Concentration of Measure

Sudeep Kamath

PRINCETON UNIVERSITY

CIRM workshop, 29 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.

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*
- We saw how the entropy method and log-Sobolev inequalities showed sub-Gaussian tails

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*
- We saw how the entropy method and log-Sobolev inequalities showed sub-Gaussian tails
- Today, we'll study the transportation method which uses a beautiful idea of coupling

Transportation lemma

Let P be a probability measure on Ω and $Z : \Omega \mapsto \mathbb{R}$ be any random variable.

The following are equivalent:

$$\log \mathbb{E} e^{\lambda(Z-\mathbb{E}Z)} \leq \frac{\lambda^2\sigma^2}{2}, \ \ \forall\, \lambda > 0$$

$$\mathbb{E}_Q Z - \mathbb{E}Z \leq \sqrt{2\sigma^2 D(Q||P)}, \ \ \forall\, Q << P$$

Proof on board.

Transportation lemma

Let P be a probability measure on Ω and $Z : \Omega \mapsto \mathbb{R}$ be any random variable.

The following are equivalent:

$$\log \mathbb{E} e^{\lambda(Z-\mathbb{E}Z)} \leq \frac{\lambda^2\sigma^2}{2}, \ \ \forall \, \lambda > 0$$

$$\mathbb{E}_Q Z - \mathbb{E}Z \leq \sqrt{2\sigma^2 D(Q||P)}, \ \ \forall \, Q << P$$

Proof on board.

Coupling

Given two probability distributions P and Q on Ω, a coupling M of P and Q is a probability distribution on $\Omega \times \Omega$ whose marginals are respectively P and Q.

Let $\mathscr{C}(P, Q)$ be the set of all couplings of P and Q

If $d_{\mathrm{TV}}(P,Q) = \sup_A |P(A) - Q(A)|$, then

$$d_{\mathrm{TV}}(P,Q)^2 \leq \frac{1}{2} D(Q||P).$$

If $d_{\mathrm{TV}}(P, Q) = \sup_A |P(A) - Q(A)|$, then

$$d_{\mathrm{TV}}(P, Q)^2 \leq \frac{1}{2} D(Q||P).$$

This is the Csiszár-Kemperman-Kullback-Pinsker inequality and follows from Hoeffding's lemma and the transportation inequality.

If $d_{\mathrm{TV}}(P,Q) = \sup_A |P(A) - Q(A)|$, then

$$d_{\mathrm{TV}}(P,Q)^2 \leq \frac{1}{2} D(Q||P).$$

This is the Csiszár-Kemperman-Kullback-Pinsker inequality and follows from Hoeffding's lemma and the transportation inequality.

This approach also shows that the constant $\frac{1}{2}$ is tight.

If $d_{\mathrm{TV}}(P, Q) = \sup_A |P(A) - Q(A)|$, then

$$d_{\mathrm{TV}}(P, Q)^2 \leq \frac{1}{2} D(Q||P).$$

This is the Csiszár-Kemperman-Kullback-Pinsker inequality and follows from Hoeffding's lemma and the transportation inequality.

This approach also shows that the constant $\frac{1}{2}$ is tight.

If $d_{\mathrm{TV}}(P,Q) = \sup_A |P(A) - Q(A)|$, then

$$d_{\mathrm{TV}}(P,Q)^2 \leq \frac{1}{2} D(Q||P).$$

This is the Csiszár-Kemperman-Kullback-Pinsker inequality and follows from Hoeffding's lemma and the transportation inequality.

This approach also shows that the constant $\frac{1}{2}$ is tight.

Let $\sigma_p^2 = \dfrac{1-2p}{2\log\frac{1-p}{p}}$ for $p \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$.

If $d_{\mathrm{TV}}(P,Q) = \sup_A |P(A) - Q(A)|$, then

$$d_{\mathrm{TV}}(P,Q)^2 \le \frac{1}{2} D(Q||P).$$

This is the Csiszár-Kemperman-Kullback-Pinsker inequality and follows from Hoeffding's lemma and the transportation inequality.

This approach also shows that the constant $\dfrac{1}{2}$ is tight.

Let $\sigma_p^2 = \dfrac{1-2p}{2\log\frac{1-p}{p}}$ for $p \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$.

Kearns-Saul '98: If $Z \in \{0,1\}$ and $Z \sim \mathrm{Ber}(p)$, then

$$\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \le \frac{\lambda^2 \sigma_p^2}{2} \ \forall\ \lambda \in \mathbb{R}$$

If $d_{\mathrm{TV}}(P,Q) = \sup_A |P(A) - Q(A)|$, then

$$d_{\mathrm{TV}}(P,Q)^2 \leq \frac{1}{2} D(Q||P).$$

This is the Csiszár-Kemperman-Kullback-Pinsker inequality and follows from Hoeffding's lemma and the transportation inequality.

This approach also shows that the constant $\frac{1}{2}$ is tight.

Let $\sigma_p^2 = \dfrac{1-2p}{2\log\frac{1-p}{p}}$ for $p \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$.

Kearns-Saul '98: If $Z \in \{0,1\}$ and $Z \sim \mathrm{Ber}(p)$, then

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

Ordentlich-Weinberger '05: If $P = \mathrm{Ber}(p), Q = \mathrm{Ber}(q)$, then

$$d_{\mathrm{TV}}(Q,P)^2 \leq 2\sigma_p^2 D(Q||P) \ \forall\ q, 0 \leq q \leq 1$$

If $d_{\mathrm{TV}}(P,Q) = \sup_A |P(A) - Q(A)|$, then

$$d_{\mathrm{TV}}(P,Q)^2 \le \frac{1}{2} D(Q||P).$$

This is the Csiszár-Kemperman-Kullback-Pinsker inequality and follows from Hoeffding's lemma and the transportation inequality.

This approach also shows that the constant $\frac{1}{2}$ is tight.

Let $\sigma_p^2 = \dfrac{1-2p}{2\log\frac{1-p}{p}}$ for $p \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$.

Kearns-Saul '98: If $Z \in \{0,1\}$ and $Z \sim \mathrm{Ber}(p)$, then

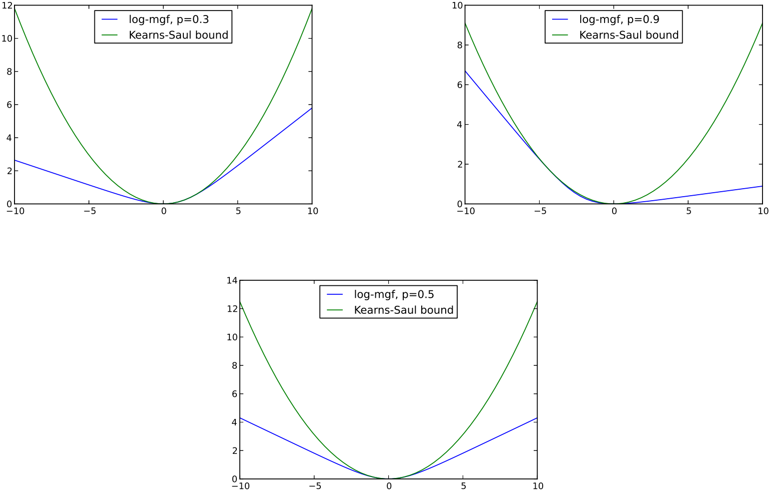
$$\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \le \frac{\lambda^2 \sigma_p^2}{2} \ \forall\ \lambda \in \mathbb{R}$$

Ordentlich-Weinberger '05: If $P = \mathrm{Ber}(p), Q = \mathrm{Ber}(q)$, then

$$d_{\mathrm{TV}}(Q,P)^2 \le 2\sigma_p^2 D(Q||P)\ \forall\ q, 0 \le q \le 1$$

The two statements above are equivalent.

