Lecture 5: Concentration of Measure and Isoperimetry

Tianpei Xie

Jan. 19th., 2023

Contents

1 The Classic Isoperimetry Inequalities

1.1 Brunn-Minkowski Inequality

- **Definition** (***Minkowski Sum of Sets***)
  Consider sets $A, B \subseteq \mathbb{R}^n$ and define ***the Minkowski sum*** of A and B as the set of all vectors in $\mathbb{R}^n$ formed by sums of elements of A and B:

  $$A + B := \{x + y : x \in A, y \in B\}$$

  Similarly, for $c \in \mathbb{R}$, let $cA = \{cx : x \in A\}$. Denote by $\mathrm{Vol}(A)$ the ***Lebesgue measure*** of a *(measurable) set* $A \subset \mathbb{R}^n$.

- **Theorem 1.1** *(**Brunn-Minkowski Inequality**) [Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  *Let $A, B \subset \mathbb{R}^n$ be **non-empty compact sets**. Then for all $\lambda \in [0, 1]$,*

  $$Vol(\lambda A + (1 - \lambda)B)^{\frac{1}{n}} \geq \lambda\, Vol(A)^{\frac{1}{n}} + (1 - \lambda)\, Vol(B)^{\frac{1}{n}}. \tag{1}$$

  Note: a convex body in $\mathbb{R}^n$ is closed and compact set.

  Proof: (***Part 1,*** $n = 1$)
  Note that if $A \subset \mathbb{R}$, and $c \geq 0$ then $\mathrm{Vol}(cA) = c\mathrm{Vol}(A)$. Thus it suffice to prove

  $$\mathrm{Vol}\,(A + B) \geq \mathrm{Vol}(A) + \mathrm{Vol}(B).$$

  To see this, observe that none of the three volumes involved changes if the sets A and B are ***translated*** arbitrarily. Since A, B are compact subsets in $\mathbb{R}$, it is closed and bounded. Let $a = \max\{a' : a' \in A\}$ and $b = \min\{b' : b' \in B\}$. Let $A' = A + \{-a\}$ and $B' = B + \{-b\}$ so that $A' \subset (-\infty, 0]$ and $B' \subset [0, +\infty)$. Also $\mathrm{Vol}(A') = \mathrm{Vol}(A)$ and $\mathrm{Vol}(B') = \mathrm{Vol}(B)$. However,

  $$\begin{aligned} A' \cup B' &\subset A' + B' \\ \Rightarrow \mathrm{Vol}(A') + \mathrm{Vol}(B') = \mathrm{Vol}(A' \cup B') &\leq \mathrm{Vol}(A' + B') \end{aligned}$$

  This prove the 1-dimensional case for *the Brunn-Minkowski inequality.* ∎

  To prove $n > 1$ case, we need the following inequalities:

- **Theorem 1.2** *(**The Prékopa-Leindler Inequality**). [Boucheron et al., 2013, Wainwright, 2019]*
  *Let $\lambda \in (0, 1)$, and let $f, g, h : \mathbb{R}^n \to [0, \infty)$ be **non-negative measurable functions** such that for all $x, y \in \mathbb{R}^n$,*

  $$h\,(\lambda x + (1 - \lambda)y) \geq f(x)^{\lambda} g(y)^{1-\lambda}.$$

  Then

  $$\int_{\mathbb{R}^n} h(x)dx \geq \left(\int_{\mathbb{R}^n} f(x)dx\right)^{\lambda} \left(\int_{\mathbb{R}^n} g(x)dx\right)^{1-\lambda}. \tag{2}$$

  Proof: The proof goes by induction with respect to the dimension n.

1. ($n = 1$ ***case***). Consider measurable non-negative functions f, g, h satisfying the condition of the theorem. By *the monotone convergence theorem*, it suffices to prove the statement for ***bounded functions*** f and g. Without loss of generality, assume that $\sup_{x\in\mathbb{R}^n} f(x) = \sup_{x\in\mathbb{R}^n} g(x) = 1$. Then

$$\int_{\mathbb{R}} f(x)dx = \int_0^1 \mathrm{Vol}\left\{x : f(x) \geq t\right\} dt$$
$$\int_{\mathbb{R}} g(x)dx = \int_0^1 \mathrm{Vol}\left\{x : g(x) \geq t\right\} dt.$$

For any fixed $t \in [0,1]$, if $f(x) \geq t$ and $g(y) \geq t$, then by the hypothesis of the theorem, $h\left(\lambda x + (1-\lambda)y\right) \geq t$. This implication may be re-written as

$$\lambda\left\{x : f(x) \geq t\right\} + (1-\lambda)\left\{x : g(x) \geq t\right\} \subset \left\{x : h(x) \geq t\right\}.$$

Thus

$$\begin{aligned}
\int_{\mathbb{R}} h(x)dx &= \int_0^\infty \mathrm{Vol}\left\{x : h(x) \geq t\right\} dt \\
&\geq \int_0^1 \mathrm{Vol}\left\{x : h(x) \geq t\right\} dt \\
&\geq \int_0^1 \mathrm{Vol}\left(\lambda\left\{x : f(x) \geq t\right\}\right) + \mathrm{Vol}\left((1-\lambda)\left\{x : g(x) \geq t\right\}\right) dt \\
&\text{(by 1-dimensional \textit{Brunn-Minkowski inequality})} \\
&\geq \lambda \int_0^1 \mathrm{Vol}\left(\left\{x : f(x) \geq t\right\}\right) dt + (1-\lambda)\int_0^1 \mathrm{Vol}\left(\left\{x : g(x) \geq t\right\}\right) dt \\
&= \lambda \int_{\mathbb{R}} f(x)dx + (1-\lambda)\int_{\mathbb{R}} g(x)dx \\
&\geq \left(\int_{\mathbb{R}} f(x)dx\right)^{\lambda}\left(\int_{\mathbb{R}} g(x)dx\right)^{1-\lambda} \quad \text{(by the \textit{arithmetic-geometric mean inequality})}
\end{aligned}$$

2. For the induction step, assume that the theorem holds for all dimensions $1, \ldots, n-1$ and let $f, g, h : \mathbb{R}^n \to [0,\infty)$, $\lambda \in (0,1)$ be such that they satisfy the assumption of the theorem. Now let $x, y \in \mathbb{R}^{n-1}$ and $a, b \in \mathbb{R}$. Then

$$h\left(\lambda\left(x,a\right) + (1-\lambda)\left(y,b\right)\right) \geq f\left((x,a)\right)^{\lambda} g((y,b))^{1-\lambda},$$

so by the inductive hypothesis

$$\int_{\mathbb{R}^{n-1}} h\left((x, \lambda a + (1-\lambda)b)\right) dx \geq \left(\int_{\mathbb{R}^{n-1}} f\left((x,a)\right) dx\right)^{\lambda}\left(\int_{\mathbb{R}^{n-1}} g((x,b))dx\right)^{1-\lambda}$$

In other words, introducing

$$F(a) := \int_{\mathbb{R}^{n-1}} f\left((x,a)\right) dx, \quad G(b) := \int_{\mathbb{R}^{n-1}} g((x,b))dx$$
$$H((\lambda a + (1-\lambda)b)) := \int_{\mathbb{R}^{n-1}} h\left((x, \lambda a + (1-\lambda)b)\right) dx.$$

We have

$$H((\lambda a+(1-\lambda)b)) \geq (F(a))^{\lambda}\,(G(b))^{1-\lambda}\,,$$

so by *Fubini's theorem* and the one-dimensional inequality, we have

$$\begin{aligned}\int_{\mathbb{R}^n} h(x)dx = \int_{\mathbb{R}} H(a)da &\geq \left(\int_{\mathbb{R}} F(a)da\right)^{\lambda}\left(\int_{\mathbb{R}} G(a)da\right)^{1-\lambda} \\ &= \left(\int_{\mathbb{R}^n} f(x)dx\right)^{\lambda}\left(\int_{\mathbb{R}^n} g(x)dx\right)^{1-\lambda}. \qquad \blacksquare\end{aligned}$$

- **Corollary 1.3** *(**Weaker Brunn-Minkowski Inequality**) [Boucheron et al., 2013, Wainwright, 2019]*
*Let $A, B \subset \mathbb{R}^n$ be **non-empty compact sets**. Then for all $\lambda \in [0,1]$,*

$$\mathit{Vol}(\lambda A+(1-\lambda)B) \geq \mathit{Vol}(A)^{\lambda}\,\mathit{Vol}(B)^{1-\lambda}. \tag{3}$$

Proof: We apply *the Prékopa-Leindler inequality* with $f(x)=\mathbb{1}\{x\in A\}$, $g(x)=\mathbb{1}\{x\in B\}$ and $h(x)=\mathbb{1}\{x\in \lambda A+(1-\lambda)B\}$. We see that

$$h(\lambda x+(1-\lambda)y)=\mathbb{1}\{\lambda x+(1-\lambda)y\in\lambda A+(1-\lambda)B\}\geq \mathbb{1}\{x\in A, y\in B\}=f(x)^{\lambda}g(y)^{1-\lambda}.$$

