Concentration of measure

SIAM Working Group - Spring 2019

Contents

1. Introduction

Abstract. In this lecture, we introduce the book's central theme, the study of random fluctuations of functions of independent random variables, along with three techniques that facilitate this study. The specific focus is on how these functions concentrate around measures of central tendency such as their mean and median. We discuss the three main methods below.

Method 1: Isoperimetric Inequalities. Suppose $(\mathcal{X}, d)$ is a metric space, X is a $\mathcal{X}$-valued RV with law $\mathbb{P}$, and $Mf(X)$ is a median of $f(X)$. Then

$$\mathbb{P}\left\{|f(X) - Mf(X)| \geq t\right\} \leq 2\alpha(t)$$

where

$$\alpha(t) := \sup_{\substack{A \in \mathcal{B}(\mathcal{X}) \\ \mathbb{P}(A) \geq \frac{1}{2}}} \mathbb{P}\{d(X, A) \geq t\}$$

and $\mathcal{B}(\mathcal{X})$ is the Borel σ-algebra on $\mathcal{X}$. Thus, bounding $\alpha(t)$ allows one to describe how $f(X)$ concentrates around its median.

Method 2: Entropy. Let $I \subseteq \mathbb{R}$ be an interval, X a I-valued RV, and $\Phi : I \to \mathbb{R}$ a convex function. then the Φ-entropy of X is

$$H_\Phi(X) := E\Phi(X) - \Phi(EX).$$

If $\Phi(x) = x\log(x)$ then we write Ent in place of H_Φ. Bounds on this entropy can be translated to bounds on the concentration of functions of random variables around their mean. For example the Gaussian logarithmic Sobolev inequality states that if $X \sim N(0, I_{n\times n})$ and $f \in C^1(\mathbb{R}^n)$ then

$$\text{Ent}\left[f^2(X)\right] \leq 2E\left[\|\nabla f(X)\|^2\right].$$

In turn, this implies that

$$\mathbb{P}\left\{f(X) - Ef(X) \geq t\right\} \leq e^{-\frac{t^2}{2}}.$$

Method 3: Transportation. Another way of deriving concentration bounds is by getting estimates of the *transportation cost* between two probability measures P and Q:

$$\min_{\mathbf{P}\in\mathcal{P}(P,Q)} \mathbb{E}_{\mathbf{P}} d(X,Y)$$

where d is some cost function, and $\mathcal{P}(P,Q)$ is the class of joint distributions of X and Y such that the marginal distribution of X is P and the marginal distribution of Y is Q.

2. Basic inequalities

Abstract. We use the Cramér-Chernoff method to establish a family of inequalities which give quantitative estimates on concentration of measure subject to certain regularity conditions. We use the machinery of the Cramér-Chernoff method, the log of the moment generating function, to define families of random variables with rapidly decaying tails, sub-Gaussian and more generally sub-Gamma random variables.

THEOREM 1 (Hoeffdinig's Inequality). *Let $X_1, \ldots, X_n$ be independent random variables such that X_i takes its values in $[a_i, b_i]$ almost surely for all $i \leq n$. Let*

$$S := \sum_{i=1}^{n} (X_i - \mathbb{E}X_i).$$

Then for every $t > 0$,

$$\mathbb{P}\{S \geq t\} \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Camé-Chernoff Method. For a random variable X, we use the fact that $\exp(\cdot)$ is an increasing function along with Markov's inequality to get the estimate

$$\mathbb{P}\{X \geq t\} = \mathbb{P}\{e^{\lambda X} \geq e^{\lambda t}\} \leq e^{-\lambda t}\mathbb{E}\left(e^{\lambda X}\right)$$

for every $\lambda > 0$. Picking the optimal λ for a fixed t yields

$$\mathbb{P}\{X \geq t\} \leq \exp\left(-\psi^*(t)\right)$$

where

$$\psi(\lambda) := \log \mathbb{E}\left(e^{\lambda X}\right) \text{ and}$$
$$\psi^*(t) := \sup_{\lambda > 0}\left(t\lambda - \psi(\lambda)\right)$$

which agrees with the Legendre transform of ψ when $t > \mathbb{E}X$.