Kearns-Saul bound plots



Transportation method for bounded differences inequality

Suppose f satisfies the bounded difference property: a change in the i^{th} co-ordinate can change f by at most c_i.

$$\begin{aligned}
\mathbb{E}_Q Z - \mathbb{E} Z =& \mathbb{E}_M[f(Y) - f(X)] \\
\leq& \mathbb{E}_M[\sum_{i=1}^{n} c_i 1_{X_i \neq Y_i}] \\
=& \sum_{i=1}^{n} c_i \mathbb{P}_M[X_i \neq Y_i] \\
\leq& \sqrt{\sum_{i=1}^{n} c_i^2} \sqrt{\sum_{i=1}^{n} \mathbb{P}_M[X_i \neq Y_i]^2}
\end{aligned}$$

Marton's transportation inequality

If $P = P_1 \times P_2 \times \ldots \times P_n$, and $Q << P$, and

$$X = (X_1, X_2, \ldots, X_n) \sim P,$$

$$Y = (Y_1, Y_2, \ldots, Y_n) \sim Q,$$

then

$$\inf_{M \in \mathscr{C}(P,Q), (X,Y) \sim M} \sum_{i=1}^{n} \mathbb{P}_M (X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q||P),$$

where $\mathscr{C}(P, Q)$ is the set of all couplings of P and Q.

Marton's transportation inequality

If $P = P_1 \times P_2 \times \ldots \times P_n$, and $Q << P$, and

$$X = (X_1, X_2, \ldots, X_n) \sim P,$$

$$Y = (Y_1, Y_2, \ldots, Y_n) \sim Q,$$

then

$$\inf_{M \in \mathscr{C}(P,Q), (X,Y) \sim M} \sum_{i=1}^{n} \mathbb{P}_M (X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q||P),$$

where $\mathscr{C}(P, Q)$ is the set of all couplings of P and Q.

$n = 1$: follows from Pinsker's inequality, since

Marton's transportation inequality

If $P = P_1 \times P_2 \times \ldots \times P_n$, and $Q << P$, and

$$X = (X_1, X_2, \ldots, X_n) \sim P,$$

$$Y = (Y_1, Y_2, \ldots, Y_n) \sim Q,$$

then

$$\inf_{M \in \mathscr{C}(P,Q), (X,Y) \sim M} \sum_{i=1}^{n} \mathbb{P}_M(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q||P),$$

where $\mathscr{C}(P, Q)$ is the set of all couplings of P and Q.

$n = 1$: follows from Pinsker's inequality, since
$$\inf_{M \in \mathscr{C}(P,Q), (X,Y) \sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P, Q)$$

Optimal coupling

$$\inf_{M \in \mathscr{C}(P,Q),(X,Y)\sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

Optimal coupling

$$\inf_{M \in \mathscr{C}(P,Q),(X,Y)\sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

$|P(A) - Q(A)|$

Optimal coupling

$$\inf_{M \in \mathscr{C}(P,Q),(X,Y)\sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

$$\begin{aligned} &|P(A) - Q(A)| \\ =&|\mathbb{E}_M[1_{X\in A} - 1_{Y\in A}]| \end{aligned}$$

Optimal coupling

$$\inf_{M\in\mathscr{C}(P,Q),(X,Y)\sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

$$\begin{aligned} &|P(A) - Q(A)| \\ =&|\mathbb{E}_M[1_{X\in A} - 1_{Y\in A}]| \\ \leq&\mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|] \end{aligned}$$

Optimal coupling

$$\inf_{M \in \mathscr{C}(P,Q),(X,Y)\sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

$$\begin{aligned}
&|P(A) - Q(A)| \\
=&|\mathbb{E}_M[1_{X\in A} - 1_{Y\in A}]| \\
\leq&\mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|] \\
=&\mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|1_{X\neq Y}]
\end{aligned}$$

Optimal coupling

$$\inf_{M \in \mathscr{C}(P,Q),(X,Y)\sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

$$\begin{aligned}
&|P(A) - Q(A)| \\
=&|\mathbb{E}_M[1_{X\in A} - 1_{Y\in A}]| \\
\leq&\mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|] \\
=&\mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|1_{X\neq Y}] \\
\leq&\mathbb{E}_M[1_{X\neq Y}]
\end{aligned}$$

Optimal coupling

$$\inf_{M \in \mathscr{C}(P,Q),(X,Y)\sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

$$\begin{aligned}
&|P(A) - Q(A)| \\
=&|\mathbb{E}_M[1_{X\in A} - 1_{Y\in A}]| \\
\leq&\mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|] \\
=&\mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|1_{X\neq Y}] \\
\leq&\mathbb{E}_M[1_{X\neq Y}] \\
=&\mathbb{P}_M[X \neq Y]
\end{aligned}$$

Optimal coupling

$$\inf_{M \in \mathscr{C}(P,Q),(X,Y)\sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

$$\begin{aligned}
&|P(A) - Q(A)| \\
=&|\mathbb{E}_M[1_{X\in A} - 1_{Y\in A}]| \\
\leq&\mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|] \\
=&\mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|1_{X\neq Y}] \\
\leq&\mathbb{E}_M[1_{X\neq Y}] \\
=&\mathbb{P}_M[X \neq Y]
\end{aligned}$$

Thus,

$$d_{\mathrm{TV}}(P,Q) \leq \mathbb{P}_M[X \neq Y]$$

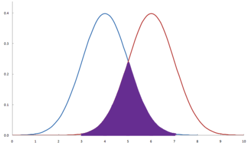
Optimal coupling

$$\inf_{M \in \mathscr{C}(P,Q), (X,Y) \sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

$$\begin{aligned}
&|P(A) - Q(A)| \\
=&|\mathbb{E}_M[1_{X \in A} - 1_{Y \in A}]| \\
\leq&\mathbb{E}_M[|1_{X \in A} - 1_{Y \in A}|] \\
=&\mathbb{E}_M[|1_{X \in A} - 1_{Y \in A}| 1_{X \neq Y}] \\
\leq&\mathbb{E}_M[1_{X \neq Y}] \\
=&\mathbb{P}_M[X \neq Y]
\end{aligned}$$

Thus,

$$d_{\mathrm{TV}}(P,Q) \leq \mathbb{P}_M[X \neq Y]$$

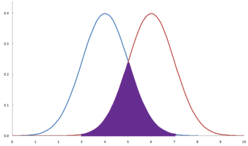


Optimal coupling

$$\inf_{M \in \mathscr{C}(P,Q),(X,Y)\sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

$$\begin{aligned}
& |P(A) - Q(A)| \\
=& |\mathbb{E}_M[1_{X\in A} - 1_{Y\in A}]| \\
\leq& \mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|] \\
=& \mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}| 1_{X\neq Y}] \\
\leq& \mathbb{E}_M[1_{X\neq Y}] \\
=& \mathbb{P}_M[X \neq Y]
\end{aligned}$$

