Concentration inequalities for occupancy models with log-concave marginals

Jay Bartroff, Larry Goldstein, and Ümit Işlak

USC

International Colloquium on Stein's Method, Concentration Inequalities, and Malliavin Calculus 2014

Thanks to Janna Beling and the Borchard Foundation

Concentration and Coupling

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

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

(Inequalities)

Concentration inequalities for occupancy models with log-concave marginals

Main idea: How to get bounded, size biased couplings for certain multivariate occupancy models, then use existing methods to get concentration inequalities

Outline

1. The models
2. Some methods for concentration inequalities
3. Our main result
4. Applications
    - Erdös-Rényi random graph
    - Germ-grain models
    - Multinomial counts
    - Multivariate hypergeometric sampling
5. Comparisons
    - McDiarmid's Inequality
    - Negative association
    - Certifiable functions

Setup

- Occupancy model $\boldsymbol{M} = (M_\alpha)$
- M_α may be
  - degree count of vertex α in an Erdös-Rényi random graph
  - # of grains containing point α in a germ-grain model
  - # of balls in box α in multinomial model
  - # balls of color α in sample from urn of colored balls
- We consider statistics like

$$Y_{ge} = \sum_{\alpha=1}^{m} \mathbf{1}\{M_\alpha \geq d\}, \quad Y_{eq} = \int \mathbf{1}\{M(x) = d\}dx$$

$$Y_{ge} = \sum_{\alpha=1}^{m} w_\alpha \mathbf{1}\{M_\alpha \geq d_\alpha\}, \quad Y_{eq} = \int w(x)\mathbf{1}\{M(x) = d(x)\}dx$$

Some Methods for Concentration Inequalities

McDiarmid's (Bounded Difference) Inequality

If

- $X_1, \ldots, X_n$ independent
- $Y = f(X_1, \ldots, X_n)$, f measurable
- there are c_i such that

$$\sup_{x_i, x_i'} \left| f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n) \right| \leq c_i,$$

then

$$P(Y - \mu \geq t) \leq \exp\left(-\frac{t^2}{2 \sum_{i=1}^{n} c_i^2} \right) \quad \text{for all } t > 0,$$

and a similar left tail bound.

Some Methods for Concentration Inequalities

Negative Association (NA)

$X_1, X_2, ..., X_m$ are NA if

$$E(f(X_i; i \in A_1)g(X_j; j \in A_2)) \leq E(f(X_i; i \in A_1))E(g(X_j; j \in A_2))$$

for any

- $A_1, A_2 \subset [m]$ disjoint,
- f, g coordinate-wise nondecreasing.

Dubashi & Ranjan 98

If $X_1, X_2, ..., X_m$ are NA indicators, then $Y = \sum_{i=1}^m X_i$ satisfies

$$P(Y - \mu \geq t) \leq \left(\frac{\mu}{\mu + t}\right)^{t+\mu} e^t \quad \text{for all } t > 0$$
$$= O(\exp(-t \log t)) \quad \text{as } t \to \infty.$$

McDiarmid & Reed 06

If $X_1, X_2, ..., X_n$ independent and $Y = f(X_1, X_2, ..., X_n)$ where f is **certifiable:**

- There is c such that changing any coordinate x_j changes the value of $f(x)$ by at most c,
- If $f(x) = s$ then there is $C \subset [n]$ with $|C| \leq as + b$ such that that $y_i = c_i \ \forall i \in C$ implies $f(y) \geq s$,

Then

$$P(Y - \mu \leq -t) \leq \exp\left(-\frac{t^2}{2c^2(a\mu + b + t/3c)}\right) \quad \text{for all } t > 0,$$
$$= O(\exp(-t)) \quad \text{as } t \to \infty.$$

A similar right tail bound.

If there is a coupling Y^s of Y with the Y-size bias distribution and $Y^s \leq Y + c$ for some $c > 0$ with probability one, then

$$\max\{P(Y-\mu \geq t), P(Y-\mu \leq -t)\} \leq b_{\mu,c}(t).$$

Ghosh & Goldstein 11: For all $t > 0$,

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

b exponential as $t \to \infty$.
Arratia & Baxendale 13:

$$b_{\mu,c}(t) = \exp\left(-\frac{\mu}{c}h\left(\frac{t}{\mu}\right)\right), \quad h(x) = (1+x)\log(1+x) - x.$$

b Poisson as $t \to \infty$.

