Concentration of measure

This lecture is drawn from:

Chernoff’s inequality

**Theorem**

Let $X_1, X_2, ..., X_n$ be jointly independent random variables with mean 0 and such that $|X_i| \leq 1$. Let

$$X := X_1 + \cdots + X_n$$

and let $\sigma = \sqrt{\text{Var}[X]}$ the standard deviation. Then for any $\lambda > 0$,

$$\text{Prob}(|X| > \lambda \sigma) \leq 2 \max(e^{-\lambda^2/4}, e^{-\lambda \sigma/2}).$$

The concentration of measure phenomenon seeks to obtain ‘Gaussian-type’ tail decay in circumstances with less independence.
Lemma

Let $X$ be a random variable with $|X| \leq 1$ and $E[X] = 0$. Then for any $-1 \leq t \leq 1$ we have $E[e^{tX}] \leq \exp(t^2 \Var[X])$.

Proof.

By Taylor expansion, $e^{tX} \leq 1 + tX + t^2 X^2$. Thus

$$E[e^{tX}] \leq 1 + t^2 \Var[X] \leq \exp(t^2 \Var[X]).$$
Chernoff’s inequality

Proof of Chernoff’s inequality.

- By symmetry it suffices to prove $\Pr(X \geq \lambda \sigma) \leq e^{-t\lambda \sigma / 2}$ where $t = \min(\lambda / 2\sigma, 1)$.
- Use $\Pr(X \geq \lambda) = \Pr(e^{tX} \geq e^{t\lambda}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t\lambda}}$.
- Thus

\[
\Pr(X \geq \lambda \sigma) \leq e^{-t\lambda \sigma} \mathbb{E}[e^{tX_1} \cdots e^{tX_n}]
\]

\[
= e^{-t\lambda \sigma} \mathbb{E}[e^{tX_1}] \cdots \mathbb{E}[e^{tX_n}]
\]

\[
\leq e^{-t\lambda \sigma} \exp(t^2(\text{Var}[X_1] + \cdots + \text{Var}[X_n]))
\]

\[
= \exp(t^2\sigma^2 - t\lambda \sigma).
\]

- The claim follows, since $t \leq \lambda / 2\sigma$. 

□
Azuma’s inequality

The following is a martingale variant of Chernoff’s bound.

**Theorem (Azuma’s inequality)**

Let $0 = X_0, X_1, \ldots, X_m$ be a martingale sequence, with $\mathcal{F}_i = \sigma(X_0, \ldots, X_i)$ and $\mathbb{E}[X_i | \mathcal{F}_{i-1}] = X_{i-1}$. Assume

$$|X_i - X_{i-1}| \leq 1$$

for all $1 \leq i \leq m$. Let $\lambda > 0$. Then

$$\text{Prob} \left[ X_m > \lambda \sqrt{m} \right] < e^{-\lambda^2/2}.$$
Azuma’s inequality

Proof.

- Set $\alpha = \lambda/\sqrt{m}$.
- Let $Y_i = X_i - X_{i-1}$, so $|Y_i| \leq 1$ and $E[Y_i|X_0, \ldots, X_{i-1}] = 0$.
- By convexity we have

  $$E[e^{\alpha Y_i}|X_0, \ldots, X_{i-1}] \leq \cosh(\alpha) \leq e^{\alpha^2/2}.$$
Azuma’s inequality

Proof.

- Setting apart one variable at a time,

\[
E[e^{aX_m}] = E \left[ \prod_{i=1}^{m} e^{\alpha Y_i} \right] \\
= E \left[ \left( \prod_{i=1}^{m-1} e^{\alpha Y_i} \right) E[e^{\alpha Y_m} | X_0, ..., X_{m-1}] \right] \\
\leq e^{\alpha^2/2} E \left[ \prod_{i=1}^{m-1} e^{\alpha Y_i} \right] \leq e^{\alpha^2 m/2}.
\]
Azuma’s inequality

Proof.

Thus

\[
\begin{align*}
\text{Prob} \left( X_m > \lambda \sqrt{m} \right) &= \text{Prob} \left( e^{\alpha X_m} > e^{\alpha \lambda \sqrt{m}} \right) \\
&< E[e^{\alpha X_m}] e^{-\alpha \lambda \sqrt{m}} \\
&\leq e^{\alpha^2 m/2 - \alpha \lambda \sqrt{m}} = e^{-\lambda^2/2}.
\end{align*}
\]
Let $n \geq 1$ be an integer and $0 < p < 1$. The random graph $G(n, p)$ is a graph on $n$ vertices $\{1, 2, \ldots, n\}$ with each edge appearing i.i.d. with probability $p$.

