Concentration of Measure

Sudeep Kamath

PRINCETON UNIVERSITY

CIRM workshop, 28 Jan 2016

Review

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 $X_1, X_2, \ldots, X_n$, under what conditions can we say typical deviations of Z are small?

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 $X_1, X_2, \ldots, X_n$, under what conditions can we say typical deviations of Z are small?

- Low variance captured by a general theorem: *tensorization of variance*

Review

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 $X_1, X_2, \ldots, X_n$, under what conditions can we say typical deviations of Z are small?

- Low variance captured by a general theorem: *tensorization of variance*
- Today, we'll see the entropy method, a general tool to show sub-Gaussian tails

Entropy

If P is a probability measure on Ω and $Z : \Omega \mapsto \mathbb{R}_{\geq 0}$, we define $\mathrm{Ent}(Z) := \mathbb{E}[Z \log Z] - (\mathbb{E}Z) \log \mathbb{E}Z$

Entropy

If P is a probability measure on Ω and $Z : \Omega \mapsto \mathbb{R}_{\geq 0}$, we define $\mathrm{Ent}(Z) := \mathbb{E}[Z \log Z] - (\mathbb{E}Z) \log \mathbb{E}Z$

$$\frac{\mathrm{Ent}(Z)}{\mathbb{E}Z} = D\left(P_{\frac{Z}{\mathbb{E}Z}} || P\right)$$

(where if $Q = P_Y$, then $Q(A) = \mathbb{E}[Y 1_A]$, $\mathbb{E}_Q[W] = \mathbb{E}[YW]$)

Entropy

If P is a probability measure on Ω and $Z : \Omega \mapsto \mathbb{R}_{\geq 0}$, we define $\mathrm{Ent}(Z) := \mathbb{E}[Z \log Z] - (\mathbb{E}Z) \log \mathbb{E}Z$

$$\frac{\mathrm{Ent}(Z)}{\mathbb{E}Z} = D\left(P_{\frac{Z}{\mathbb{E}Z}} || P\right)$$

(where if $Q = P_Y$, then $Q(A) = \mathbb{E}[Y 1_A], \;\; \mathbb{E}_Q[W] = \mathbb{E}[YW]$)

Tensorization of entropy

Entropy

If P is a probability measure on Ω and $Z : \Omega \mapsto \mathbb{R}_{\geq 0}$, we define $\mathrm{Ent}(Z) := \mathbb{E}[Z \log Z] - (\mathbb{E}Z) \log \mathbb{E}Z$

$$\frac{\mathrm{Ent}(Z)}{\mathbb{E}Z} = D\left(P_{\frac{Z}{\mathbb{E}Z}} || P\right)$$

(where if $Q = P_Y$, then $Q(A) = \mathbb{E}[Y 1_A], \;\; \mathbb{E}_Q[W] = \mathbb{E}[YW]$)

Tensorization of entropy

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

Entropy

If P is a probability measure on Ω and $Z : \Omega \mapsto \mathbb{R}_{\geq 0}$, we define $\mathrm{Ent}(Z) := \mathbb{E}[Z \log Z] - (\mathbb{E}Z) \log \mathbb{E}Z$

$$\frac{\mathrm{Ent}(Z)}{\mathbb{E}Z} = D\left(P_{\frac{Z}{\mathbb{E}Z}} || P\right)$$

(where if $Q = P_Y$, then $Q(A) = \mathbb{E}[Y 1_A], \ \ \mathbb{E}_Q[W] = \mathbb{E}[YW]$)

Tensorization of entropy

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

$$X^{(i)} := (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n), \mathrm{Ent}^{(i)}(Z) := \mathrm{Ent}(Z|X^{(i)})$$

Entropy

If P is a probability measure on Ω and $Z : \Omega \mapsto \mathbb{R}_{\geq 0}$, we define $\mathrm{Ent}(Z) := \mathbb{E}[Z \log Z] - (\mathbb{E}Z) \log \mathbb{E}Z$

$$\frac{\mathrm{Ent}(Z)}{\mathbb{E}Z} = D\left(P_{\frac{Z}{\mathbb{E}Z}} || P\right)$$

(where if $Q = P_Y$, then $Q(A) = \mathbb{E}[Y 1_A], \;\; \mathbb{E}_Q[W] = \mathbb{E}[YW]$)

Tensorization of entropy

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

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

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

Log Moment Generating Functions (log m.g.f.)

With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

Log Moment Generating Functions (log m.g.f.)

With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

- $\psi'(\lambda) = \dfrac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]}$

Log Moment Generating Functions (log m.g.f.)

With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

- $\psi'(\lambda) = \frac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]} = \mathbb{E}_Q[Z]$

Log Moment Generating Functions (log m.g.f.)

With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

- $\psi'(\lambda) = \dfrac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]} = \mathbb{E}_Q[Z]$ where $Q = P_{e^{\lambda Z}/\mathbb{E}e^{\lambda Z}}$

Log Moment Generating Functions (log m.g.f.)

With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

- $\psi'(\lambda) = \dfrac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]} = \mathbb{E}_Q[Z]$ where $Q = P_{e^{\lambda Z}/\mathbb{E}e^{\lambda Z}}$
- $\psi'(0) = \mathbb{E}Z = 0$

Log Moment Generating Functions (log m.g.f.)