Thus the hypothesis of *the Prékopa-Leindler inequality* holds. $\blacksquare$

- **Proof:** ($n>1$ ***case for Brunn-Minkowski Inequality***). First observe that it suffices to prove that for all *nonempty compact sets* A and B,

$$\mathrm{Vol}\,(A+B)^{\frac{1}{n}} \geq \mathrm{Vol}(A)^{\frac{1}{n}}+\mathrm{Vol}(B)^{\frac{1}{n}}$$

since $\mathrm{Vol}\,(cA)^{1/n} = c\mathrm{Vol}\,(A)^{1/n}$ for any $c\in\mathbb{R}$ and $A\subset\mathbb{R}^n$. Also notice that we may assume that $\mathrm{Vol}(A), \mathrm{Vol}(B)>0$ because otherwise the inequality holds trivially. Defining $A'=\mathrm{Vol}(A)^{-\frac{1}{n}}A$ and $B'=\mathrm{Vol}(B)^{-\frac{1}{n}}B$, we have $\mathrm{Vol}(A')=\mathrm{Vol}(B')=1$. By *weaker Brunn-Minkowski inequality*, for $\lambda\in(0,1)$,

$$\mathrm{Vol}\left(\lambda A'+(1-\lambda)B'\right)\geq 1.$$

Finally, we apply this *inequality* with the choice

$$\lambda=\frac{\mathrm{Vol}(A)^{\frac{1}{n}}}{\mathrm{Vol}(A)^{\frac{1}{n}}+\mathrm{Vol}(B)^{\frac{1}{n}}}$$

obtaining

$$\begin{aligned}\mathrm{Vol}\left(\frac{\mathrm{Vol}(A)^{\frac{1}{n}}A'}{\mathrm{Vol}(A)^{\frac{1}{n}}+\mathrm{Vol}(B)^{\frac{1}{n}}}+\frac{\mathrm{Vol}(B)^{\frac{1}{n}}B'}{\mathrm{Vol}(A)^{\frac{1}{n}}+\mathrm{Vol}(B)^{\frac{1}{n}}}\right)&\geq 1\\ \Rightarrow \mathrm{Vol}\left(\frac{A}{\mathrm{Vol}(A)^{\frac{1}{n}}+\mathrm{Vol}(B)^{\frac{1}{n}}}+\frac{B}{\mathrm{Vol}(A)^{\frac{1}{n}}+\mathrm{Vol}(B)^{\frac{1}{n}}}\right)&\geq 1\\ \Rightarrow \mathrm{Vol}\left(\frac{A+B}{\mathrm{Vol}(A)^{\frac{1}{n}}+\mathrm{Vol}(B)^{\frac{1}{n}}}\right)&\geq 1\\ \Rightarrow \frac{\mathrm{Vol}(A+B)}{\left(\mathrm{Vol}(A)^{\frac{1}{n}}+\mathrm{Vol}(B)^{\frac{1}{n}}\right)^n}&\geq 1\end{aligned}$$

which proves the theorem. $\blacksquare$

Figure 5.1 Isoperimetric inequality in $\mathbb{R}^n$ states that among all sets A of given volume, the Euclidean balls minimize the volume of the ε-neighborhood A_ε.

Figure 1: Isoperimetry in $\mathbb{R}^n$ [Vershynin, 2018]

1.2 The Blowup of Sets and Classical Isoperimetry Theorem

- **Definition** (***Blowup of Sets***)
  For any $t > 0$, and any (measurable) sets $A \subset \mathbb{R}^n$, ***the t-blowup (or, t-enlargement)*** *of* A is defined by
  $$A_t := \{x \in \mathbb{R}^n : d(x, A) < t\} = A + t\,B$$
  where $B = \{x \in \mathbb{R}^n : d(0, x) < 1\}$ is an *open unit ball* and $d(x, A) = \inf_{y \in A} d(x, y)$.

- **Definition** (***Surface Area of Sets***)
  let $A \subset \mathbb{R}^n$ be a measurable set and denote by $\mathrm{Vol}(A)$ its *Lebesgue measure*. *The* ***surface area*** *of* A is defined by
  $$\mathrm{Vol}(\partial A) = \lim_{t \to 0} \frac{\mathrm{Vol}(A_t) - \mathrm{Vol}(A)}{t}.$$
  provided that the limit exists. Here A_t denotes *the t-blowup* of A.

- **Remark** (***Isoperimetry Theorem***)
  The classical isoperimetric theorem in $\mathbb{R}^n$ states that, among all sets with ***a given volume***, ***the Euclidean unit ball minimizes the surface area***. This theoerm can be formally stated as below:

- **Theorem 1.4** *(**Isoperimetry Theorem**) [Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  Let $A \subset \mathbb{R}^n$ be such that $Vol(A) = Vol(B)$ where $B := \{x \in \mathbb{R}^n : d(0, x) < 1\}$ is an unit ball. Then for any $t > 0$,
  $$Vol(A_t) \ge Vol(B_t) \tag{4}$$
  Moreover, if $Vol(\partial A)$ exists, then
  $$Vol(\partial A) \ge Vol(\partial B). \tag{5}$$

  Proof: By *the Brunn-Minkowski inequality*,
  $$\begin{aligned}\mathrm{Vol}(A_t)^{1/n} = \mathrm{Vol}(A + tB)^{1/n} &\ge \mathrm{Vol}(A)^{1/n} + t\mathrm{Vol}(B)^{1/n} \\ &= (1 + t)\mathrm{Vol}(B)^{1/n} \\ &= \mathrm{Vol}(B_t)^{1/n},\end{aligned}$$

establishing the first statement. The second follows simply because

$$\mathrm{Vol}(A_t) - \mathrm{Vol}(A) \geq \mathrm{Vol}(B)((1+t)^n - 1) \geq nt\mathrm{Vol}(B)$$

where $(1+t)^n \geq 1 + nt$ for $t \geq 0$. Thus $\mathrm{Vol}(\partial A) \geq n\mathrm{Vol}(B)$. The isoperimetric theorem now follows from the fact that $\mathrm{Vol}(\partial B) = n\mathrm{Vol}(B)$. ∎

2 Concentration via Isoperimetry

2.1 Levy's Inequalities

- **Remark** We can generalize the classical isoperimety problem to a probability space $(\mathcal{X}, \mathcal{B}[\mathcal{X}], \mathbb{P})$ where $\mathcal{X}$ is a *metric space* with metric d, $\mathcal{B}[\mathcal{X}]$ is the Borel σ-algrebra and $\mathbb{P}$ is a probability measure on $\mathcal{B}[\mathcal{X}]$. Let $B := \{x \in \mathbb{R}^n : d(0,x) < 1\}$. The classical isoperimetry problem aims at finding the set $A^* \subset \mathcal{X}$ that ***minimizes the surface area***

$$\mathbb{P}(\partial A) = \lim_{t \to 0} \frac{\mathbb{P}(A_t) - \mathbb{P}(A)}{t}$$

  This is equivalent to find subset A in $\mathcal{X}$ with ***minimal t-blowup*** for given p, and for all $t > 0$

$$A^* := \inf_{A \subset \mathcal{X} : \mathbb{P}(A) \geq p} \mathbb{P}(A_t), \quad \forall t > 0$$

  where

$$A_t = A + tB = \{x \in \mathcal{X} : \exists y \in A \text{ s.t. } d(x,y) < t\} = \left\{x \in \mathcal{X} : \inf_{y \in A} d(x,y) := d(x,A) < t\right\}.$$

  We write the definition formally.

- **Definition** (***Isoperimetry Problem***) [Boucheron et al., 2013]
  Given a ***metric space*** $\mathcal{X}$ with corresponding *distance* d, consider ***the measure space*** formed by $\mathcal{X}$, *the σ-algebra* of all ***Borel sets of*** $\mathcal{X}$, and a probability measure $\mathbb{P}$. Let X be a *random variable* taking values in $\mathcal{X}$, distributed according to $\mathbb{P}$.

  The isoperimetric problem in this case is the following: given $p \in (0,1)$ and $t > 0$, ***determine the sets*** A with $\mathbb{P}[X \in A] \geq p$ for which *the measure*

$$\mathbb{P}[d(X,A) \geq t]$$

  is ***maximal***.

- **Remark** (***Isoperimetric Inequalities***)
  Even though the exact solution is only known in a few special cases, useful *bounds* for $\mathbb{P}[d(X,A) \geq t]$ can be derived under remarkably general circumstances. *Such bounds are usually referred to as* ***isoperimetric inequalities***.

- **Definition** (***Concentration Function***) [Boucheron et al., 2013, Wainwright, 2019]
  The concentration function $\alpha : [0, \infty) \to \mathbb{R}_+$ associated with ***metric measure space*** $((\mathcal{X}, d), \mathbb{P})$ is given by

$$\alpha_{\mathbb{P},(\mathcal{X},d)}(t) := \sup_{A \subset \mathcal{X} :\, \mathbb{P}(A) \geq \frac{1}{2}} \mathbb{P}[d(X,A) \geq t] = \sup_{A \subset \mathcal{X} :\, \mathbb{P}(A) \geq \frac{1}{2}} \mathbb{P}(A_t^c)$$

where $A_t := A + tB = \{x \in \mathcal{X} : d(x, A) < t\}$ is *the t-blowup* of $A \subset \mathcal{X}$. We simply denote it as $\alpha(t)$.

Thus the optimal A^* for isoperimetry problem is the one that attains the $\alpha(t) = \mathbb{P}(A_t^c)$.

