Stein's Method: Distributional Approximation and Concentration of Measure

Larry Goldstein
University of Southern California

36^{th} Midwest Probability Colloquium, 2014

Concentration of Measure

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

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

Log Sobolev Inequality

For a probability measure μ on $\mathbb{R}^n$ and f a non-negative real valued measurable function, let

$$\mathrm{Ent}_\mu(f) = E_\mu[f \log f] - E_\mu[f] \log E_\mu[f].$$

If

$$\mathrm{Ent}_\mu(f^2) \le 2CE_\mu \|\nabla f\|^2,$$

then for every 1-Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}$,

$$P_\mu(F \ge EF + r) \le e^{-r^2/2C}$$

Herbst (Unpublished), Gross (1975), Ledoux and Talagrand (1991), Talagrand (1995), Ledoux (1999)

Transportation Method

For μ and ν probability measures, define the L^p transport distance

$$W_p^d(\nu,\mu) = \inf \left(\int \int d(x,y)^p d\pi(x,y) \right)^{1/p},$$

we say μ satisfies the L^p-transportation cost inequality if for some C such that for al ν

$$W_p^d(\mu,\nu) \le \sqrt{2CH(\nu|\mu)} \quad \text{where} \quad H(\nu|\mu) = E_\nu \log \frac{d\nu}{d\mu}.$$

Transportation Method

The measure μ satisfies the L^1 transportation cost inequality if and only if for all Lipschitz functions F

$$E_\mu e^{\lambda(F-EF)} \le \exp\left(\frac{\lambda^2}{2} C\|F\|_{\mathrm{Lip}}^2\right) \quad \text{for all } \lambda \in \mathbb{R},$$

in which case

$$\mu(F - EF > r) \le \exp\left(-\frac{r^2}{2C\|F\|_{\mathrm{Lip}}^2}\right) \quad \text{for all } r \in \mathbb{R}.$$

Bobkov and Götze (1999), Marton (1996), Talagrand (1996)

Aside: Improved Log Sobolev

For γ standard Gaussian measure in $\mathbb{R}^n$, and a probability measure ν with $d\nu = hd\gamma$ classical LSI can be written

$$H(\nu|\gamma) = \int_{\mathbb{R}^n} h \log h d\gamma \leq \frac{1}{2} \int_{\mathbb{R}^n} \frac{\|\nabla h\|^2}{h} d\gamma = \frac{1}{2} I(\nu|\gamma).$$

Ledoux, Nourdin and Peccati (2014) show

$$H(\nu|\gamma) \leq \frac{1}{2} S^2(\nu|\gamma) \log \left(1 + \frac{I(\nu|\gamma)}{S^2(\nu|\gamma)}\right)$$

where the 'Stein discrepancy' is given by

$$S(\nu|\gamma) = \left(\int_{\mathbb{R}^d} \|\tau_\nu - I_n\|_{\mathrm{HS}}^2 d\nu\right)^{1/2},$$

with τ_ν a (multivariate) Stein coefficient,

$$\int_{\mathbb{R}^d} \mathbf{x} \cdot \nabla \phi(x) d\nu = \int_{\mathbb{R}^d} \langle \tau_\nu, \mathrm{Hess}(\phi) \rangle d\nu.$$

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 $f : \mathbb{R}^n \to \mathbb{R}$

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

are bounded by c_i, then

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

Hoeffding (1963), Azuma (1967), McDiarmid (1989)

Longest Common Subsequence Problem

Let $L(m,n)$ be the length of the longest common subsequence between $(X_1,\ldots,X_m)$ and $(X_{m+1},\ldots,X_{m+n})$, two sequences of independent letters of lengths m and n from some discrete alphabet.

As changing one letter can change the longest common subsequence by at most one, $L(m,n)$ attains the two sided tail bound $2\exp\left(-t^2/2(n+m)\right)$ about its expectation.