With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

- $\psi'(\lambda) = \dfrac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]} = \mathbb{E}_Q[Z]$ where $Q = P_{e^{\lambda Z}/\mathbb{E}e^{\lambda Z}}$
- $\psi'(0) = \mathbb{E}Z = 0$
- $\psi''(\lambda) = \dfrac{\mathbb{E}[e^{\lambda Z}]\mathbb{E}[Z^2 e^{\lambda Z}] - (\mathbb{E}[Ze^{\lambda Z}])^2}{(\mathbb{E}[e^{\lambda Z}])^2}$

Log Moment Generating Functions (log m.g.f.)

With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

- $\psi'(\lambda) = \dfrac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]} = \mathbb{E}_Q[Z]$ where $Q = P_{e^{\lambda Z}/\mathbb{E}e^{\lambda Z}}$
- $\psi'(0) = \mathbb{E}Z = 0$
- $\psi''(\lambda) = \dfrac{\mathbb{E}[e^{\lambda Z}]\mathbb{E}[Z^2 e^{\lambda Z}] - (\mathbb{E}[Ze^{\lambda Z}])^2}{(\mathbb{E}[e^{\lambda Z}])^2}$

  $= \mathbb{E}_Q[Z^2] - (\mathbb{E}_Q[Z])^2$

Log Moment Generating Functions (log m.g.f.)

With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

- $\psi'(\lambda) = \dfrac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]} = \mathbb{E}_Q[Z]$ where $Q = P_{e^{\lambda Z}/\mathbb{E}e^{\lambda Z}}$
- $\psi'(0) = \mathbb{E}Z = 0$
- $\psi''(\lambda) = \dfrac{\mathbb{E}[e^{\lambda Z}]\mathbb{E}[Z^2 e^{\lambda Z}] - (\mathbb{E}[Ze^{\lambda Z}])^2}{(\mathbb{E}[e^{\lambda Z}])^2}$

  $= \mathbb{E}_Q[Z^2] - (\mathbb{E}_Q[Z])^2 = \mathrm{Var}_Q(Z)$

Log Moment Generating Functions (log m.g.f.)

With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

- $\psi'(\lambda) = \dfrac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]} = \mathbb{E}_Q[Z]$ where $Q = P_{e^{\lambda Z}/\mathbb{E}e^{\lambda Z}}$
- $\psi'(0) = \mathbb{E}Z = 0$
- $\psi''(\lambda) = \dfrac{\mathbb{E}[e^{\lambda Z}]\mathbb{E}[Z^2 e^{\lambda Z}] - (\mathbb{E}[Ze^{\lambda Z}])^2}{(\mathbb{E}[e^{\lambda Z}])^2}$

  $= \mathbb{E}_Q[Z^2] - (\mathbb{E}_Q[Z])^2 = \mathrm{Var}_Q(Z)$

  Therefore, $\psi(\cdot)$ is convex!

Log Moment Generating Functions (log m.g.f.)

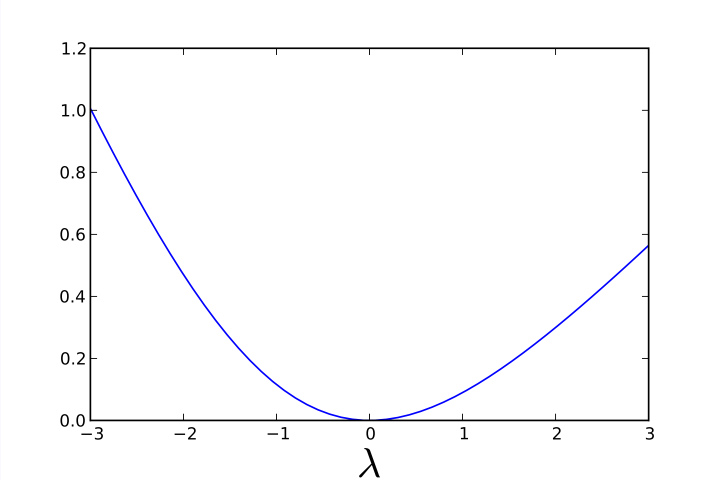
With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

- $\psi'(\lambda) = \dfrac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]} = \mathbb{E}_Q[Z]$ where $Q = P_{e^{\lambda Z}/\mathbb{E}e^{\lambda Z}}$
- $\psi'(0) = \mathbb{E}Z = 0$
- $\psi''(\lambda) = \dfrac{\mathbb{E}[e^{\lambda Z}]\mathbb{E}[Z^2 e^{\lambda Z}] - (\mathbb{E}[Ze^{\lambda Z}])^2}{(\mathbb{E}[e^{\lambda Z}])^2}$

  $= \mathbb{E}_Q[Z^2] - (\mathbb{E}_Q[Z])^2 = \mathrm{Var}_Q(Z)$

  Therefore, $\psi(\cdot)$ is convex!
- $\psi''(0) = \mathrm{Var}(Z)$

Log Moment Generating Functions (log m.g.f.)



Log Moment Generating Functions (log m.g.f.)