3. Bounding the variance

Abstract. We extensively utilize the *Efron-Stein* inequality, which gives a bound on the variance of random variables in terms of the mean of their conditional variances. This simple bound has surprising consequences when applied to specific types of functions of random variables and allows us to almost effortlessly obtain sharp bounds on the variance. This is particularly the case for functions satisfying the *bounded difference*, *self-bounding* and *configuration* properties. Other applications include proving Poincaré type inequalities for Gaussian RV's and deriving exponential tail bounds.

THEOREM 2 (Efron-Stein Inequality). *Let $X_1, \dots, X_n$ be independent random variables taking values in $\mathcal{X}$ and let $f : \mathcal{X}^n \to \mathbb{R}$ be given. Set $Z = f(X_1, \dots, X_n)$ and assume $Z \in L^2$. Then we have that*

$$\mathbb{VAR}(Z) \leq \sum_{i=1}^{n} \mathbb{E}[(Z - \mathbb{E}^{(i)}[Z])^2] =: \sum_{i=1}^{n} \mathbb{E}[\mathbb{VAR}^{(i)}[Z]]$$

where

$$\mathbb{E}^{(i)}[Z] := \mathbb{E}[Z|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]$$

i.e. $\mathbb{E}^{(i)}$ *the expectation operator conditioned on*

$$X^{(i)} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n),$$

and

$$\mathbb{VAR}^{i}[Z] := \mathbb{E}^{(i)}[(Z - \mathbb{E}^{(i)}[Z])^2].$$

Fubini Trick. If we write $Z = f(X_1, \dots, X_n)$ as before and $\mathbb{E}_i[Z] := \mathbb{E}[Z|X_1, \dots, X_i]$ then by Fubini we have that

$$\mathbb{E}_i[Z] = \int_{\mathcal{X}^{n-i}} f(X_1, \dots, X_i, x_{i+1}, \dots, x_n) d\mu_{i+1}(x_{i+1}) \dots d\mu_n(x_n)$$

and similarly

$$\mathbb{E}^{(i)}[Z] = \int_{\mathcal{X}} f(X_1, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_n) d\mu_i(x_i)$$

so that we have the identity

$$\mathbb{E}_i[\mathbb{E}^{(i)}[Z]] = \mathbb{E}_{i-1}[Z], \quad \forall i \geq 1$$

where we use the convention that $\mathbb{E}_0[Z] = \mathbb{E}[Z]$. Now if we let $\Delta_i := \mathbb{E}_i[Z] - \mathbb{E}_{i-1}[Z]$ then by this Fubini observation it follows that

$$Z - \mathbb{E}[Z] = \sum_{i=1}^{n} \Delta_i = \sum_{i=1}^{n} \mathbb{E}^{i}[Z - E^{(i)}[Z]]$$

Since for $j > i$ we have that

$$\mathbb{E}[\Delta_i \Delta_j] = \mathbb{E}[\Delta_i \mathbb{E}_i[\Delta_j]] = 0$$

it follows that, using Jensen's inequality,

$$\begin{aligned}\mathbb{VAR}(Z) = \mathbb{E}[(Z - \mathbb{E}[Z])^2] = \sum_{i=1}^{n} \mathbb{E}[\Delta_i^2] + 2\sum_{j>i} \mathbb{E}[\Delta_i \Delta_j] &= \sum_{i=1}^{n} \mathbb{E}[(\mathbb{E}_{i=1}[Z - \mathbb{E}^{i}[Z]])^2] \\ &\leq \sum_{i=1}^{n} \mathbb{E}[(Z - \mathbb{E}^{(i)}[Z])^2]\end{aligned}$$

which gives us Efron-Stein.