Main Result

$$\boldsymbol{M} = (M_\alpha)_{\alpha\in[m]}, \quad M_\alpha \text{ lattice log-concave}$$

$$Y_{ge} = \sum_{\alpha\in[m]} w_\alpha \mathbf{1}\{M_\alpha \geq d_\alpha\}, \quad Y_{ne} = \sum_{\alpha\in[m]} w_\alpha \mathbf{1}\{M_\alpha \neq d_\alpha\}.$$

Main Result (in words)

1. If $\boldsymbol{M}$ is bounded from below and can be **closely coupled** to a version $\boldsymbol{M}'$ having the same distribution conditional on $M'_\alpha = M_\alpha + 1$, then there is a bounded size biased coupling $Y^s_{ge} \leq Y_{ge} + C$ and the above concentration inequalities hold.
2. If $\boldsymbol{M}$ is non-degenerate at (d_α) and can be **closely coupled** to a version $\boldsymbol{M}'$ having the same distribution conditional on $M'_\alpha \neq d_\alpha$, then there is a bounded size biased coupling $Y^s_{ne} \leq Y_{ne} + C'$ and the above concentration inequalities hold.

Main Result

A few more details on Part 1

$\boldsymbol{M} = f(\mathcal{U})$ where

- $\mathcal{U}$ is some collection of random variables
- f is measurable

Closely coupled means given $\mathcal{U}_k \sim \mathcal{L}(\mathcal{V}_k) := \mathcal{L}(\mathcal{U} | M_\alpha \geq k)$, there is coupling $\mathcal{U}_k^+$ and constant B such that

$$\mathcal{L}(\mathcal{U}_k^+ | \mathcal{U}_k) = \mathcal{L}(\mathcal{V}_k | M_{k,\alpha}^+ = M_{k,\alpha} + 1) \quad \text{and} \quad Y_{k,ge,\neq\alpha}^+ \leq Y_{k,ge,\neq\alpha} + B,$$

where $Y_{k,ge,\neq\alpha} = \sum_{\beta\neq\alpha} \mathbf{1}(M_{k,\beta} \geq d_\beta)$.

The constant is

$$C = |\boldsymbol{w}|(B|\boldsymbol{d}| + 1)$$

where $|\boldsymbol{w}| = \max w_\alpha$, $|\boldsymbol{d}| = \max d_\alpha$.

Part 2 is similar.

Main Result

Main Ingredients in Proof

Incrementing Lemma

If M is lattice log-concave then there is $\pi(x, d) \in [0, 1]$ such that if

$$M' \sim \mathcal{L}(M | M \geq d) \quad \text{and} \quad B | M' \sim \text{Bern}(\pi(M', d)),$$

then

$$M' + B \sim \mathcal{L}(M | M \geq d + 1).$$

- Extension of Goldstein & Penrose 10 for M Binomial, $d = 0$
- Analogous versions for

$$\mathcal{L}(M | M \leq d) \hookrightarrow \mathcal{L}(M | M \leq d - 1)$$
$$\mathcal{L}(M) \hookrightarrow \mathcal{L}(M | M \neq d)$$

where $\hookrightarrow$ means "coupled to"

Main Result

Main Ingredients in Proof

Mixing Lemma (Goldstein & Rinnott 96)

A nonnegative linear combination of Bernoullis with positive mean can be size biased by

1. choosing a summand with probability proportional to its mean,
2. replacing chosen summand by 1, and
3. modifying other summands to have correct conditional distribution.

Main Result

Main Steps in Proof of Part 1

1. Induction on k: Given $\mathcal{U}_k, \mathcal{U}_k^+$, let

$$\mathcal{U}_{k+1} = \begin{cases} \mathcal{U}_k^+ & \text{with probability } \pi(M_{k,\alpha}, k) \\ \mathcal{U}_k & \text{otherwise.} \end{cases}$$

   $\mathcal{U}_{k+1}$ has correct distribution by **Incrementing Lemma.**

2. Using $k = d_\alpha$ case of induction and **Mixing Lemma**, mixing $Y_{ge}^\alpha = f(\mathcal{U}_{d_\alpha})$ with probabilities $\propto w_\alpha P(M_\alpha \geq d_\alpha)$ yields size biased

$$Y_{ge}^s \leq Y_{ge} + |\boldsymbol{w}|(B|\boldsymbol{d}| + 1).$$

Application 1: Erdös-Rényi random graph

- m vertices
- Independent edges with probability $p_{\alpha,\beta} = p_{\beta,\alpha} \in [0, 1)$.
- Constructing $\mathcal{U}_k^+$ from $\mathcal{U}_k$:
  1. Selection non-neighbor β of α with probability

$$\propto \frac{p_{\alpha,\beta}}{1 - p_{\alpha,\beta}}$$

  2. Add edge connecting β to α
- This affects at most 1 other vertex so $B = 1$ and

$$Y_{ge}^s \leq Y_{ge} + |\boldsymbol{w}|(|\boldsymbol{d}| + 1).$$

Application 1: Erdös-Rényi random graph

- Applying this to $Y_{is} = m - Y_{ge}$ with $d_\alpha = 1$:

$$P(Y_{is} - \mu_{is} \leq -t) = P(Y_{ge} - \mu_{ge} \geq t) \leq \exp\left(\frac{-t^2}{4(m - \mu_{is} + t/3)}\right)$$