With $\mathbb{E}Z = 0$, let $\psi(\lambda) := \log \mathbb{E}e^{\lambda Z}, \quad \lambda \in \mathbb{R}$

- $\psi'(\lambda) = \dfrac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]} = \mathbb{E}_Q[Z]$ where $Q = P_{e^{\lambda Z}/\mathbb{E}e^{\lambda Z}}$
- $\psi'(0) = \mathbb{E}Z = 0$
- $\psi''(\lambda) = \dfrac{\mathbb{E}[e^{\lambda Z}]\mathbb{E}[Z^2 e^{\lambda Z}] - (\mathbb{E}[Ze^{\lambda Z}])^2}{(\mathbb{E}[e^{\lambda Z}])^2}$

  $= \mathbb{E}_Q[Z^2] - (\mathbb{E}_Q[Z])^2 = \mathrm{Var}_Q(Z)$

  Therefore, $\psi(\cdot)$ is convex!
- $\psi''(0) = \mathrm{Var}(Z)$

Herbst argument

$$\text{If } \frac{\mathrm{Ent}(e^{\lambda Y})}{\mathbb{E}[e^{\lambda Y}]} \leq \frac{\lambda^2\sigma^2}{2} \ \forall\ \lambda > 0,$$

$$\text{then } \log \mathbb{E}[e^{\lambda(Y-\mathbb{E}Y)}] \leq \frac{\lambda^2\sigma^2}{2} \ \forall\ \lambda > 0$$

Herbst condition equivalent to sub-Gaussianity

$$\text{If } \log \mathbb{E}[e^{\lambda(Y-\mathbb{E}Y)}] \leq \frac{\lambda^2\sigma^2}{2} \ \forall\ \lambda > 0,$$

$$\text{then } \frac{\mathrm{Ent}(e^{\lambda Y})}{\mathbb{E}[e^{\lambda Y}]} \leq \frac{\lambda^2(4\sigma^2)}{2} \ \forall\ \lambda > 0.$$

Herbst argument

$$\text{If } \frac{\mathrm{Ent}(e^{\lambda Y})}{\mathbb{E}[e^{\lambda Y}]} \leq \frac{\lambda^2\sigma^2}{2} \ \forall\ \lambda > 0,$$

$$\text{then } \log \mathbb{E}[e^{\lambda(Y-\mathbb{E}Y)}] \leq \frac{\lambda^2\sigma^2}{2} \ \forall\ \lambda > 0$$

Herbst condition equivalent to sub-Gaussianity

$$\text{If } \log \mathbb{E}[e^{\lambda(Y-\mathbb{E}Y)}] \leq \frac{\lambda^2\sigma^2}{2} \ \forall\ \lambda > 0,$$

$$\text{then } \frac{\mathrm{Ent}(e^{\lambda Y})}{\mathbb{E}[e^{\lambda Y}]} \leq \frac{\lambda^2(4\sigma^2)}{2} \ \forall\ \lambda > 0.$$

Note: $\frac{\mathrm{Ent}(e^{\lambda Y})}{\mathbb{E}[e^{\lambda Y}]} = D\left(P_{e^{\lambda Y}/\mathbb{E}[e^{\lambda Y}]} || P\right)$. (Proof on board)

Hoeffding's lemma

Hoeffding's lemma

If $Z \in [a, b]$, then Z is $\big((b-a)^2/4\big)$-sub-Gaussian,

i.e. $\psi(\lambda) = \log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \le \frac{\lambda^2}{2} \frac{(b-a)^2}{4} \;\; \forall \;\; \lambda \in \mathbb{R}$

Hoeffding's lemma

If $Z \in [a,b]$, then Z is $\big((b-a)^2/4\big)$-sub-Gaussian,

i.e. $\psi(\lambda) = \log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \le \dfrac{\lambda^2}{2} \dfrac{(b-a)^2}{4} \;\; \forall \;\; \lambda \in \mathbb{R}$

Proof 1

$$\psi''(\lambda)$$

Hoeffding's lemma

If $Z \in [a, b]$, then Z is $\big((b-a)^2/4\big)$-sub-Gaussian,

i.e. $\psi(\lambda) = \log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{\lambda^2}{2} \frac{(b-a)^2}{4} \quad \forall \quad \lambda \in \mathbb{R}$

Proof 1

$$\psi''(\lambda) = \mathrm{Var}_Q(Z)$$

Hoeffding's lemma

If $Z \in [a, b]$, then Z is $\big((b-a)^2/4\big)$-sub-Gaussian,

$$\text{i.e.} \quad \psi(\lambda) = \log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{\lambda^2}{2} \frac{(b-a)^2}{4} \quad \forall \quad \lambda \in \mathbb{R}$$

Proof 1

$$\psi''(\lambda) = \mathrm{Var}_Q(Z) \leq (b-a)^2/4$$

Hoeffding's lemma

If $Z \in [a, b]$, then Z is $\big((b-a)^2/4\big)$-sub-Gaussian,

i.e. $\psi(\lambda) = \log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \dfrac{\lambda^2}{2} \dfrac{(b-a)^2}{4} \;\; \forall \;\; \lambda \in \mathbb{R}$

Proof 1

$$\psi''(\lambda) = \mathrm{Var}_Q(Z) \leq (b-a)^2/4$$

Proof 2

$$\frac{\mathrm{Ent}(e^{\lambda Y})}{\mathbb{E}[e^{\lambda Y}]} = \lambda\psi'(\lambda) - \lambda = \int_0^\lambda \theta\psi''(\theta)\, d\theta \leq \frac{\lambda^2}{2} \frac{(b-a)^2}{4}$$

Then, Herbst argument.