- **Example** (***Concentration Function of Lebesgue Measure in $\mathbb{R}^n$ and Isoperimetric Inequality***)
  Note that the volume of a *t-ball in* $\mathbb{R}^n$ is

$$\mathrm{Vol}(tB) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)} t^n \equiv c_n t^n$$

  Thus the radius of ball B with the same volume of A is

$$r := \left(\frac{\mathrm{Vol}(A)}{c_n}\right)^{\frac{1}{n}}.$$

  The classical isoperimetric inequality states that

$$\begin{aligned} \mathrm{Vol}(A_t)) &\geq \left((r+t)\mathrm{Vol}(B)^{1/n}\right)^n \\ \Leftrightarrow \mathrm{Vol}(A_t) &\geq c_n \left(\left(\frac{\mathrm{Vol}(A)}{c_n}\right)^{\frac{1}{n}} + t\right)^n \\ \Leftrightarrow \left(\frac{\mathrm{Vol}(A_t)}{c_n}\right)^{\frac{1}{n}} &\geq \left(\frac{\mathrm{Vol}(A)}{c_n}\right)^{\frac{1}{n}} + t \end{aligned} \tag{6}$$

  Define ***the isoperimetric function*** of *the Lebesgue measure space* $(\mathbb{R}^n, \mu)$ as

$$\lambda(u) := \left(\frac{u}{c_n}\right)^{\frac{1}{n}}$$

  so *the classical isoperimetric inequality* is equivalent to *the concentration of Lebesgue measure*

$$\lambda\left(\mu(A_t)\right) \geq \lambda\left(\mu(A)\right) + t.$$

- **Theorem 2.1** ***(Levy's Inequalities)****[Boucheron et al., 2013, Wainwright, 2019]*
  For any Lipschitz function $f : \mathcal{X} \to \mathbb{R}$,

$$\begin{aligned} \mathbb{P}\left\{f(X) \geq \mathit{Med}(f(X)) + t\right\} \leq \alpha_{\mathbb{P}}(t) \\ \mathbb{P}\left\{f(X) \leq \mathit{Med}(f(X)) - t\right\} \leq \alpha_{\mathbb{P}}(t). \end{aligned} \tag{7}$$

  where $Med(f(X))$ is ***the median*** *of $f(X)$, i.e.*

$$\mathbb{P}\left\{f(X) \leq \mathit{Med}(f(X)\right\} \geq \frac{1}{2}, \quad \mathit{and} \;\; \mathbb{P}\left\{f(X) \geq \mathit{Med}(f(X)\right\} \geq \frac{1}{2}.$$

  Proof: Consider the set $A = \{x : f(x) \leq \mathrm{Med}(f(X))\}$. By the definition of a *median*, $\mathbb{P}(A) \geq \frac{1}{2}$. On the other hand, by *the Lipschitz property* of f, for any $x, y \in \mathcal{X}$,

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

So for all $y \in A$, $f(y) \leq \text{Med}(f(X))$

$$\begin{aligned} f(x) - \text{Med}(f(X)) &\leq f(x) - f(y) \leq d(x, y) \\ \Rightarrow f(x) - \text{Med}(f(X)) &\leq \inf_{y \in A} d(x, y) := d(x, A). \end{aligned}$$

Equivalently,

$$\begin{aligned} A_t := \{x \in \mathcal{X} : d(x, A) < t\} &\subseteq \{x \in \mathcal{X} : f(x) < \text{Med}(f(X)) + t\} \\ \mathbb{P}(A_t^c) &\geq \mathbb{P}\{f(X) \geq \text{Med}(f(X)) + t\} \end{aligned}$$

The first inequality now follows from the definition of the concentration function. The second inequality follows from the first by considering f. ∎

- **Remark** For L-Lipschitz function f, the inequality becomes

$$\mathbb{P}\{f(X) - \text{Med}(f(X)) \geq t\} \leq \alpha\left(\frac{t}{L}\right), \quad \mathbb{P}\{f(X) - \text{Med}(f(X)) \leq -t\} \leq \alpha\left(\frac{t}{L}\right).$$

- **Theorem 2.2** ***(Converse of Levy's Inequalities)****[Boucheron et al., 2013, Wainwright, 2019]*
If $\beta : \mathbb{R}_+ \to [0, 1]$ is a function such that for ***every Lipschitz function*** $f : \mathcal{X} \to \mathbb{R}$

$$\mathbb{P}\{f(X) - \mathit{Med}(f(X)) \geq t\} \leq \beta(t). \tag{8}$$

then $\beta(t) \geq \alpha_{\mathbb{P}}(t)$.

Proof: Note that for any $A \subset \mathcal{X}$, the function f_A defined by $f_A(x) = d(x, A)$ is *Lipschitz* since

$$|f_A(x) - f_A(y)| = |d(x, A) - d(y, A)| \leq d(x, y).$$

Also, if $P(A) \geq 1/2$, then 0 is a median of $f_A(X)$, since

$$\mathbb{P}\{f_A(x) \leq 0\} = \mathbb{P}\{d(X, A) \leq 0\} = \mathbb{P}(A) \geq \frac{1}{2}.$$

Therefore

$$\alpha(t) := \sup_{A \subset \mathcal{X} : \mathbb{P}(A) \geq 1/2} \mathbb{P}\{f_A(x) - \text{Med}(f_A(X)) \geq t\} \leq \beta(t). \quad \blacksquare$$

- **Proposition 2.3** ***(Levy's Inequalities for Mean)****[Boucheron et al., 2013, Wainwright, 2019]*
If $\beta : \mathbb{R}_+ \to [0, 1]$ is a function such that for ***every Lipschitz function*** $f : \mathcal{X} \to \mathbb{R}$

$$\mathbb{P}\{f(X) - \mathbb{E}[f(X)] \geq t\} \leq \beta(t). \tag{9}$$

then $\beta(t) \geq \alpha_{\mathbb{P}}(t/2)$.

- **Remark** (***Isoperimetric Inequalities*** $\Leftrightarrow$ ***Concentration of Lipschitz Functions***)
The first result points out that *isoperimetric inequalities* (more precisely, ***upper bounds*** *for* ***the concentration function***) imply *concentration of Lipschitz functions.*

The converse shows that *concentration of Lipschitz functions* implies an *isoperimetric inequality.* In other word, *among all upper bounds* of $\mathbb{P}(A_t^c)$ for fixed A_t,

- **Corollary 2.4** *(**Concentration of Measure on Hamming Metric Space**) [Boucheron et al., 2013]*
  Consider independent random variables $Z_1, \ldots, Z_n$ *taking their values in a* (measurable) set $\mathcal{X}$ *and denote the vector of these variables by* $Z = (Z_1, \ldots, Z_n)$ *taking its value in* $\mathcal{X}^n$. *For an arbitrary (measurable) set* $A \subset \mathcal{X}^n$, *we write* $\mathbb{P}(A) = \mathbb{P}(Z \in A)$. *The* ***Hamming distance*** $d_H(x, y)$ *between the vectors* $x, y \in \mathcal{X}^n$ *is defined as* ***the number of coordinates*** *in which* x *and* y ***differ***. *Then for any* $t > 0$,

$$\mathbb{P}\left\{d_H(x, A) \geq \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}} + t\right\} \leq \exp\left(-\frac{2t^2}{n}\right) \tag{10}$$

  Proof: As we shown in previous proof, $f_A(x) = d_H(x, A)$ is a Lipschitz function with respect to Hamming distance d_H. It follows from the definition that

$$\sup_{x \in \mathcal{X}^n, y_i \in \mathcal{X}} \left|f_A(x) - f_A(\widetilde{x}^{(i)})\right| \leq d_H(x, \widetilde{x}^{(i)}) = 1$$

  where $\widetilde{x}^{(i)} = (x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)$, so f_A has the bounded difference property. By bounded difference inequality,

$$\mathbb{P}\left\{\mathbb{E}\left[f_A(Z)\right] - f_A(Z) \geq t\right\} \leq \exp\left(-\frac{2t^2}{n}\right).$$

  Taking $t = \mathbb{E}\left[f_A(Z)\right] = \mathbb{E}\left[d_H(Z, A)\right]$, the left-hand side becomes $\mathbb{P}\left\{f_A(Z) \leq 0\right\} = \mathbb{P}\left\{d_H(Z, A) \leq 0\right\} = \mathbb{P}(A)$. Then the inequality becomes

$$\mathbb{P}(A) \leq \exp\left(-\frac{2}{n}\left(\mathbb{E}\left[d_H(Z, A)\right]\right)^2\right)$$
$$\Rightarrow \mathbb{E}\left[d_H(Z, A)\right] \leq \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}}.$$

  Then, by using the bounded difference inequality again, we obtain

$$\mathbb{P}\left\{d_H(Z, A) \geq \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}} + t\right\} \leq \mathbb{P}\left\{d_H(Z, A) \geq \mathbb{E}\left[d_H(Z, A)\right] + t\right\} \leq \exp\left(-\frac{2t^2}{n}\right). \quad \blacksquare$$

- **Remark** (***Equivalent Form***)
  From above isoperimetric inequality,

$$\mathbb{P}\left\{d_H(x, A) \geq \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}} + t\right\} \leq \exp\left(-\frac{2t^2}{n}\right)$$

