Concentration of Measure

Sudeep Kamath

PRINCETON UNIVERSITY

CIRM workshop, 25 Jan 2016

What is concentration?

“A random variable that depends in a smooth way on many independent random variables (but not too much on any of them) is essentially constant.”
- M. Talagrand, 1996.

What is concentration?

"A random variable that depends in a smooth way on many independent random variables (but not too much on any of them) is essentially constant."
- M. Talagrand, 1996.

If Z is a function of many independent variables $X_1, X_2, \ldots, X_n$, how large are typical deviations of Z?

What is concentration?

"A random variable that depends in a smooth way on many independent random variables (but not too much on any of them) is essentially constant."
- M. Talagrand, 1996.

If Z is a function of many independent variables $X_1, X_2, \ldots, X_n$, how large are typical deviations of Z?

Goal: Quantify by bounding for $t > 0$,

$$\mathbb{P}\left[|Z - \mathbb{E}Z| \geq t\right]$$

Applications

Concentration of measure has far-reaching consequences in

- Pure and applied probability,
- High-dimensional statistics,
- Functional analysis,
- Computer science,
- Machine learning,
- Statistical physics,
- Complex graphs and networks,
- Information theory, communication and coding theory.

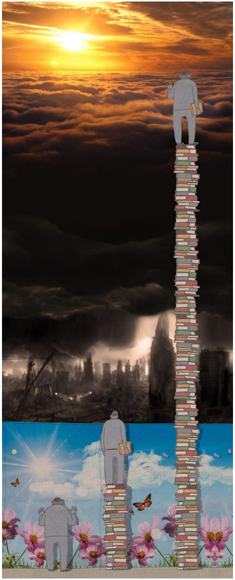
Approaches for Proving Concentration

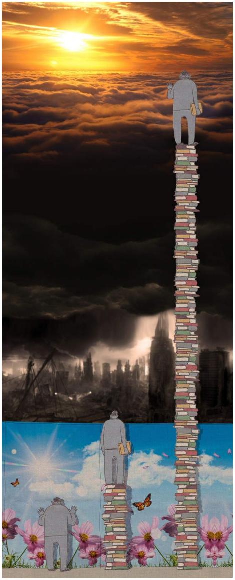
- *The martingale approach*: Hoeffding (1963), Azuma (1967), Milman and Schechtman (1986), Shamir and Spencer (1987) and McDiarmid (1989, 1998), Sipser and Spielman (1996), Richardson and Urbanke (2001)
- *Talagrand's inequalities for product measures*: Talagrand (1996).
- *Entropy method and log-Sobolev inequalities*: Ledoux (1996), Massart (1998), Lugosi et al. (1999, 2001)
- *Transportation method*: Ahlswede, Gács and Körner (1976), Marton (1986, 1996, 1997), Dembo (1997), Villani (2003, 2008)
- *Stein's method of exchangeable pairs*: Chatterjee (2007), Chatterjee and Dey (2010), Goldstein et al. (2011, 2014)

Approaches for Proving Concentration

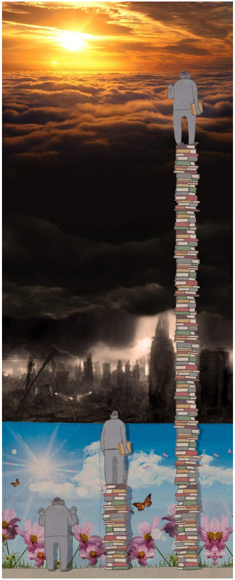
- *The martingale approach*: Hoeffding (1963), Azuma (1967), Milman and Schechtman (1986), Shamir and Spencer (1987) and McDiarmid (1989, 1998), Sipser and Spielman (1996), Richardson and Urbanke (2001)
- *Talagrand's inequalities for product measures*: Talagrand (1996).
- *Entropy method and log-Sobolev inequalities*: Ledoux (1996), Massart (1998), Lugosi et al. (1999, 2001)
- *Transportation method*: Ahlswede, Gács and Körner (1976), Marton (1986, 1996, 1997), Dembo (1997), Villani (2003, 2008)
- *Stein's method of exchangeable pairs*: Chatterjee (2007), Chatterjee and Dey (2010), Goldstein et al. (2011, 2014)

We will focus on the entropy method and transportation method where information theoretic methods shine.