Bounded differences inequality

Bounded differences inequality

Bounded differences inequality (McDiarmid's inequality)

Suppose

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

If $X = (X_1, X_2, \ldots, X_n)$ has independent components, then $Z = f(X_1, X_2, \ldots, X_n)$ is σ^2- sub-Gaussian with

$$\sigma^2 = \frac{1}{4} \sum_{i=1}^{n} c_i^2.$$

Therefore, $\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{\lambda^2 \sigma^2}{2} \ \forall\ \lambda \in \mathbb{R}$ and

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq 2e^{-\frac{t^2}{2\sigma^2}} \ \forall t > 0$$

Bounded differences inequality: proof

Bounded differences inequality: proof

Proof of bounded differences inequality

This can be proved by the entropy method as follows.

$$
\begin{aligned}
\mathrm{Ent}(e^{\lambda Z}) &\le \sum_{i=1}^{n} \mathbb{E}[\mathrm{Ent}^{(i)}(e^{\lambda Z})] \\
&\le \sum_{i=1}^{n} \mathbb{E}\left[\frac{\lambda^2}{2}\frac{c_i^2}{4}\mathbb{E}^{(i)}[e^{\lambda Z}]\right] \\
&= \frac{\lambda^2}{2}\left(\sum_{i=1}^{n}\frac{c_i^2}{4}\right)\mathbb{E}[e^{\lambda Z}]
\end{aligned}
$$

Herbst argument completes the proof.

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 $c_i = 1$

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

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

But more: $\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq 2e^{-2t^2/n}$

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.

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

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

But more: $\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq 2e^{-t^2 \log^2 n/(2n)}$

$x^{(i)}$-dependent bounded differences

Suppose

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \leq c_i(x^{(i)}).$$

If $X = (X_1, X_2, \ldots, X_n)$ has independent components, then $Z = f(X_1, X_2, \ldots, X_n)$ is σ^2- sub-Gaussian with

$$\sigma^2 = \frac{1}{4} \sup_x \left(\sum_{i=1}^{n} c_i^2(x^{(i)}) \right).$$

Therefore, $\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \dfrac{\lambda^2 \sigma^2}{2} \ \forall \ \lambda \in \mathbb{R}$ and

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq 2e^{-\frac{t^2}{2\sigma^2}} \ \forall t > 0$$

Proof of $x^{(i)}$-dependent bounded differences inequality

$$
\begin{aligned}
\mathrm{Ent}(e^{\lambda Z}) &\leq \sum_{i=1}^{n} \mathbb{E}[\mathrm{Ent}^{(i)}(e^{\lambda Z})] \\
&\leq \sum_{i=1}^{n} \mathbb{E}\left[\frac{\lambda^2}{2}\frac{c_i^2(X^{(i)})}{4}\mathbb{E}^{(i)}[e^{\lambda Z}]\right] \\
&\leq \sum_{i=1}^{n} \mathbb{E}\left[\frac{\lambda^2}{2}\mathbb{E}^{(i)}\left[\frac{c_i^2(X^{(i)})}{4}e^{\lambda Z}\right]\right] \\
&\leq \frac{\lambda^2}{2}\sigma^2\mathbb{E}[e^{\lambda Z}]
\end{aligned}
$$

Herbst argument completes the proof.

Proof of $x^{(i)}$-dependent bounded differences inequality

$$
\begin{aligned}
\mathrm{Ent}(e^{\lambda Z}) &\le \sum_{i=1}^{n} \mathbb{E}[\mathrm{Ent}^{(i)}(e^{\lambda Z})] \\
&\le \sum_{i=1}^{n} \mathbb{E}\left[\frac{\lambda^2}{2}\frac{c_i^2(X^{(i)})}{4}\mathbb{E}^{(i)}[e^{\lambda Z}]\right] \\
&\le \sum_{i=1}^{n} \mathbb{E}\left[\frac{\lambda^2}{2}\mathbb{E}^{(i)}\left[\frac{c_i^2(X^{(i)})}{4}e^{\lambda Z}\right]\right] \\
&\le \frac{\lambda^2}{2}\sigma^2\mathbb{E}[e^{\lambda Z}]
\end{aligned}
$$

Herbst argument completes the proof.

Proof does not work if $c_i(x^{(i)})$ is replaced by $c_i(x)$.

Proof of $x^{(i)}$-dependent bounded differences inequality

$$
\begin{aligned}
\operatorname{Ent}(e^{\lambda Z}) &\leq \sum_{i=1}^{n} \mathbb{E}[\operatorname{Ent}^{(i)}(e^{\lambda Z})] \\
&\leq \sum_{i=1}^{n} \mathbb{E}\left[\frac{\lambda^2}{2}\frac{c_i^2(X^{(i)})}{4}\mathbb{E}^{(i)}[e^{\lambda Z}]\right] \\
&\leq \sum_{i=1}^{n} \mathbb{E}\left[\frac{\lambda^2}{2}\mathbb{E}^{(i)}\left[\frac{c_i^2(X^{(i)})}{4}e^{\lambda Z}\right]\right] \\
&\leq \frac{\lambda^2}{2}\sigma^2\mathbb{E}[e^{\lambda Z}]
\end{aligned}
$$

Herbst argument completes the proof.

Proof does not work if $c_i(x^{(i)})$ is replaced by $c_i(x)$.