  Denote $u := \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}}$. By change of variable, for any $t \geq u$,

$$\mathbb{P}\left\{d_H(x, A) \geq t\right\} \leq \exp\left(-\frac{2\,(t-u)^2}{n}\right).$$

  On the one hand, if $t \leq 2u = \sqrt{-2n\log\mathbb{P}(A)}$, then $\mathbb{P}(A) \leq \exp(-t^2/(2n))$. On the other hand, since $(t-u)^2 \geq t^2/4$ for $t \geq 2u = \sqrt{-2n\log\mathbb{P}(A)}$. the inequality above implies

$\mathbb{P}\{d_H(x, A) \geq t\} \leq \exp(-t^2/(2n))$. Thus, for all $t > 0$, we have ***the concentration of measure in Hamming metric space***:

$$\mathbb{P}(A)\mathbb{P}\{d_H(x, A) \geq t\} \leq \min\{\mathbb{P}(A), \mathbb{P}\{d_H(x, A) \geq t\}\} \leq \exp\left(-\frac{t^2}{2n}\right) \quad (11)$$

- **Remark** (***Concentration of Measure***)
  To interpret the result in (10), we see that on the left-hand side we have the measure of the set of points whose Hamming distance is at least $t + \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}}$ away from A. This inequality means that for A with ***small measure*** $\mathbb{P}(A)$, the measure of points whose ***Hamming distance*** from A *is less than* $O(\sqrt{n})$ is ***extremely large***. In other words, ***product measure*** *on Hamming metric space are* ***concentrated*** *on* ***extremely small sets***. This phenomenon is called "***concentration of measure***".

- **Example** (***Bounded Difference Property*** $\Leftrightarrow$ ***Lipschitz Condition w.r.t. Hamming Distance***)
  Note that any function with ***bounded difference property*** is ***Lipschitz function*** with respect to ***Hamming distance***.

$$\begin{aligned}
&\sup_{x\in\mathcal{X}^n, y_i\in\mathcal{X}} |f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)| \\
&\leq c_i = c_i\, d_H((x_1, \ldots, x_n), (x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)), \quad 1 \leq i \leq n \\
\Rightarrow |f(x) - f(y)| &= \left|\sum_{i=1}^{n}(f(x_{(i-1)}) - f(x_{(i)}))\right| \\
&\leq \sum_{i=1}^{n}\left|f(x_{(i-1)}) - f(x_{(i)})\right| \\
&\leq \sum_{i=1}^{n} c_i \mathbb{1}\left\{x_{(i-1)}[i] \neq x_{(i)}[i]\right\} \\
&= d_{H,c}(x, y)
\end{aligned}$$

  where $x_{(i)}$ is replicate of $x_{(i-1)}$ except for i-th component, which is replaced by y_i. Note that $x_{(0)} = x$ and $x_{(n)} = y$. Therefore, *the bounded difference inequality* can be seen as *an isoperimetry inequality* for *Lipschitz function with respect to Hamming distance*.

$$\mathbb{P}\{f(Z) - \mathbb{E}[f(Z)] \geq t\} \leq \exp\left(-\frac{2t^2}{n}\right)$$

2.2 Isoperimetric Inequalities on the Unit Sphere

- **Definition** (***Spherical Cap and its t-Blowup***)
  Let $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\| = 1\}$ be the $(n-1)$-*dimensional* ***unit sphere***. The ***intersection*** of a ***half-space*** and $\mathbb{S}^{n-1}$ is called a ***spherical cap***. In particular, for some $y \in \mathbb{R}^n$, denote the associated spherical cap as

$$H_y := \left\{x \in \mathbb{S}^{n-1} : \langle x\,,\, y\rangle \leq 0\right\}$$

  With some simple geometry, it can be shown that its t-*blowup* corresponds to the set

$$H_y^t := \left\{x \in \mathbb{S}^{n-1} : \langle x\,,\, y\rangle < \sin(t)\right\}$$

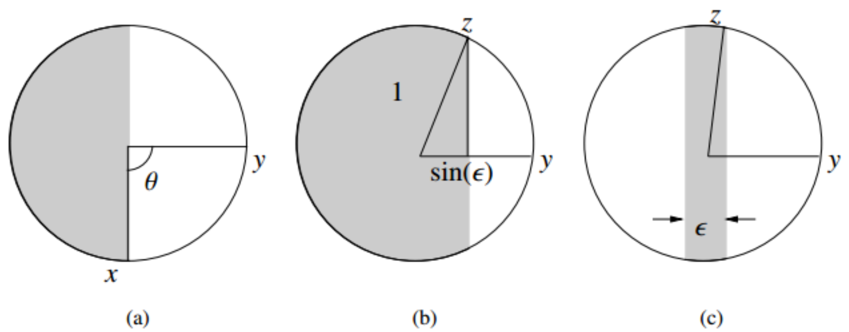

Figure 3.1 (a) Idealized illustration of the sphere $\mathbb{S}^{n-1}$. Any vector $y \in \mathbb{S}^{n-1}$ defines a hemisphere $H_y = \{x \in \mathbb{S}^{n-1} \mid \langle x, y\rangle \leq 0\}$, corresponding to those vectors whose angle $\theta = \arccos\langle x, y\rangle$ with y is at least $\pi/2$ radians. (b) The ϵ-enlargement of the hemisphere H_y. (c) A central slice $T_y(\epsilon)$ of the sphere of width ϵ.

Figure 2: spherical cap and t-blowup. [Wainwright, 2019]

- **Theorem 2.5** ***(Isoperimetry Theorem on Unit Sphere)*** *[Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  Let A be a subset of the sphere $\mathbb{S}^{n-1}$, and let σ denote the ***normalized area*** *on that sphere. Let $t > 0$. Then, among all sets $A \subset \mathbb{S}^{n-1}$ with given area $\sigma(A)$, the* ***spherical caps minimize the area of the neighborhood*** *$\sigma(A_t)$, where*

$$A_t := \left\{x \in \mathbb{S}^{n-1} : \exists y \in A \text{ such that } \|x - y\| < t\right\}$$

- **Remark** Define a *metric* ρ on sphere $\mathbb{S}^{n-1}$ as

$$\rho(x, y) := \arccos(\langle x\,,\,y\rangle)$$

  Thus $(\mathbb{S}^{n-1}, \rho)$ is a ***metric space***. Let $\mathbb{P}$ be uniform distribution on $\mathbb{S}^{n-1}$ so that $((\mathbb{S}^{n-1}, \rho), \mathbb{P})$ is a probability space.

- **Proposition 2.6** ***(Isoperimetric Inequalities for Uniform Distribution over Sphere)*** *[Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  Let $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\| = 1\}$ be the $(n-1)$-dimensional ***unit sphere****. For any $t \in [0, 1]$,*

$$\alpha_{\mathbb{S}^{n-1}}(t) \leq c \exp\left(-\frac{nt^2}{2}\right) \tag{12}$$

  for some constant c.

  Proof: Consider spherical cap

$$C(y, 0) := \left\{x \in \mathbb{S}^{n-1} : \langle x\,,\,y\rangle \geq 0\right\}$$

  and its *t-blowup*

$$C(y, t) := \left\{x \in \mathbb{S}^{n-1} : \langle x\,,\,y\rangle \geq t\right\}.$$

  According to *the isoperimetry theorem on unit sphere*, the concentration function for uniform distribution over $\mathbb{S}^{n-1}$

$$\alpha_{\mathbb{S}^{n-1}}(t) = \mathbb{P}(C(y, t)).$$

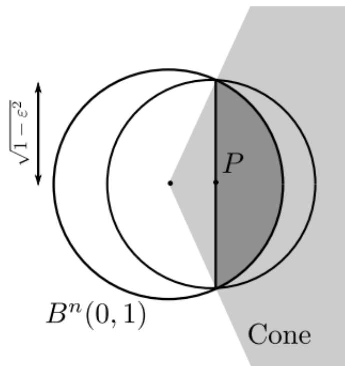

Figure 2: Small ε.

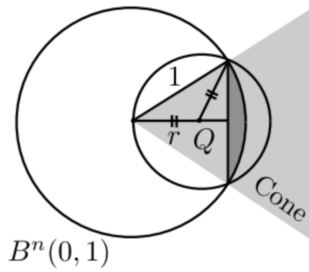

Figure 3: Large ε. By the congruence $\frac{1/2}{r} = \frac{\varepsilon}{1}$.