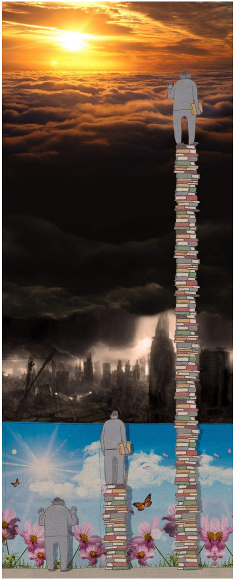




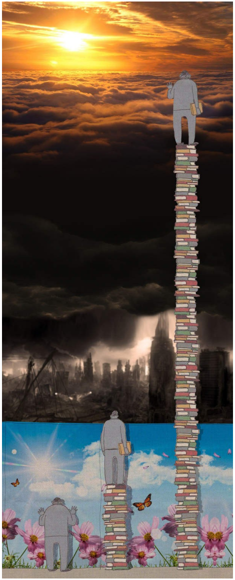
Undergraduate (“informal”)	



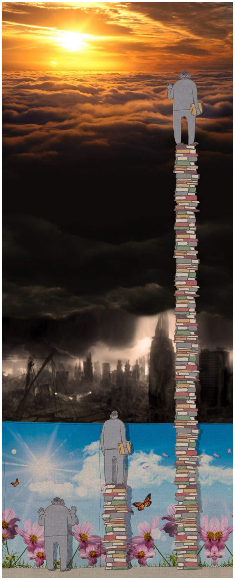
Undergraduate ("informal")	**Graduate** (formal probability)



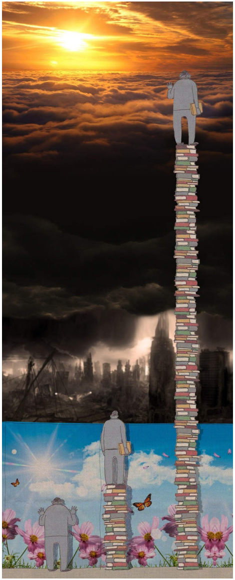
Undergraduate ("informal")	**Graduate** (formal probability)
Probability density,	Radon-Nikodym derivative,



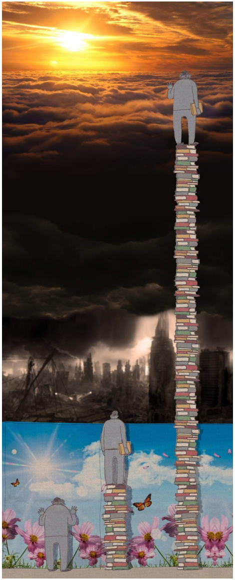
Undergraduate ("informal")	**Graduate** (formal probability)
Probability density, Riemann integral	Radon-Nikodym derivative, Lebesgue integral



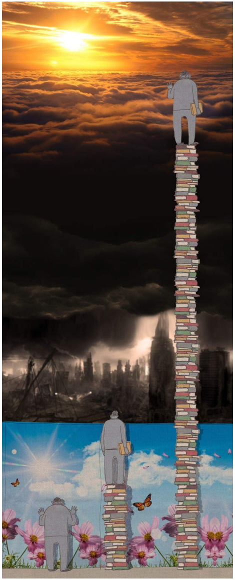
Undergraduate (“informal”)	**Graduate** (formal probability)
Probability density, Riemann integral	Radon-Nikodym derivative, Lebesgue integral
Conditional probability	Regular conditional probability



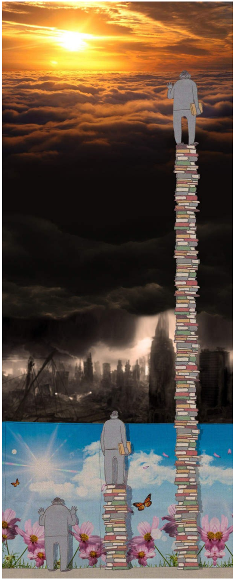
Undergraduate ("informal")	**Graduate** (formal probability)
Probability density, Riemann integral	Radon-Nikodym derivative, Lebesgue integral
Conditional probability	Regular conditional probability
$\mathbb{E}[X\|Y]$ conditioning on random variables	$\mathbb{E}[X\|\sigma(Y)]$ conditioning on σ-fields



Undergraduate ("informal")	**Graduate** (formal probability)
Probability density, Riemann integral	Radon-Nikodym derivative, Lebesgue integral
Conditional probability	Regular conditional probability
$\mathbb{E}[X\|Y]$ conditioning on random variables	$\mathbb{E}[X\|\sigma(Y)]$ conditioning on σ-fields
Convergence in distribution	Also: almost sure, in L^1, in probability



Undergraduate ("informal")	**Graduate** (formal probability)
Probability density, Riemann integral	Radon-Nikodym derivative, Lebesgue integral
Conditional probability	Regular conditional probability
$\mathbb{E}[X\|Y]$ conditioning on random variables	$\mathbb{E}[X\|\sigma(Y)]$ conditioning on σ-fields
Convergence in distribution	Also: almost sure, in L^1, in probability
None	Monotone and dominated convergence theorem



Undergraduate ("informal")	**Graduate** (formal probability)
Probability density, Riemann integral	Radon-Nikodym derivative, Lebesgue integral
Conditional probability	Regular conditional probability
$\mathbb{E}[X\|Y]$ conditioning on random variables	$\mathbb{E}[X\|\sigma(Y)]$ conditioning on σ-fields
Convergence in distribution	Also: almost sure, in L^1, in probability
None	Monotone and dominated convergence theorem
None	Non-measurable subsets of $\mathbb{R}$

Why this tutorial?

Why this tutorial?

- Many results easy to appreciate from the undergraduate view

Why this tutorial?

