Concentration of Measure for the Analysis of Randomised Algorithms

Devdatt P. Dubhashi   Alessandro Panconesi

February 4, 2008

Contents

Chapter 1

Isoperimetric Inequalities and Concentration

1.1 Isoperimetric inequalities

Everyone has heard about the mother of all isoperimetric inequalities:

Of all planar geometric figures with a given perimeter,
the circle has the largest possible area. (1.1)

An abstract form of isoperimetric inequalities is usually formulated in the setting of a space (Ω, P, d) that is simultaneously equipped with a probability measure P and a metric d. We will call such a space a MM-space. Since our applications usually involve finite sets Ω and discrete distributions on them, we will not specify any more conditions (as would usually be done in a mathematics book).

Given $A \subseteq \Omega$, the *t-neighbourhood* of A is the subset $A_t \subseteq \Omega$ defined by

$$A_t := \{x \in \Omega \mid d(x, A) \leq t\}. \tag{1.2}$$

Here, by definition,

$$d(x, A) := \min_{y \in A} d(x, y).$$

An abstract isoperimetric inequality in such a MM-space (Ω, P, d) asserts that

There is a "special" family of subsets $\mathcal{B}$ such that for any $A \subseteq \Omega$,
for all $B \in \mathcal{B}$ with $P(B) = P(A)$, $P(A_t) \leq P(B_t)$. (1.3)

To relate this to (1.1), take the underlying space to be the Euclidean plane with Lebesgue measure and Euclidean distance, and the family $\mathcal{B}$ to be balls in the plane. By letting $t \to 0$, an abstract isoperimetric inequality yields (1.1).

Often an abstract isoperimetric inequality is stated in the following form:

Assertion 1.1 *In a space* (Ω, P, d), *for any* $A \subseteq \Omega$,

$$P(A)P(\overline{A_t}) \leq g(t) \tag{1.4}$$

Such a result is often proved in two steps:

1. Prove an abstract isoperimetric inequality in the form (1.3) for s suitable family $\mathcal{B}$.
2. Explicitly compute $P(B)$ for $B \in \mathcal{B}$ to determine g.

(In § 1.4, there is an exception to this rule: the function g there is bounded from above directly.)

1.2 Isoperimetry and Concentration

An isoperimetric inequality such as (1.4) implies measure concentration if the function g decays sufficiently fast to zero as $t \to \infty$. Thus, if $A \subseteq \Omega$ satisfies $\Pr(A) \geq 1/2$, then (1.4) implies $\Pr(A_t) \geq 1 - 2g(t)$. If g goes sufficiently fast to 0, then $\Pr(A_t) \to 1$. Thus

> *"Almost all the meausre is concentrated around any subset of measure at least a half"*!

1.2.1 Concentration of Lipschitz functions

It also yields concentration of Lipschitz functions on a space (Ω, d, P). Let f be a Lipschitz function on Ω with constant 1, that is,

$$|f(x) - f(y)| \leq d(x, y).$$

A *median Lévy Mean* of f is areal number $M[f]$ such that

$$P(f \geq M[f]) \geq 1/2, \quad \text{and} \quad P(f \leq M[f]) \geq 1/2.$$

Exercise 1.2 *Let (Ω, P) be a probability space and let f be a real-valued function on Ω. Define*

$$med(f) := \sup\{t \mid P[f \leq t] \leq 1/2\}.$$

Show that:

$$P[f < med(f)], \quad P[f > med(f)] \quad \leq \quad 1/2.$$

Set

$$A := \{x \in \Omega \mid f(x) \leq M[f]\}.$$

Then, by defintiion of a median, $\Pr(A) \geq 1/2$. Note that since f is Lipschitz,

$$\{x \mid f(x) > M[f] + t\} \subseteq \overline{A_t},$$

and hence,

$$\Pr[f(x) > M[f] + t] \leq \Pr(\overline{A_t}) \leq 2g(t) \to 0.$$

Exercise 1.3 *Show that (1.4) also implies a similar bound on*

$$\Pr[f(x) > M[f] - t].$$

.