We will see later that the transportation method works in this case.

Logarithmic Sobolev inequalities

Logarithmic Sobolev inequalities

- A log-Sobolev inequality is of the form

  $$\text{“entropy}(f^2) \lesssim\ c\, \mathbb{E}[\|\text{gradient}(f)\|^2]\text{”}$$

Logarithmic Sobolev inequalities

- A log-Sobolev inequality is of the form

$$\text{“entropy}(f^2) \lesssim\ c\,\mathbb{E}[\|\text{gradient}(f)\|^2]\text{”}$$

- Such inequalities are closely associated with mixing in Markov processes

Logarithmic Sobolev inequalities

- A log-Sobolev inequality is of the form

$$\text{“entropy}(f^2) \lesssim \; c\,\mathbb{E}[\|\text{gradient}(f)\|^2]\text{”}$$

- Such inequalities are closely associated with mixing in Markov processes

Bernoulli log-Sobolev inequality

Let $X = (X_1, X_2, \ldots, X_n)$ have independent components equiprobable on $\{-1, +1\}$. Let $\tilde{X}^{(i)}$ be the vector X whose i^{th} co-ordinate is re-sampled independently.

Logarithmic Sobolev inequalities

- A log-Sobolev inequality is of the form

$$\text{“entropy}(f^2) \lesssim \; c\,\mathbb{E}[\|\text{gradient}(f)\|^2]\text{”}$$

- Such inequalities are closely associated with mixing in Markov processes

Bernoulli log-Sobolev inequality

Let $X = (X_1, X_2, \ldots, X_n)$ have independent components equiprobable on $\{-1, +1\}$. Let $\tilde{X}^{(i)}$ be the vector X whose i^{th} co-ordinate is re-sampled independently.
Then, for any $f : \{-1, +1\}^n \mapsto \mathbb{R}$, we have

Logarithmic Sobolev inequalities

- A log-Sobolev inequality is of the form

$$\text{“entropy}(f^2) \lesssim \; c\, \mathbb{E}[\|\text{gradient}(f)\|^2]\text{”}$$

- Such inequalities are closely associated with mixing in Markov processes

Bernoulli log-Sobolev inequality

Let $X = (X_1, X_2, \ldots, X_n)$ have independent components equiprobable on $\{-1, +1\}$. Let $\tilde{X}^{(i)}$ be the vector X whose i^{th} co-ordinate is re-sampled independently.
Then, for any $f : \{-1, +1\}^n \mapsto \mathbb{R}$, we have

$$\mathrm{Ent}(f^2) \leq 2\mathscr{E}(f),$$

Logarithmic Sobolev inequalities

- A log-Sobolev inequality is of the form

$$\text{“entropy}(f^2) \lesssim\ c\, \mathbb{E}[\|\text{gradient}(f)\|^2]\text{”}$$

- Such inequalities are closely associated with mixing in Markov processes

Bernoulli log-Sobolev inequality

Let $X = (X_1, X_2, \ldots, X_n)$ have independent components equiprobable on $\{-1, +1\}$. Let $\tilde{X}^{(i)}$ be the vector X whose i^{th} co-ordinate is re-sampled independently.
Then, for any $f : \{-1, +1\}^n \mapsto \mathbb{R}$, we have

$$\mathrm{Ent}(f^2) \leq 2\mathscr{E}(f),$$

where $\mathscr{E}(f) = \dfrac{1}{2}\mathbb{E}\left[\sum_{i=1}^{n}(f(X) - f(\tilde{X}^{(i)}))^2\right].$

- Tensorization of variance says $\mathrm{Var}(g) \leq \mathscr{E}(g)$

- Tensorization of variance says $\mathrm{Var}(g) \leq \mathscr{E}(g)$
- Log-Sobolev inequality says $\mathrm{Ent}(f^2) \leq 2\mathscr{E}(f)$

- Tensorization of variance says $\mathrm{Var}(g) \leq \mathscr{E}(g)$
- Log-Sobolev inequality says $\mathrm{Ent}(f^2) \leq 2\mathscr{E}(f)$
- Former follows from the latter by the choice $f = 1 + \epsilon g$ and driving ϵ to zero.

- Tensorization of variance says $\mathrm{Var}(g) \leq \mathscr{E}(g)$
- Log-Sobolev inequality says $\mathrm{Ent}(f^2) \leq 2\mathscr{E}(f)$
- Former follows from the latter by the choice $f = 1 + \epsilon g$ and driving ϵ to zero.

Proof of Bernoulli log-Sobolev inequality

$n = 1$ case: Calculus exercise

$$\frac{a^2}{2}\log a^2 + \frac{b^2}{2}\log b^2 - \frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2} \leq \frac{1}{2}(a-b)^2$$

- Tensorization of variance says $\mathrm{Var}(g) \leq \mathscr{E}(g)$
- Log-Sobolev inequality says $\mathrm{Ent}(f^2) \leq 2\mathscr{E}(f)$
- Former follows from the latter by the choice $f = 1 + \epsilon g$ and driving ϵ to zero.