Thus,
$$d_{\mathrm{TV}}(P,Q) \leq \mathbb{P}_M[X \neq Y]$$



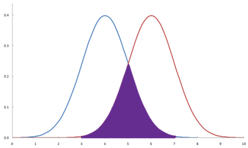
If $d_{TV}(P,Q) = V$, purple area $= 1 - V$

Optimal coupling

$$\inf_{M \in \mathscr{C}(P,Q),(X,Y)\sim M} \mathbb{P}_M(X \neq Y) = d_{\mathrm{TV}}(P,Q)$$

$$
\begin{aligned}
& |P(A) - Q(A)| \\
=& |\mathbb{E}_M[1_{X\in A} - 1_{Y\in A}]| \\
\leq& \mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}|] \\
=& \mathbb{E}_M[|1_{X\in A} - 1_{Y\in A}| 1_{X\neq Y}] \\
\leq& \mathbb{E}_M[1_{X\neq Y}] \\
=& \mathbb{P}_M[X \neq Y]
\end{aligned}
$$

Thus,
$d_{\mathrm{TV}}(P,Q) \leq \mathbb{P}_M[X \neq Y]$



If $d_{TV}(P,Q) = V$, purple area $= 1 - V$

$$
\begin{aligned}
& M(x,y) \\
=& (1-V) \cdot \frac{\min\{p(x), q(x)\} 1_{x=y}}{1-V} \\
& + V \cdot \frac{(p(x) - q(x))_+}{V} \cdot \frac{(q(y) - p(y))_+}{V}
\end{aligned}
$$

Sketch of induction argument for $n > 1$:

Sketch of induction argument for $n > 1$:

Suppose

$$\inf_{M_i \in \mathscr{C}(P_i, Q_i), (X_i, Y_i) \sim M_i} \mathbb{P}_{M_i}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q_i || P_i) \ \forall \ \ Q_i << P_i.$$

Sketch of induction argument for $n > 1$:

Suppose

$$\inf_{M_i \in \mathscr{C}(P_i, Q_i), (X_i, Y_i) \sim M_i} \mathbb{P}_{M_i}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q_i || P_i) \ \forall \ \ Q_i << P_i.$$

Given $Q = Q_{Y_1, Y_2}(y_1, y_2)$, generate X_1, Y_1, X_2, Y_2 as follows:

Sketch of induction argument for $n > 1$:
Suppose

$$\inf_{M_i \in \mathscr{C}(P_i, Q_i), (X_i, Y_i) \sim M_i} \mathbb{P}_{M_i}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q_i || P_i) \ \forall \ \ Q_i << P_i.$$

Given $Q = Q_{Y_1, Y_2}(y_1, y_2)$, generate X_1, Y_1, X_2, Y_2 as follows:

- Let M_1 be an optimal coupling corresponding to P_1 and Q_{Y_1}. Generate (X_1, Y_1) according to M_1, say (x_1, y_1).

Sketch of induction argument for $n > 1$:

Suppose

$$\inf_{M_i \in \mathscr{C}(P_i, Q_i), (X_i, Y_i) \sim M_i} \mathbb{P}_{M_i}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q_i || P_i) \ \forall \ \ Q_i << P_i.$$

Given $Q = Q_{Y_1, Y_2}(y_1, y_2)$, generate X_1, Y_1, X_2, Y_2 as follows:

- Let M_1 be an optimal coupling corresponding to P_1 and Q_{Y_1}. Generate (X_1, Y_1) according to M_1, say (x_1, y_1).
- Let M_2 be an optimal coupling corresponding to P_2 and $Q_{Y_2|Y_1}(y_2|y_1)$. Generate (X_2, Y_2) according to M_2, say (x_2, y_2).

Sketch of induction argument for $n > 1$:

Suppose

$$\inf_{M_i \in \mathscr{C}(P_i, Q_i), (X_i, Y_i) \sim M_i} \mathbb{P}_{M_i}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q_i || P_i) \ \forall \ \ Q_i << P_i.$$

Given $Q = Q_{Y_1,Y_2}(y_1, y_2)$, generate X_1, Y_1, X_2, Y_2 as follows:

- Let M_1 be an optimal coupling corresponding to P_1 and Q_{Y_1}. Generate (X_1, Y_1) according to M_1, say (x_1, y_1).
- Let M_2 be an optimal coupling corresponding to P_2 and $Q_{Y_2|Y_1}(y_2|y_1)$. Generate (X_2, Y_2) according to M_2, say (x_2, y_2).

(X_1, X_2) and (Y_1, Y_2) have right marginals.

Sketch of induction argument for $n > 1$:

Suppose

$$\inf_{M_i \in \mathscr{C}(P_i,Q_i),(X_i,Y_i)\sim M_i} \mathbb{P}_{M_i}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q_i||P_i) \ \forall \ \ Q_i << P_i.$$

Given $Q = Q_{Y_1,Y_2}(y_1, y_2)$, generate X_1, Y_1, X_2, Y_2 as follows:

- Let M_1 be an optimal coupling corresponding to P_1 and Q_{Y_1}. Generate (X_1, Y_1) according to M_1, say (x_1, y_1).
- Let M_2 be an optimal coupling corresponding to P_2 and $Q_{Y_2|Y_1}(y_2|y_1)$. Generate (X_2, Y_2) according to M_2, say (x_2, y_2).

(X_1, X_2) and (Y_1, Y_2) have right marginals.

$$P_M(X_1 \neq Y_1)^2 \leq D(Q_1||P_1)$$

Sketch of induction argument for $n > 1$:

Suppose

$$\inf_{M_i \in \mathscr{C}(P_i, Q_i), (X_i, Y_i) \sim M_i} \mathbb{P}_{M_i}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q_i || P_i) \ \forall \ \ Q_i << P_i.$$

Given $Q = Q_{Y_1, Y_2}(y_1, y_2)$, generate X_1, Y_1, X_2, Y_2 as follows:

- Let M_1 be an optimal coupling corresponding to P_1 and Q_{Y_1}. Generate (X_1, Y_1) according to M_1, say (x_1, y_1).
- Let M_2 be an optimal coupling corresponding to P_2 and $Q_{Y_2|Y_1}(y_2|y_1)$. Generate (X_2, Y_2) according to M_2, say (x_2, y_2).

(X_1, X_2) and (Y_1, Y_2) have right marginals.

$$P_M(X_1 \neq Y_1)^2 \leq D(Q_1 || P_1)$$

$$P_M(X_2 \neq Y_2 | Y_1 = y_1)^2 \leq D(Q_{Y_2|Y_1}(\cdot | y_1) || P_1)$$

Sketch of induction argument for $n > 1$:

Suppose

$$\inf_{M_i \in \mathscr{C}(P_i,Q_i),(X_i,Y_i)\sim M_i} \mathbb{P}_{M_i}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q_i||P_i) \ \forall \ \ Q_i << P_i.$$

Given $Q = Q_{Y_1,Y_2}(y_1, y_2)$, generate X_1, Y_1, X_2, Y_2 as follows:

- Let M_1 be an optimal coupling corresponding to P_1 and Q_{Y_1}. Generate (X_1, Y_1) according to M_1, say (x_1, y_1).
- Let M_2 be an optimal coupling corresponding to P_2 and $Q_{Y_2|Y_1}(y_2|y_1)$. Generate (X_2, Y_2) according to M_2, say (x_2, y_2).

