Math 639: Lecture 22

Concentration of measure

Bob Hough

May 5, 2017

Concentration of measure

This lecture is drawn from:

- M. Ledoux. *The concentration of measure phenomenon*. AMS 89, 2001.
- M. Ledoux and M. Talagrand. *Probability in Banach spaces*. Springer, 1991.
- N. Alon and J. Spencer. *The probabilistic method*. Wiley, 2016.

Chernoff's inequality

Theorem

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

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

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

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

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

Chernoff's inequality

Lemma

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

Proof.

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

$$\mathrm{E}[e^{tX}] \leqslant 1 + t^2 \operatorname{Var}[X] \leqslant \exp(t^2 \operatorname{Var}[X]).$$

$\square$

Chernoff's inequality

Proof of Chernoff's inequality.

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

$$\begin{aligned}
\mathrm{Prob}(X \geqslant \lambda\sigma) &\leqslant e^{-t\lambda\sigma}\,\mathrm{E}[e^{tX_1}\cdots e^{tX_n}] \\
&= e^{-t\lambda\sigma}\,\mathrm{E}[e^{tX_1}]\cdots\mathrm{E}[e^{tX_n}] \\
&\leqslant e^{-t\lambda\sigma}\exp(t^2(\mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n])) \\
&= \exp(t^2\sigma^2 - t\lambda\sigma).
\end{aligned}$$

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

$\square$

Azuma's inequality

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

Theorem (Azuma's inequality)

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

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

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

$$\mathrm{Prob}\left[X_m > \lambda\sqrt{m}\right] < e^{-\lambda^2/2}.$$

Azuma's inequality

Proof.

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

$$\mathrm{E}[e^{\alpha Y_i} | X_0, ..., X_{i-1}] \leqslant \cosh(\alpha) \leqslant e^{\alpha^2/2}.$$

□

Azuma's inequality

Proof.

- Setting apart one variable at a time,

$$\begin{aligned}
\mathrm{E}[e^{aX_m}] &= \mathrm{E}\left[\prod_{i=1}^{m} e^{\alpha Y_i}\right] \\
&= \mathrm{E}\left[\left(\prod_{i=1}^{m-1} e^{\alpha Y_i}\right) \mathrm{E}[e^{\alpha Y_m} | X_0, ..., X_{m-1}]\right] \\
&\leqslant e^{\alpha^2/2}\,\mathrm{E}\left[\prod_{i=1}^{m-1} e^{\alpha Y_i}\right] \leqslant e^{\alpha^2 m/2}.
\end{aligned}$$

□

Proof.

- Thus

$$\begin{aligned}\mathrm{Prob}\left(X_m > \lambda\sqrt{m}\right) &= \mathrm{Prob}\left(e^{\alpha X_m} > e^{\alpha\lambda\sqrt{m}}\right)\\ &< \mathrm{E}[e^{\alpha X_m}]e^{-\alpha\lambda\sqrt{m}}\\ &\leqslant e^{\alpha^2 m/2-\alpha\lambda\sqrt{m}} = e^{-\lambda^2/2}.\end{aligned}$$

□

Edge exposure martingale

- Let $n \geqslant 1$ be an integer and $0 < p < 1$. The random graph $G(n, p)$ is a graph on n vertices $\{1, 2, ..., n\}$ with each edge appearing i.i.d. with probability p.
- Let $m = \binom{n}{2}$ and let the potential edges be $e_1, ..., e_m$.
- Let f be a function on graphs, and define a martingale $X_0, X_1, X_2, ...$ 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 $e_1, ..., e_i$ belongs to G, then taking the expectation of $f(G)$ where the remaining edges are random.

Vertex exposure martingale

- Let f be a function on graphs as before, and let $X_1 = \mathrm{E}[f(G)]$ when G is sampled from $G(n,p)$
- Define martingale $X_1, ..., X_n$ by letting X_i be the conditional expectation in which all edges between vertices $j, k \leqslant i$ are deterministic, and all other edges are random.