- Many results easy to appreciate from the undergraduate view
- Non-asymptotic results: easy to use

Why this tutorial?

- Many results easy to appreciate from the undergraduate view
- Non-asymptotic results: easy to use
- Use basic information-theoretic ideas

Why this tutorial?

- Many results easy to appreciate from the undergraduate view
- Non-asymptotic results: easy to use
- Use basic information-theoretic ideas

Role of information theory

"The emphasis put on information theoretic methods is one main feature of the exposition and there is considerable benefit in this approach for a number of fundamental results [...]" - M. Ledoux, foreword to 'Concentration Inequalities' by Boucheron, Lugosi, Massart.

Roadmap

Roadmap

- Monday: Variance bounds

Roadmap

- Monday: Variance bounds
- Tuesday: Information inequalities

Roadmap

- Monday: Variance bounds
- Tuesday: Information inequalities
- Thursday: Entropy method and log-Sobolev inequalities

Roadmap

- Monday: Variance bounds
- Tuesday: Information inequalities
- Thursday: Entropy method and log-Sobolev inequalities
- Friday: Transportation method

Roadmap

- Monday: Variance bounds
- Tuesday: Information inequalities
- Thursday: Entropy method and log-Sobolev inequalities
- Friday: Transportation method

Thanks to Ramon van Handel, Igal Sason, Max Raginsky

Roadmap

- Monday: Variance bounds
- Tuesday: Information inequalities
- Thursday: Entropy method and log-Sobolev inequalities
- Friday: Transportation method

Thanks to Ramon van Handel, Igal Sason, Max Raginsky

Reference: 'Concentration Inequalities' by Boucheron, Lugosi, Massart

Roadmap

- Monday: Variance bounds
- Tuesday: Information inequalities
- Thursday: Entropy method and log-Sobolev inequalities
- Friday: Transportation method

Thanks to Ramon van Handel, Igal Sason, Max Raginsky

Reference: 'Concentration Inequalities' by Boucheron, Lugosi, Massart

Slides available on my homepage:
`http://www.princeton.edu/~sukamath/concentration.pdf`

Variance bounds

Variance bounds

Say Z is a function of independent random variables $X_1, X_2, \ldots, X_n$. An upper bound on $\mathrm{Var}(Z)$ gives tail bounds as:

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq \frac{\mathrm{Var}(Z)}{t^2}$$

Variance bounds

Say Z is a function of independent random variables $X_1, X_2, \ldots, X_n$. An upper bound on $\mathrm{Var}(Z)$ gives tail bounds as:

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq \frac{\mathrm{Var}(Z)}{t^2}$$

Probability of Z being within 10 standard deviations, i.e. $t = 10\sqrt{\mathrm{Var}(Z)}$ of $\mathbb{E}Z$ is at least 99%

Variance bounds

Say Z is a function of independent random variables $X_1, X_2, \ldots, X_n$. An upper bound on $\mathrm{Var}(Z)$ gives tail bounds as:

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq \frac{\mathrm{Var}(Z)}{t^2}$$

Probability of Z being within 10 standard deviations, i.e. $t = 10\sqrt{\mathrm{Var}(Z)}$ of $\mathbb{E}Z$ is at least 99%

Trivial example

Let $Z = X_1 + X_2 + \ldots + X_n$ where $\{X_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.) with finite variance. Then,

$$\mathbb{E}Z = n\mathbb{E}X_1 \qquad \mathrm{Var}(Z) = n\,\mathrm{Var}(X_1)$$

Say Z is a function of independent random variables $X_1, X_2, \ldots, X_n$. An upper bound on $\mathrm{Var}(Z)$ gives tail bounds as:

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq \frac{\mathrm{Var}(Z)}{t^2}$$

Probability of Z being within 10 standard deviations, i.e. $t = 10\sqrt{\mathrm{Var}(Z)}$ of $\mathbb{E}Z$ is at least 99%

Trivial example

Let $Z = X_1 + X_2 + \ldots + X_n$ where $\{X_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.) with finite variance. Then,

$$\mathbb{E}Z = n\mathbb{E}X_1 \qquad \mathrm{Var}(Z) = n\,\mathrm{Var}(X_1)$$

Mean $= \Theta(n)$, Standard Deviation $= O(\sqrt{n})$.

Variance bounds: sharper truths

Variance bounds: sharper truths

Spectral norm of a random matrix

Populate an $m \times n$ matrix A by independent entries, each taking values in $[0, 1]$. The random variable $Z = \|A\|$ satisfies

Variance bounds: sharper truths

Spectral norm of a random matrix

Populate an $m \times n$ matrix A by independent entries, each taking values in $[0, 1]$. The random variable $Z = \|A\|$ satisfies

$$\mathrm{Var}(Z) \leq 1$$

Variance bounds: sharper truths

Spectral norm of a random matrix

Populate an $m \times n$ matrix A by independent entries, each taking values in $[0, 1]$. The random variable $Z = \|A\|$ satisfies

$$\mathrm{Var}(Z) \leq 1$$

Plug-in entropy estimation

Let Z be the estimate of entropy of an unknown distribution defined by the entropy of the empirical distribution from drawing n independent samples.