Let $m = \binom{n}{2}$ and let the potential edges be $e_1, \ldots, e_m$.

Let $f$ be a function on graphs, and define a martingale $X_0, X_1, X_2, \ldots$ by setting $X_0$ to be the expectation of $f(G)$ when graph $G$ is sampled from $G(n, p)$.

Let $X_i$ be determined by deciding whether $e_1, \ldots, e_i$ belongs to $G$, then taking the expectation of $f(G)$ where the remaining edges are random.
Let $f$ be a function on graphs as before, and let $X_1 = \mathbb{E}[f(G)]$ when $G$ is sampled from $G(n, p)$

Define martingale $X_1, \ldots, X_n$ by letting $X_i$ be the conditional expectation in which all edges between vertices $j, k \leq i$ are deterministic, and all other edges are random.
The chromatic number $\chi(G)$ of a graph $G$ is the least number of colors needed to color the vertices of $G$ so that no edge is monochromatic.

**Theorem (Shamir and Spencer, 1987)**

Let $n \geq 1$ and $0 < p < 1$. Set $c = E[\chi(G)]$ when $G$ is sampled from $G(n, p)$. Then

$$\text{Prob} \left[ |\chi(G) - c| > \lambda \sqrt{n - 1} \right] < 2e^{-\lambda^2/2}.$$
Proof.

- Let $f(G) = \chi(G)$ be the chromatic number, and let $c = X_1, X_2, ..., X_n$ be the corresponding vertex exposure martingale.
- The bounded difference condition applies, since a single vertex can be given a new color.
- Hence the result follows from Azuma’s inequality.
Azuma’s inequality variant

The following slight generalization of Azuma’s inequality is sometimes useful.

**Theorem (Azuma’s inequality variant)**

Let $0 = X_0, X_1, \ldots, X_m$ be a martingale sequence, with differences $Y_i = X_i - X_{i-1}$. Assume that $\|Y_i\|_\infty < \infty$. Let

$$a = \left( \sum_{i=1}^{m} \|Y_i\|_\infty^2 \right)^{\frac{1}{2}}.$$

Let $\lambda > 0$. Then

$$\text{Prob}[|X_m| > \lambda] < 2e^{-\lambda^2/(2a^2)}.$$

The proof is essentially the same.
Khintchine’s inequality

Let $\epsilon_1, ..., \epsilon_n$ be i.i.d. Rademacher random variables ($\pm 1$ with equal probability) and let $\alpha_1, ..., \alpha_n$ be real constants. By independence,

$$E \left[ \left| \sum_{i=1}^{n} \epsilon_i \alpha_i \right|^2 \right] = \sum_{i=1}^{n} \alpha_i^2.$$  

Khintchine’s inequality gives the following approximate orthogonality in $L^p$.

**Theorem (Khintchine’s inequality)**

For any $0 < p < \infty$, there exist positive finite constants $A_p$ and $B_p$ depending on $p$ only such that for any finite sequence $(\alpha_i)$ of real numbers,

$$A_p \| \alpha_i \|_2 \leq \left( E \left| \sum_i \epsilon_i \alpha_i \right|^p \right)^{\frac{1}{p}} \leq B_p \| \alpha_i \|_2.$$
Khintchine’s inequality

Proof.

- Rescale so $\sum_i \alpha_i^2 = 1$.
- By the variant of Azuma,

$$E \left| \sum_i \epsilon_i \alpha_i \right|^p = \int_0^\infty \text{Prob} \left( \left| \sum_i \epsilon_i \alpha_i \right| > t \right) dt^p$$

$$\leq 2 \int_0^\infty \exp(-t^2/2) dt^p = B_p^p.$$
Proof.

By Jensen, it suffices to prove the lower bound $p < 2$

$$1 = E \left| \sum_i \epsilon_i \alpha_i \right|^2 = E \left( \left| \sum_i \epsilon_i \alpha_i \right|^{2p/3} \left| \sum_i \epsilon_i \alpha_i \right|^{2-2p/3} \right)$$

$$\leq \left( E \left| \sum_i \epsilon_i \alpha_i \right|^p \right)^{2/3} \left( E \left| \sum_i \epsilon_i \alpha_i \right|^{6-2p} \right)^{1/3}$$

$$\leq \left( E \left| \sum_i \epsilon_i \alpha_i \right|^p \right)^{2/3} B^{2-2p/3}_{6-2p}. $$
Metric examples

Definition

Let \((X, d)\) be a finite metric space. We say \((X, d)\) has \emph{length} at most \(\ell\) if there exists