- Ghosh, Goldstein, & Raič 11 studied Y_{is} using an unbounded size biased coupling

$$P(Y_{is} - \mu_{is} \leq -t) \leq \exp\left(\frac{-t^2}{4\mu_{is}}\right)$$

- New bound
  - an improvement for $t \leq 6\mu_{is} - 3m$
  - applicable for all d_α

Application 2: Germ-Grain Models

- Used in forestry, wireless sensor networks, material science, ...
- Germs $U_\alpha \sim f_\alpha$ strictly positive on $[0, r)^p$
- Grains $B_\alpha =$ closed ball of radius ρ_α centered at U_α
- $d : [0, r)^p \to \{0, 1, \ldots, m\} =$ # of intersections we're interested in at $x \in [0, r)^p$
- Choice of r relative to p, ρ_α guarantees nontrivial distribution of

$$\begin{aligned} M(x) &= \text{\# of grains containing at point } x \in [0, r)^p \\ &= \sum_{\alpha \in [m]} \mathbf{1}\{x \in B_\alpha\} \\ Y_{ge} &= \int_{[0,r)^p} w(x)\mathbf{1}\{M(x) \geq d(x)\}dx \\ &= \text{(weighted) volume of } d\text{-way intersections of grains} \end{aligned}$$

Application 2: Germ-Grain Models

Main ideas in proof

Different approach:

1. Generate U_0 independent of $U_1, \ldots, U_m$
2. Compute $\mathcal{U}_0, \ldots, \mathcal{U}_{d(U_0)}$ and set $Y_{ge}^s = Y_{ge}(M_{d(U_0)})$
3. Y_{ge}^s has size bias distribution by **Conditional Lemma** with $A = \{M(U_0) \geq d(U_0)\}$:

Conditional Lemma (Goldstein & Penrose 10)

If $P(A) \in (0, 1) < 1$ and $Y = P(A|\mathcal{F})$, then Y^s has the Y-size bias distribution if $\mathcal{L}(Y^s) = \mathcal{L}(Y|A)$.

Application 2: Germ-Grain Models

Main ideas in proof

Argument: Generate $U_0 \sim w(x)/\int w$. Given $\mathcal{U}_k \sim \mathcal{L}(\mathcal{U}_0 | M(U_0) \geq k)$, with probability $\pi(M_k(U_0), k)$ choose germ β with probability

$$\propto \frac{p_\beta(U_0)}{1 - p_\beta(U_0)}, \quad \text{where} \quad p_\beta(x) = P(x \in U_\beta),$$

from germs whose grains do not contain U_0, replace it with $U'_\beta \sim P_{U_0}$ to get $\mathcal{U}_{k+1}$, where

$$P_{U_0}(V) = P(U_\beta \in V | D(U_\beta, U_0) \leq \rho_\beta).$$

Otherwise $\mathcal{U}_{k+1} = \mathcal{U}_k$.

- Volume increase replacing U_β by U'_β at most $\nu_p |\boldsymbol{\rho}|^p$ ($\nu_p =$ vol. of unit ball)
- Volume increase between $\mathcal{U}_0$ and $\mathcal{U}_{d(U_0)}$ at most $\nu_p |\boldsymbol{\rho}|^p |\boldsymbol{d}|$
- Y^s_{ge} increases Y_{ge} by at most $\nu_p |\boldsymbol{\rho}|^p |\boldsymbol{d}| |\boldsymbol{w}|$

Application 3: Multinomial Counts

- n balls independently into m boxes
- Applications in species trapping, linguistics, ...
- # empty boxes proved asymptotically normal by Weiss 58, Rényi 62 in uniform case
- Englund 81: L^∞ bound for # of empty cells, uniform case
- Dubashi & Ranjan 98: Concentration inequality via NA
- Penrose 09: L^∞ bound for # of isolated balls, uniform and nonuniform cases
- Bartroff & Goldstein 13: L^∞ bound for all $d \geq 2$, uniform case

Application 3: Multinomial Counts

$$
\begin{aligned}
p_{\alpha,j} &= \text{ prob. that ball } j \in [n] \text{ falls in box } \alpha \in [m] \\
M_\alpha &= \#\text{ balls in box } \alpha \\
&= \sum_{j\in[n]} \mathbf{1}\{\text{ball } j \text{ falls in box } \alpha\}
\end{aligned}
$$

Constructing $\mathcal{U}_k^+$ from $\mathcal{U}_k$: Choose ball $j \notin$ box α with probability

$$\propto \frac{p_{\alpha,j}}{1-p_{\alpha,j}}$$

and add it to box α.