Variance bounds: sharper truths

Spectral norm of a random matrix

Populate an $m \times n$ matrix A by independent entries, each taking values in $[0, 1]$. The random variable $Z = \|A\|$ satisfies

$$\mathrm{Var}(Z) \leq 1$$

Plug-in entropy estimation

Let Z be the estimate of entropy of an unknown distribution defined by the entropy of the empirical distribution from drawing n independent samples.

$$\mathrm{Var}(Z) \leq \frac{\log^2 n}{n}$$

High-level idea

High-level idea

- Obtain a bound for a function of many random variables by bounds for functions of each individual random variable

High-level idea

- Obtain a bound for a function of many random variables by bounds for functions of each individual random variable
- Not obvious this is possible

High-level idea

- Obtain a bound for a function of many random variables by bounds for functions of each individual random variable
- Not obvious this is possible
- When it is, we say the quantity *tensorizes*

High-level idea

- Obtain a bound for a function of many random variables by bounds for functions of each individual random variable
- Not obvious this is possible
- When it is, we say the quantity *tensorizes*
- Quantities that tensorize behave well in high dimension

High-level idea

- Obtain a bound for a function of many random variables by bounds for functions of each individual random variable
- Not obvious this is possible
- When it is, we say the quantity *tensorizes*
- Quantities that tensorize behave well in high dimension
- Variance is such a quantity!

Tensorization of variance

Tensorization of variance

Let $Z = f(X_1, X_2, \ldots, X_n)$ where $X_1, X_2, \ldots, X_n$ are independent random variables.

Tensorization of variance

Let $Z = f(X_1, X_2, \ldots, X_n)$ where $X_1, X_2, \ldots, X_n$ are independent random variables.

$X^{(i)} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$

Tensorization of variance

Let $Z = f(X_1, X_2, \ldots, X_n)$ where $X_1, X_2, \ldots, X_n$ are independent random variables.

$$X^{(i)} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n), \qquad \mathbb{E}^{(i)}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|X^{(i)}]$$

Tensorization of variance

Let $Z = f(X_1, X_2, \ldots, X_n)$ where $X_1, X_2, \ldots, X_n$ are independent random variables.

$$X^{(i)} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n), \qquad \mathbb{E}^{(i)}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|X^{(i)}]$$

$$\mathrm{Var}^{(i)}(Z) := \mathrm{Var}(Z|X^{(i)})$$

Tensorization of variance

Let $Z = f(X_1, X_2, \ldots, X_n)$ where $X_1, X_2, \ldots, X_n$ are independent random variables.

$$X^{(i)} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n), \qquad \mathbb{E}^{(i)}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|X^{(i)}]$$

$$\mathrm{Var}^{(i)}(Z) := \mathrm{Var}(Z|X^{(i)})$$

$$g_i(x^{(i)}) = \mathrm{Var}(f(x_1, \ldots, X_i, \ldots, x_n))$$

Tensorization of variance

Let $Z = f(X_1, X_2, \ldots, X_n)$ where $X_1, X_2, \ldots, X_n$ are independent random variables.

$$X^{(i)} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n), \qquad \mathbb{E}^{(i)}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|X^{(i)}]$$

$$\mathrm{Var}^{(i)}(Z) := \mathrm{Var}(Z|X^{(i)})$$

$$g_i(x^{(i)}) = \mathrm{Var}(f(x_1, \ldots, X_i, \ldots, x_n)) \implies \mathrm{Var}^{(i)}(Z) = g_i(X^{(i)})$$

Tensorization of variance

Let $Z = f(X_1, X_2, \ldots, X_n)$ where $X_1, X_2, \ldots, X_n$ are independent random variables.

$$X^{(i)} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n), \qquad \mathbb{E}^{(i)}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|X^{(i)}]$$

$$\mathrm{Var}^{(i)}(Z) := \mathrm{Var}(Z|X^{(i)})$$

$$g_i(x^{(i)}) = \mathrm{Var}(f(x_1, \ldots, X_i, \ldots, x_n)) \implies \mathrm{Var}^{(i)}(Z) = g_i(X^{(i)})$$

Tensorization of variance (Efron-Stein-Steele inequality)

$$\mathrm{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}[\mathrm{Var}^{(i)}(Z)]$$

Tensorization of variance

Let $Z = f(X_1, X_2, \ldots, X_n)$ where $X_1, X_2, \ldots, X_n$ are independent random variables.

$$X^{(i)} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n), \qquad \mathbb{E}^{(i)}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|X^{(i)}]$$

$$\mathrm{Var}^{(i)}(Z) := \mathrm{Var}(Z|X^{(i)})$$

$$g_i(x^{(i)}) = \mathrm{Var}(f(x_1, \ldots, X_i, \ldots, x_n)) \implies \mathrm{Var}^{(i)}(Z) = g_i(X^{(i)})$$

Tensorization of variance (Efron-Stein-Steele inequality)

$$\mathrm{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}[\mathrm{Var}^{(i)}(Z)] = \sum_{i=1}^{n} \mathbb{E}[(Z - \mathbb{E}^{(i)} Z)^2]$$

Recall: if $Y \in [a, b]$, then $\mathrm{Var}(Y) \leq$

Recall: if $Y \in [a, b]$, then $\mathrm{Var}(Y) \leq \inf_u \mathbb{E}[(Y-u)^2]$

Recall: if $Y \in [a,b]$, then $\mathrm{Var}(Y) \leq \inf_u \mathbb{E}[(Y-u)^2] \leq \frac{(b-a)^2}{4}$

from $u = \frac{1}{2}(a+b)$.