- an \emph{increasing sequence}

\[
\{X\} = \mathcal{X}^0, \mathcal{X}^1, \ldots, \mathcal{X}^n = \{\{x\}\}_{x \in X}
\]

of partitions of \(X\), with \(\mathcal{X}^i\) a refinement of \(\mathcal{X}^{i-1}\)

- positive numbers \(a_1, \ldots, a_n\), with \(\ell = \left(\sum_{i=1}^{n} a_i^2\right)^{\frac{1}{2}}\), such that if

\[
\mathcal{X}^i = \{A^i_j\}_{1 \leq j \leq m}
\]

then for all \(A^i_j, A^i_k\) contained in some \(A^{i-1}_p\) there exists a bijection \(\phi: A^i_j \rightarrow A^i_k\) such that \(d(x, \phi(x)) \leq a_i\) for all \(x \in A^i_j\).

The length of a metric space is always at most its diameter.
Theorem

Let \((X, d)\) be a finite metric space of length at most \(\ell\), and let \(\mu\) be the uniform probability measure on \(X\). For every 1-Lipschitz function \(F\) on \((X, d)\) and every \(r \geq 0\),

\[
\mu \left( \left\{ F \geq \int F d\mu + r \right\} \right) \leq e^{-r^2/2\ell^2}.
\]
Proof.

- Let \( \mathcal{F}_i \) be the \( \sigma \)-field generated by \( \mathcal{X}_i \), and set \( F_i = \mathbb{E}[F|\mathcal{F}_i] \), which is a martingale sequence with \( F_0 = \int F d\mu \).
- Let \( B = A^i_j \), \( C = A^i_k \) be distinct atoms of \( \mathcal{F}_i \) contained in a single atom \( A^i_{p-1} \) of \( \mathcal{F}_{i-1} \).
- Thus \( F_i \) is constant on \( B, C \), and

\[
F_i|_C = \frac{1}{|C|} \sum_{x \in C} F(x) = \frac{1}{|B|} \sum_{x \in B} F(\phi(x))
\]

so that \( |F_i|_C - F_i|_B| \leq a_i \) by the 1-Lipschitz property.
- The conclusion follows from the variant of Azuma’s inequality.
Metric examples

- Consider the symmetric group $\mathfrak{S}_n$ on $n$ letters, given the metric, for $\sigma, \pi \in \mathfrak{S}_n$,

$$d(\sigma, \pi) = \frac{1}{n} \# \{ i : \sigma(i) \neq \pi(i) \}.$$ 

- Let $\mathcal{X}_i$ be the partition consisting of sets

$$A_{j_1, \ldots, j_i} = \{ \sigma \in \mathfrak{S}_n : \sigma(1) = j_1, \ldots, \sigma(i) = j_i \}.$$ 

- If $B, C \in \mathcal{X}_i$ satisfy $B, C \subset A \in \mathcal{X}_{i-1}$ then $B$ and $C$ differ only at place $i$, given by $j_i, j'_i$, say.

- Let $\phi$ be the relabeling that swaps $j_i$ and $j'_i$ in the image of the permutation, so that we may take all $a_i = \frac{2}{n}$ and $\ell = \frac{2}{\sqrt{n}}$. Note that the diameter is 1.
Metric examples

We obtain the following corollary for the symmetric group.

**Theorem**

Let $\mu$ be the uniform probability measure on $(\mathfrak{S}_n, d)$. For any 1-Lipschitz function $F$ on $(\mathcal{F}_n, d)$ and any $r \geq 0$,

$$
\mu \left( \left\{ F \geq \int F d\mu + r \right\} \right) \leq e^{-nr^2/8}.
$$
Example

Let $F(\sigma)$ be the number of transpositions $(i, j)$ required to reach permutation $\sigma$ from the identity. $F$ is $n$-Lipschitz, as may be seen by moving one coordinate into correct position at a time. Hence

$$\mu \left( \left\{ F \geq \int Fd\mu + r \right\} \right) \leq e^{-r^2/8n}$$

so $F$ is concentrated at a scale of $\sqrt{n}$ about its mean.
Consider (finite) probability spaces \((\Omega_i, \Sigma_i, \mu_i)_{i=1}^n\) with product measure \(P = \mu_1 \otimes \cdots \otimes \mu_n\) on \(X = \Omega_1 \times \cdots \times \Omega_n\).