The chromatic number of a random graph

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 $n \geqslant 1$ *and* $0 < p < 1$. *Set* $c = \mathrm{E}[\chi(G)]$ *when* G *is sampled from* $G(n, p)$. *Then*

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

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.

$\square$

Azuma's inequality variant

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

Theorem (Azuma's inequality variant)

Let $0 = X_0, X_1, ..., X_m$ *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*

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

The proof is essentially the same.

Khintchine's inequality

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

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

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 (α_i) of real numbers,

$$A_p \|\alpha_i\|_2 \leqslant \left(\mathrm{E}\left|\sum_i \epsilon_i \alpha_i\right|^p\right)^{\frac{1}{p}} \leqslant B_p \|\alpha_i\|_2.$$

Khintchine's inequality

Proof.

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

$$\mathrm{E}\left|\sum_i \epsilon_i \alpha_i\right|^p = \int_0^\infty \mathrm{Prob}\left(\left|\sum_i \epsilon_i \alpha_i\right| > t\right) dt^p$$
$$\leqslant 2\int_0^\infty \exp(-t^2/2) dt^p = B_p^p.$$

$\square$

Khintchine's inequality

Proof.

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

$$
\begin{aligned}
1 = \mathrm{E}\left|\sum_i \epsilon_i \alpha_i\right|^2 &= \mathrm{E}\left(\left|\sum_i \epsilon_i\alpha_i\right|^{2p/3}\left|\sum_i \epsilon_i\alpha_i\right|^{2-2p/3}\right)\\
&\leqslant \left(\mathrm{E}\left|\sum_i \epsilon_i\alpha_i\right|^p\right)^{2/3}\left(\mathrm{E}\left|\sum_i \epsilon_i\alpha_i\right|^{6-2p}\right)^{1/3}\\
&\leqslant \left(\mathrm{E}\left|\sum_i \epsilon_i\alpha_i\right|^p\right)^{2/3} B_{6-2p}^{2-2p/3}.
\end{aligned}
$$

□

Metric examples

Definition

Let (X,d) be a finite metric space. We say (X,d) has *length* at most ℓ if there exists

- an *increasing sequence*

$$\{X\} = \mathscr{X}^0, \mathscr{X}^1, ..., \mathscr{X}^n = \{\{x\}\}_{x\in X}$$

  of partitions of X, with $\mathscr{X}^i$ a refinement of $\mathscr{X}^{i-1}$

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

$$\mathscr{X}^i = \{A_j^i\}_{1\leqslant j\leqslant m}$$

  then for all A_j^i, A_k^i contained in some A_p^{i-1} there exists a bijection $\phi: A_j^i \to A_k^i$ such that $d(x,\phi(x)) \leqslant a_i$ for all $x \in A_j^i$.

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

Metric examples

Theorem

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

$$\mu\left(\left\{F \geqslant \int F d\mu + r\right\}\right) \leqslant e^{-r^2/2\ell^2}.$$

Metric examples

Proof.

- Let $\mathscr{F}_i$ be the σ-field generated by $\mathscr{X}^i$, and set $F_i = \mathrm{E}[F|\mathscr{F}_i]$, 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 $\mathscr{F}_i$ contained in a single atom A^{i-1}_p of $\mathscr{F}_{i-1}$.
- 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 $|F_i|_C - F_i|_B| \leqslant a_i$ by the 1-Lipschitz property.
- The conclusion follows from the variant of Azuma's inequality.

$\square$

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

We obtain the following corollary for the symmetric group.

Theorem

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

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

Metric examples

Example

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

$$\mu\left(\left\{F \geqslant \int F d\mu + r\right\}\right) \leqslant e^{-r^2/8n}$$

so F is concentrated at a scale of $\sqrt{n}$ about its mean.