(X_1, X_2) and (Y_1, Y_2) have right marginals.

$$P_M(X_1 \neq Y_1)^2 \leq D(Q_1||P_1)$$

$$P_M(X_2 \neq Y_2|Y_1 = y_1)^2 \leq D(Q_{Y_2|Y_1}(\cdot|y_1)||P_1)$$

Average over y_1, apply Jensen on LHS and chain rule of relative entropy on RHS:

$$\sum_{i=1}^{2} \mathbb{P}_M(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q||P)$$

The most powerful concentration inequality (in this tutorial)

The most powerful concentration inequality (in this tutorial)

Suppose $f(y) - f(x) \le \sum_{i=1}^{n} c_i(x) 1_{x_i \neq y_i}$ and $\sum_{i=1}^{n} c_i^2(x) \le \sigma^2$.

The most powerful concentration inequality (in this tutorial)

Suppose $f(y) - f(x) \leq \sum_{i=1}^{n} c_i(x) 1_{x_i \neq y_i}$ and $\sum_{i=1}^{n} c_i^2(x) \leq \sigma^2$.

If $Z = f(X_1, \ldots, X_n)$ and $X_1, X_2 \ldots, X_n$ are independent, then Z is σ^2-sub-Gaussian:

The most powerful concentration inequality (in this tutorial)

Suppose $f(y) - f(x) \leq \sum_{i=1}^{n} c_i(x) 1_{x_i \neq y_i}$ and $\sum_{i=1}^{n} c_i^2(x) \leq \sigma^2$.

If $Z = f(X_1, \ldots, X_n)$ and $X_1, X_2 \ldots, X_n$ are independent, then Z is σ^2-sub-Gaussian:

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

The most powerful concentration inequality (in this tutorial)

Suppose $f(y) - f(x) \leq \sum_{i=1}^{n} c_i(x) 1_{x_i \neq y_i}$ and $\sum_{i=1}^{n} c_i^2(x) \leq \sigma^2$.

If $Z = f(X_1, \ldots, X_n)$ and $X_1, X_2 \ldots, X_n$ are independent, then Z is σ^2-sub-Gaussian:

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

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

The most powerful concentration inequality (in this tutorial)

Suppose $f(y) - f(x) \le \sum_{i=1}^{n} c_i(x) 1_{x_i \ne y_i}$ and $\sum_{i=1}^{n} c_i^2(x) \le \sigma^2$.

If $Z = f(X_1, \ldots, X_n)$ and $X_1, X_2 \ldots, X_n$ are independent, then Z is σ^2-sub-Gaussian:

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

Then, $\mathbb{P}[Z - \mathbb{E}Z \ge t], \mathbb{P}[Z - \mathbb{E}Z \le -t] \le e^{-\frac{t^2}{2\sigma^2}}, \ \forall t > 0$

Prove sub-Gaussianity by the equivalent transportation description:

$$\mathbb{E}_Q Z - \mathbb{E}Z \le \sqrt{2\sigma^2 D(Q||P)}, \ \ \forall \ \ Q << P$$

$\mathbb{E}_Q Z - \mathbb{E} Z$

$$\mathbb{E}_Q Z - \mathbb{E} Z = \mathbb{E}_M[f(Y) - f(X)]$$

$$\begin{aligned}\mathbb{E}_Q Z - \mathbb{E} Z &= \mathbb{E}_M[f(Y) - f(X)] \\ &\leq \mathbb{E}_M[\sum_{i=1}^{n} c_i(X) 1_{X_i \neq Y_i}]\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}_Q Z - \mathbb{E} Z &= \mathbb{E}_M[f(Y) - f(X)] \\
&\leq \mathbb{E}_M[\sum_{i=1}^{n} c_i(X) 1_{X_i \neq Y_i}] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[c_i(X) \mathbb{P}_M[X_i \neq Y_i | X]\right]
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_Q Z - \mathbb{E} Z &= \mathbb{E}_M[f(Y) - f(X)] \\
&\le \mathbb{E}_M[\sum_{i=1}^{n} c_i(X) 1_{X_i \ne Y_i}] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[c_i(X) \mathbb{P}_M[X_i \ne Y_i | X]\right] \\
&= \sum_{i=1}^{n} \sqrt{\mathbb{E}[c_i(X)^2]} \sqrt{\mathbb{E}\left[\mathbb{P}_M[X_i \ne Y_i | X]^2\right]}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_Q Z - \mathbb{E} Z &= \mathbb{E}_M[f(Y) - f(X)] \\
&\leq \mathbb{E}_M[\sum_{i=1}^n c_i(X) 1_{X_i \neq Y_i}] \\
&= \sum_{i=1}^n \mathbb{E}\left[c_i(X) \mathbb{P}_M[X_i \neq Y_i | X]\right] \\
&= \sum_{i=1}^n \sqrt{\mathbb{E}[c_i(X)^2]} \sqrt{\mathbb{E}\left[\mathbb{P}_M[X_i \neq Y_i | X]^2\right]} \\
&\leq \sqrt{\sum_{i=1}^n \mathbb{E} c_i(X)^2} \sqrt{\sum_{i=1}^n \mathbb{E}[\mathbb{P}_M[X_i \neq Y_i | X]^2]}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_Q Z - \mathbb{E} Z &= \mathbb{E}_M[f(Y) - f(X)] \\
&\leq \mathbb{E}_M[\sum_{i=1}^{n} c_i(X) 1_{X_i \neq Y_i}] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[c_i(X) \mathbb{P}_M[X_i \neq Y_i | X]\right] \\
&= \sum_{i=1}^{n} \sqrt{\mathbb{E}[c_i(X)^2]} \sqrt{\mathbb{E}\left[\mathbb{P}_M[X_i \neq Y_i | X]^2\right]} \\
&\leq \sqrt{\sum_{i=1}^{n} \mathbb{E} c_i(X)^2} \sqrt{\sum_{i=1}^{n} \mathbb{E}[\mathbb{P}_M[X_i \neq Y_i | X]^2]} \\
&\leq \sqrt{\sigma^2} \sqrt{\sum_{i=1}^{n} \mathbb{E}\left[\mathbb{P}_M[X_i \neq Y_i | X]^2\right]}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_Q Z - \mathbb{E} Z &= \mathbb{E}_M[f(Y) - f(X)] \\
&\leq \mathbb{E}_M[\sum_{i=1}^n c_i(X) 1_{X_i \neq Y_i}] \\
&= \sum_{i=1}^n \mathbb{E}\left[c_i(X) \mathbb{P}_M[X_i \neq Y_i | X]\right] \\
&= \sum_{i=1}^n \sqrt{\mathbb{E}[c_i(X)^2]} \sqrt{\mathbb{E}\left[\mathbb{P}_M[X_i \neq Y_i | X]^2\right]} \\
&\leq \sqrt{\sum_{i=1}^n \mathbb{E} c_i(X)^2} \sqrt{\sum_{i=1}^n \mathbb{E}[\mathbb{P}_M[X_i \neq Y_i | X]^2]} \\
&\leq \sqrt{\sigma^2} \sqrt{\sum_{i=1}^n \mathbb{E}\left[\mathbb{P}_M[X_i \neq Y_i | X]^2\right]}
\end{aligned}
$$

Need to show the second quantity in $\sqrt{}$ is $\leq 2D(Q||P)$

Follow a similar procedure:

Follow a similar procedure:

$$n = 1 : \inf_{M_i \in \mathscr{C}(P_i, Q_i)} \mathbb{E}\left[\mathbb{P}_{M_i}[X_i \neq Y_i | X_i]^2\right] = \mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right]$$

Follow a similar procedure:

$$n = 1 : \inf_{M_i \in \mathscr{C}(P_i, Q_i)} \mathbb{E}\left[\mathbb{P}_{M_i}[X_i \neq Y_i | X_i]]^2\right] = \mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right]$$

RHS is also called Marton's divergence.