Consider weighted Hamming metrics. Let \(a = (a_1, \ldots, a_n) \in \mathbb{R}_+^n\),

\[
|a|^2 = \sum_{i=1}^{n} a_i^2
\]

and

\[
d_a(x, y) = \sum_{i=1}^{n} a_i \mathbf{1}(x_i \neq y_i).
\]
Talagrand’s inequality

- Given a non-empty set $A \subseteq X$ and $x \in X$ define a distance
  \[ D_A(x) = \sup_{|a|=1} d_a(x, A). \]

- Let
  \[ U_A(x) = \{ s = (s_i)_{1 \leq i \leq n} \in \{0, 1\}^n : \exists y \in A, y_i = x_i \text{ if } s_i = 0 \}. \]

- Let $V_A(x)$ be the convex hull in $[0, 1]^n$ of $U_A(x)$. Note that
  $0 \in V_A(x)$ if and only if $x \in A$. 

Talagrand’s inequality

Lemma

We have

\[ D_A(x) = d(0, V_A(x)) = \inf_{y \in V_A(x)} |y|. \]
Talagrand’s inequality

Proof.

- If $d(0, V_A(x)) \leq r$, there exists $z \in V_A(x)$ with $|z| \leq r$. Let $a \in \mathbb{R}^n_+$ with $|a| = 1$. Then

$$\inf_{y \in V_A(x)} a \cdot y \leq a \cdot z \leq |z| \leq r.$$ 

- Since

$$\inf_{y \in V_A(x)} a \cdot y = \inf_{s \in U_A(x)} a \cdot s = d_a(x, A)$$

this proves $D_A(x) \leq r$. 

Bob Hough

Math 639: Lecture 22
May 5, 2017  27 / 61
Talagrand’s inequality

Proof.

- To prove the reverse direction, let \( z \in V_A(x) \) such that \( |z| = d(0, V_A(x)) > 0 \) and let \( a = \frac{z}{|z|} \).
- Let \( y \in V_A(x) \). Then for \( \theta \in [0, 1] \), \( \theta y + (1 - \theta)z \in V_A(x) \) so
  \[
  |z + \theta(y - z)|^2 = |\theta y + (1 - \theta)z|^2 \geq |z|^2.
  \]
- Letting \( \theta \to 0 \), \( (y - z) \cdot z \geq 0 \), so
  \[
  a \cdot y \geq |z| = d(0, V_A(x)).
  \]
- Hence
  \[
  D_A(x) \geq d_a(x, A) = \inf_{y \in V_A(x)} a \cdot y \geq d(0, V_A(x)).
  \]
Theorem (Talagrand’s inequality)

For every measurable non-empty subset $A$ of $X = \Omega^1 \times \cdots \times \Omega^n$, and every product probability $P$ on $X$,

$$
\int e^{D_A^2/4} \, dP \leq \frac{1}{P(A)}.
$$

In particular, for every $r \geq 0$,

$$
P(\{D_A \geq r\}) \leq \frac{e^{-r^2/4}}{P(A)}.
$$
Proof.

- Without loss of generality, let \((\Omega, \Sigma, \mu)\) be a prob. space and let \(P = \mu^n\) be the \(n\)-fold product on \(X = \Omega^n\).
- The proof is by induction. The case \(n = 1\) amounts to the inequality

\[ P(A)(1 - P(A)) \leq \frac{1}{4} < e^{-1/4}. \]

- To make the inductive step, let \(A \in \Omega^{n+1}\) and let \(B\) be the projection to \(\Omega^n\), forgetting the last coordinate.
- For \(\omega \in \Omega\) let \(A(\omega)\) be the section of \(A\) along \(\omega\).
Proof.

Given \( x \in \Omega^n \) and \( \omega \in \Omega \), write \( z = (x, \omega) \).

If \( s \in U_{A(\omega)} \) then \((s, 0) \in U_A(z)\). If \( t \in U_B(x) \) then \((t, 1) \in U_A(z)\).

Hence if \( \xi \in V_{A(\omega)}(x) \) and \( \zeta \in V_B(x) \) and \( 0 \leq \theta \leq 1 \) then
\[ (\theta \xi + (1 - \theta)\zeta, 1 - \theta) \in V_A(z). \]

By convexity,
\[ D_A(z)^2 \leq (1 - \theta)^2 + |\theta \xi + (1 - \theta)\zeta|^2 \]
\[ \leq (1 - \theta)^2 + \theta |\xi|^2 + (1 - \theta) |\zeta|^2. \]

So
\[ D_A(z)^2 \leq (1 - \theta)^2 + \theta D_{A(\omega)}(x)^2 + (1 - \theta) D_B(x)^2. \]
Talagrand’s inequality