Recall: if $Y \in [a, b]$, then $\mathrm{Var}(Y) \leq \inf_u \mathbb{E}[(Y-u)^2] \leq \frac{(b-a)^2}{4}$

from $u = \frac{1}{2}(a+b)$.

Simplest application: Bounded differences inequality

Suppose

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \leq c_i.$$

Recall: if $Y \in [a, b]$, then $\mathrm{Var}(Y) \leq \inf_u \mathbb{E}[(Y-u)^2] \leq \frac{(b-a)^2}{4}$

from $u = \frac{1}{2}(a+b)$.

Simplest application: Bounded differences inequality

Suppose

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \leq c_i.$$

Then, $\mathrm{Var}(f(X)) \leq \sum_{i=1}^{n} \mathbb{E}[\mathrm{Var}^{(i)}(Z)] \leq \frac{1}{4} \sum_{i=1}^{n} c_i^2$

Recall: if $Y \in [a,b]$, then $\operatorname{Var}(Y) \le \inf_u \mathbb{E}[(Y-u)^2] \le \frac{(b-a)^2}{4}$

from $u = \frac{1}{2}(a+b)$.

Simplest application: Bounded differences inequality

Suppose

$$|f(x_1,\dots,x_i,\dots,x_n) - f(x_1,\dots,x_i',\dots,x_n)| \le c_i.$$

Then, $\operatorname{Var}(f(X)) \le \sum_{i=1}^n \mathbb{E}[\operatorname{Var}^{(i)}(Z)] \le \frac{1}{4}\sum_{i=1}^n c_i^2$

Tight if $f(X) = \sum_{i=1}^n X_i$ with X_i equiprobable on $\{-1,+1\}$

Bin packing problem

Bin packing problem

Let $X_1, X_2, \ldots, X_n \in [0, 1]$ be i.i.d.

Bin packing problem

Let $X_1, X_2, \ldots, X_n \in [0, 1]$ be i.i.d.

Let Z be the minimum number of bins in to which they can be packed so that each bin adds to at most 1.

Bin packing problem

Let $X_1, X_2, \ldots, X_n \in [0,1]$ be i.i.d.

Let Z be the minimum number of bins in to which they can be packed so that each bin adds to at most 1.

Changing one X_i changes $Z = f(X)$ by at most $c_i = 1$

Bin packing problem

Let $X_1, X_2, \ldots, X_n \in [0,1]$ be i.i.d.

Let Z be the minimum number of bins in to which they can be packed so that each bin adds to at most 1.

Changing one X_i changes $Z = f(X)$ by at most $c_i = 1$

Therefore, $\mathrm{Var}(Z) \leq \frac{1}{4}\sum_{i=1}^{n} c_i^2 = \frac{n}{4}$.

Bin packing problem

Let $X_1, X_2, \ldots, X_n \in [0,1]$ be i.i.d.

Let Z be the minimum number of bins in to which they can be packed so that each bin adds to at most 1.

Changing one X_i changes $Z = f(X)$ by at most $c_i = 1$

Therefore, $\mathrm{Var}(Z) \leq \frac{1}{4} \sum_{i=1}^{n} c_i^2 = \frac{n}{4}$.

However, $\mathbb{E}Z \geq \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mathbb{E}X_1$.

Bin packing problem

Let $X_1, X_2, \ldots, X_n \in [0,1]$ be i.i.d.

Let Z be the minimum number of bins in to which they can be packed so that each bin adds to at most 1.

Changing one X_i changes $Z = f(X)$ by at most $c_i = 1$

$$\text{Therefore, } \operatorname{Var}(Z) \leq \frac{1}{4}\sum_{i=1}^{n} c_i^2 = \frac{n}{4}\ .$$

$$\text{However, } \mathbb{E}Z \geq \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mathbb{E}X_1.$$

Standard deviation $= O(\sqrt{n})$, Mean $= \Theta(n)$.