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

Longest Common Subsequence Problem

Let $L(m, n)$ be the length of the longest common subsequence between $(X_1, \ldots, X_m)$ and $(X_{m+1}, \ldots, X_{m+n})$, two sequences of independent letters of lengths m and n from some discrete alphabet.

As changing one letter can change the longest common subsequence by at most one, $L(m, n)$ attains the two sided tail bound $2\exp\left(-t^2/2(n+m)\right)$ about its expectation.

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

Longest Common Subsequence Problem

Let $L(m,n)$ be the length of the longest common subsequence between $(X_1,\ldots,X_m)$ and $(X_{m+1},\ldots,X_{m+n})$, two sequences of independent letters of lengths m and n from some discrete alphabet.

As changing one letter can change the longest common subsequence by at most one, $L(m,n)$ attains the two sided tail bound $2\exp\left(-t^2/2(n+m)\right)$ about its expectation.

Though the distribution of $L(m,n)$ is intractable (even the constant $c=\lim_{m\to\infty} L(m,m)/m$ is famously unknown for fair coin tossing), 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 $\sum_{i=1}^{n} \alpha_i(x)^2 \leq c^2$ for some constant c. Then for $X_1, \ldots, X_n$, i.i.d. $\mathcal{U}([0,1]^d)$,

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

where M_n is the median of $L(X_1, \ldots, X_n)$.
Applications, e.g. Steiner Tree, Traveling Salesman Problem. Need to construct weights $\alpha_i(x)$,, and bound their sum of squares.

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 $t \geq 0$,

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

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

McDiarmid and Reed (2006)

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 when 'good' couplings exist (small perturbation).

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 when 'good' couplings exist (small perturbation).

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 when 'good' couplings exist (small perturbation).

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 when 'good' couplings exist (small perturbation).

Stein Couplings

We say the triple (G, W, W') is a Stein coupling if for all $f \in \mathcal{F}$,

$$E[G(f(W) - f(W'))] = E[Wf(W)].$$

Chen and Röllin (2010), for normal approximation; exchangeable pair and size bias are special cases.

For $f(w) = e^{\theta w}$ and $m(\theta) = E[e^{\theta W}]$, right hand side is

$$E[Wf(W)] = E[We^{\theta W}] = m'(\theta).$$

Obtain differential inequality for $m(\theta)$: Herbst argument.

Exchangeable Pair Couplings

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

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

with

$$\Delta(\mathbf{X}) \leq bf(\mathbf{X}) + c \quad \text{where} \quad \Delta(\mathbf{X}) = \frac{1}{2}\mathbb{E}[|(f(\mathbf{X}) - f(\mathbf{X}'))F(\mathbf{X}, \mathbf{X}')|\big|\mathbf{X}].$$

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

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

Subgaussian left tail bound. No independence assumption.
Chatterjee (2006), Chatterjee and Dey (2010).

Exchangeable Pair Couplings

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

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

with

$$\Delta(\mathbf{X}) \le bf(\mathbf{X}) + c \quad \text{where} \quad \Delta(\mathbf{X}) = \frac{1}{2}\mathbb{E}[|(f(\mathbf{X}) - f(\mathbf{X}'))F(\mathbf{X}, \mathbf{X}')| \big| \mathbf{X}].$$

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

$$\mathbb{P}(Y \ge t) \le 2\exp\left(-\frac{t^2}{2c + 2bt}\right).$$

Subgaussian left tail bound. No independence assumption.
Chatterjee (2006), Chatterjee and Dey (2010).