Proof. By Hölder’s inequality and the induction hypothesis, for fixed $\omega \in \Omega$,

$$
\int_{\Omega^n} e^{D_A(x,\omega)^2/4} dP(x) \leq e^{(1-\theta)^2} \left( \int_{\Omega^n} e^{D_A(\omega)^2/4} dP \right)^{\theta} \left( \int_{\Omega^n} e^{D_B^2/4} dP \right)^{1-\theta} \\
\leq e^{(1-\theta)^2} \left( \frac{1}{P(A(\omega)))} \right)^{\theta} \left( \frac{1}{P(B)} \right)^{1-\theta} \\
= \frac{1}{P(B)} e^{(1-\theta)^2} \left( \frac{P(A(\omega)))}{P(B)} \right)^{-\theta}.
$$

Use $\inf_{\theta \in [0,1]} e^{(1-\theta)^2} u^{-\theta} \leq 2 - u$, so

$$
\int_{\Omega^n} e^{D_A(x,\omega)^2/4} dP(x) \leq \frac{1}{P(B)} \left( 2 - \frac{P(A(\omega)))}{P(B)} \right).
$$
Proof.

- Use $u(2 - u) \leq 1$ and integrate in $\omega$ to find

$$
\int_{\Omega^{n+1}} e^{D_A(x,\omega)^2/4} dP(x) d\mu(\omega) \leq \frac{1}{P(B)} \left( 2 - \frac{P \otimes \mu(A)}{P(B)} \right)
$$

$$
\leq \frac{1}{P \otimes \mu(A)}.
$$
Consider points $x_1, \ldots, x_n \in [0, 1]$.

Denote by $L_n(x_1, \ldots, x_n) = L_n(x)$ the length of the longest increasing subsequence, that is, the largest $p$ so that there exist $i_1 < i_2 < \cdots < i_p$ with

$$x_{i_1} < x_{i_2} < \cdots < x_{i_p}.$$

When $U_1, \ldots, U_n$ are i.i.d. uniform on $[0, 1]$, $L_n(U_1, \ldots, U_n)$ has the same distribution as the longest increasing sequence in a random permutation.
Given $s \geq 0$, let $A = A_s = \{ x \in [0, 1]^n : L_n(x) \leq s \}$. We have

$$s \geq L_n(x) - D_A(x) \sqrt{L_n(x)}.$$  

In particular,

$$D_A(x) \geq \frac{u}{\sqrt{s + u}}$$

whenever $L_n(x) \geq s + u$.  


Longest increasing subsequence

Proof.

- Let $l \subseteq \{1, 2, \ldots, n\}$ with $|l| = L_n(x)$ such that if $i, j \in l$ with $i < j$ then $x_i < x_j$.
- Choose a supported on $l$ with value $a|l| \equiv \frac{1}{\sqrt{L_n(x)}}$ to find that there exists $y \in A$ such that $J = \{i \in l : y_i \neq x_i\}$ satisfies
  $$|J| \leq D_A \sqrt{L_n(x)}.$$  

- It follows that $(x_i)_{i \in l \setminus J}$ is an increasing subsequence of $y$, which proves the first part of the lemma.
- The second part of the lemma follows from $D_A \geq \frac{L_n(x) - s}{\sqrt{L_n(x)}}$ since $u \mapsto \frac{u-s}{\sqrt{u}}$ is increasing in $u \geq s$. 

\[ \square \]
Longest increasing subsequence

**Theorem**

Let $m_n$ be a median of $L_n = L_n(U_1, ..., U_n)$, so $P(L_n > m_n) \leq 1/2$ and $P(L_n < m_n) \leq 1/2$. For every $r \geq 0$, 

\[
P(\{L_n \geq m_n + r\}) \leq 2e^{-r^2/4(m_n+r)}
\]

\[
P(\{L_n \leq m_n - r\}) \leq 2e^{-r^2/4m_n}
\]

so, in particular, for $0 \leq r \leq m_n$, 

\[
P(\{|L_n - m_n| \geq r\}) \leq 4e^{-r^2/8m_n}.
\]
**Proof.**

- Let $A = \{ x : L_n(x) \leq m_n \}$ and let $B = \{ x : L_n(x) \geq m_n + r \}$.
- By Talagrand’s inequality,
  \[
  \int_B e^{D_A^2/4} \leq \frac{1}{P(A)} \leq 2.
  \]
- $D_A \geq \frac{r}{\sqrt{m_n+r}}$ on $B$, the first bound follows.
- Now let $A = \{ x : L_n(x) \leq m_n - r \}$ and $B = \{ x : L_n(x) \geq m_n \}$ so that $D_A(x) \geq \frac{r}{\sqrt{m_n}}$ on $B$, so
  \[
  \frac{1}{2} \leq P(B) \leq \frac{e^{-r^2/4m_n}}{P(A)}.
  \]
Lipschitz functions