Talagrand's inequality

- 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 $a = (a_1, ..., a_n) \in \mathbb{R}_+^n$,

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

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 $A \subset X$ and $x \in X$ define a distance

$$D_A(x) = \sup_{|a|=1} d_a(x, A).$$

- Let

$$U_A(x) = \{s = (s_i)_{1 \leqslant i \leqslant n} \in \{0,1\}^n : \exists y \in A, y_i = x_i \text{ if } s_i = 0\}.$$

- 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 $d(0, V_A(x)) \leqslant r$, there exists $z \in V_A(x)$ with $|z| \leqslant r$. Let $a \in \mathbb{R}_+^n$ with $|a| = 1$. Then
$$\inf_{y \in V_A(x)} a \cdot y \leqslant a \cdot z \leqslant |z| \leqslant r.$$
- 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 $D_A(x) \leqslant r$.

$\square$

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
$$|z + \theta(y-z)|^2 = |\theta y + (1-\theta)z|^2 \geqslant |z|^2.$$
- Letting $\theta \to 0$, $(y-z)\cdot z \geqslant 0$, so
$$a \cdot y \geqslant |z| = d(0, V_A(x)).$$
- Hence
$$D_A(x) \geqslant d_a(x, A) = \inf_{y \in V_A(x)} a \cdot y \geqslant d(0, V_A(x)).$$

$\square$

Talagrand's inequality

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,

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

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

$$P(\{D_A \geqslant r\}) \leqslant \frac{e^{-r^2/4}}{P(A)}.$$

Talagrand's inequality

Proof.

- Without loss of generality, let (Ω, Σ, μ) 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

$$P(A)(1 - P(A)) \leqslant \frac{1}{4} < e^{-1/4}.$$

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

$\square$

Talagrand's inequality

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 $0 \leqslant \theta \leqslant 1$ then $(\theta\xi + (1-\theta)\zeta, 1-\theta) \in V_A(z)$.
- By convexity,

$$\begin{aligned} D_A(z)^2 &\leqslant (1-\theta)^2 + |\theta\xi + (1-\theta)\zeta|^2 \\ &\leqslant (1-\theta)^2 + \theta|\xi|^2 + (1-\theta)|\zeta|^2. \end{aligned}$$

so

$$D_A(z)^2 \leqslant (1-\theta)^2 + \theta D_{A(\omega)}(x)^2 + (1-\theta) D_B(x)^2.$$

$\square$

Talagrand's inequality

Proof.

- By Hölder's inequality and the induction hypothesis, for fixed $\omega \in \Omega$,

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

- Use $\inf_{\theta \in [0,1]} e^{\frac{(1-\theta)^2}{4}} u^{-\theta} \leqslant 2 - u$, so

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

□

Talagrand's inequality

Proof.

- Use $u(2-u) \leqslant 1$ and integrate in ω to find

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

□

Longest increasing subsequence

- Consider points $x_1, ..., x_n \in [0, 1]$.
- Denote by $L_n(x_1, ..., x_n) = L_n(x)$ the length of the longest increasing subsequence, that is, the largest p so that there exist $i_1 < i_2 < ... < i_p$ with

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

- When $U_1, ..., U_n$ are i.i.d. uniform on $[0, 1]$, $L_n(U_1, ..., U_n)$ has the same distribution as the longest increasing sequence in a random permutation.

Longest increasing subsequence

Lemma

Given $s \geqslant 0$, *let* $A = A_s = \{x \in [0,1]^n : L_n(x) \leqslant s\}$. *We have*

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

In particular,

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

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

Longest increasing subsequence

Proof.

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

$$|J| \leqslant D_A \sqrt{L_n(x)}.$$

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

□

Longest increasing subsequence

Theorem

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

$$P(\{L_n \geqslant m_n + r\}) \leqslant 2e^{-r^2/4(m_n+r)}$$
$$P(\{L_n \leqslant m_n - r\}) \leqslant 2e^{-r^2/4m_n}$$

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

$$P(\{|L_n - m_n| \geqslant r\}) \leqslant 4e^{-r^2/8m_n}.$$

Longest increasing subsequence

Proof.