Plug-in entropy estimation

Plug-in entropy estimation

Entropy of a distribution $p = (p_1, p_2, \ldots, p_k)$ is defined as

$$H(p) = \sum_{r=1}^{k} p_r \log \frac{1}{p_r}$$

Plug-in entropy estimation

Entropy of a distribution $p = (p_1, p_2, \ldots, p_k)$ is defined as

$$H(p) = \sum_{r=1}^{k} p_r \log \frac{1}{p_r}$$

Let $X_1, X_2, \ldots, X_n$ be independent samples from p

Plug-in entropy estimation

Entropy of a distribution $p = (p_1, p_2, \ldots, p_k)$ is defined as

$$H(p) = \sum_{r=1}^{k} p_r \log \frac{1}{p_r}$$

Let $X_1, X_2, \ldots, X_n$ be independent samples from p

Let $\hat{p}_r = \frac{1}{n}|\{i : X_i = r\}|$

Plug-in entropy estimation

Entropy of a distribution $p = (p_1, p_2, \ldots, p_k)$ is defined as

$$H(p) = \sum_{r=1}^{k} p_r \log \frac{1}{p_r}$$

Let $X_1, X_2, \ldots, X_n$ be independent samples from p

Let $\hat{p}_r = \frac{1}{n}|\{i : X_i = r\}|, \qquad Z = \sum_{r=1}^{k} \hat{p}_r \log \frac{1}{\hat{p}_r}$

Plug-in entropy estimation

Entropy of a distribution $p = (p_1, p_2, \ldots, p_k)$ is defined as

$$H(p) = \sum_{r=1}^{k} p_r \log \frac{1}{p_r}$$

Let $X_1, X_2, \ldots, X_n$ be independent samples from p

Let $\hat{p}_r = \frac{1}{n}|\{i : X_i = r\}|, \qquad Z = \sum_{r=1}^{k} \hat{p}_r \log \frac{1}{\hat{p}_r}$

A change in any one co-ordinate X_i affects two of the $\hat{p}_r$'s.

Plug-in entropy estimation

Entropy of a distribution $p = (p_1, p_2, \ldots, p_k)$ is defined as

$$H(p) = \sum_{r=1}^{k} p_r \log \frac{1}{p_r}$$

Let $X_1, X_2, \ldots, X_n$ be independent samples from p

Let $\hat{p}_r = \frac{1}{n}|\{i : X_i = r\}|, \qquad Z = \sum_{r=1}^{k} \hat{p}_r \log \frac{1}{\hat{p}_r}$

A change in any one co-ordinate X_i affects two of the $\hat{p}_r$'s.

$$\left| a \log \frac{1}{a} - b \log \frac{1}{b} \right| \leq \frac{\log n}{n} \text{ if } |a - b| = \frac{1}{n}.$$

Plug-in entropy estimation

Entropy of a distribution $p = (p_1, p_2, \ldots, p_k)$ is defined as

$$H(p) = \sum_{r=1}^{k} p_r \log \frac{1}{p_r}$$

Let $X_1, X_2, \ldots, X_n$ be independent samples from p

Let $\hat{p}_r = \frac{1}{n}|\{i : X_i = r\}|, \qquad Z = \sum_{r=1}^{k} \hat{p}_r \log \frac{1}{\hat{p}_r}$

A change in any one co-ordinate X_i affects two of the $\hat{p}_r$'s.

$$\left| a \log \frac{1}{a} - b \log \frac{1}{b} \right| \leq \frac{\log n}{n} \text{ if } |a - b| = \frac{1}{n}.$$

Thus, $\mathrm{Var}(Z) \leq \sum_{i=1}^{n} c_i^2/4 = (\log^2 n)/n$

Plug-in entropy estimation

Entropy of a distribution $p = (p_1, p_2, \ldots, p_k)$ is defined as

$$H(p) = \sum_{r=1}^{k} p_r \log \frac{1}{p_r}$$

Let $X_1, X_2, \ldots, X_n$ be independent samples from p

Let $\hat{p}_r = \frac{1}{n}|\{i : X_i = r\}|, \qquad Z = \sum_{r=1}^{k} \hat{p}_r \log \frac{1}{\hat{p}_r}$

A change in any one co-ordinate X_i affects two of the $\hat{p}_r$'s.

$$\left| a \log \frac{1}{a} - b \log \frac{1}{b} \right| \le \frac{\log n}{n} \text{ if } |a - b| = \frac{1}{n}.$$

Thus, $\mathrm{Var}(Z) \le \sum_{i=1}^{n} c_i^2/4 = (\log^2 n)/n$

But Z is not really concentrated at $H(p)$ unless $n \gtrsim k$.

Plug-in entropy estimation

Entropy of a distribution $p = (p_1, p_2, \ldots, p_k)$ is defined as

$$H(p) = \sum_{r=1}^{k} p_r \log \frac{1}{p_r}$$

Let $X_1, X_2, \ldots, X_n$ be independent samples from p

Let $\hat{p}_r = \frac{1}{n}|\{i : X_i = r\}|, \qquad Z = \sum_{r=1}^{k} \hat{p}_r \log \frac{1}{\hat{p}_r}$

A change in any one co-ordinate X_i affects two of the $\hat{p}_r$'s.

$$\left| a \log \frac{1}{a} - b \log \frac{1}{b} \right| \leq \frac{\log n}{n} \text{ if } |a - b| = \frac{1}{n}.$$

Thus, $\mathrm{Var}(Z) \leq \sum_{i=1}^{n} c_i^2/4 = (\log^2 n)/n$

But Z is not really concentrated at $H(p)$ unless $n \gtrsim k$.
For $n << k$, Z is concentrated but somewhere else!