Follow a similar procedure:

$$n = 1 : \inf_{M_i \in \mathscr{C}(P_i, Q_i)} \mathbb{E}\left[\mathbb{P}_{M_i}[X_i \neq Y_i | X_i]^2\right] = \mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right]$$

RHS is also called Marton's divergence.

Optimal coupling identical to that for total variation distance!

Follow a similar procedure:

$$n = 1: \inf_{M_i \in \mathscr{C}(P_i, Q_i)} \mathbb{E}\left[\mathbb{P}_{M_i}[X_i \neq Y_i | X_i]^2\right] = \mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right]$$

RHS is also called Marton's divergence.

Optimal coupling identical to that for total variation distance!

$$\mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right] \leq 2D(Q_i || P_i)$$

Follow a similar procedure:

$$n = 1 : \inf_{M_i \in \mathscr{C}(P_i, Q_i)} \mathbb{E}\left[\mathbb{P}_{M_i}[X_i \neq Y_i | X_i]]^2\right] = \mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right]$$

RHS is also called Marton's divergence.

Optimal coupling identical to that for total variation distance!

$$\mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right] \leq 2D(Q_i || P_i)$$

Induction argument also applies to give a coupling M so that

$$\sum_{i=1}^{n} \mathbb{E}\left[\mathbb{P}_M[X_i \neq Y_i | X]^2\right] \leq 2D(Q || P)$$

Follow a similar procedure:

$$n = 1 : \inf_{M_i \in \mathscr{C}(P_i, Q_i)} \mathbb{E}\left[\mathbb{P}_{M_i}[X_i \neq Y_i | X_i]^2\right] = \mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right]$$

RHS is also called Marton's divergence.

Optimal coupling identical to that for total variation distance!

$$\mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right] \leq 2D(Q_i || P_i)$$

Induction argument also applies to give a coupling M so that

$$\sum_{i=1}^{n} \mathbb{E}\left[\mathbb{P}_M[X_i \neq Y_i | X]^2\right] \leq 2D(Q || P)$$

Similarly, we can show $\displaystyle\sum_{i=1}^{n} \mathbb{E}_M\left[\mathbb{P}_M[X_i \neq Y_i | Y]^2\right] \leq 2D(Q || P)$

Follow a similar procedure:

$$n = 1 : \inf_{M_i \in \mathscr{C}(P_i, Q_i)} \mathbb{E}\left[\mathbb{P}_{M_i}[X_i \neq Y_i | X_i]]^2\right] = \mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right]$$

RHS is also called Marton's divergence.

Optimal coupling identical to that for total variation distance!

$$\mathbb{E}\left[\left(1 - \frac{q_i(X)}{p_i(X)}\right)_+^2\right] \leq 2D(Q_i || P_i)$$

Induction argument also applies to give a coupling M so that

$$\sum_{i=1}^{n} \mathbb{E}\left[\mathbb{P}_M[X_i \neq Y_i | X]^2\right] \leq 2D(Q || P)$$

Similarly, we can show $\sum_{i=1}^{n} \mathbb{E}_M\left[\mathbb{P}_M[X_i \neq Y_i | Y]^2\right] \leq 2D(Q || P)$

This gives BOTH left and right tails!

Spectral norm of a random matrix

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

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

Spectral norm of a random matrix

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

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

Previously, we showed $\mathrm{Var}(Z) \leq 1$.

Spectral norm of a random matrix

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

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

Previously, we showed $\mathrm{Var}(Z) \leq 1$.

Now, we get $\log \mathbb{E} e^{\lambda(Z-\mathbb{E}Z)} \leq \dfrac{\lambda^2}{2} \ \forall\ \lambda \in \mathbb{R}$

Spectral norm of a random matrix

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

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

Previously, we showed $\mathrm{Var}(Z) \leq 1$.

Now, we get $\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{\lambda^2}{2} \ \forall \ \lambda \in \mathbb{R}$

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

Largest eigenvalue of a random matrix

Let A be an $n \times n$ symmetric matrix with independent entries $X_{ij}, 1 \leq i \leq j \leq n$ independent, $-1 \leq X_{ij} \leq 1$.

Largest eigenvalue of a random matrix

Let A be an $n \times n$ symmetric matrix with independent entries $X_{ij}, 1 \leq i \leq j \leq n$ independent, $-1 \leq X_{ij} \leq 1$.

Let $Z = \lambda_{\max}(A) = \max_{\|w\|=1} w^{\mathrm{T}} A w = u^{\mathrm{T}} A u$

for some u that depends on the X_{ij}'s.

Largest eigenvalue of a random matrix

Let A be an $n \times n$ symmetric matrix with independent entries $X_{ij}, 1 \le i \le j \le n$ independent, $-1 \le X_{ij} \le 1$.

Let $Z = \lambda_{\max}(A) = \max_{\|w\|=1} w^{\mathrm{T}} A w = u^{\mathrm{T}} A u$
for some u that depends on the X_{ij}'s.

Previously, we showed $\mathrm{Var}(Z) \le 16$

Largest eigenvalue of a random matrix

Let A be an $n \times n$ symmetric matrix with independent entries $X_{ij}, 1 \leq i \leq j \leq n$ independent, $-1 \leq X_{ij} \leq 1$.

Let $Z = \lambda_{\max}(A) = \max_{\|w\|=1} w^{\mathrm{T}} A w = u^{\mathrm{T}} A u$
for some u that depends on the X_{ij}'s.

Previously, we showed $\mathrm{Var}(Z) \leq 16$

Now, we get $\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \dfrac{\lambda^2(16)}{2} \ \forall \ \lambda \in \mathbb{R}$.

Largest eigenvalue of a random matrix

Let A be an $n \times n$ symmetric matrix with independent entries $X_{ij}, 1 \leq i \leq j \leq n$ independent, $-1 \leq X_{ij} \leq 1$.

Let $Z = \lambda_{\max}(A) = \max_{\|w\|=1} w^{\mathrm{T}} A w = u^{\mathrm{T}} A u$ for some u that depends on the X_{ij}'s.

Previously, we showed $\mathrm{Var}(Z) \leq 16$

Now, we get $\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{\lambda^2 (16)}{2} \ \forall \ \lambda \in \mathbb{R}$.

So, $\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq 2e^{-t^2/32} \ \forall t > 0$

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

For $\alpha \in \mathbb{R}_+^n, \|\alpha\|_2 = 1$, define $d_\alpha(x, x') := \sum_{i=1}^{n} \alpha_i 1_{x \neq x'_i}$

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

For $\alpha \in \mathbb{R}^n_+, \|\alpha\|_2 = 1$, define $d_\alpha(x, x') := \sum_{i=1}^{n} \alpha_i 1_{x \neq x'_i}$

Let $d_\alpha(x, A) = \inf_{x' \in A} d_\alpha(x, x')$

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

For $\alpha \in \mathbb{R}^n_+, \|\alpha\|_2 = 1$, define $d_\alpha(x, x') := \sum_{i=1}^{n} \alpha_i 1_{x \neq x'_i}$

Let $d_\alpha(x, A) = \inf_{x' \in A} d_\alpha(x, x')$

Define $d_T(x, A) = \sup_{\alpha \geq 0, \|\alpha\|_2 = 1} d_\alpha(x, A)$

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

For $\alpha \in \mathbb{R}^n_+, \|\alpha\|_2 = 1$, define $d_\alpha(x, x') := \sum_{i=1}^{n} \alpha_i 1_{x \neq x'_i}$

Let $d_\alpha(x, A) = \inf_{x' \in A} d_\alpha(x, x')$

Define $d_T(x, A) = \sup_{\alpha \geq 0, \|\alpha\|_2 = 1} d_\alpha(x, A)$

Let $A_t = \{x : d_T(x, A) \leq t\}$.