- Let $A = \{x : L_n(x) \leqslant m_n\}$ and let $B = \{x : L_n(x) \geqslant m_n + r\}$.
- By Talagrand's inequality,

$$\int_B e^{D_A^2/4} \leqslant \frac{1}{P(A)} \leqslant 2.$$

- $D_A \geqslant \frac{r}{\sqrt{m_n+r}}$ on B, the first bound follows.
- Now let $A = \{x : L_n(x) \leqslant m_n - r\}$ and $B = \{x : L_n(x) \geqslant m_n\}$ so that $D_A(x) \geqslant \frac{r}{\sqrt{m_n}}$ on B, so

$$\frac{1}{2} \leqslant P(B) \leqslant \frac{e^{-r^2/4m_n}}{P(A)}.$$

$\square$

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 $y \in X$,

$$F(x) \leqslant F(y) + d_a(x, y).$$

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 $P(F \geqslant m_F), P(F \leqslant m_F) \geqslant \frac{1}{2}$. Then, for every $r \geqslant 0$,

$$P(\{|F - m_F| \geqslant r\}) \leqslant 4e^{-r^2/4}.$$

Talagrand's inequality for Lipschitz functions

Proof.

- Let $A = \{F \leqslant m_F\}$.
- By the 1-Lipschitz property, for each x there exists $a = a(x)$ such that

$$F(x) \leqslant m_F + d_a(x, A) \leqslant m_F + D_A(x).$$

- Hence, by Talagrand's inequality,

$$P(\{F \geqslant m_F + r\}) \leqslant P(\{D_A \geqslant r\}) \leqslant \frac{e^{-r^2/4}}{P(A)} \leqslant 2e^{-r^2/4}.$$

- 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
$$Z = \sup_{t \in \mathscr{T}} \sum_{i=1}^{n} t_i Y_i$$
where $\mathscr{T}$ is a finite family of vectors $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$.
- Let $\sigma = \sup_{t \in \mathscr{T}} |t|_2$.

Suprema of linear functionals

- Let $X = [0,1]^n$ with P the product measure of the laws of the Y_i, and, for $x \in X$, $F(x) = \sup_{t \in \mathscr{T}} \sum_{i=1}^n t_i x_i$.
- Given $x = (x_1, ..., x_n) \in X$, let $t = t(x)$ achieve the supremum of $F(x)$. Then, for all $y \in X$,

$$\begin{aligned} F(x) = \sum_{i=1}^n t_i x_i &\leqslant \sum_{i=1}^n t_i y_i + \sum_{i=1}^n |t_i||x_i - y_i| \\ &\leqslant F(y) + \sigma \sum_{i=1}^n \frac{|t_i|}{\sigma} \mathbf{1}(x_i \neq y_i). \end{aligned}$$

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

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 $\mathscr{T}$ be a finite family of linear functionals on $\mathbb{R}^n$ bounded in ℓ^2 by σ, and let

$$Z = \sup_{t \in \mathscr{T}} \sum_{i=1}^n t_i Y_i.$$

Let m_Z be a median of Z. Then, for every $r \geqslant 0$,

$$P(\{|Z - m_Z| \geqslant r\}) \leqslant 4e^{-r^2/4\sigma^2}.$$

First passage percolation

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 $\mathscr{T}$ be a set of subsets of E. Given $T \in \mathscr{T}$, let $Y_T = \sum_{e \in T} Y_e$. Define

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

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

$$P(\{|Z_{\mathscr{T}} - m| \geqslant r\}) \leqslant 4e^{-r^2/4D}.$$

The set $\mathscr{T}$ could be taken to be a collection of paths connecting a pair of vertices x, y. $Z_{\mathscr{T}}$ 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 $\mathbb{R}^N$,

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

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 $t \geqslant 0$,

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

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

Concentration in Gauss space

Proof.