Warning

Warning

- We have shown bounds on deviation of Z from $\mathbb{E}Z$

Warning

- We have shown bounds on deviation of Z from $\mathbb{E}Z$
- But say nothing about $\mathbb{E}Z$ itself!!

Warning

- We have shown bounds on deviation of Z from $\mathbb{E}Z$
- But say nothing about $\mathbb{E}Z$ itself!!
- Estimating magnitude and fluctuations are two quite distinct problems

Warning

- We have shown bounds on deviation of Z from $\mathbb{E}Z$
- But say nothing about $\mathbb{E}Z$ itself!!
- Estimating magnitude and fluctuations are two quite distinct problems
- We have a general theorem for bounding fluctuations and elementary ideas can often bound sensitivity to coordinates, even if the function itself is complicated

Warning

- We have shown bounds on deviation of Z from $\mathbb{E}Z$
- But say nothing about $\mathbb{E}Z$ itself!!
- Estimating magnitude and fluctuations are two quite distinct problems
- We have a general theorem for bounding fluctuations and elementary ideas can often bound sensitivity to coordinates, even if the function itself is complicated
- No such general principle for estimating $\mathbb{E}Z$

Warning

- We have shown bounds on deviation of Z from $\mathbb{E}Z$
- But say nothing about $\mathbb{E}Z$ itself!!
- Estimating magnitude and fluctuations are two quite distinct problems
- We have a general theorem for bounding fluctuations and elementary ideas can often bound sensitivity to coordinates, even if the function itself is complicated
- No such general principle for estimating $\mathbb{E}Z$
- Can estimate $\mathbb{E}Z$ from Monte Carlo methods if Z is concentrated

Recall that $\mathrm{Var}(Z) = \inf_u \mathbb{E}[(Z-u)^2]$

Recall that $\mathrm{Var}(Z) = \inf_u \mathbb{E}[(Z-u)^2]$

So, $\mathrm{Var}^{(i)}(Z) = \inf_{f_i(x^{(i)})} \mathbb{E}^{(i)}[(Z - f_i(X^{(i)}))^2]$

Recall that $\mathrm{Var}(Z) = \inf_u \mathbb{E}[(Z-u)^2]$

So, $\mathrm{Var}^{(i)}(Z) = \inf_{f_i(x^{(i)})} \mathbb{E}^{(i)}[(Z - f_i(X^{(i)}))^2]$

Let $Z_i = f_i(X^{(i)})$ for any function f_i.

Recall that $\mathrm{Var}(Z) = \inf_u \mathbb{E}[(Z-u)^2]$

So, $\mathrm{Var}^{(i)}(Z) = \inf_{f_i(x^{(i)})} \mathbb{E}^{(i)}[(Z - f_i(X^{(i)}))^2]$

Let $Z_i = f_i(X^{(i)})$ for any function f_i.

Then, $\mathrm{Var}^{(i)}(Z) \leq \mathbb{E}^{(i)}[(Z - Z_i)^2]$

Recall that $\mathrm{Var}(Z) = \inf_u \mathbb{E}[(Z-u)^2]$

So, $\mathrm{Var}^{(i)}(Z) = \inf_{f_i(x^{(i)})} \mathbb{E}^{(i)}[(Z - f_i(X^{(i)}))^2]$

Let $Z_i = f_i(X^{(i)})$ for any function f_i.

Then, $\mathrm{Var}^{(i)}(Z) \leq \mathbb{E}^{(i)}[(Z - Z_i)^2]$

Variant : “guess functions”

$$\mathrm{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}[\mathrm{Var}^{(i)}(Z)] \leq \sum_{i=1}^{n} \mathbb{E}[(Z - Z_i)^2]$$

Convex Lipschitz functions

Suppose $f : [a, b]^n \mapsto \mathbb{R}$ is convex, differentiable and L-Lipschitz, i.e. $|f(x) - f(y)| \leq L\|x - y\|_2 \; \forall \; x, y$.

Convex Lipschitz functions

Suppose $f : [a,b]^n \mapsto \mathbb{R}$ is convex, differentiable and L-Lipschitz, i.e. $|f(x) - f(y)| \leq L\|x-y\|_2 \ \forall \ x, y$.

Choose $Z_i = \inf_{x_i} f(X_1, X_2, \ldots, x_i, \ldots, X_n)$ with $\inf$ attained at X_i'

Convex Lipschitz functions

Suppose $f : [a,b]^n \mapsto \mathbb{R}$ is convex, differentiable and L-Lipschitz, i.e. $|f(x) - f(y)| \leq L\|x - y\|_2 \;\forall\; x, y$.

Choose $Z_i = \inf_{x_i} f(X_1, X_2, \ldots, x_i, \ldots, X_n)$ with $\inf$ attained at X_i'

$$Z_i \geq Z + \frac{\partial f}{\partial x_i}(X)(X_i' - X_i)$$

Convex Lipschitz functions

Suppose $f : [a,b]^n \mapsto \mathbb{R}$ is convex, differentiable and L-Lipschitz, i.e. $|f(x) - f(y)| \leq L\|x - y\|_2 \ \forall \ x, y$.