Definition

Let $X = \Omega_1 \times \cdots \times \Omega_n$. We say that a function $F : X \to \mathbb{R}$ is 1-Lipschitz in the sense of Talagrand, if for every $x \in X$ there exists $a = a(x)$ such that, for every $y \in X$,

$$F(x) \leq F(y) + d_a(x, y).$$
Talagrand’s inequality for Lipschitz functions

**Theorem**

Let $P$ be a product probability measure on the space $X = \Omega_1 \times \cdots \times \Omega_n$, and let $F : X \to \mathbb{R}$ be 1-Lipschitz in the sense of Talagrand. Let $m_F$ be a median for $F$, so that $P(F \geq m_F), P(F \leq m_F) \geq \frac{1}{2}$. Then, for every $r \geq 0$,

$$P(\{|F - m_F| \geq r\}) \leq 4e^{-r^2/4}.$$
Proof.

- Let $A = \{ F \leq m_F \}$.
- By the 1-Lipschitz property, for each $x$ there exists $a = a(x)$ such that
  \[ F(x) \leq m_F + d_a(x, A) \leq m_F + D_A(x). \]
- Hence, by Talagrand’s inequality,
  \[ P(\{ F \geq m_F + r \}) \leq P(\{ D_A \geq r \}) \leq \frac{e^{-r^2/4}}{P(A)} \leq 2e^{-r^2/4}. \]
- To bound the lower tail, argue similarly, replacing $m_F$ with $m_F - r$. 
Suprema of linear functionals

- Let $Y_1, \ldots, Y_n$ be independent random variables taking values in $[0, 1]$
- Let
  \[ Z = \sup_{t \in \mathcal{T}} \sum_{i=1}^{n} t_i Y_i \]
  where $\mathcal{T}$ is a finite family of vectors $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$.
- Let $\sigma = \sup_{t \in \mathcal{T}} |t|_2$. 

Bob Hough  
Math 639: Lecture 22  
May 5, 2017  
42 / 61
Suprema of linear functionals

- Let $X = [0, 1]^n$ with $P$ the product measure of the laws of the $Y_i$, and, for $x \in X$, $F(x) = \sup_{t \in \mathcal{F}} \sum_{i=1}^{n} t_i x_i$.

- Given $x = (x_1, \ldots, x_n) \in X$, let $t = t(x)$ achieve the supremum of $F(x)$. Then, for all $y \in X$,

$$F(x) = \sum_{i=1}^{n} t_i x_i \leq \sum_{i=1}^{n} t_i y_i + \sum_{i=1}^{n} |t_i| |x_i - y_i|$$

$$\leq F(y) + \sigma \sum_{i=1}^{n} \frac{|t_i|}{\sigma} 1(x_i \neq y_i).$$

It follows that $\sigma^{-1} F$ is 1-Lipschitz in the sense of Talagrand, by choosing $a = a(x) = \sigma^{-1}(|t_1|, \ldots, |t_n|)$. 
Suprema of linear functionals

We obtain the following corollary.

**Corollary**

Let \( \{Y_i\}_{i=1}^{n} \) be independent random variables taking values in [0, 1], let \( \mathcal{T} \) be a finite family of linear functionals on \( \mathbb{R}^n \) bounded in \( \ell^2 \) by \( \sigma \), and let

\[
    Z = \sup_{t \in \mathcal{T}} \sum_{i=1}^{n} t_i Y_i.
\]

Let \( m_Z \) be a median of \( Z \). Then, for every \( r \geq 0 \),

\[
P(\{|Z - m_Z| \geq r\}) \leq 4e^{-r^2/4\sigma^2}.
\]
Theorem

Let $G = (V, E)$ be a graph. Let $(Y_e)_{e \in E}$ be i.i.d. random variables (passage times) taking values in $[0, 1]$. Let $\mathcal{T}$ be a set of subsets of $E$. Given $T \in \mathcal{T}$, let $Y_T = \sum_{e \in T} Y_e$. Define

$$Z_{\mathcal{T}} = \inf_{T \in \mathcal{T}} Y_T = \inf_{T \in \mathcal{T}} \sum_{e \in T} Y_e.$$ 

Let $D = \sup_{T \in \mathcal{T}} |T|$ and let $m$ be a median of $Z_{\mathcal{T}}$. Then, for each $r > 0$,

$$P(\{|Z_{\mathcal{T}} - m| \geq r\}) \leq 4e^{-r^2/4D}.$$ 

The set $\mathcal{T}$ could be taken to be a collection of paths connecting a pair of vertices $x, y$. $Z_{\mathcal{T}}$ is then the lowest cost path among these.
Further applications