Exchangeable Pair Couplings

With $m(\theta) = E[e^{\theta f(\mathbf{X})}]$, we have

$$m'(\theta) = \frac{1}{2} E\left[\left(e^{\theta f(\mathbf{X})} - e^{\theta f(\mathbf{X}')}\right) F(\mathbf{X}, \mathbf{X}')\right]$$

Apply

$$|e^x - e^y| \leq \frac{1}{2}|x - y|\,|e^x + e^y|$$

to obtain, taking $\theta \geq 0$ for the right tail bound,

$$\begin{aligned} |m'(\theta)| &\leq \frac{\theta}{4} E\left(e^{\theta f(\mathbf{X})} + e^{\theta f(\mathbf{X}')}\right) |\left(f(\mathbf{X}) - f(\mathbf{X}')\right) F(\mathbf{X}, \mathbf{X}')| \\ &= \frac{\theta}{2} E\left(e^{\theta f(\mathbf{X})}\Delta(\mathbf{X}) + e^{\theta f(\mathbf{X}')}\Delta(\mathbf{X}')\right) = E\left[\theta e^{\theta f(\mathbf{X})}\Delta(\mathbf{X})\right] \\ &\leq \theta E\left(bf(\mathbf{X}) + c\right) e^{\theta f(\mathbf{X})} = b\theta m'(\theta) + c\theta m(\theta). \end{aligned}$$

Curie Weiss Model

Consider the complete graph on n vertices $V = \{1, \ldots, n\}$ with Hamiltonian

$$H_h(\sigma) = \frac{1}{n} \sum_{j<k} \sigma_j \sigma_k + h \sum_{i \in V} \sigma_j$$

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

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

Let

$$m = \frac{1}{n} \sum_{j \in V} \sigma_j \quad \text{and} \quad m_i = \frac{1}{n} \sum_{j:j \neq i} \sigma_j.$$

Curie Weiss Concentration

Take $h = 0$ for simplicity. Then

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

The magnetization m is concentrated about the roots of the equation

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

Seems not possible to use concentration results that would require expressing m in terms of independent random variables.

Curie Weiss Concentration

Take $h = 0$ for simplicity. Then

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

The magnetization m is concentrated about the roots of the equation

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

Seems not possible to use concentration results that would require expressing m in terms of independent random variables.

Curie Weiss Concentration

Take $h = 0$ for simplicity. Then

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

The magnetization m is concentrated about the roots of the equation

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

Seems not possible to use concentration results that would require expressing m in terms of independent random variables.

Curie Weiss Concentration

Choose $v \in V$ uniformly and sample σ'_v from the conditional distribution of σ_v given $\sigma_j, j \notin N_v$. Then the configurations (X, X') are exchangeable. Now let

$$F(X, X') = \sum_{i=1}^{n} (\sigma_i - \sigma'_i) = \sigma_v - \sigma'_v.$$

Then $F(X, X')$ is anti-symmetric, and

$$f(X) = E[F(X, X')|X] = \frac{1}{n} \sum_{i=1}^{n} \left(\sigma_i - E(\sigma'_i|X)\right) \approx m - \tanh(\beta m),$$

since $E(\sigma'_i|X) = P(\sigma_i = 1|\sigma_j, j \neq i) - P(\sigma_i = -1|\sigma_j, j \neq i)$, and

$$P(\sigma_i = 1|\sigma_j, j \neq i) = \frac{e^{\beta m_i}}{e^{\beta m_i} + e^{-\beta m_i}},$$

so

$$E(\sigma_i|\sigma_j, j \neq i) = \frac{e^{\beta m_i} - e^{-\beta m_i}}{e^{\beta m_i} + e^{-\beta m_i}} = \tanh(\beta m_i).$$

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

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

When Y is a sum of (possibly dependent) non-trivial indictors, then we may form Y^s by choosing one proportional to is expectation and setting it to one, and for the remainder, sampling from the conditional distribution of the others given that the chosen one now takes the value one.

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

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

When Y is a sum of (possibly dependent) non-trivial indictors, then we may form Y^s by choosing one proportional to is expectation and setting it to one, and for the remainder, sampling from the conditional distribution of the others given that the chosen one now takes the value one.