Exercise 1.4 *Show that it suffices to impose a one-sided condition on f:*

$$f(x) \leq f(y) + d(x, y),$$

or

$$f(x) \geq f(y) - d(x, y).$$

to obtain two-sided concentration around a Lévy Mean.

Usually one has a concentration around the expectation. In Problem 1.15 you are asked to check that if the concentration is strong enough, concentration around the expectation or a median are essentially equivalent.

To get a quantitative bound on how good the concentration is, one needs to look at the behaviour of g in (1.4). Let (Ω, P, d) be a MM-space, and let

$$D := \max\{d(x, y) \mid x, y \in \Omega\}.$$

For $0 < \epsilon < 1$, let

$$\alpha(\Omega, \epsilon) := \max\{1 - P(A_{\epsilon D}) \mid P(A) \geq 1/2\}.$$

So a space with small $\alpha(\Omega, \epsilon)$ is one in which there is measure concentration around sets of measure at least 1/2.

A family of spaces $(\Omega_n, d_n, P_n), n \geq 1$ is called

- a *Lévy family* if
$$\lim_{n\to\infty} \alpha(\Omega_n, \epsilon) = 0.$$
- a *concentrated Lévy family* if there are constants $C_1, C_2 > 0$ such that
$$\alpha(\Omega_n, \epsilon) \leq C_1 \exp\left(-C_2 \epsilon \sqrt{n}\right).$$
- a *normal Lévy family* if there are constants $C_1, C_2 > 0$ such that
$$\alpha(\Omega_n, \epsilon) \leq C_1 \exp\left(-C_2 \epsilon^2 n\right).$$

1.3 Examples: Classical and Discrete

1.3.1 Euclidean Space with Lebesgue Measure

Consider Euclidean space R^n with the Eucledean metric and Lebesgue measure μ.

Theorem 1.5 (Isoperimetry for Euclidean Space) *For any compact subset $A \subseteq R^n$, and any $t \geq 0$,*
$$\mu(A_t) \geq \mu(B_t),$$
where B is a ball with $\mu(B) = \mu(A)$.

In Problem 1.16 you are asked to prove this using the famous Brunn-Minkowski inequality.

1.3.2 The Euclidean Sphere

For the sphere S^{n-1} with the usual Eucledean metric inherited from R^n, a r-ball is a sphereical cap i.e. an intersection of S^{n-1} with a half-space.

Theorem 1.6 (Isoperimetry for Euclidean Sphere) *For any measurable $A \subseteq S^{n-1}$, and any $t \geq 0$,*
$$\Pr(A_t) \geq \Pr(C_t),$$
where C is a spherical cap with $\Pr(C) = \Pr(A)$.

A calculation for spherical caps then yields:

Theorem 1.7 (Measure Concentration on the Sphere) *Let $A \subseteq S^{n-1}$ be a meqsurable set with* $\Pr(A) \geq 1/2$. *Then,*

$$P(A_t) \geq 1 - 2e^{-t^2 n/2}.$$

Note that the Sphere S^{n-1} has diameter 2 so this inequality shows that the faimily of spheres $\{S^{n-1} \mid n \geq 1\}$ is a normal Lévy family.

1.3.3 Euclidean Space with Gaussian Measure

Consider R^n with the Eucledean metric and the n-dimensional Gaussian measure γ:

$$\gamma(A) := (2\pi)^{-n/2} \int_A e^{-||x||^2/2} dx.$$

This is a probability distribution on R^n corresponding to the n-dimensional normal distribution. Let $Z_1, \ldots, Z_n$ be i.i.d. variables with the normal distribution $N(0, 1)$ i.e. for any real z,

$$\Pr[Z_i \leq z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt.$$

Then the vector $(Z_1, \cdots, Z_n)$ is distributed according to the measure .γ. The distribution γ is spherically symmetric: the density function depends only on the distance from the origin.