Talagrand’s method may also be used to prove concentration for the traveling salesman problem, and minimum length spanning tree for random collections of points in $[0, 1]^2$. 
Concentration in Gauss space

**Definition**

Denote $\gamma_N(dx) = (2\pi)^{-N/2} \exp(-|x|^2/2)dx$ the Gaussian measure on $\mathbb{R}^N$. Define the usual Lipschitz norm of a real function $f$ on $\mathbb{R}^N$,

$$\|f\|_{\text{Lip}} = \sup \left\{ \left| \frac{f(x) - f(y)}{|x - y|} \right| : x, y \in \mathbb{R}^N \right\}.$$

We say a function is Lipschitz if it has finite Lipschitz norm.
Concentration in Gauss space

**Theorem**

*Given Lipschitz function $f$, let*

$$E_f = \int_{\mathbb{R}^N} f(x) d\gamma_N.$$  

*For any $t \geq 0$,*

$$\gamma_N(|f - E_f| > t) \leq 2 \exp\left(-2t^2/\pi^2\|f\|_{\text{Lip}}^2\right).$$

*With more care, the constant $\frac{2}{\pi^2}$ can be replaced with $\frac{1}{2}$ in the exponential.*
Proof.

- Let $f$ Lipschitz on $\mathbb{R}^N$, so $f$ is a.e. differentiable and satisfies $|\nabla f| \leq \|f\|_{\text{Lip}}$.
- Shifting by a constant, assume $\int f d\gamma_N = 0$.
- By convexity

$$
\gamma_N(f > t) \leq \exp(-\lambda t) \int \exp(\lambda f) d\gamma_N \\
\leq \exp(-\lambda t) \int \int \exp[\lambda(f(x) - f(y))] d\gamma_N(x) d\gamma_N(y).
$$
Proof.

- Given $x, y \in \mathbb{R}^n$, let

$$x(\theta) = x \sin \theta + y \cos \theta, \quad x'(\theta) = x \cos \theta - y \sin \theta$$

so that

$$f(x) - f(y) = \int_0^{\pi/2} \frac{d}{d\theta} f(x(\theta)) d\theta = \int_0^{\pi/2} \langle \nabla f(x(\theta)), x'(\theta) \rangle d\theta.$$ 

- By Jensen, $\gamma_N(f > t)$ is bounded by

$$\exp(-\lambda t) \frac{2}{\pi} \int_0^{\pi/2} \int \int \exp \left[ \frac{\lambda \pi}{2} \langle \nabla f(x(\theta)), x'(\theta) \rangle \right] d\gamma_N(x) d\gamma_N(y) d\theta$$
Concentration in Gauss space

Proof.

- For fixed $\theta$, the distribution of $(x(\theta), x'(\theta))$ is the same as the distribution of $x, y$. Hence

$$\gamma_N(f > t) \leq \exp(-\lambda t) \int \int \exp\left[\frac{\lambda \pi}{2} \langle \nabla f(x), y \rangle \right] d\gamma_N(x) d\gamma_N(y)$$

$$\leq \exp(-\lambda t) \int \exp\left(\frac{\lambda^2 \pi^2}{8} |\nabla f|^2 \right) d\gamma_n$$

$$\leq \exp\left(-\lambda t + \frac{\lambda^2 \pi^2}{8} \|f\|_{\text{Lip}}^2 \right).$$

- Choose $\lambda = \frac{4t}{\pi^2 \|f\|_{\text{Lip}}^2}$ to obtain

$$\gamma_N(f > t) \leq \exp(-2t^2/\pi^2 \|f\|_{\text{Lip}}^2).$$
Definition

Given a probability space \((\Omega, \Sigma, \mu)\) and a non-negative measurable \(f\), define it's entropy

\[
\operatorname{Ent}_\mu(f) = \int f \log fd\mu - \int fd\mu \log \int fd\mu
\]

where \(\int f (\log 1 + f) d\mu < \infty\) and \(\infty\) otherwise.

This is homogeneous of degree 1.
Definition

We say a Borel probability measure \( \mu \) on \( \mathbb{R}^n \) satisfies a *logarithmic Sobolev inequality* with constant \( C > 0 \) if, for all smooth enough functions \( f \),

\[
\text{Ent}_{\mu}(f^2) \leq 2C \int |\nabla f|^2 d\mu.
\]
Abbreviate $\gamma$ the Gaussian measure on $\mathbb{R}^n$.

**Theorem**

*For every smooth enough function $f$ on $\mathbb{R}^n$,*

$$\text{Ent}_\gamma(f^2) \leq 2 \int |\nabla f|^2 d\gamma.$$
Log Sobolev inequalities

Proof.