Choose $Z_i = \inf_{x_i} f(X_1, X_2, \ldots, x_i, \ldots, X_n)$ with $\inf$ attained at X_i'

$$Z_i \geq Z + \frac{\partial f}{\partial x_i}(X)(X_i' - X_i) \qquad\qquad 0 \leq Z - Z_i \leq -\frac{\partial f}{\partial x_i}(X)(X_i' - X_i)$$

Convex Lipschitz functions

Suppose $f : [a,b]^n \mapsto \mathbb{R}$ is convex, differentiable and L-Lipschitz, i.e. $|f(x) - f(y)| \leq L\|x-y\|_2 \ \forall \ x, y$.

Choose $Z_i = \inf_{x_i} f(X_1, X_2, \ldots, x_i, \ldots, X_n)$ with $\inf$ attained at X_i'

$$Z_i \geq Z + \frac{\partial f}{\partial x_i}(X)(X_i' - X_i) \qquad 0 \leq Z - Z_i \leq -\frac{\partial f}{\partial x_i}(X)(X_i' - X_i)$$

$$0 \leq (Z - Z_i)^2 \leq \left|\frac{\partial f}{\partial x_i}(X)\right|^2 (b-a)^2$$

Convex Lipschitz functions

Suppose $f : [a,b]^n \mapsto \mathbb{R}$ is convex, differentiable and L-Lipschitz, i.e. $|f(x) - f(y)| \leq L\|x - y\|_2 \; \forall \; x, y$.

Choose $Z_i = \inf_{x_i} f(X_1, X_2, \ldots, x_i, \ldots, X_n)$ with $\inf$ attained at X_i'

$$Z_i \geq Z + \frac{\partial f}{\partial x_i}(X)(X_i' - X_i) \qquad\qquad 0 \leq Z - Z_i \leq -\frac{\partial f}{\partial x_i}(X)(X_i' - X_i)$$

$$0 \leq (Z - Z_i)^2 \leq \left|\frac{\partial f}{\partial x_i}(X)\right|^2 (b-a)^2$$

$$\mathrm{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}[(Z - Z_i)^2] \leq L^2(b-a)^2$$

Convex Lipschitz functions

Suppose $f : [a, b]^n \mapsto \mathbb{R}$ is convex, differentiable and L-Lipschitz, i.e. $|f(x) - f(y)| \leq L\|x - y\|_2 \; \forall \; x, y$.

Choose $Z_i = \inf_{x_i} f(X_1, X_2, \dots, x_i, \dots, X_n)$ with $\inf$ attained at X_i'

$$Z_i \geq Z + \frac{\partial f}{\partial x_i}(X)(X_i' - X_i) \qquad 0 \leq Z - Z_i \leq -\frac{\partial f}{\partial x_i}(X)(X_i' - X_i)$$

$$0 \leq (Z - Z_i)^2 \leq \left|\frac{\partial f}{\partial x_i}(X)\right|^2 (b - a)^2$$

$$\mathrm{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}[(Z - Z_i)^2] \leq L^2(b - a)^2$$

Differentiability assumption unnecessary: convolve f with a smooth kernel.

Spectral norm of a random matrix

Populate an $m \times n$ matrix A by independent entries, each taking values in $[0, 1]$.

Spectral norm of a random matrix

Populate an $m \times n$ matrix A by independent entries, each taking values in $[0, 1]$.

The function $f : [0, 1]^{m \times n} \mapsto \mathbb{R}$, given by

$$Z = f(A) = \|A\| = \sup_{\|v\|_2 = 1} \|Av\|_2$$

Spectral norm of a random matrix

Populate an $m \times n$ matrix A by independent entries, each taking values in $[0, 1]$.

The function $f : [0, 1]^{m \times n} \mapsto \mathbb{R}$, given by

$$Z = f(A) = \|A\| = \sup_{\|v\|_2 = 1} \|Av\|_2$$

is convex and 1-Lipschitz

Spectral norm of a random matrix

Populate an $m \times n$ matrix A by independent entries, each taking values in $[0, 1]$.

The function $f : [0, 1]^{m \times n} \mapsto \mathbb{R}$, given by

$$Z = f(A) = \|A\| = \sup_{\|v\|_2 = 1} \|Av\|_2$$

is convex and 1-Lipschitz
(hint: spectral norm $\leq$ Frobenius norm)

Spectral norm of a random matrix

Populate an $m \times n$ matrix A by independent entries, each taking values in $[0, 1]$.

The function $f : [0, 1]^{m \times n} \mapsto \mathbb{R}$, given by

$$Z = f(A) = \|A\| = \sup_{\|v\|_2=1} \|Av\|_2$$

is convex and 1-Lipschitz
(hint: spectral norm $\leq$ Frobenius norm)

With $a = 0, b = 1, L = 1$ in previous result, we get

$$\mathrm{Var}(Z) \leq 1.$$