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

For $\alpha \in \mathbb{R}_+^n, \|\alpha\|_2 = 1$, define $d_\alpha(x, x') := \sum_{i=1}^{n} \alpha_i 1_{x \neq x_i'}$

Let $d_\alpha(x, A) = \inf_{x' \in A} d_\alpha(x, x')$

Define $d_T(x, A) = \sup_{\alpha \geq 0, \|\alpha\|_2 = 1} d_\alpha(x, A)$

Let $A_t = \{x : d_T(x, A) \leq t\}$.

Then, $P(A)P(A_t^c) \leq e^{-t^2/4} \; \forall \; t > 0$

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

For $\alpha \in \mathbb{R}_+^n$, $\|\alpha\|_2 = 1$, define $d_\alpha(x, x') := \sum_{i=1}^{n} \alpha_i 1_{x \neq x'_i}$

Let $d_\alpha(x, A) = \inf_{x' \in A} d_\alpha(x, x')$

Define $d_T(x, A) = \sup_{\alpha \geq 0, \|\alpha\|_2 = 1} d_\alpha(x, A)$

Let $A_t = \{x : d_T(x, A) \leq t\}$.

Then, $P(A)P(A_t^c) \leq e^{-t^2/4} \; \forall \; t > 0$

Proof on board!

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

For $\alpha \in \mathbb{R}_+^n, \|\alpha\|_2 = 1$, define $d_\alpha(x, x') := \sum_{i=1}^n \alpha_i 1_{x \neq x'_i}$

Let $d_\alpha(x, A) = \inf_{x' \in A} d_\alpha(x, x')$

Define $d_T(x, A) = \sup_{\alpha \geq 0, \|\alpha\|_2 = 1} d_\alpha(x, A)$

Let $A_t = \{x : d_T(x, A) \leq t\}$.

Then, $P(A)P(A_t^c) \leq e^{-t^2/4} \; \forall \; t > 0$

Proof on board!

If $V(x, A) = \left\{ \left(1_{x_i \neq x'_i} \right)_{i=1}^n : x' \in A \right\}$

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

For $\alpha \in \mathbb{R}_+^n, \|\alpha\|_2 = 1$, define $d_\alpha(x, x') := \sum_{i=1}^{n} \alpha_i 1_{x \neq x'_i}$

Let $d_\alpha(x, A) = \inf_{x' \in A} d_\alpha(x, x')$

Define $d_T(x, A) = \sup_{\alpha \geq 0, \|\alpha\|_2 = 1} d_\alpha(x, A)$

Let $A_t = \{x : d_T(x, A) \leq t\}$.

Then, $P(A)P(A_t^c) \leq e^{-t^2/4} \; \forall \; t > 0$

Proof on board!

If $V(x, A) = \left\{ \left(1_{x_i \neq x'_i} \right)_{i=1}^{n} : x' \in A \right\} \subseteq [0, 1]^n$,

Talagrand's convex distance inequality

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a product measure.

For $\alpha \in \mathbb{R}_+^n$, $\|\alpha\|_2 = 1$, define $d_\alpha(x, x') := \sum_{i=1}^n \alpha_i 1_{x \neq x_i'}$

Let $d_\alpha(x, A) = \inf_{x' \in A} d_\alpha(x, x')$

Define $d_T(x, A) = \sup_{\alpha \geq 0, \|\alpha\|_2 = 1} d_\alpha(x, A)$

Let $A_t = \{x : d_T(x, A) \leq t\}$.

Then, $P(A)P(A_t^c) \leq e^{-t^2/4} \; \forall \; t > 0$

Proof on board!

If $V(x, A) = \left\{ \left(1_{x_i \neq x_i'} \right)_{i=1}^n : x' \in A \right\} \subseteq [0, 1]^n$,

then $d_T(x, A)$ is the minimum Euclidean distance of the origin from the convex hull of $V(x, A)$ (hence the name 'convex distance').

Application: Longest increasing subsequence problem

Let $X_1, X_2, \ldots, X_n$ be independent, each drawn uniformly from $[0, 1]$.

A sequence $i_1 < i_2 < \ldots < i_r$ constitutes an increasing subsequence if $X_{i_1} < X_{i_2} < \ldots < X_{i_r}$.

What is the behavior of the length $Z = f(X_1, X_2, \ldots, X_n)$ of the longest increasing subsequence?

Erdös-Szekeres theorem (1935)

If m^2+1 people of different heights stand in a line, there exists among them either a monotonically increasing subsequence of length $m+1$ or a monotonically decreasing subsequence of length $m+1$

Erdös-Szekeres theorem (1935)

If $m^2 + 1$ people of different heights stand in a line, there exists among them either a monotonically increasing subsequence of length $m + 1$ or a monotonically decreasing subsequence of length $m + 1$

Therefore, $\mathbb{E}Z \geq \frac{\lfloor \sqrt{n-1} \rfloor}{2}$

$$\mathbb{P}[Z \geq k]$$

$$
\begin{aligned}
&\mathbb{P}[Z \geq k] \\
= &\mathbb{P}[\text{at least one increasing subsequence of length} = k]
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}[Z \geq k] \\
&= \mathbb{P}[\text{at least one increasing subsequence of length} = k] \\
&\leq \mathbb{E}[\text{number of increasing subsequences of length } k]
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}[Z \geq k] \\
&= \mathbb{P}[\text{at least one increasing subsequence of length} = k] \\
&\leq \mathbb{E}[\text{number of increasing subsequences of length } k] \\
&= \frac{1}{k!}\binom{n}{k}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}[Z \geq k] \\
&= \mathbb{P}[\text{at least one increasing subsequence of length} = k] \\
&\leq \mathbb{E}[\text{number of increasing subsequences of length } k] \\
&= \frac{1}{k!}\binom{n}{k} \\
&\leq \frac{n^k}{(k/e)^{2k}}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}[Z \geq k] \\
&= \mathbb{P}[\text{at least one increasing subsequence of length} = k] \\
&\leq \mathbb{E}[\text{number of increasing subsequences of length } k] \\
&= \frac{1}{k!}\binom{n}{k} \\
&\leq \frac{n^k}{(k/e)^{2k}}
\end{aligned}
$$

Choose $k = 2e\sqrt{n}$ to get

$$
\begin{aligned}
&\mathbb{P}[Z \geq k] \\
&= \mathbb{P}[\text{at least one increasing subsequence of length} = k] \\
&\leq \mathbb{E}[\text{number of increasing subsequences of length } k] \\
&= \frac{1}{k!}\binom{n}{k} \\
&\leq \frac{n^k}{(k/e)^{2k}}
\end{aligned}
$$