Proof of Bernoulli log-Sobolev inequality

$n = 1$ case: Calculus exercise

$$\frac{a^2}{2}\log a^2 + \frac{b^2}{2}\log b^2 - \frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2} \leq \frac{1}{2}(a-b)^2$$

$n > 1$: Tensorization of entropy

- Tensorization of variance says $\mathrm{Var}(g) \le \mathscr{E}(g)$
- Log-Sobolev inequality says $\mathrm{Ent}(f^2) \le 2\mathscr{E}(f)$
- Former follows from the latter by the choice $f = 1 + \epsilon g$ and driving ϵ to zero.

Proof of Bernoulli log-Sobolev inequality

$n = 1$ case: Calculus exercise

$$\frac{a^2}{2}\log a^2 + \frac{b^2}{2}\log b^2 - \frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2} \le \frac{1}{2}(a-b)^2$$

$n > 1$: Tensorization of entropy

By replacing f with $e^{\lambda f/2}$ and using Herbst argument, we can show concentration inequalities on the Boolean hypercube

Gaussian log-Sobolev inequality

Gaussian log-Sobolev inequality

For $f : \{-1, +1\}^n \mapsto \mathbb{R}$,
$$\mathrm{Var}(f) \leq \mathscr{E}(f)$$

Gaussian log-Sobolev inequality

For $f : \{-1, +1\}^n \mapsto \mathbb{R}$,
$\mathrm{Var}(f) \leq \mathscr{E}(f)$

implies

Gaussian log-Sobolev inequality

For $f : \{-1, +1\}^n \mapsto \mathbb{R}$,
$$\mathrm{Var}(f) \leq \mathscr{E}(f)$$

implies

Gaussian Poincaré

For $f : \mathbb{R}^n \mapsto \mathbb{R}$ differentiable and $X \sim \mathscr{N}(0, I_n)$, we have
$$\mathrm{Var}(f(X)) \leq \mathbb{E}\left[\|\nabla f(X)\|^2\right]$$

Gaussian log-Sobolev inequality

For $f : \{-1,+1\}^n \mapsto \mathbb{R}$,
$\mathrm{Var}(f) \leq \mathscr{E}(f)$

For $f : \{-1,+1\}^n \mapsto \mathbb{R}$,
$\mathrm{Ent}(f^2) \leq 2\mathscr{E}(f)$

implies

Gaussian Poincaré

For $f : \mathbb{R}^n \mapsto \mathbb{R}$
differentiable and
$X \sim \mathscr{N}(0, I_n)$, we have
$\mathrm{Var}(f(X)) \leq \mathbb{E}\left[\|\nabla f(X)\|^2\right]$

Gaussian log-Sobolev inequality

For $f : \{-1,+1\}^n \mapsto \mathbb{R}$,
$\mathrm{Var}(f) \leq \mathscr{E}(f)$

implies

Gaussian Poincaré

For $f : \mathbb{R}^n \mapsto \mathbb{R}$ differentiable and $X \sim \mathcal{N}(0, I_n)$, we have
$\mathrm{Var}(f(X)) \leq \mathbb{E}\left[\|\nabla f(X)\|^2\right]$

For $f : \{-1,+1\}^n \mapsto \mathbb{R}$,
$\mathrm{Ent}(f^2) \leq 2\mathscr{E}(f)$

implies

Gaussian log-Sobolev inequality

For $f : \{-1,+1\}^n \mapsto \mathbb{R}$,
$\mathrm{Var}(f) \leq \mathscr{E}(f)$

implies

Gaussian Poincaré

For $f : \mathbb{R}^n \mapsto \mathbb{R}$ differentiable and $X \sim \mathscr{N}(0, I_n)$, we have
$\mathrm{Var}(f(X)) \leq \mathbb{E}\left[\|\nabla f(X)\|^2\right]$

For $f : \{-1,+1\}^n \mapsto \mathbb{R}$,
$\mathrm{Ent}(f^2) \leq 2\mathscr{E}(f)$

implies

Gaussian log-Sobolev

For $f : \mathbb{R}^n \mapsto \mathbb{R}$ differentiable and $X \sim \mathscr{N}(0, I_n)$, we have
$\mathrm{Ent}(f(X)^2) \leq 2\mathbb{E}\left[\|\nabla f(X)\|^2\right]$

Gaussian concentration

Tsirelson-Ibragimov-Sudakov inequality

If $X \sim \mathscr{N}(0, I_n)$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is L-Lipschitz

Tsirelson-Ibragimov-Sudakov inequality

If $X \sim \mathscr{N}(0, I_n)$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is L-Lipschitz

i.e. $|f(x) - f(y)| \leq L\|x - y\|_2 \ \forall \ x, y,$

Tsirelson-Ibragimov-Sudakov inequality

If $X \sim \mathcal{N}(0, I_n)$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is L-Lipschitz

i.e. $|f(x) - f(y)| \leq L\|x - y\|_2 \; \forall \; x, y$,

then $Z = f(X)$ is L^2-sub-Gaussian.