4. Basic information inequalities

Abstract. In this lecture we lay the groundwork for the 'entropy method'. In particular we discuss basic properties of the Shannon entropy (relation to conditioning and 'Chain Rule'), we sketch the proofs of Han's inequality and its variant for relative entropies, and we prove sub-additivity of the entropy for discrete random variables (which is an exact analogue of the Efron-Stein inequality where entropies replace variances).

THEOREM 3 (Sub-Additivity of Entropy). *Let* $\Phi(x) := x \log x$ *for* $x > 0$*, let* $X_1, \ldots, X_n$ *be independent random variables, and let* $Y = f(X_1, \ldots, X_n)$ *be a nonnegative measurable function of these variables such that* $\Phi(Y) = Y \log Y$ *is integrable.*

Denote by

- $\mathrm{Ent}(Y)$ *the* entropy *of* Y *defined by*

$$\mathrm{Ent}(Y) := \mathbb{E}\Phi(Y) - \Phi(\mathbb{E}Y),$$

and,

- *for every* $1 \leq i \leq n$*,* $\mathrm{Ent}^{(i)}(Y)$ *the* conditional entropy *of* Y *given* $X^{(i)}$ *defined by*

$$\mathrm{Ent}^{(i)} := \mathbb{E}^{(i)}\Phi(Y) - \Phi(\mathbb{E}^{(i)}Y).$$

Then

$$\mathrm{Ent}(Y) \leq \mathbb{E}\sum_{i=1}^{n} \mathrm{Ent}^{(i)}(Y).$$

Entropy as a Kullback-Leibler divergence (discrete case). For any discrete probability distributions P and Q, if P is absolutely continuous with respect to Q then the Kullback-Leibler divergence from P to Q is defined as

$$D(P||Q) := \sum_x p(x) \log \frac{p(x)}{q(x)}.$$

Recall that this can be interpreted as the additional string length required (on average) to encode a string generated by P using a code designed for Q. This is why P and Q are sometimes referred to as the 'reference' and 'target' distributions respectively.

Now consider a random variable $Z = f(X)$ such that $\mathbb{E}Z = 1$, let P be the disribution of X, and let Q be defined via $q(x) = f(x)p(x)$. Then

$$D(Q||P) = \mathrm{Ent}\, Z.$$

Indeed

$$D(Q||P) = \sum_x f(x)p(x)\frac{f(x)p(x)}{p(x)} = \sum_x f(x)p(x) \log f(x) = \mathbb{E}Z Log Z = \mathrm{Ent}\, Z$$

since $\Phi(\mathbb{E}Z) = \Phi(1) = 0$.

5. Logarithmic Sobolev inequalities

Abstract. The main goal of this chapter is to develop analogues to the Efron-Stein inequality. In particular we will see the similarity between the Efron-Stein inequality and the logarithmic Sobolev inequality, and between the Gaussian Poincaré inequality and the Gaussian logarithmic Sobolev inequality. In each case, not only do these inequalities have structural similarities, but the latter ones are generalizations of the former ones (in some sense). With this similarity in mind, we can generalize the concept of entropy to Φ-entropies, which will be described in Chapter 14. Although the powerful inequalities proved in this chapter only apply to a restricted class of random variables, they will be generalized in Chapter 6.

THEOREM 4 (Logarithmic Sobolev inequality for the symmetric Bernoulli distribution). *Let $f : \{-1,1\}^n \to \mathbb{R}$ be an arbitrary real-valued function defined on the n-dimensional binary hypercube $\{-1,1\}^n$ and assume that X is uniformly distributed over $\{-1,1\}^n$. Then*

$$\mathrm{Ent}(f^2) \leq 2\,\mathcal{E}(f)$$

where

$$\mathcal{E}(f) = \frac{1}{4}\,\mathbb{E}\left[\sum_{i=1}^{n}\Big(f(X) - f(\overline{X}^{(i)})\Big)^2\right] = \frac{1}{2}\,\mathbb{E}\left[\sum_{i=1}^{n}\Big(f(X) - f(\overline{X}^{(i)})\Big)_+^2\right]$$

and where

$$\overline{X}^{(i)} = (X_1,\, \ldots,\, X_{i-1},\, -X_i,\, X_{i+1},\, \ldots,\, X_n)$$

is obtained by flipping the i-th component of X while leaving the others intact.