The isoperimetric inequality for Gaussian measure asserts that among all subsets A with a given $\gamma(A)$, a half space has the smallest possible measure of the t-neighbourhood. By a simple calculation, this yields,

Theorem 1.8 (Gaussian Measure Concentration) *Let $A \subseteq R^n$ be measurable and satisfy $\gamma(A) \geq 1/2$. Then $\gamma(A_t) \geq 1 - e^{-t^2/2}$.*

1.3.4 The Hamming Cube

Consider the *Hamming cube* $Q_n := \{0, 1\}^n$ with uniform measure and the *Hamming metric*:

$$d(x, y) := |\{i \in [n] \mid x_i \neq y_i\}.$$

A r-ball in this space is $B^r := \{x \mid d(x, 0) \le r\}$ i.e. the set of all 0/1 sequences that has at most r 1s. Clearly

$$\Pr(B^r) = \frac{1}{2^n} \sum_{0 \le i \le r} \binom{n}{i}.$$

Note that the t-neighbourhood of a r-ball is a $r+t$-ball: $B^r_t = B^{r+t}$.

Theorem 1.9 (Harper's Isoperimetric inequality) *If $A \subseteq Q_n$ satisfies $\Pr(A) \ge \Pr(B^r)$, then $\Pr(A_t) \ge \Pr(B^{r+t})$.*

Corollary 1.10 (Measure Concentration for the Hamming Cube) *Let $A \subseteq Q_n$ be such that $\Pr(A) \ge 1/2$. Then $\Pr(A_t) \ge 1 - e^{-2t^2/n}$.*

Since the diameter of Q_n is n, this shows that the family of cubes $\{Q^n \mid n \ge 1\}$ is a normal Lévy family.

Exercise 1.11 *Use the CH bound to deduce Corollary 1.10 from Harper's isoperimetric inequality.*

Exercise 1.12 *Deduce the Chernoff bound for iid variables corresponding to fair coin flips from Corollary 1.10.*

1.4 Martingales and Isoperimetric inequalities

In § 1.2 we saw that an isoperimetric inequality yields the method of bounded differences i.e. concentration for Lipschitz functions. In this section we see that conversely, isoperimetric inequalities can be derived via the method of bounded differences. So, isoperimetric inequalities and the concentration of Lipschitz functions are essentially equivalent.

Consider the space $\{0,1\}^n$ with the uniform measure (which is also the product measure with $p = 1/2$ in each co–ordinate) and the Hamming metric, d_H. Let A be a subset of size at least 2^{n-1} so that $\mu(A) \ge 1/2$. Consider the function $f(x) := d_H(x, A)$, the Hamming distance of x to A. Surely f is Lipshitz. Let $X_1, \ldots, X_n$ be independent and uniformly distributed in $\{0,1\}$. Then, by applying the method of bounded differences,

$$\mu[f > \mathtt{E}[f] + t], \mu[f < \mathtt{E}[f] - t] \le e^{\frac{-t^2}{2n}}.$$

In particular,

$$\begin{aligned} 1/2 &\leq \mu(A) \\ &= \mu(f=0) \\ &\leq \mu(f < \mathtt{E}[f] - \mathtt{E}[f]) \\ &\leq e^{\frac{-\mathtt{E}[f]^2}{2n}}. \end{aligned}$$

Thus $\mathtt{E}[f] \leq t_0 := \sqrt{2 \ln 2n}$. Finally then,

$$\mu(A_t) \geq 1 - \exp\left(\frac{-(t-t_0)^2}{2n}\right).$$

Consider now a weighted verison: the space is $\{0,1\}^n$ with the uniform measure, but the metric is given by

$$d_\alpha(x,y) := \sum_{x_i \neq y_i} \alpha_i,$$

for fixed non=negative reals $\alpha_i, i \in [n]$.

Exercise 1.13 *Show that*

$$\mu(A_t) \geq 1 - \exp\left(\frac{-(t-t_0)^2}{2\sum_i \alpha_i^2}\right).$$

Exercise 1.14 *Check that the result of the previous exercise holds in arbitrary product spaces with arbitrary product distributions and a weighted Hamming metric.*

In the next chapter we will see a powerful extension of this inequality.