Bounded Coupling implies Concentration Inequality

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 G. (2011), improvement by Arratia and Baxendale (2013)

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

Bounded Coupling implies Concentration Inequality

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 G. (2011), improvement by Arratia and Baxendale (2013)

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

Proof of Upper Tail Bound

For $\theta \geq 0$,

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

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

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

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

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

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 $v \notin \mathcal{V}_v$. 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

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

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 $v \notin \mathcal{V}_v$. 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

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

The sum

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

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

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

Choose $V = v$ proportional to EX_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 over $\mathcal{V}_v$, and let $\mathbf{X}^v$ be the indicators of local maxima on this new configuration. Then Y^s, the number of local maxima on $\mathbf{X}^I$, has the Y-size bias distribution.

Making the value at v larger could not have made more local maxima among $\mathcal{V}_v$, but could have among $\mathcal{V}_w$, so

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

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

Choose $V = v$ proportional to EX_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 over $\mathcal{V}_v$, and let $\mathbf{X}^v$ be the indicators of local maxima on this new configuration. Then Y^s, the number of local maxima on $\mathbf{X}^I$, has the Y-size bias distribution.

Making the value at v larger could not have made more local maxima among $\mathcal{V}_v$, but could have among $\mathcal{V}_w$, so

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

Bounded Difference Inequality

Changing value at single vertex w can at most change number of local maxima f by size of neighborhood $|\mathcal{V}_w|$, so f is a bounded difference function and so satisfies

$$\mathbb{P}(Y - \mu \geq t) \leq \exp\left(-\frac{t^2}{\sum_{w \in \mathcal{V}} |\mathcal{V}_w|^2}\right).$$

If neighborhood sizes are constant c, say, behaves like $\exp(-t^2/c^2 n)$.

Size bias bound has Poisson tails, and can show is smaller than

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

Replaces n by μ. Can have function of n^2 variables, e.g. color edges in a graph, count number of 'monochromatic' vertices.

Bounded Difference Inequality

Changing value at single vertex w can at most change number of local maxima f by size of neighborhood $|\mathcal{V}_w|$, so f is a bounded difference function and so satisfies

$$\mathbb{P}(Y - \mu \geq t) \leq \exp\left(-\frac{t^2}{\sum_{w \in \mathcal{V}} |\mathcal{V}_w|^2}\right).$$

If neighborhood sizes are constant c, say, behaves like $\exp(-t^2/c^2 n)$.
Size bias bound has Poisson tails, and can show is smaller than

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

Replaces n by μ. Can have function of n^2 variables, e.g. color edges in a graph, count number of 'monochromatic' vertices.

Bounded Difference Inequality

Changing value at single vertex w can at most change number of local maxima f by size of neighborhood $|\mathcal{V}_w|$, so f is a bounded difference function and so satisfies

$$\mathbb{P}(Y - \mu \geq t) \leq \exp\left(-\frac{t^2}{\sum_{w \in \mathcal{V}} |\mathcal{V}_w|^2}\right).$$

If neighborhood sizes are constant c, say, behaves like $\exp(-t^2/c^2 n)$.
Size bias bound has Poisson tails, and can show is smaller than

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

Replaces n by μ. Can have function of n^2 variables, e.g. color edges in a graph, count number of 'monochromatic' vertices.

Self Bounding and Configuration Functions

Consider a collection of 'hereditary' sets $\Pi_k \subset \Omega^k, k = 0, \ldots, n$, that is $(x_1, \ldots, x_k) \in \Pi_k$ implies $(x_{i_1}, \ldots, x_{i_j}) \in \Pi_j$ for any $1 \le i_1 < \ldots < i_{i_j} \le k$. Consider the function $f(\mathbf{x})$ 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 at $\mathbf{x}$ 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 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