Figure 3: proof for upper bound of area of spherical cap (left) for small t (right) for large t

Note that $\mathbb{P}\left(C(y,0)\right) \leq 1/2$. In order to bound the concentration function from above, consider for small $t \in [0, 1/\sqrt{2}]$,

$$\begin{aligned}
\alpha_{\mathbb{S}^{n-1}}(t) = \mathbb{P}\left(C(y,t)\right) &= \frac{\mathrm{Vol}(B^n(0,1) \cap \mathrm{Cone})}{\mathrm{Vol}(B^n(0,1))} \\
&\leq \frac{\mathrm{Vol}(B^n(P, \sqrt{1-t^2}))}{\mathrm{Vol}(B^n(0,1))} \\
&= (\sqrt{1-t^2})^n \\
&\leq \exp\left(-\frac{nt^2}{2}\right)
\end{aligned}$$

For $t \in [1/\sqrt{2}, 1)$, it is enough to consider a different auxiliary ball which includes the set $\mathrm{Cone} \cap B^n(0,1)$. We obtain

$$\begin{aligned}
\alpha_{\mathbb{S}^{n-1}}(t) = \mathbb{P}\left(C(y,t)\right) &\leq \frac{\mathrm{Vol}(B^n(Q,r))}{\mathrm{Vol}(B^n(0,1))} \\
&= r^n = \left(\frac{1}{2t}\right)^n \\
&\leq \exp\left(-\frac{nt^2}{2}\right)
\end{aligned}$$

where the last inequality is from $e^{x^2/2} \leq 2x$ for $x \in [1/\sqrt{2}, 1]$. Due to convexity, this is only to be checked at the boundary of our interval $[1/\sqrt{2}, 1]$, $\blacksquare$

- By Levy's inequality, we have the following proposition

  Proposition 2.7 ***(Lipschitz Function on*** $\mathbb{S}^{n-1}$*) [Wainwright, 2019]*
  For any 1-Lipschitz function f defined on the sphere $\mathbb{S}^{n-1}$, we have the two-sided bound

$$\mathbb{P}\left\{|f(Z) - \mathit{Med}(f(Z))| \geq t\right\} \leq \sqrt{2\pi} \exp\left(-\frac{nt^2}{2}\right) \tag{13}$$

Moreover, replacing median by the mean, we have

$$\mathbb{P}\left\{|f(Z) - \mathbb{E}\left[f(Z)\right]| \ge t\right\} \le 2\sqrt{2\pi}\exp\left(-\frac{nt^2}{8}\right) \tag{14}$$

- **Exercise 2.8** ***(The Blow-Up Phenomenon)***
  Let A be a subset of the sphere $\sqrt{n}\mathbb{S}^{n-1}$ such that

  $$\mathbb{P}\left(A\right) > 2\exp\left(-cs^2\right) \;\textit{ for some } s > 0;$$

  1. *Prove that $\mathbb{P}(A_s) > 1/2$.*
  2. *Deduce from this that for any $t \ge s$,*

     $$\mathbb{P}\left(A_{2t}\right) > 1 - 2\exp\left(-ct^2\right).$$

     Here $c > 0$ is the absolute constant in upper bound of concentration function.

- **Remark** (***Zero-One Law for Independent Variables***) [Vershynin, 2018]
  The blow-up phenomenon we just saw may be quite *counter-intuitive* at first sight. How can an exponentially small set A undergo such a dramatic transition to an exponentially large set A_{2t} under such a small perturbation $2t$? (Remember that t can be much smaller than the radius $\sqrt{n}$ of the sphere.)

  However perplexing this may seem, this is *a typical phenomenon in* ***high dimensions***. It is *reminiscent* of ***zero-one laws*** in *probability theory*, which basically state that *events that are determined by many random variables* tend to have *probabilities either zero or one.*

2.3 Gaussian Isoperimetric Inequalities and Concentration of Gaussian Measure

- **Remark** (***Gaussian Isoperimetric Problem***)
  The Gaussian isoperimetric problem is to determine which (Borel) sets A have ***minimal Gaussian boundary measure*** among all sets in $\mathbb{R}^n$ with a *given probability* p.

  The Gaussian isoperimetric theorem states the beautiful fact that ***the extremal sets*** are ***linear half-spaces*** *in all dimensions* and *for all* p.

- **Definition** (***Gaussian Isoperimetric Function***)
  Denote *the cumulative distribution function of standard Normal distribution*:

  $$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt := \int_{-\infty}^{x} \varphi(t)dt$$

  where $\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = (\Phi(x))'$ is *the probability density function* of standard normal distribution. $\Phi^{-1}(x)$ is *the quantile function of normal distribution.*

  Define ***the Gaussian isoperimetric function*** as

  $$\gamma(x) := \varphi\left(\Phi^{-1}(x)\right), \quad x \in (0,1).$$

  Also we define $\gamma(0) = \gamma(1) = 0$.

- **Remark** Note that

$$\begin{aligned} x &= \Phi(\Phi^{-1}(x)) \\ \Rightarrow 1 &= \varphi(\Phi^{-1}(x))(\Phi^{-1}(x))' = \gamma(x)(\Phi^{-1}(x))' \\ \Leftrightarrow 1/\gamma(x) &= (\Phi^{-1}(x))'. \end{aligned}$$

  The quantity $1/\gamma(x) = (\Phi^{-1}(x))'$ is known as ***quantile-density function*** *of normal distribution.*

- **Proposition 2.9** *(**Basic Property of the Gaussian Isoperimetric Function**) [Boucheron et al., 2013]*
  The Gaussian isoperimetric function γ satisfies:

  1.

  $$\gamma'(x) = -\Phi^{-1}(x), \quad \textit{for all } x \in (0,1),$$

  2.

  $$\gamma(x)\gamma''(x) = -1, \quad \textit{for all } x \in (0,1),$$

  3. $(\gamma')^2$ *is convex over* $(0,1)$.

  Proof: 1. See that

  $$\begin{aligned} \varphi'(x) &= \frac{1}{\sqrt{2\pi}}(-x)e^{-\frac{x^2}{2}} = (-x)\varphi(x) \\ \varphi''(x) &= (x^2-1)\varphi(x) \end{aligned}$$

  Thus

  $$\gamma(x)' = \left(\varphi\left(\Phi^{-1}(x)\right)\right)' = \frac{d\varphi}{dy}(\Phi^{-1}(x))\frac{d\Phi^{-1}}{dx}(x) = (-\Phi^{-1}(x))\left(\Phi^{-1}(x)\right)'\gamma(x) = -\Phi^{-1}(x),$$

  since $\left(\Phi^{-1}(x)\right)'\gamma(x) = 1$, we have the result.

  2.

  $$\begin{aligned} \gamma''(x) &= (\gamma'(x))' = -\left(\Phi^{-1}(x)\right)' = -\frac{1}{\gamma(x)} \\ \gamma(x)\gamma''(x) &= -1 \end{aligned}$$

  3. Since $\gamma > 0$ for $x \in (0,1)$,

  $$\gamma''(x) = -\frac{1}{\gamma(x)} < 0.$$

  Thus $\gamma(x)$ is concave function in $(0,1)$. ∎

- **Lemma 2.10** *(**Asymptotic Behavior of Gaussian Isoperimetric Function**) [Boucheron et al., 2013]*
  For all $x \in [0, 1/2]$,

  $$x\sqrt{\frac{1}{2}\log\frac{1}{x}} \le \gamma(x) \le x\sqrt{2\log\frac{1}{x}}.$$

Moreover,

$$\lim_{x \to 0} \frac{\gamma(x)}{x\sqrt{2 \log \frac{1}{x}}} = 1$$

- **Proposition 2.11** ***(Bobkov's Inequality)*** *[Boucheron et al., 2013]*
Suppose Z is uniformly distributed over $\{-1, 1\}^n$. Then for all $n \geq 1$ and for all functions $f : \{-1, 1\}^n \to [0, 1]$,

$$\gamma\left(\mathbb{E}\left[f(Z)\right]\right) \leq \mathbb{E}\left[\sqrt{\gamma(f(Z))^2 + \|\nabla f(Z)\|_2^2}\right] \tag{15}$$

- **Proposition 2.12** ***(Bobkov's Gaussian Inequality)*** *[Boucheron et al., 2013]*
Let $Z := (Z_1, \ldots, Z_n)$ be a vector of ***independent standard Gaussian*** *random variables. Let $f : \mathbb{R}^n \to [0, 1]$ be a differentiable function with gradient ∇f. Then*

$$\gamma\left(\mathbb{E}\left[f(X)\right]\right) \leq \mathbb{E}\left[\sqrt{\gamma(f(X))^2 + \|\nabla f(X)\|_2^2}\right] \tag{16}$$

where $\gamma = \varphi \circ \Phi^{-1}$ is ***the Gaussian isoperimetric function****.*

- **Theorem 2.13** ***(Gaussian Isoperimetric Theorem)*** *[Boucheron et al., 2013] [Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
Let $\mathbb{P}$ be the ***standard Gaussian measure*** *on $\mathbb{R}^n$ and let $A \subset \mathbb{R}^n$ be a Borel set. Then*

$$\liminf_{t \to 0} \frac{\mathbb{P}(A_t) - \mathbb{P}(A)}{t} \geq \gamma\left(\mathbb{P}(A)\right), \tag{17}$$

where $A_t := \{x : d(x, A) < t\}$ be the t-blowup of A. Moreover, if A is a ***half-space*** *defined by $A := \{x \in \mathbb{R}^n : x_1 \leq z\}$, then*

$$\liminf_{t \to 0} \frac{\mathbb{P}(A_t) - \mathbb{P}(A)}{t} = \gamma\left(\mathbb{P}(A)\right) = \varphi(z), \tag{18}$$

where $\gamma := \varphi \circ \Phi^{-1}$ is ***the Gaussian isoperimetric function****.*

- **Proposition 2.14** ***(Differentiablity of Measure of t-Blowup)*** *[Boucheron et al., 2013]*
If A is a ***finite union*** *of* ***open*** *balls in $\mathbb{R}^n$, then $\mathbb{P}(A_t)$ is a* ***differentiable*** *function of $t > 0$.*

- Next we describe ***an equivalent version*** of ***the Gaussian isoperimetric theorem*** in the manner of *measure concentration*:

  Theorem 2.15 ***(Gaussian Concentration Theorem)*** *[Boucheron et al., 2013] [Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  Let $\mathbb{P}$ be the ***standard Gaussian measure*** *on $\mathbb{R}^n$ and let $A \subset \mathbb{R}^n$ be a Borel set. Then for all $t \geq 0$,*

$$\begin{aligned} \mathbb{P}(A_t) &\geq \Phi\left(\Phi^{-1}\left(\mathbb{P}(A)\right) + t\right). \\ \Leftrightarrow \Phi^{-1}(\mathbb{P}(A_t)) &\geq \Phi^{-1}\left(\mathbb{P}(A)\right) + t \end{aligned} \tag{19}$$

  Equality holds if A is a ***half-space****.*

Proof: We call *a Borel set* $A \subset \mathbb{R}^n$ ***smooth*** if $\mathbb{P}(A_t)$ is a *differentiable function* of t on $(0, \infty)$.

