alpha)^2.$$

Hence, in this asymptotic Y satisfies the right tail concentration inequality with constants A and B satisfying

$$A \asymp \frac{2(1 - e^{-\alpha})}{e^{-\alpha} - e^{-2\alpha}(\alpha^2 - \alpha + 1)} \quad \text{and} \quad B \asymp \frac{1}{\sqrt{n}g(\alpha)}.$$

Non-Isolated Balls, Concentration

For positive functions f and h depending on n write $f \asymp h$ when $\lim_{n\to\infty} f/h = 1$.

If m and n both go to infinity such that $n/m \to \alpha \in (0,\infty)$, then with $g(\alpha)^2 = e^{-\alpha} - e^{-2\alpha}(\alpha^2 - \alpha + 1) > 0$, the mean and variance of Y satisfy

$$\mu \asymp n(1-e^{-\alpha}) \quad \text{and } \sigma^2 \asymp ng(\alpha)^2.$$

Hence, in this asymptotic Y satisfies the right tail concentration inequality with constants A and B satisfying

$$A \asymp \frac{2(1-e^{-\alpha})}{e^{-\alpha} - e^{-2\alpha}(\alpha^2-\alpha+1)} \quad \text{and} \quad B \asymp \frac{1}{\sqrt{n}g(\alpha)}.$$

Similar phenomenon using the zero bias coupling

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

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

Under existence of MGF in a certain interval, when a coupling satisfying $|Y^* - Y| \leq C$ exists,

$$P(Y \geq t) \leq \exp\left(-\frac{t^2}{2(\sigma^2 + Ct)}\right), \tag{1}$$

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

May improve on size bias coupling

Binomial case, without loss of generality $p \in (0, 1/2]$. Size bias bound:

$$\mathbb{P}(Y - np \geq t) \leq \begin{cases} \exp(-\frac{t^2}{2(np+t)}) & 0 \leq t \leq n(1-2p) \\ \exp(-\frac{t^2}{2n(1-p)}) & t > n(1-2p), \end{cases}$$

Zero bias bound:

$$\mathbb{P}(Y - np \geq t) \leq \exp\left(-\frac{t^2}{2np(1-p) + 2t}\right)$$

This inequality improves on the size bias bound for $0 \leq t \leq n(1-2p)$, and the second zero bias bound for $n(1-2p) < t < 2n(1-p)(1-2p)$ for $p \in (0, 1/2]$,

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.

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 $E\sigma_A^2 < E\mu_A$.

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. Close, in particular bounded couplings imply concentration of measure, and central limit behavior.

Unbounded couplings can also be handled but seemingly yet only on a case by case basis – 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. Close, in particular bounded couplings imply concentration of measure, and central limit behavior.

Unbounded couplings can also be handled but seemingly yet only on a case by case basis – 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. Close, in particular bounded couplings imply concentration of measure, and central limit behavior.

Unbounded couplings can also be handled but seemingly yet only on a case by case basis – 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. Close, in particular bounded couplings imply concentration of measure, and central limit behavior.

Unbounded couplings can also be handled but seemingly yet only on a case by case basis – e.g., the number of isolated vertices in the Erdös-Rényi random graph (Ghosh, Goldstein and Raič).

Work in Progress

Joint with Jay Bartroff and Ümit Işlak, for occupancy models with log concave marginal distributions, by generalizing the bounded size bias coupling of Penrose and Goldstein, we produce concentration bounds for multinomial allocation models, random graphs, multivariate hypergeometric sampling and germ-grain models in stochastic geometry.

References

Goldstein, L, and Işlak, Ü. (2013) Concentration inequalities via zero bias couplings. *preprint*.

Ghosh, S., and Goldstein, L. (2011) Concentration of Measures via Size Biased Couplings. *Prob. Theory and Related Fields,* **149** pp. 271-278.

Ghosh, S., and Goldstein, L. (2011) Applications of size biased couplings for concentration of measures. *Electronic Communications in Probability*, **16** pp. 70-83.

Ghosh, S., Goldstein, L. and Raič, M. (2011) Concentration of measure for the number of isolated vertices in the Erdös-Rényi random graph by size bias couplings. *Statistics and Probability Letters*, **81**, pp.1565-1570.