The central limit theorem kicks in when we extend Theorem 4 to a random variable X with a standard normal distribution: for any ε_i which are uniformly distributed on $\{-1,1\}$ (i.e. Rademacher random variables)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i \to X$$

as $n \to \infty$. We thus have:

THEOREM 5 (Gaussian logarithmic Sobolev inequality). *Let $X = (X_1,\, \ldots,\, X_n)$ be a vector of n independent standard normal random variables and let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function. Then*

$$\mathrm{Ent}(f^2) \leq 2\|\nabla f(X)\|^2.$$

Herbst's argument. Herbst's argument is the main tool which will lead us to concentration inequalities. By putting $g_\lambda(x) := e^{\lambda f(x)/2}$ we can derive a differential inequality satsfied by the entropy. Solving this inequality leads to a concentration result. To be precise: for $Z = f(x)$,

$$\mathrm{Ent}(g_\lambda^2(X)) = \mathrm{Ent}(e^{\lambda f(X)}) = \lambda\mathbb{E}[Ze^{\lambda Z}] - \mathbb{E}e^{\lambda Z}\log\mathbb{E}e^{\lambda Z}.$$

Let $F(\lambda) = \mathbb{E}e^{\lambda Z}$ so that we can express $\mathbb{E}[Ze^{\lambda Z}] = F'(\lambda)$. Therefore

$$\mathrm{Ent}(g_\lambda^2(X)) = \lambda F'(\lambda) - F(\lambda)\log F(\lambda).$$

Tsirelson, Ibragimov and Sudakov used Herbst's argument to prove exponential tail inequalites for L-Lipschitz functions of independent Gaussian random variables. It should be noted that the resulting concentration bounds are dimension-free. This feature allows us to extend the results from previous chapters.

6. Entropy method

Abstract. In this chapter, we will explore a method of proof called the *entropy method*. The main idea is that one will try to use the convexity of the entropy and Herbst's argument in order to prove the desired inequalities. Important results obtained via this method include an exponential inequality for self-bounded functions and the exponential Efron-Stein inequality.

Herbst's agument. Let $Z = f(X_1, \ldots, X_n)$ where the X_i's are independent random variables (not necessarily identically distributed). Denoting $\varphi(\lambda) := \log \mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}$ one can see that

$$\frac{\mathrm{Ent}(e^{\lambda Z})}{\mathbb{E}e^{\lambda Z}} = \lambda\varphi'(\lambda) - \varphi(\lambda)\,.$$

Herbst's argument allows us to use an differential inequality to prove that if

$$\frac{\mathrm{Ent}(e^{\lambda Z})}{\mathbb{E}e^{\lambda Z}} \leq C\lambda^2\,,$$

then

$$\log \mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} \leq C\lambda^2\,.$$

Indeed, if

$$\frac{\mathrm{Ent}(e^{\lambda Z})}{\mathbb{E}e^{\lambda Z}} \leq C\lambda^2\,,$$

then

$$\lambda\varphi'(\lambda) - \varphi(\lambda) = \frac{\mathrm{Ent}(e^{\lambda Z})}{\mathbb{E}e^{\lambda Z}} \leq C\lambda^2$$

and hence

$$\left(\frac{\varphi(\lambda)}{\lambda}\right)' = \frac{\varphi'(\lambda)}{\lambda} - \frac{\varphi(\lambda)}{\lambda^2} \leq C^2.$$

Integrating this inequality gives us that $\log \mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} = \varphi(\lambda) \leq C\lambda^2$.