1. Observe that if A is *smooth*, then

$$\begin{aligned}\frac{d}{dt}\Phi^{-1}(\mathbb{P}(A_t)) &= \left[(\Phi^{-1})'(\mathbb{P}(A_t))\right]\frac{d}{dt}\mathbb{P}(A_t)\\ &= \frac{1}{\gamma(\mathbb{P}(A_t))}\frac{d}{dt}\mathbb{P}(A_t)\\ &\geq \frac{1}{\frac{d}{dt}\mathbb{P}(A_t)}\left(\frac{d}{dt}\mathbb{P}(A_t)\right) = 1\end{aligned}$$

   The last inequality is due to *the Gaussian isoperimetric inequality*

$$\frac{d}{dt}\mathbb{P}(A_t) \geq \liminf_{s\to 0}\frac{\mathbb{P}(A_{t+s}) - \mathbb{P}(A_t)}{s} \geq \gamma\left(\mathbb{P}(A_t)\right).$$

   Therefore, by integration

$$\begin{aligned}\Phi^{-1}(\mathbb{P}(A_t)) &= \Phi^{-1}(\mathbb{P}(A)) + \int_0^t \frac{d}{ds}\Phi^{-1}(\mathbb{P}(A_s))ds\\ &\geq \Phi^{-1}(\mathbb{P}(A)) + \int_0^t ds = \Phi^{-1}(\mathbb{P}(A)) + t.\end{aligned}$$

   Hence, the theorem holds for all smooth sets. The remaining work is to extend this to all *Borel sets.*

2. Note first that if $\mathbb{P}(A) = 0$, the theorem is automatically satisfied and therefore we may focus on Borel sets A with *positive probability.* By Proposition 2.14, the concentration property holds for ***any finite union of open balls***.

3. Now let A be *any Borel set* with $\mathbb{P}(A) > 0$. Let $0 < \epsilon < t$. Then by ***Vitali's covering theorem***, there exists a *countable* collection of *disjoint open balls* $\{B_1, B_2, \ldots\}$, all intersecting A and *diameter at most* ϵ, such that $P(A - \bigcup_{n=1}^{\infty} B_n) = 0$. But then

$$\begin{aligned}\mathbb{P}(A_t) &\geq \mathbb{P}\left(\bigcup_{n=1}^{\infty}(B_n)_{t-\epsilon}\right)\\ &= \lim_{n\to\infty}\mathbb{P}\left(\bigcup_{i=1}^{n}(B_i)_{t-\epsilon}\right)\\ &\geq \lim_{n\to\infty}\Phi\left(\Phi^{-1}\left(\mathbb{P}\left(\bigcup_{i=1}^{n}(B_i)_{t-\epsilon}\right)\right) + t - \epsilon\right)\\ &= \Phi\left(\Phi^{-1}\left(\mathbb{P}\left(\bigcup_{i=1}^{\infty}(B_i)_{t-\epsilon}\right)\right) + t - \epsilon\right)\\ &\geq \Phi\left(\Phi^{-1}\left(\mathbb{P}\left(A\right)\right) + t - \epsilon\right)\end{aligned}$$

   The argument is completed by taking ϵ to 0. ∎

- **Remark** (***Gaussian Concentration Theorem*** $\equiv$ ***Gaussian Isoperimetric Theorem***)
  The Gaussian concentration theorem is equivalent to *the Gaussian isoperimetric theorem* since

$$\begin{aligned}\liminf_{t\to 0}\frac{\mathbb{P}(A_t)-\mathbb{P}(A)}{t} &\ge \liminf_{t\to 0}\frac{\Phi\left(\Phi^{-1}\left(\mathbb{P}(A)\right)+t\right)-\Phi\left(\Phi^{-1}\left(\mathbb{P}(A)\right)\right)}{t}\\ &= \Phi'(\Phi^{-1}(\mathbb{P}(A)))\\ &= \varphi(\Phi^{-1}(\mathbb{P}(A)))\\ &= \gamma(\mathbb{P}(A)).\end{aligned}$$

- **Exercise 2.16** ***(From Isoperimetry to Concentration)*** *[Boucheron et al., 2013]*
  Assume that a probability distribution $\mathbb{P}$ on $\mathbb{R}^n$ satisfies, for all Borel sets $A \subset \mathbb{R}^n$,

$$\liminf_{t\to 0}\frac{\mathbb{P}(A_t)-\mathbb{P}(A)}{t} \ge c\, f\left(F^{-1}(\mathbb{P}(A))\right),$$

  where $c \in (0,1]$ is a constant, F is a continously differentiable distribution function and $f = F'$ is its derivative. Prove that for all Borel set A and all $t \ge 0$,

$$\mathbb{P}(A_t) \ge F\left(F^{-1}(\mathbb{P}(A)) + c\,t\right).$$

- As a direct consequence of *the Gaussian isoperimetric inequality*, we have the improved result for *Gaussian concentration inequality*:

  Theorem 2.17 ***(Gaussian Concentration Inequality, Sharp Bound)*** *[Boucheron et al., 2013, Wainwright, 2019]*
  Let $Z = (Z_1, \ldots, Z_n)$ be a vector of n ***independent standard normal*** *random variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ denote an L-**Lipschitz function**. Then, for all $t > 0$,*

$$\mathbb{P}\left\{f(Z) - Med(f(Z)) \ge t\right\} \le 1 - \Phi\left(\frac{t}{L}\right). \tag{20}$$

  where $\Phi(t)$ is the cumulative distribution function of standard normal random variable.

- **Remark** Note that by ***Gordon's inequality***

$$1-\Phi(t) \le \left(\frac{1}{\sqrt{2\pi}}\right)\frac{1}{t}e^{-\frac{t^2}{2}} = \frac{1}{t}\varphi(t)$$

  The Gaussian concentration inequality fails to capture the corrective factor t^{-1}. The inequality above cannot be improved in general as for $f(x) = n^{-1/2}\sum_{i=1}^{n} x_i$, equality is achieved for all $t > 0$.

2.4 Convex Distance Inequality

- **Definition** (***Weighted Hamming Distance***)
  Given $\alpha = (\alpha_1, \ldots, \alpha_n)$, where $\alpha_i \ge 0$, ***the weighed Hamming distance*** between $x, y \in \mathcal{X}^n$ is defined as

$$d_\alpha(x,y) = \sum_{i=1}^{n}\alpha_i \mathbb{1}\left\{x_i \neq y_i\right\}.$$

- **Remark** (***Measure Concentration in Weighted Hamming Distance Space***)
  Similar to the inequality (10), for *metric measure space* $\mathcal{X}^n$ with respect to *weighted Hamming distance*, we have the measure concentration inequality for $A \subset \mathcal{X}^n$

$$\mathbb{P}\left\{d_\alpha(x, A) \geq \sqrt{\frac{\|\alpha\|_2}{2} \log \frac{1}{\mathbb{P}(A)}} + t\right\} \leq \exp\left(-\frac{2t^2}{\|\alpha\|_2}\right)$$

  where $\|\alpha\|_2 = \sqrt{\sum_{i=1}^n \alpha_i^2}$. Assume $\|\alpha\|_2 = 1$

$$\mathbb{P}\left\{d_\alpha(x, A) \geq \sqrt{\frac{1}{2} \log \frac{1}{\mathbb{P}(A)}} + t\right\} \leq \exp\left(-2t^2\right)$$

  Following the same argument, we can find *an equivalent form* as in (11)

$$\sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2 = 1} \mathbb{P}(A)\mathbb{P}\left\{d_\alpha(x, A) \geq t\right\} \leq \sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2 = 1} \min\left\{\mathbb{P}(A), \mathbb{P}\left\{d_\alpha(x, A) \geq t\right\}\right\} \leq \exp\left(-\frac{t^2}{2}\right)$$

  A key contribution for ***convex distance inequality*** is that the above inequality remains true even if the ***supremum*** is taken ***within the probability***; i.e.

$$\mathbb{P}(A)\mathbb{P}\left\{\sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2 = 1} d_\alpha(x, A) \geq t\right\} \leq \exp\left(-\frac{t^2}{4}\right).$$