Tsirelson-Ibragimov-Sudakov inequality

If $X \sim \mathcal{N}(0, I_n)$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is L-Lipschitz

i.e. $|f(x) - f(y)| \leq L\|x - y\|_2 \ \forall\ x, y$,

then $Z = f(X)$ is L^2-sub-Gaussian.

$$\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{\lambda^2 L^2}{2}, \quad \lambda \in \mathbb{R}$$

Tsirelson-Ibragimov-Sudakov inequality

If $X \sim \mathcal{N}(0, I_n)$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is L-Lipschitz

i.e. $|f(x) - f(y)| \leq L\|x - y\|_2 \; \forall \; x, y$,

then $Z = f(X)$ is L^2-sub-Gaussian.

$$\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{\lambda^2 L^2}{2}, \quad \lambda \in \mathbb{R}$$

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq 2e^{-t^2/(2L^2)}, \; \forall t > 0$$

Tsirelson-Ibragimov-Sudakov inequality

If $X \sim \mathcal{N}(0, I_n)$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is L-Lipschitz

i.e. $|f(x) - f(y)| \leq L\|x - y\|_2 \ \forall \ x, y$,

then $Z = f(X)$ is L^2-sub-Gaussian.

$$\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{\lambda^2 L^2}{2}, \quad \lambda \in \mathbb{R}$$

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq 2e^{-t^2/(2L^2)}, \ \forall t > 0$$

Note: tight if f is linear.

Proof of Gaussian concentration

Suppose f is differentiable and apply the Gaussian log-Sobolev inequality to $e^{\lambda f/2}$.

$$\mathrm{Ent}((e^{\lambda f/2})^2) \leq 2\mathbb{E}\left[\|\nabla e^{\lambda f/2}\|^2\right]$$

Proof of Gaussian concentration

Suppose f is differentiable and apply the Gaussian log-Sobolev inequality to $e^{\lambda f/2}$.

$$\mathrm{Ent}((e^{\lambda f/2})^2) \leq 2\mathbb{E}\left[\|\nabla e^{\lambda f/2}\|^2\right]$$
$$\mathrm{Ent}(e^{\lambda f}) \leq \frac{\lambda^2}{2}\mathbb{E}\left[e^{\lambda f}\|\nabla f\|^2\right]$$

Proof of Gaussian concentration

Suppose f is differentiable and apply the Gaussian log-Sobolev inequality to $e^{\lambda f/2}$.

$$\begin{aligned}
\mathrm{Ent}((e^{\lambda f/2})^2) &\leq 2\mathbb{E}\left[\|\nabla e^{\lambda f/2}\|^2\right] \\
\mathrm{Ent}(e^{\lambda f}) &\leq \frac{\lambda^2}{2}\mathbb{E}\left[e^{\lambda f}\|\nabla f\|^2\right] \\
&= \frac{\lambda^2 L^2}{2}\mathbb{E}\left[e^{\lambda f}\right]
\end{aligned}$$

Proof of Gaussian concentration

Suppose f is differentiable and apply the Gaussian log-Sobolev inequality to $e^{\lambda f/2}$.

$$\begin{aligned}
\mathrm{Ent}((e^{\lambda f/2})^2) &\leq 2\mathbb{E}\left[\|\nabla e^{\lambda f/2}\|^2\right] \\
\mathrm{Ent}(e^{\lambda f}) &\leq \frac{\lambda^2}{2}\mathbb{E}\left[e^{\lambda f}\|\nabla f\|^2\right] \\
&= \frac{\lambda^2 L^2}{2}\mathbb{E}\left[e^{\lambda f}\right]
\end{aligned}$$

Herbst argument completes the proof.

Proof of Gaussian concentration

Suppose f is differentiable and apply the Gaussian log-Sobolev inequality to $e^{\lambda f/2}$.

$$
\begin{aligned}
\mathrm{Ent}((e^{\lambda f/2})^2) &\leq 2\mathbb{E}\left[\|\nabla e^{\lambda f/2}\|^2\right] \\
\mathrm{Ent}(e^{\lambda f}) &\leq \frac{\lambda^2}{2}\mathbb{E}\left[e^{\lambda f}\|\nabla f\|^2\right] \\
&= \frac{\lambda^2 L^2}{2}\mathbb{E}\left[e^{\lambda f}\right]
\end{aligned}
$$

Herbst argument completes the proof.

If not differentiable, convolve f with a smooth kernel.

Application: Johnson-Lindenstrauss

Application: Johnson-Lindenstrauss

Johnson-Lindenstrauss lemma

Let $y_1, y_2, \ldots, y_n$ be n points in $\mathbb{R}^n$. For every $0 < \epsilon < 1$, there is a linear map $T : \mathbb{R}^n \mapsto \mathbb{R}^k$ where $k \geq \dfrac{24 \log n}{\epsilon^2}$ so that for every $1 \leq i, j \leq n$,

$$(1-\epsilon)\|Ty_i - Ty_j\| \leq \|y_i - y_j\| \leq (1+\epsilon)\|Ty_i - Ty_j\|$$

Let T have entries i.i.d. $\mathcal{N}(0, 1/k)$.

Application: Johnson-Lindenstrauss

Johnson-Lindenstrauss lemma

Let $y_1, y_2, \ldots, y_n$ be n points in $\mathbb{R}^n$. For every $0 < \epsilon < 1$, there is a linear map $T : \mathbb{R}^n \mapsto \mathbb{R}^k$ where $k \geq \dfrac{24 \log n}{\epsilon^2}$ so that for every $1 \leq i, j \leq n$,

$$(1-\epsilon)\|Ty_i - Ty_j\| \leq \|y_i - y_j\| \leq (1+\epsilon)\|Ty_i - Ty_j\|$$

Let T have entries i.i.d. $\mathcal{N}(0, 1/k)$. Fix $y \in \mathbb{R}^n$. The function $f(T) = \|Ty\|$ is L-Lipschitz with $L = \|y\|$.