Choose $k = 2e\sqrt{n}$ to get

$$
\mathbb{E}Z \leq \mathbb{P}[Z \leq k]k + \mathbb{P}[Z \geq k]n
$$

$$
\begin{aligned}
&\mathbb{P}[Z \geq k] \\
&= \mathbb{P}[\text{at least one increasing subsequence of length} = k] \\
&\leq \mathbb{E}[\text{number of increasing subsequences of length } k] \\
&= \frac{1}{k!}\binom{n}{k} \\
&\leq \frac{n^k}{(k/e)^{2k}}
\end{aligned}
$$

Choose $k = 2e\sqrt{n}$ to get

$$
\begin{aligned}
\mathbb{E}Z &\leq \mathbb{P}[Z \leq k]k + \mathbb{P}[Z \geq k]n \\
&\leq 2e\sqrt{n} + O(e^{-c\sqrt{n}})
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}[Z \geq k] \\
&= \mathbb{P}[\text{at least one increasing subsequence of length} = k] \\
&\leq \mathbb{E}[\text{number of increasing subsequences of length } k] \\
&= \frac{1}{k!}\binom{n}{k} \\
&\leq \frac{n^k}{(k/e)^{2k}}
\end{aligned}
$$

Choose $k = 2e\sqrt{n}$ to get

$$
\begin{aligned}
\mathbb{E}Z &\leq \mathbb{P}[Z \leq k]k + \mathbb{P}[Z \geq k]n \\
&\leq 2e\sqrt{n} + O(e^{-c\sqrt{n}})
\end{aligned}
$$

Thus, $\mathbb{E}Z = \Theta(\sqrt{n})$.

$Z = f(X_1, X_2, \ldots, X_n)$ where the function f has the property that a change in any one co-ordinate can alter the value of f by at most 1

$Z = f(X_1, X_2, \ldots, X_n)$ where the function f has the property that a change in any one co-ordinate can alter the value of f by at most 1

Thus, by bounded differences inequality, Z is σ^2-sub-Gaussian with

$$\sigma^2 = \frac{n}{4}$$

$Z = f(X_1, X_2, \ldots, X_n)$ where the function f has the property that a change in any one co-ordinate can alter the value of f by at most 1

Thus, by bounded differences inequality, Z is σ^2-sub-Gaussian with

$$\sigma^2 = \frac{n}{4}$$

However, this is not satisfactory, since this gives typical deviation $O(\sqrt{n})$ and we know $\mathbb{E}Z = \Theta(\sqrt{n})$

$Z = f(X_1, X_2, \ldots, X_n)$ where the function f has the property that a change in any one co-ordinate can alter the value of f by at most 1

Thus, by bounded differences inequality, Z is σ^2-sub-Gaussian with $\sigma^2 = \dfrac{n}{4}$

However, this is not satisfactory, since this gives typical deviation $O(\sqrt{n})$ and we know $\mathbb{E}Z = \Theta(\sqrt{n})$

Talagrand's convex distance inequality can show sub-Gaussian tail bounds with typical deviation $O(n^{1/4})$

Fix any b and t

Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$

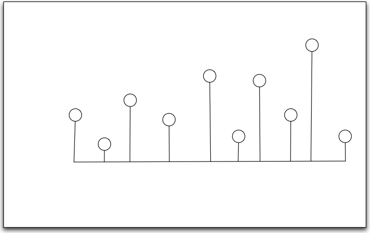
Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

Pick any sequence $x \in B$, so now $f(x) \geq b$.

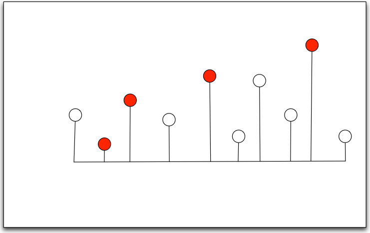
Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

Pick any sequence $x \in B$, so now $f(x) \geq b$.



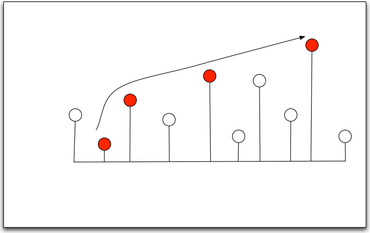
Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

Pick any sequence $x \in B$, so now $f(x) \geq b$.



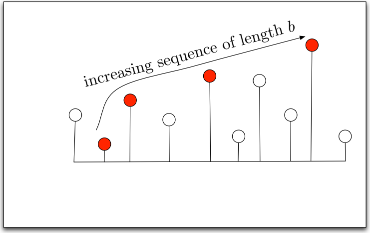
Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

Pick any sequence $x \in B$, so now $f(x) \geq b$.



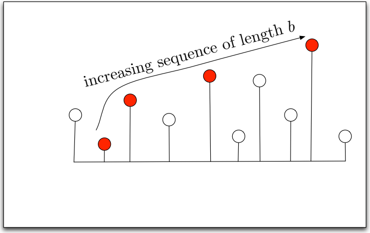
Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

Pick any sequence $x \in B$, so now $f(x) \geq b$.



Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

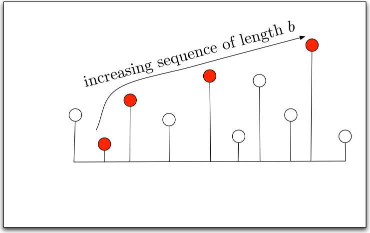
Pick any sequence $x \in B$, so now $f(x) \geq b$.



Suppose $d_T(x, A) \leq t$

Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

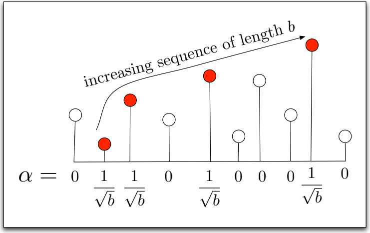
Pick any sequence $x \in B$, so now $f(x) \geq b$.



Suppose $d_T(x, A) \leq t$ i.e. $\sup_{\alpha \geq 0 : \|\alpha\|_2 = 1} d_\alpha(x, A) \leq t$.

Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

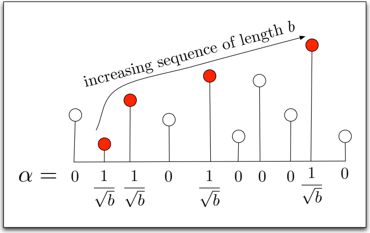
Pick any sequence $x \in B$, so now $f(x) \geq b$.



Suppose $d_T(x, A) \leq t$ i.e. $\sup_{\alpha \geq 0 : \|\alpha\|_2 = 1} d_\alpha(x, A) \leq t$.

Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

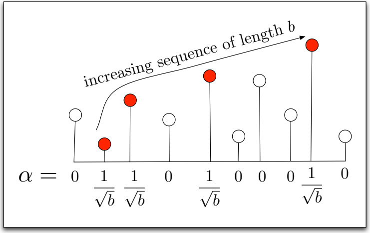
Pick any sequence $x \in B$, so now $f(x) \geq b$.



Suppose $d_T(x, A) \leq t$ i.e. $\sup_{\alpha \geq 0 : \|\alpha\|_2 = 1} d_\alpha(x, A) \leq t$.

$\exists\ y \in A$ such that $\sum_{i \in red} \frac{1}{\sqrt{b}} 1_{x_i \neq y_i} \leq t$

Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.
Pick any sequence $x \in B$, so now $f(x) \geq b$.