- **Definition** (***Convex Distance***)
  For any $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$, ***the convex distance*** of x from the set A by

$$d_T(x, A) := \sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2 = 1} d_\alpha(x, A)$$

- **Theorem 2.18** *(**Convex Distance Inequality**) [Boucheron et al., 2013]*
  For any subset $A \subset \mathcal{X}^n$ and $t > 0$,

$$\mathbb{P}(A)\mathbb{P}\left\{d_T(X, A) \geq t\right\} = \mathbb{P}(A)\mathbb{P}\left\{\sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2 = 1} d_\alpha(X, A) \geq t\right\} \leq \exp\left(-\frac{t^2}{4}\right). \tag{21}$$

- With convex distance inequality, we can improve *the concentration bound* for *convex Lipschitz functions*. First, we relate convex distance with the minimal distance to convex set

  Lemma 2.19 *(**Convex Distance vs. Distance to Convex Set**) [Boucheron et al., 2013]*
  *Let $A \subset [0, 1]^n$ be a **convex set** and let $x = (x_1, \ldots, x_n) \in [0, 1]^n$. Then*

$$d(x, A) := \inf_{y \in A} \|x - y\|_2 \leq d_T(x, A). \tag{22}$$

- **Theorem 2.20** *(**Concentration of Quasi-Convex Lipschitz Functions**) [Boucheron et al., 2013]*

*Let $Z := (Z_1, \ldots, Z_n)$ be independent random variables taking values in the interval $[0,1]$ and let $f : [0,1]^n \to \mathbb{R}$ be a **quasi-convex function**; that is*

$$\{z : f(z) \le s\} \text{ is convex set for all } s \in \mathbb{R}.$$

Moreover, f is Lipschitz function satisfying

$$|f(x) - f(y)| \le \|x - y\| \quad \text{for all } x, y \in [0,1]^n.$$

Then $X = f(Z_1, \ldots, Z_n)$ satisfies, for all $t > 0$,

$$\mathbb{P}\{f(Z) \ge Med(f(Z)) + t\} \le 2\exp\left(-\frac{t^2}{4}\right), \tag{23}$$
$$\mathbb{P}\{f(Z) \le Med(f(Z)) - t\} \le 2\exp\left(-\frac{t^2}{4}\right).$$

Proof: For some $s \in \mathbb{R}$, define the set $A_s = \{z : f(z) \le s\} \subset [0,1]^n$. Because of *quasi-convexity*, A_s is *convex*. By *the Lipschitz property* and *Lemma* 2.19, for all $z \in [0,1]^n$,

$$f(z) \le s + d(z, A) \le s + d_T(z, A).$$

So by *convex distance inequality*,

$$\mathbb{P}\{f(Z) \le s\}\,\mathbb{P}\{f(Z) \ge s + t\} \le \exp\left(-\frac{t^2}{4}\right)$$

Take $s = \text{Med}(f(Z))$ to get the *upper tail inequality* and $s = \text{Med}(f(Z)) - t$ to get *the lower tail inequality.* ∎

2.5 Edge Isoperimetric Inequality on the Binary Hypercube

- **Remark** (***Binary Hypercube as Nearest Neighbor Graph with respect to Hamming Distance***)
  Consider binary hypercube $\{-1,1\}^n$ with *Hamming distance metric*

  $$d_H(x,y) = \sum_{i=1}^{n} \mathbb{1}\{x_i \ne y_i\}$$

  The elements x of the binary n-cube may be considered as ***vertices*** *of a graph* $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ in which two elements x and x' of $\{-1,1\}^n$ are ***adjacent*** *if and only if* ***their Hamming distance is*** 1; i.e.

  $$\mathcal{E} = \{(x,y) \in \{-1,1\}^n \times \{-1,1\}^n : d_H(x,y) = 1\}.$$

  The graph structure has $|\mathcal{V}| := N = 2^n$ *vertices* and $|\mathcal{E}| = n2^{n-1}$ *undirected edges.* Its ***density*** (the ratio between the number of edges and the number of vertices) is thus $n/2 = (\log_2 N)/2$.

- **Remark** (***Maximum Density of Subgraph***)
  A remarkable property of the binary n-cube is that *for any subset $A \subseteq \{-1,1\}^n$, the* ***density*** *of the subgraph induced by A is at most* $(\log_2 |A|)/2$. Note that ***equality*** is achieved if the graph induced by A is a ***lower-dimensional hypercube***, since if A is a hypercube of dimension $d \le n$, then the subgraph induced by A has 2^d vertices and $\mathcal{E}(A) = d2^{d-1}$ edges.

Theorem 2.21 ***(Maximum Density of Subgraph)*** *[Boucheron et al., 2013]*
Let A be a subset of $\{-1,1\}^n$. Let $\mathcal{E}(A)$ denote the set of edges of the subgraph induced by A, that is, the collection of (unordered) pairs (x,x') with $x,x' \in A$ such that $d_H(x,x')=1$. Then

$$|\mathcal{E}(A)| \le \frac{|A|}{2}\log_2(|A|). \tag{24}$$

Proof: Define the random vector $X=(X_1\,,\ldots,X_n)$ taking values in $\{-1,1\}^n$ such that X has ***the uniform distribution over*** A. Denote by $\mathcal{P}$ the probability mass function of X. The Shannon entropy of X is clearly $\log_2|A|$. Writing $X_{(-i)}=(X_1\,,\ldots,X_{i-1},X_{i+1}\,,\ldots,X_n)$, and using the definition of *conditional entropy*, we have

$$H(X|X_{(-i)}) = H(X) - H(X_{(-i)}) = H(X_i|X_{(-i)}) = -\sum_{x\in A}\mathcal{P}(x)\log(\mathcal{P}(x_i|x_{(-i)}))$$

By definition $\mathcal{P}(x)=1/|A|$ for $x\in A$ and

$$\mathcal{P}(x_i|x_{(-i)}) = \begin{cases} \frac{1}{2} & \widetilde{x}^{(i)}\in A \\ 1 & \text{o.w.}\end{cases}$$

where $\widetilde{x}^{(i)}=(x_1\,,\ldots,x_{i-1},-x_i,x_{i+1}\,,\ldots,x_n)$. Thus

$$\begin{aligned} H(X)-H(X_{(-i)}) &= -\sum_{x\in A}\mathcal{P}(x)\log(\mathcal{P}(x_i|x_{(-i)})) \\ &= -\frac{1}{|A|}\left\{\sum_{x\in A,\widetilde{x}^{(i)}\in A}\log_2\left(\frac{1}{2}\right) + \sum_{x\in A,\widetilde{x}^{(i)}\notin A}\log_2(1)\right\} \\ &= \frac{\log_2(2)}{|A|}\sum_{x\in A}\mathbb{1}\left\{x\in A,\widetilde{x}^{(i)}\in A\right\} \end{aligned}$$

and therefore

$$\sum_{i=1}^n\left(H(X)-H(X_{(-i)})\right) \le \frac{\log_2(2)}{|A|}\sum_{i=1}^n\sum_{x\in A}\mathbb{1}\left\{x\in A,\widetilde{x}^{(i)}\in A\right\} = \frac{\log_2(2)2\,|\mathcal{E}(A)|}{|A|}$$

Thus, *Han's inequality* implies

$$H(X)=\log_2|A| \ge \sum_{i=1}^n\left(H(X)-H(X_{(-i)})\right) = \frac{2\,|\mathcal{E}(A)|}{|A|}. \quad \blacksquare$$

- **Definition** (***Influence of Binary Variable with respect to Set***)
  Let the binary random vector $X=(X_1\,,\ldots,X_n)$ be *uniformly distributed* over $\{-1,1\}^n$ and denote by $\widetilde{X}^{(i)}=(X_1\,,\ldots,X_{i-1},-X_i,X_{i+1}\,,\ldots,X_n)$ the vector obtained by ***flipping the i-th bit*** *of* X. For any $A\subseteq\{-1,1\}^n$, ***the influence of the i-th variable*** is defined by

$$\begin{aligned} I_i(A) &= \mathbb{P}\left\{\mathbb{1}\{X\in A\}\neq\mathbb{1}\left\{\widetilde{X}^{(i)}\in A\right\}\right\} \\ &= \mathbb{P}\left\{(X\in A\wedge\widetilde{X}^{(i)}\notin A)\vee(X\notin A\wedge\widetilde{X}^{(i)}\in A)\right\} \end{aligned}$$

If $\mathbb{1}\{X \in A\} \neq \mathbb{1}\left\{\widetilde{X}^{(i)} \in A\right\}$, then the i-th variable is said to be ***pivotal*** for A. Thus, the influence $I_i(A)$ is just ***the probability that the*** i***-th variable is pivotal for*** A.