The entropy method. In this chapter, a central theme is that one can deduce from the convexity (or sub-additivity) of the entropy a modified logarithmic Sobolev inequality:

$$\mathrm{Ent}\left(e^{\lambda Z}\right) = \lambda\mathbb{E}(Ze^{\lambda Z}) - \mathbb{E}(e^{\lambda Z})\log\mathbb{E}(e^{\lambda Z}) \leq \sum_{i=1}^{n}\mathbb{E}(e^{\lambda Z}\phi(-\lambda(Z-Z_i)))\,,$$

where $\phi(x) := e^x - x - 1$ and $Z_i := f\left(X^{(i)}\right) = f\left(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\right)$. Combining this with Herbst's argument allows us to prove all sorts of exponential inequalities.

Theorem 6 (Exponential self-bounding inequality). *Let $X_1, \ldots, X_n$ be independent random variables taking values in $\mathcal{X}$ and let $f : \mathcal{X} \to [0,\infty)$ satisfy the self-bounding property, i.e. there exist functions $f_i : \mathcal{X}^{n-1} \to \mathbb{R}$ such that, for all $x_1, \ldots, x_n \in \mathcal{X}$ and all $i = 1, \ldots, n$,*

$$0 \leq f\left(x_1, \ldots, x_n\right) - f_i\left(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\right) \leq 1$$

and

$$\sum_{i=1}^{n}\left(f\left(x_1, \ldots, x_n\right) - f_i\left(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\right)\right) \leq f\left(x_1, \ldots, x_n\right).$$

Then, for any $\lambda \in \mathbb{R}$ and for $Z := f\left(X_1, \ldots, X_n\right)$,

$$\log \mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} \leq \phi(\lambda)\mathbb{E}Z\,.$$

where $\phi(x) := e^x - x - 1$.

For example, $f = ||\cdot||_{l^p}^p$ is a self-bounding function on $[0,1]^n$ for any $1 \le p < \infty$ since $f - f_i = |x_i|^p \in [0,1]$ and $\sum_i (f - f_i) = \sum_i |x|^i = |x_i|^p = f$. Similarly, on any bounded set, appropriate multiples of $||\cdot||_{l^p}^p$ are self-bounding functions.

THEOREM 7 (Exponential Efron-Stein inequality). *Let* $X_1, \dots, X_n$ *be independent random variables and let* $Z = f(X_1, \dots, X_n)$. *For each* $i = 1, \dots, n$ *let* X_i' *be an independent copy of* X_i *and let* $X_i' := f(X_1, \dots, X_i', \dots, X_n)$. *Define* $\mathbb{E}' := \mathbb{E}[\,\cdot\,|X]$ *(i.e. integrating out the variables* $X_1', \dots, X_n'$*) and*

$$V^+ := \sum_{i=1}^{n} \mathbb{E}'\left[(Z - Z_i')_+^2\right]$$

and let $\theta, \lambda > 0$ *such that* $\theta\lambda < 1$ *and* $\mathbb{E}e^{\lambda V^+/\theta} < \infty$. *Then*

$$\log \mathbb{E}e^{\lambda(Z - \mathbb{E}Z)} \le \frac{\lambda\theta}{1 - \lambda\theta} \log \mathbb{E}e^{\lambda V^+/\theta}.$$

7. Concentration and Isoperimetry

Abstract. In this chapter, we discuss the relationship between isoperimeter problems and concentration of measure. In particular, we introduce the concentration function and Lévy's inequalities. We also discuss Talagrand's convex distance inequality. The key to this proof involves the idea of a self bounding function, and that the squared convex distance is self bounding.

THEOREM 8 (Lévy's inequalities). *Given a metric space* $(\mathcal{X}, d)$ *and a random variable* X *on* $\mathcal{X}$ *distributed according to the measure* μ, *we define the concentration function*

$$\alpha(t) := \sup_{A \subset \mathcal{X},\ \mu(A) \ge 1/2} \mathbb{P}\{d(X, A) \ge t\}$$

For any 1-Lipschitz function $f : \mathcal{X} \to \mathbb{R}$, *we have*

$$\mathbb{P}\{f(X) \ge Mf(X) + t\} \le \alpha(t) \text{ and } \mathbb{P}\{f(X) \le Mf(X) - t\} \le \alpha(t),$$

where $Mf(X)$ *denotes any median of* $f(X)$. *Conversely, if* $\beta : (0,\infty) \to [0,1]$ *is a function such that for every 1-Lipschitz function* $f : \mathcal{X} \to \mathbb{R}$ *we have*

$$\mathbb{P}\{f(X) \ge Mf(X) + t\} \le \beta(t),$$

then $\beta(t) \le \alpha(t)$.