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

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

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

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

Multivariate Concentration

Let $\mathbf{Y}$ have mean $\boldsymbol{\mu}$ and variances $\boldsymbol{\sigma}^2$. Size bias in direction i,

$$E[Y_i g(\mathbf{Y})] = E[Y_i]E[g(\mathbf{Y}^i)].$$

If $\|\mathbf{Y}^i - \mathbf{Y}\|_2 \leq K$ then (mgf, operations componentwise)

$$P\left(\frac{\mathbf{Y}-\boldsymbol{\mu}}{\boldsymbol{\sigma}} \geq \mathbf{t}\right) \leq \exp\left(-\frac{\|\mathbf{t}\|_2^2}{2(K_1 + K_2\|\mathbf{t}\|_2)}\right),$$

for

$$K_1 = \frac{2K}{\sigma_{(1)}}\|\frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}\|_2 \quad \text{and} \quad K_2 = \frac{K}{2\sigma_{(1)}}.$$

Applications, e.g. counting patterns in permutations, Işlak and Ghosh (2013).

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 $|Y^* - Y| \leq c$ a.s., then (mgf),

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

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 $|Y^* - Y| \leq c$ a.s., then (mgf),

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

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 $|Y^* - Y| \leq c$ a.s., then (mgf),

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

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.

Dependence. G. and Işlak (2014).

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.

Dependence. G. and Işlak (2014).

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.

Dependence. G. and Işlak (2014).

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.

Dependence. G. and Işlak (2014).

Combinatorial CLT, Exchangeable Pair Coupling

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

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

and under this same 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

Let $(\mathbf{Z},\mathbf{Z}')$ be exchangeable, (X,X') a pair of $d \times d$ Hermitian matrices,

$$E[K(\mathbf{Z},\mathbf{Z}')|\mathbf{Z}] = X, \quad K(\mathbf{Z},\mathbf{Z}') = -K(\mathbf{Z}',\mathbf{Z})$$

and

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

If there exists $s>0$ such that

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

then for all $t \geq 0$,

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

Paulin, Mackey and Tropp 2013.

Matrix Concentration Inequalities

Proof by differential inequality for trace moment generating function $m(\theta) = E\mathrm{tr}[e^{\theta X}] \geq E\lambda_{\max}(e^{\theta X})$. Real valued inequality used earlier

$$|e^x - e^y| \leq \frac{1}{2}|x-y|\,|e^x + e^y|$$

replaced by matrix inequality, holding for all Hermitian A, B, C and $s > 0$,

$$\mathrm{tr}[C(e^A - e^B)] \leq \frac{1}{4}\mathrm{tr}\left[(s(A-B)^2 + s^{-1}C^2)(e^A + e^B)\right].$$

Can obtain 'bounded difference' inequality, and handle Dobrushin type dependence.

Summary

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

Many concentration of measure results require independence.

Stein type couplings, posses $E[Wf(W)]$ term for use in Herbst type arguments.

Couplings, like perturbations, can measure departures from independence.

Bounded or otherwise well behaved Stein couplings imply concentration of measure, and central limit behavior.

Summary

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

Many concentration of measure results require independence.

Stein type couplings, posses $E[Wf(W)]$ term for use in Herbst type arguments.

Couplings, like perturbations, can measure departures from independence.

Bounded or otherwise well behaved Stein couplings imply concentration of measure, and central limit behavior.

Summary

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

Many concentration of measure results require independence.

Stein type couplings, posses $E[Wf(W)]$ term for use in Herbst type arguments.

Couplings, like perturbations, can measure departures from independence.

Bounded or otherwise well behaved Stein couplings imply concentration of measure, and central limit behavior.

Summary

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

Many concentration of measure results require independence.

Stein type couplings, posses $E[Wf(W)]$ term for use in Herbst type arguments.

Couplings, like perturbations, can measure departures from independence.

Bounded or otherwise well behaved Stein couplings imply concentration of measure, and central limit behavior.

Summary

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

Many concentration of measure results require independence.

Stein type couplings, posses $E[Wf(W)]$ term for use in Herbst type arguments.

Couplings, like perturbations, can measure departures from independence.

Bounded or otherwise well behaved Stein couplings imply concentration of measure, and central limit behavior.