The total influence is defined by the *sum* of *individual influences*

$$I(A) := \sum_{i=1}^{n} I_i(A)$$

- **Definition** (***Edge Boundary of Subset***)
  Let A be a subset of $\{-1,1\}^n$. Let $\mathcal{E}(A)$ denote the set of edges of the subgraph induced by A. ***The edge boundary*** of A, $\partial\mathcal{E}(A)$, is define as

  $$\partial\mathcal{E}(A) := \{(x,y) : x \in A, y \in A^c, d_H(x,y) = 1\}.$$

  Thus *the total number of edges connects to all of vertices in* A can be computed as

  $$n\,|A| = 2\,|\mathcal{E}(A)| + |\partial\mathcal{E}(A)| \tag{25}$$

  where each vertex connects to exactly n edges, and every edge with both endpoints in A is counted twice. Also we have that

  $$I(A) := \frac{2\,|\partial\mathcal{E}(A)|}{2^n}.$$

- **Theorem 2.22** ***(Edge Isoperimetric Theorem of Binary Hypercube)*** *[Boucheron et al., 2013]*
  For any $A \subset \{-1,1\}^n$*, let* $\mathbb{P}(A)$ *denote* $\mathbb{P}\{X \in A\} = |A|/2^n$*. Then*

  $$I(A) \geq 2\mathbb{P}(A)\log_2\left(\frac{1}{\mathbb{P}(A)}\right) \tag{26}$$

  By theorem on maximum density of subgraph, we see that

  $$|\mathcal{E}(A)| \leq \frac{|A|}{2}\log_2(|A|).$$

  Using the formula (25), we have inequality:

  $$\begin{aligned} n\,|A| - |\partial\mathcal{E}(A)| &\leq |A|\log_2(|A|) \\ \Rightarrow |\partial\mathcal{E}(A)| &\geq |A|\,(n - \log_2(|A|)) \\ &= 2^n\mathbb{P}(A)(n - \log_2(2^n\mathbb{P}(A))) = 2^n\mathbb{P}(A)(-\log\mathbb{P}(A)) \end{aligned}$$

  Finally, note that

  $$I(A) := \frac{2\,|\partial\mathcal{E}(A)|}{2^n} \geq 2\mathbb{P}(A)(-\log\mathbb{P}(A)) \quad \blacksquare$$

2.6 Vertex Isoperimetric Inequality on the Binary Hypercube

- **Definition** (***Vertex Boundary of Graph***) [Boucheron et al., 2013]
Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $A \subset \mathcal{V}$ be a set of its vertices. ***The vertex boundary*** of A is defined as *the set of those vertices*, ***not in*** A, which are ***connected*** to *some vertex in* $\mathcal{V}$ by *an edge*. We denote *the vertex boundary of* A by $\partial V(A)$.

- **Definition** (***Vertex Isoperimetric Problem of Graph***) [Boucheron et al., 2013]
The vertex isoperimetric problem in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is to determine *the sets* $\mathcal{A} \subset \mathcal{V}$ *of* ***a given cardinality*** whose ***vertex boundary*** contains a ***minimal number of vertices***.

- **Remark** (***Binary Hypercube as Nearest Neighbor Graph with Hamming Distance***)
Consider the graph as binary hypercube $\{-1, +1\}^n$ in which two vertices are connected by an edge if and only if their ***Hamming distance*** *equals* 1. Define the norm as the Hamming distance to $-1^n = (-1\,, \ldots,\, -1)$

$$\|x\|_H := \sum_{i=1}^{n} \mathbb{1}\{x_i = 1\} = d_H(x, -1^n)$$

- **Definition** (***Simplicial Order***)
We define the so-called ***simplicial order*** *of the elements* of the binary hypercube. We say that $x = (x_1\,, \ldots, x_n) \in \{-1, 1\}^n$ ***precedes*** $y = (y_1\,, \ldots, y_n) \in \{-1, 1\}^n$ *in the simplicial order* if either $\|x\|_H < \|y\|_H$ (where $\|x\|_H := \sum_{i=1}^{n} \mathbb{1}\{x_i = 1\} = d_H(x, -1^n)$) or $\|x\|_H = \|y\|_H$ and $x_i = 1$ and $y_i = -1$ for the smallest i for which $x_i \neq y_i$. That is

$$x \prec y$$
$$\Leftrightarrow \{(x, y) : \|x\|_H < \|y\|_H \;\vee (\|x\|_H = \|y\|_H \wedge (x_i = 1 \wedge y_i = -1, \text{ where } i = \min\{k : x_k \neq x_k\}))\}$$

In other words, the vector with ***less*** 1's ***precedes*** the vector with more 1's. If the number of 1's are the same, then the first 1's on the leftmost position is preferred.

- **Example** For $n = 3$, $(-1, -1, -1) \prec (1, -1, -1) \prec (-1, 1, -1) \prec (-1, -1, 1) \prec (1, 1, -1) \prec (1, -1, 1) \prec (-1, 1, 1) \prec (1, 1, 1)$

Theorem 2.23 ***(Harp's Vertex Isoperimetric Theorem)*** *[Boucheron et al., 2013]*
For $N = 1\,, \ldots, 2^n$, let S_N denote the set of ***first N elements*** *of $\{-1, +1\}^n$ in the* ***simplicial order****. For any subset $A \subset \{-1, +1\}^n$, where $|A| = N$,*

$$|\partial V(A)| \geq |\partial V(S_N)|$$

- **Remark** Note that if $N = \sum_{i=0}^{k} \binom{n}{i}$, for $k = 0,\, \ldots, n$, then

$$S_N = \{x \in \{-1, +1\}^n : d_H(x, -1^n) \leq k\} = B_H(-1^n, k)$$

In other words, S_N is a ***Hamming ball*** *centered at the vector* $-1^n = (-1\,, \ldots,\, -1)$.

- **Definition** (***t-Blowup of Set A in Binary Hypercube***)
For any $A \subset \{-1, +1\}^n$ and $x \in \{-1, +1\}^n$, let $d_H(x, A) = \min_{y \in A} d_H(x, y)$ be the *Hamming distance* of x to the set A. Also, denote by

$$A_t := \{x \in \{-1, +1\}^n : d_H(x, A) < t\}$$

the *t-blowup of the set A*, that is, the set of points whose Hamming distance from A is at most t.

- **Remark** (***Harp's Vertex Isoperimetric Theorem*** $\Leftrightarrow$ ***Classical Isoperimetric Theorem in*** $\{-1,1\}^n$)
  The fact that among all sets with *a* ***given volume***, ***balls minimize the surface area*** is in close analogy with *the classical isoperimetric theorem.*

  Observe that if S_N is a *Hamming ball with radius* k, i.e. $N = \sum_{i=0}^{k}\binom{n}{i}$, then $S_N \cup \partial V(S_N)$ is *the Hamming ball of radius* $k+1$. This implies that for any set $A \subset \{-1,+1\}^n$ with $|A| \geq \sum_{i=0}^{k}\binom{n}{i}$, we have

  $$|A \cup \partial V(A)| \geq \sum_{i=0}^{k+1}\binom{n}{i}.$$

  By iterating this argument, we obtain the following simple *consequence of Harper's theorem.*

  Corollary 2.24 *(**Isoperimetric Inequality in Binary Hypercube**) [Boucheron et al., 2013]*
  Let $A \subset \{-1,+1\}^n$ such that $|A| \geq \sum_{i=0}^{k}\binom{n}{i}$. Then for any $t = 1, 2, \ldots, n-k+1$,

  $$|A_t| \geq \sum_{i=0}^{k+1-t}\binom{n}{i}. \tag{27}$$

  In particular, if $|A|/2^n \geq 1/2$ then we may take $k = \lfloor n/2 \rfloor$ in the corollary above and

  $$\frac{|A_t|}{2^n} \geq \mathbb{P}\{X < \mathbb{E}[X] + t\} \geq 1 - \exp\left(-\frac{2t^2}{n}\right) \tag{28}$$

  *where $X \sim Ber(1/2)$ is a **symmetric Bernoulli** random variable taking values in $\{-1,+1\}$ with $\mathbb{P}\{X = 1\} = \mathbb{P}\{X = -1\} = 1/2$. The last inequality comes from Hoeffding's inequality.*

- **Remark** (***Concentration of Uniform Measure on Binary Hypercube***)
  Consider any set A containing at least half of the points of $\{-1,+1\}^n$. According to the corollary above, *the fraction of those points* which ***cannot be obtained*** by *changing* ***at most*** $c\sqrt{n}$ bits of some point in A is ***at most*** e^{2c^2}. In other words, ***an immense majority*** *of the points in $\{-1,+1\}^n$ is* ***within Hamming distance of the order of*** $\sqrt{n}$ ***of*** A

- **Definition** (***Monotone Set***)
  A set $A \subset \{-1,+1\}^n$ is ***monotone*** if $\mathbb{1}\{x \in A\} \geq \mathbb{1}\{y \in A\}$ for all $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\{-1,+1\}^n$ such that $x_i \geq y_i$ ***for all*** i.

- **Theorem 2.25** *(**Isoperimetric Theorem for Bernoulli Distribution**) [Boucheron et al., 2013]*
  *Let $k \in \{0, \ldots, n\}$ and let $S = \{x \in \{-1,+1\}^n : \|x\| < t\}$ be a **Hamming ball of radius** k. $\mathbb{P}$ is Bernoulli distribution on $\{-1,1\}^n$ with parameter p such that*

  $$\mathbb{P}(x) := p^{\|x\|}(1-p)^{1-\|x\|}.$$

  *If $A \subset \{-1,+1\}^n$ is a **monotone set** such that $\mathbb{P}(A) \geq \mathbb{P}(S)$ then*

  $$\mathbb{P}(\partial V(A)) \geq \mathbb{P}(\partial V(S)).$$

References

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.