Application: Johnson-Lindenstrauss

Johnson-Lindenstrauss lemma

Let $y_1, y_2, \ldots, y_n$ be n points in $\mathbb{R}^n$. For every $0 < \epsilon < 1$, there is a linear map $T : \mathbb{R}^n \mapsto \mathbb{R}^k$ where $k \geq \dfrac{24 \log n}{\epsilon^2}$ so that for every $1 \leq i, j \leq n$,

$$(1-\epsilon)\|Ty_i - Ty_j\| \leq \|y_i - y_j\| \leq (1+\epsilon)\|Ty_i - Ty_j\|$$

Let T have entries i.i.d. $\mathcal{N}(0, 1/k)$. Fix $y \in \mathbb{R}^n$. The function $f(T) = \|Ty\|$ is L-Lipschitz with $L = \|y\|$.

By Gaussian concentration, we have

$$\mathbb{P}[|\|Ty\| - \mathbb{E}\|Ty\|| \geq \epsilon\|y\|] \leq 2e^{-k\epsilon^2/2}$$

Application: Johnson-Lindenstrauss

Johnson-Lindenstrauss lemma

Let $y_1, y_2, \ldots, y_n$ be n points in $\mathbb{R}^n$. For every $0 < \epsilon < 1$, there is a linear map $T : \mathbb{R}^n \mapsto \mathbb{R}^k$ where $k \geq \dfrac{24 \log n}{\epsilon^2}$ so that for every $1 \leq i, j \leq n$,

$$(1-\epsilon)\|Ty_i - Ty_j\| \leq \|y_i - y_j\| \leq (1+\epsilon)\|Ty_i - Ty_j\|$$

Let T have entries i.i.d. $\mathcal{N}(0, 1/k)$. Fix $y \in \mathbb{R}^n$. The function $f(T) = \|Ty\|$ is L-Lipschitz with $L = \|y\|$.

By Gaussian concentration, we have

$$\mathbb{P}[|\|Ty\| - \mathbb{E}\|Ty\|| \geq \epsilon\|y\|] \leq 2e^{-k\epsilon^2/2}$$

A typical problem with using concentration results is that there is no general theory to describe where concentration happens.

Application: Johnson-Lindenstrauss

Johnson-Lindenstrauss lemma

Let $y_1, y_2, \ldots, y_n$ be n points in $\mathbb{R}^n$. For every $0 < \epsilon < 1$, there is a linear map $T : \mathbb{R}^n \mapsto \mathbb{R}^k$ where $k \geq \dfrac{24 \log n}{\epsilon^2}$ so that for every $1 \leq i, j \leq n$,

$$(1-\epsilon)\|Ty_i - Ty_j\| \leq \|y_i - y_j\| \leq (1+\epsilon)\|Ty_i - Ty_j\|$$

Let T have entries i.i.d. $\mathscr{N}(0, 1/k)$. Fix $y \in \mathbb{R}^n$. The function $f(T) = \|Ty\|$ is L-Lipschitz with $L = \|y\|$.

By Gaussian concentration, we have

$$\mathbb{P}[|\|Ty\| - \mathbb{E}\|Ty\|| \geq \epsilon\|y\|] \leq 2e^{-k\epsilon^2/2}$$

A typical problem with using concentration results is that there is no general theory to describe where concentration happens.

$\mathbb{E}\|Ty\| \leq \sqrt{\mathbb{E}\left[\|Ty\|^2\right]}$ and $\mathrm{Var}(\|Ty\|) = \mathbb{E}\left[\|Ty\|^2\right] - (\mathbb{E}\|Ty\|)^2 \leq L^2$

Application: Johnson-Lindenstrauss

Johnson-Lindenstrauss lemma

Let $y_1, y_2, \ldots, y_n$ be n points in $\mathbb{R}^n$. For every $0 < \epsilon < 1$, there is a linear map $T : \mathbb{R}^n \mapsto \mathbb{R}^k$ where $k \geq \dfrac{24 \log n}{\epsilon^2}$ so that for every $1 \leq i, j \leq n$,

$$(1-\epsilon)\|Ty_i - Ty_j\| \leq \|y_i - y_j\| \leq (1+\epsilon)\|Ty_i - Ty_j\|$$

Let T have entries i.i.d. $\mathcal{N}(0, 1/k)$. Fix $y \in \mathbb{R}^n$. The function $f(T) = \|Ty\|$ is L-Lipschitz with $L = \|y\|$.

By Gaussian concentration, we have

$$\mathbb{P}[|\|Ty\| - \mathbb{E}\|Ty\|| \geq \epsilon\|y\|] \leq 2e^{-k\epsilon^2/2}$$

A typical problem with using concentration results is that there is no general theory to describe where concentration happens.

$\mathbb{E}\|Ty\| \leq \sqrt{\mathbb{E}\left[\|Ty\|^2\right]}$ and $\mathrm{Var}(\|Ty\|) = \mathbb{E}\left[\|Ty\|^2\right] - (\mathbb{E}\|Ty\|)^2 \leq L^2$

Then, apply union bound

Summary

- Entropy method: a general tool to show sub-Gaussian tails for functions of many independent random variables
- It is especially powerful in conjunction with log-Sobolev inequalities
- We'll see another general tool, the transportation method next time