Let \((P_t)_{t \geq 0}\) denote the Ornstein-Uhlenbeck semigroup, which has integral representation

\[
P_t f(x) = \int f(e^{-t}x + (1 - e^{-2t})y) d\gamma(y), \quad t \geq 0, x \in \mathbb{R}^n.
\]

Let \(f\) be smooth and non-negative, satisfying \(\epsilon \leq f \leq 1/\epsilon\).

Since \(P_0 f = f\) and \(\lim_{t \to \infty} P_t f = \int f d\gamma\),

\[
\text{Ent}_\gamma(f) = -\int_0^\infty \frac{d}{dt} \left( \int P_t f \log P_t f d\gamma \right) dt
\]
Log Sobolev inequalities

Proof.

- We have $P_t = e^{tL}$ where $L = \Delta - x \cdot \nabla$. The second order differential operator $L$ satisfies, for smooth $f, g$,

$$
\int f(Lg) d\gamma = - \int \nabla f \cdot \nabla g d\gamma.
$$

- Hence

$$
\frac{d}{dt} \int P_t f \log P_t f d\gamma = \int LP_t f \log P_t f d\gamma + \int LP_t f d\gamma
$$

$$
= - \int \frac{|\nabla P_t f|^2}{P_t f} d\gamma.
$$
Proof.

- Calculate, from the integral representation,
  \[ \nabla P_t f = e^{-t} P_t(\nabla f) \Rightarrow |\nabla P_t f| \leq e^{-t} P_t(|\nabla f|). \]

- By Cauchy-Schwarz,
  \[ P_t(|\nabla f|)^2 \leq P_t(f) P_t \left( \frac{|\nabla f|^2}{f} \right). \]

- Combining these steps,
  \[ \text{Ent}_\gamma(f) \leq \int_0^\infty e^{-2t} \left( \int P_t \left( \frac{|\nabla f|^2}{f} \right) d\gamma \right) dt = \frac{1}{2} \int \frac{|\nabla f|^2}{f} d\gamma. \]
  
  The conclusion follows on replacing \( f \) with \( f^2 \) and letting \( \epsilon \downarrow 0 \).
We can now use the Log Sobolev inequality satisfied by Gaussian measure to obtain the sharper constant in Gaussian concentration.

**Theorem**

*Let $F$ be a 1-Lipschitz function on $\mathbb{R}^n$. Then*

$$
\gamma \left( \left\{ F \geq \int F d\gamma + r \right\} \right) \leq e^{-r^2/2}.
$$
Log Sobolev inequalities

The following argument is due to Herbst.

Proof.

- Let $F$ be a 1-Lipschitz function, satisfying $|\nabla F| \leq \|F\|_{\text{Lip}} = 1$ a.e.
- Assume, as we may, that $\int Fd\gamma = 0$.
- Consider $f^2 = e^{\lambda F - \lambda^2/2}$. We have

$$
\int |\nabla f|^2 d\gamma = \frac{\lambda^2}{4} \int |\nabla F|^2 e^{\lambda F - \lambda^2/2} d\gamma \leq \frac{\lambda^2}{4} \int e^{\lambda F - \lambda^2/2} d\gamma.
$$

\square
Log Sobolev inequalities

Proof.

Let $\Lambda(\lambda) = \int e^{\lambda F - \frac{\lambda^2}{2}} d\gamma$. By log-Sob,

$$\int \left[ \lambda F - \frac{\lambda^2}{2} \right] e^{\lambda F - \frac{\lambda^2}{2}} d\gamma - \Lambda(\lambda) \log \Lambda(\lambda) \leq \frac{1}{2} \lambda^2 \Lambda(\lambda).$$

which rearranges to

$$\lambda \Lambda'(\lambda) \leq \Lambda(\lambda) \log \Lambda(\lambda) \iff \lambda \frac{\Lambda'(\lambda)}{\Lambda(\lambda)} \leq \log \Lambda(\lambda).$$

It follows that $H(\lambda) = \frac{\log \Lambda(\lambda)}{\lambda}$ if $\lambda > 0$, $H(0) = \frac{\Lambda'(0)}{\Lambda(0)} = \int F d\gamma = 0$ satisfies $H'(\lambda) \leq 0$. Hence $\Lambda(\lambda) \leq 1$. 
Proof.

- We’ve checked, for all $\lambda$,

$$
\int e^{\lambda F} \, d\gamma \leq e^{\frac{\lambda^2}{2}}
$$

- Hence $P(F \geq r) \leq e^{-\lambda r + \lambda^2/2}$. Choosing $\lambda = r$ proves the claim.