1.5 Bibliographic Notes

Ledoux [4][Chapter 1] has a thorough discussion of isoperimetric inequalities and concentration. The vexed issue of concentration around the mean or the median is addressed in Prop. 1.7 and the following discussion there. See also McDiarmid [9]. Examples of isoperimetric inequalities in different spaces are discussed in Ledoux [4][§2.1]. Matousek [8][Chapter 14] has a nice discussion and many examples.

1.6 Problems

Problem 1.15 [Expectation versus Median] In this problem, we check that concentration around the expectation or a median are essentially equivalent.

(a) Let $\Omega_n, P_n, d_n), n \geq 1$ be a normal Lévy family. let Ω_n have diameter D_n. Show that if f is a 1-Lipschitz function on Ω_n, then for some constant $c > 0$,

$$|M[f] - \mathtt{E}[f]| \leq c \frac{D_n}{\sqrt{n}}.$$

(b) Deduce that if $f : S^{n-1} \to R$ is 1-Lipschitz, then for some constant $c > 0$,

$$|M[f] - \mathtt{E}[f]| \leq c \frac{1}{\sqrt{n}}.$$

(c) Deduce that if $f : Q^n \to R$ is 1-Lipschitz, then for some constant $c > 0$,

$$|M[f] - \mathtt{E}[f]| \leq c\sqrt{n}.$$

▽

Problem 1.16 [Brunn-Minkowski] Recall the famous **Brunn-Minkowski** inequality: for any non-emty compact subsets $A, B \subseteq R^n$,

$$\mathrm{vol}^{1/n}(A) + \mathrm{vol}^{1/n}(B) \leq \mathrm{vol}^{1/n}(A + B).$$

Deduce the isoperimetric inequality for R^n with Lebesgue measure and Euclidean distance form this. (HINT: Note that $A_t = A + tB$ where B is a ball of unit radius.) ▽

Problem 1.17 [Measure Concentration in Expander Graphs] The **edge expansion** or **conductance** $\Phi(G)$ of a graph $G = (V, E)$ is defined by:

$$\Phi(G) := \min \left\{ \frac{e(A, V \setminus A)}{|A|} \mid \emptyset \neq A \subseteq V, |A| \leq |V|/2 \right\}.$$

where $e(A, B)$ denotes the number of edges with one endpoint in A and the other in B. Regard G as a MM-space by G with the usual graph distance metric and equipped with the uniform measure P on V. Suppose $\Phi := \Phi(G) > 0$, and that the maximum degree of a vertex in G is Δ. Prove the following measure concentration inequality: if $A \subseteq V$ satisfies $P(A) \geq 1/2$, then $P(A_t) \geq 1 - \frac{1}{2}e^{-t\Phi/\Delta}$. (A constant degree **expander graph** G satisfies $\Phi(G) \geq c_1$ and $\Delta \leq c_2$ for constants $c_1, c_2 > 0$.) ▽

Problem 1.18 [Concentration for Permutations] Apply the average method of bounded differences to establish an isoperimetric inequality for the space of all permutations with the uniform measure and transposition distance. ▽

Problem 1.19 [Measure Concentration and Length] Schectmann, generalizing Maurey, introduced the notion of **length** in a finite metric space (Ω, d). Say that (Ω, d) has length at most ℓ if there are constants $c_1, \cdots, c_n > 0$ with $\sqrt{\sum_i c_i^2} = \ell$ and a sequence of partitions $P_0 \preceq \cdots \preceq P_n$ of Ω with P_0 trivial, P_n discrete and such that whenever we have sets $A, B \in P_k$ with $A \cup B \subseteq C \in P_{k-1}$, then $|A| = |B|$ and there is a bijection $\phi : A \to B$ with $d(x, \phi(x)) \leq c_k$ for all $x \in A$.