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

$$\begin{aligned}\gamma_N(f > t) &\leqslant \exp(-\lambda t) \int \exp(\lambda f) d\gamma_N \\ &\leqslant \exp(-\lambda t) \int \int \exp[\lambda(f(x) - f(y))] d\gamma_N(x) d\gamma_N(y).\end{aligned}$$

□

Concentration in Gauss space

Proof.

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

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

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

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

□

Concentration in Gauss space

Proof.

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

$$\gamma_N(f > t) \leqslant \exp(-\lambda t) \iint \exp\left[\frac{\lambda\pi}{2}\langle \nabla f(x), y\rangle\right] d\gamma_N(x) d\gamma_N(y)$$
$$\leqslant \exp(-\lambda t) \int \exp\left(\frac{\lambda^2\pi^2}{8}|\nabla f|^2\right) d\gamma_n$$
$$\leqslant \exp\left(-\lambda t + \frac{\lambda^2\pi^2}{8}\|f\|_{\mathrm{Lip}}^2\right).$$

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

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

□

Log Sobolev inequalities

Definition

Given a probability space (Ω, Σ, μ) and a non-negative measurable f, define it's entropy

$$\mathrm{Ent}_\mu(f) = \int f \log f d\mu - \int f d\mu \log \int f d\mu$$

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

This is homogeneous of degree 1.

Log Sobolev inequalities

Definition

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

$$\mathrm{Ent}_\mu(f^2) \leqslant 2C \int |\nabla f|^2 d\mu.$$

Log Sobolev inequalities

Abbreviate γ the Gaussian measure on $\mathbb{R}^n$.

Theorem

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

$$\mathrm{Ent}_\gamma(f^2) \leqslant 2 \int |\nabla f|^2 d\gamma.$$

Log Sobolev inequalities

Proof.

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

$$P_t f(x) = \int f(e^{-t}x + (1-e^{-2t})y)d\gamma(y), \qquad t \geqslant 0, x \in \mathbb{R}^n.$$

- Let f be smooth and non-negative, satisfying $\epsilon \leqslant f \leqslant 1/\epsilon$.
- Since $P_0 f = f$ and $\lim_{t\to\infty} P_t f = \int f d\gamma$,

$$\mathrm{Ent}_\gamma(f) = -\int_0^\infty \frac{d}{dt}\left(\int P_t f \log P_t f d\gamma\right) dt$$

□

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,

$$\nabla P_t f = e^{-t} P_t(\nabla f) \Rightarrow |\nabla P_t f| \leqslant e^{-t} P_t(|\nabla f|).$$

- By Cauchy-Schwarz,

$$P_t(|\nabla f|)^2 \leqslant P_t(f) P_t \left(\frac{|\nabla f|^2}{f} \right).$$

- Combining these steps,

$$\mathrm{Ent}_\gamma(f) \leqslant \int_0^\infty e^{-2t} \left(\int P_t \left(\frac{|\nabla f|^2}{f} \right) d\gamma \right) dt = \frac{1}{2} \int \frac{|\nabla f|^2}{f} d\gamma.$$

The conclusion follows on replacing f with f^2 and letting $\epsilon \downarrow 0$.

$\square$

Log Sobolev inequalities

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

$$\gamma\left(\left\{F \geqslant \int F d\gamma + r\right\}\right) \leqslant e^{-r^2/2}.$$

Log Sobolev inequalities

The following argument is due to Herbst.

Proof.

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

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

$\square$

Log Sobolev inequalities

Proof.

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

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

  which rearranges to

$$\lambda \Lambda'(\lambda) \leqslant \Lambda(\lambda) \log \Lambda(\lambda) \;\Leftrightarrow\; \lambda \frac{\Lambda'(\lambda)}{\Lambda(\lambda)} \leqslant \log \Lambda(\lambda).$$

- 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 $H'(\lambda) \leqslant 0$. Hence $\Lambda(\lambda) \leqslant 1$.

$\square$

Proof.

- We've checked, for all λ,

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

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

$\square$