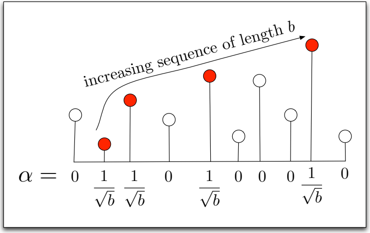
Suppose $d_T(x, A) \leq t$ i.e. $\sup_{\alpha \geq 0 : \|\alpha\|_2 = 1} d_\alpha(x, A) \leq t$.

$$\exists\ y \in A \text{ such that } \sum_{i \in red} \frac{1}{\sqrt{b}} 1_{x_i \neq y_i} \leq t$$

Impossible since for each $y \in A$, $f(y) < b - t\sqrt{b}$.

Fix any b and t, let $A = \{y : f(y) < b - t\sqrt{b}\}$, let $B = \{x : f(x) \geq b\}$.

Pick any sequence $x \in B$, so now $f(x) \geq b$.



Suppose $d_T(x, A) \leq t$ i.e. $\sup_{\alpha \geq 0 : \|\alpha\|_2 = 1} d_\alpha(x, A) \leq t$.

$$\exists\ y \in A \text{ such that } \sum_{i \in red} \frac{1}{\sqrt{b}} 1_{x_i \neq y_i} \leq t$$

Impossible since for each $y \in A$, $f(y) < b - t\sqrt{b}$.

Thus, $d_T(x, A) > t$ and so, $B \subseteq A_t^c$.

By Talagrand's inequality, $\mathbb{P}[Z \le b - t\sqrt{b}]\mathbb{P}[Z \ge b] \le e^{-t^2/4}$

By Talagrand's inequality, $\mathbb{P}[Z \le b - t\sqrt{b}]\mathbb{P}[Z \ge b] \le e^{-t^2/4}$

Choose $b = \mathbb{M}Z$ so $\mathbb{P}[Z \ge \mathbb{M}Z] \ge 1/2, \mathbb{P}[Z \le \mathbb{M}Z] \ge 1/2$

By Talagrand's inequality, $\mathbb{P}[Z \leq b - t\sqrt{b}]\mathbb{P}[Z \geq b] \leq e^{-t^2/4}$

Choose $b = \mathbb{M}Z$ so $\mathbb{P}[Z \geq \mathbb{M}Z] \geq 1/2, \mathbb{P}[Z \leq \mathbb{M}Z] \geq 1/2$

Then, $\mathbb{P}[Z - \mathbb{M}Z \leq -t\sqrt{\mathbb{M}Z}] \leq 2e^{-t^2/4}$

By Talagrand's inequality, $\mathbb{P}[Z \le b - t\sqrt{b}]\mathbb{P}[Z \ge b] \le e^{-t^2/4}$

Choose $b = \mathbb{M}Z$ so $\mathbb{P}[Z \ge \mathbb{M}Z] \ge 1/2, \mathbb{P}[Z \le \mathbb{M}Z] \ge 1/2$

Then, $\mathbb{P}[Z - \mathbb{M}Z \le -t\sqrt{\mathbb{M}Z}] \le 2e^{-t^2/4}$

Similarly, $\mathbb{M}Z = b - t\sqrt{b}$ gives upper tail bounds

By Talagrand's inequality, $\mathbb{P}[Z \le b - t\sqrt{b}]\mathbb{P}[Z \ge b] \le e^{-t^2/4}$

Choose $b = \mathbb{M}Z$ so $\mathbb{P}[Z \ge \mathbb{M}Z] \ge 1/2, \mathbb{P}[Z \le \mathbb{M}Z] \ge 1/2$

Then, $\mathbb{P}[Z - \mathbb{M}Z \le -t\sqrt{\mathbb{M}Z}] \le 2e^{-t^2/4}$

Similarly, $\mathbb{M}Z = b - t\sqrt{b}$ gives upper tail bounds

Arguments we made earlier to show $c_1\sqrt{n} \le \mathbb{E}Z \le c_2\sqrt{n}$ also show $c_1\sqrt{n} \le \mathbb{M}Z \le c_2\sqrt{n}$

By Talagrand's inequality, $\mathbb{P}[Z \leq b - t\sqrt{b}]\mathbb{P}[Z \geq b] \leq e^{-t^2/4}$

Choose $b = \mathbb{M}Z$ so $\mathbb{P}[Z \geq \mathbb{M}Z] \geq 1/2, \mathbb{P}[Z \leq \mathbb{M}Z] \geq 1/2$

Then, $\mathbb{P}[Z - \mathbb{M}Z \leq -t\sqrt{\mathbb{M}Z}] \leq 2e^{-t^2/4}$

Similarly, $\mathbb{M}Z = b - t\sqrt{b}$ gives upper tail bounds

Arguments we made earlier to show $c_1\sqrt{n} \leq \mathbb{E}Z \leq c_2\sqrt{n}$ also show $c_1\sqrt{n} \leq \mathbb{M}Z \leq c_2\sqrt{n}$

Therefore, typical deviation $= O(n^{1/4})$ with sub-Gaussian tail bounds

By Talagrand's inequality, $\mathbb{P}[Z \leq b - t\sqrt{b}]\mathbb{P}[Z \geq b] \leq e^{-t^2/4}$

Choose $b = \mathbb{M}Z$ so $\mathbb{P}[Z \geq \mathbb{M}Z] \geq 1/2, \mathbb{P}[Z \leq \mathbb{M}Z] \geq 1/2$

Then, $\mathbb{P}[Z - \mathbb{M}Z \leq -t\sqrt{\mathbb{M}Z}] \leq 2e^{-t^2/4}$

Similarly, $\mathbb{M}Z = b - t\sqrt{b}$ gives upper tail bounds

Arguments we made earlier to show $c_1\sqrt{n} \leq \mathbb{E}Z \leq c_2\sqrt{n}$ also show $c_1\sqrt{n} \leq \mathbb{M}Z \leq c_2\sqrt{n}$

Therefore, typical deviation $= O(n^{1/4})$ with sub-Gaussian tail bounds

Actually, $Z = 2\sqrt{n} + O(n^{1/6})$ and the limiting distribution of $\dfrac{Z - 2\sqrt{n}}{n^{1/6}}$ is known (Baik, Deift, Johansson, 1999)

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.

Summary

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.

- Tensorization of variance property: bounds on variance

Summary

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.

- Tensorization of variance property: bounds on variance
- Entropy method: a general tool to show sub-Gaussian tails (in conjunction with log-Sobolev inequalities)

Summary

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.

- Tensorization of variance property: bounds on variance
- Entropy method: a general tool to show sub-Gaussian tails (in conjunction with log-Sobolev inequalities)
- Transportation method: uses idea of coupling to show sub-Gaussian tails

Summary

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.

- Tensorization of variance property: bounds on variance
- Entropy method: a general tool to show sub-Gaussian tails (in conjunction with log-Sobolev inequalities)
- Transportation method: uses idea of coupling to show sub-Gaussian tails
- Talagrand's convex distance inequality: extremely powerful inequality for product measures

Summary

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.

- Tensorization of variance property: bounds on variance
- Entropy method: a general tool to show sub-Gaussian tails (in conjunction with log-Sobolev inequalities)
- Transportation method: uses idea of coupling to show sub-Gaussian tails
- Talagrand's convex distance inequality: extremely powerful inequality for product measures

(All slides available on my webpage)