(a) Show that the discrete Hamming Cube Q_n with the Hamming metric has length at most $\sqrt{n}$ by considering the partitions induced by the equivalence relations $x \equiv_k y$ iff $X_i = y_i, i \leq k$ for $0 \leq k \leq n$.

(b) Let $\alpha := (\alpha_1, \cdots, \alpha_n) \geq 0$. Show that the discrete Hamming Cube Q_n with the weighted Hamming metric $d_\alpha(x, y) := \sum_{x_i \neq y_i} \alpha_i$ has length at most $\|\alpha\|_2$.

(c) Show that the group of permutations S_n equipped with the usual transporsition metric has small length.

(d) Show that Lipschitz functions on a finite metric space of small length are strongly concentrated around their mean. when the space is equppied with the uniform measure:

Theorem 1.20 *Let (Ω, d) be a finite metric space of length at most ℓ, and let f be a Lipschitz function i.e. $|f(x) - f(y)| \leq d(x, y)$ for all $x, y \in \Omega$. Then, if P is the uniform measure on Ω,*

$$P\left(f \geq E[f] + a\right), P\left(f \leq E[f] - a\right) \leq e^{-a^2/2\ell^2}.$$

(e) Generalize to the case when P is not the uniform distribution by requiring that the map $\phi : A \to B$ above is measure preserving. Show that a similar result holds for the concentration of Lipschitz functions with this condition.

▽

Problem 1.21 [Diameter, Laplace Functional and Concentration] Let (Ω, P, d) be a MM-space. The *Laplace functional*, $E = E_{\Omega,P,d}$ is defined by:

$$E(\lambda); = \sup\{\mathtt{E}[e^{\lambda f}] \mid f : \Omega \to R \text{ is 1-Lipschitz and } \mathtt{E}[f] = 0\}.$$

(a) Show that if $E(\lambda) \leq e^{a\lambda^2/2}$ for some $a > 0$, then $\texttt{Pr}[|f - Ef| > t] \leq e^{-t^2/2a}$. (HINT: recall basic Chernoff bound argument!)

(b) Show that the Laplace functional is sub-additive under products: let $(\Omega_i, P_i, d_i), i = 1, 2$ be two spaces, and let (Ω, P, d) be the product space with $\Omega := \Omega_1 \times \Omega_2$, $P := P_1 \times P_2$ and $d := d_1 + d_2$. Then

$$E_{\Omega,P,d} \leq E_{\Omega_1,P_1,d_1} \cdot E_{\Omega_2,P_2,d_2}.$$

(c) If (Ω, d) has diameter at most 1, show that $E(\lambda) \leq e^{-\lambda^2/2}$. (HINT: First note that by Jensen's inequality, $e^{\texttt{E}[f]} \leq \texttt{E}[e^f]$, hence if $\texttt{E}[f] = 0$, then $\texttt{E}[e^{-f}] \geq 1$. Now, let f be 1-Lipschitz, and let X and Y be two independent variables distributed according to P. Then,

$$\begin{aligned}
\texttt{E}[e^{\lambda f(X)}] &\leq \texttt{E}[e^{\lambda f(X)}]\texttt{E}[e^{-\lambda f(Y)}] \\
&= \texttt{E}[e^{\lambda(f(X)-f(Y))}] \\
&= \texttt{E}\left[\sum_{i\geq 0} \frac{\lambda^i \left(f(X) - f(Y)\right)^i}{i!}\right] \\
&= \sum_{i\geq 0} \texttt{E}\left[\frac{\lambda^i \left(f(X) - f(Y)\right)^i}{i!}\right]
\end{aligned}$$

Argue that the terms for odd i vanish and bound the terms for even i by using the Lipschitz condition on f.

(d) Deduce the Chernoff-Hoeffding bound from (b) and (c).

$\bigtriangledown$

Bibliography

[1] Michel Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Society, 2001.

[2] Jiří Matoušek. *Lectures on Discrete Geometry*. Springer, 2002.

[3] Colin McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.