$Y^s_{ge,\neq\alpha} \leq Y_{ge,\neq\alpha}$ so $B = 0$, thus $Y^s_{ge} \leq Y_{ge} + |\boldsymbol{w}|$

Application 4: Multivariate Hypergeometric Sampling

- Urn with $n = \sum_{\alpha \in [m]} n_\alpha$ colored balls, n_α balls of color α
- Sample of size s drawn without replacement
- $M_\alpha =$ # balls in sample of color α
- Applications in sampling (and subsampling) theory, gambling, coupon-collector problems

Constructing $\mathcal{U}_k^+$ from $\mathcal{U}_k$: Select non-α colored ball in sample with probability

$$\propto \frac{n_{\alpha(j)}/n}{1 - n_{\alpha(j)}/n}, \quad \alpha(j) = \text{color of ball } j$$

and replace it with α-colored ball

$Y^s_{ge,\neq\alpha} \leq Y_{ge,\neq\alpha}$ so $B = 0$, thus $Y^s_{ge} \leq Y_{ge} + |\boldsymbol{w}|$

Comparison 1: McDiarmid's Inequality

If

- $X_1, \ldots, X_n$ independent
- $Y = f(X_1, \ldots, X_n)$, f measurable
- there are c_i such that

$$\sup_{x_i, x_i'} \left| f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n) \right| \leq c_i,$$

then

$$P(Y - \mu \geq t) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^n c_i^2} \right) \quad \text{for all } t > 0,$$

and a similar left tail bound.

Comparison 1: McDiarmid's Inequality

Erdös-Rényi random graph

m vertices, probability p of edge,

$$Y_{ge} = f(X_1, \ldots, X_{\binom{m}{2}}), \quad X_i = \mathbf{1}\{\text{edge between vertex pair } i\},$$

f has bounded differences with $c_i = 2$

$$\text{McDiarmid} \Rightarrow P(Y_{eq} - \mu_{ge} \leq -t) \leq \exp\left(\frac{-t^2}{4m(m-1)}\right)$$

$$\begin{aligned}\text{Size-bias} \Rightarrow P(Y_{eq} - \mu_{ge} \leq -t) &\leq \exp\left(\frac{-t^2}{2(d+1)\mu_{ge}}\right) \\ &\leq \exp\left(\frac{-t^2}{2m(d+1)}\right)\end{aligned}$$

since $\mu_{ge} \leq m$.

Comparison 2: Negative Association

$X_1, X_2, ..., X_m$ are NA if

$$E(f(X_i; i \in A_1)g(X_j; j \in A_2)) \leq E(f(X_i; i \in A_1))E(g(X_j; j \in A_2))$$

for any

- $A_1, A_2 \subset [m]$ disjoint,
- f, g coordinate-wise nondecreasing.

Dubashi & Ranjan 98

If $X_1, X_2, ..., X_m$ are NA indicators, then $Y = \sum_{i=1}^m X_i$ satisfies

$$\begin{aligned} P(Y - \mu \geq t) &\leq \left(\frac{\mu}{\mu + t}\right)^{t+\mu} e^t \quad \text{for all } t > 0 \\ &= O(\exp(-t \log t)) \quad \text{as } t \to \infty. \end{aligned}$$

Comparison 2: Negative Association

Both NA and our method yield same order bound for Y_{ge} in

- Multinomial counts
- Multivariate hypergeometric sampling

but NA cannot be applied to:

- Y_{ne} in multinomial counts
- Y_{ne} in multivariate hypergeometric sampling
- Y_{ge} or Y_{ne} in Erdös-Rényi random graph
- Y_{ge} or Y_{ne} in germ-grain models

Comparison 3: Certifiable Functions

McDiarmid & Reed 06

If $X_1, X_2, ..., X_n$ independent and $Y = f(X_1, X_2, ..., X_n)$ where f is **certifiable:**

- There is c such that changing any coordinate x_j changes the value of $f(x)$ by at most c,
- If $f(x) = s$ then there is $C \subset [n]$ with $|C| \leq as + b$ such that that $y_i = c_i \ \forall i \in C$ implies $f(y) \geq s$,

Then

$$P(Y - \mu \leq -t) \leq \exp\left(-\frac{t^2}{2c^2(a\mu + b + t/3c)}\right) \quad \text{for all } t > 0,$$
$$= O(\exp(-t)) \quad \text{as } t \to \infty.$$

A similar right tail bound.

Comparison 3: Certifiable Functions

Asymptotically $O(e^{-t})$.

- Best possible rate via log Sobolev inequalities(?)

Multinomial Occupancy: We showed $C = |\boldsymbol{w}|$ so if $w_\alpha = 1$,

$$P(Y_{ge} - \mu_{ge} \leq -t) \leq \exp\left(\frac{-t^2}{2\mu_{ge}}\right).$$

Similar for right tail, Y_{ne}

Merci pour votre attention!