THEOREM 9 (Talagrand's convex distance inequality). *Consider independent random variables* $X_1, \dots, X_n$ *taking values in* $\mathcal{X}$ *and denote their vector as* $X = (X_1, \dots, X_n) \in \mathcal{X}^n$. *For* $\alpha \in [0,\infty)^n$ *we define the weighted Hamming distance to a set* $A \subset \mathcal{X}^n$ *by*

$$d_\alpha(x, A) = \inf_{y \in A} \sum_{i : x_i \ne y_i} |\alpha_i|$$

The convex distance to a set A *is given by*

$$d_T(x, A) := \sup_{\alpha \in [0,\infty)^n, \|\alpha\| = 1} d_\alpha(x, A) \tag{1}$$

where $\|\cdot\|$ *denotes the Euclidean norm. We have*

$$\mathbb{P}(A)\,\mathbb{P}(d_T(X, A) \ge t) \le e^{-t^2/4}.$$

Gaussian isoperimetric inequality. Lévy's inequalities provide us with the connection between concentration of measure type results and isoperimetric problems. In particular, we can get isoperimetric results very cheaply by using our concentration theorems. We will demonstrate a way to get a very sharp isoperimetric inequality for the Gaussian measure.

If X is a standard Gaussian vector then for any 1-Lipschitz $f : \mathbb{R}^n \to \mathbb{R}$ and $t > 0$,

$$\begin{aligned}\mathbb{P}\{f(X) \geq Mf(X) + t\} &= \mathbb{P}\{f(X) \geq Ef(X) + (Mf(X) - Ef(X) + t)\} \\ &\leq \exp\left\{\frac{1}{2}(Mf(X) - Ef(X) + t)^2\right\},\end{aligned}$$

where we have used a concentration result for $f(X)$. Now, for any random variable Z we have

$$|MZ - EZ| \leq \sqrt{\mathbb{VAR}(Z)},$$

so in particular we have

$$Mf(X) - Ef(X) \leq \sqrt{\mathbb{VAR}(f(X))} \leq 1$$

and so, by Lévy's inequality,

$$\alpha(t) \leq e^{-\frac{1}{2}(1+t)^2}.$$

8. The Transportation Method

Abstract. This chapter takes on a different approach than previous chapters, by using tools from the theory of *optimal transport* to derive Sub-Gaussian concentration bounds. We will derive similar bounds to those we have seen using the Entropy method; most notably proving versions of the Bounded Differences and Talagrand convex distance results seen before. Though most of the examples that will be discussed give rise to similar bounds proved using other methods, this method is not just an alternative – in many cases it leads to concentration of measure results for which other methods to derive them are not known.

THEOREM 10 (Marton's Transportation Inequality). *Let P, Q be measures on $\mathcal{X}^n$ with $Q << P$ and write $\mathcal{P}(P,Q)$ for the set of all joint probability measures of P and Q (i.e. measures on $\mathcal{X}^n \times \mathcal{X}^n$ that preserve the marginals P and Q). Letting $X = (X_1, \ldots, X_n) \sim P$ and $Y = (Y_1, \ldots, Y_n) \sim Q$ we have that*

$$\min_{\mathbb{P} \in \mathcal{P}(P,Q)} \sum_{i=1}^{n} \mathbb{P}^2(X_i \neq Y_i) \leq \frac{1}{2} D(Q||P).$$

THEOREM 11 (Marton's Conditional Transportation Inequality). *Adopt the notation of Theorem 10. Then*

$$\min_{\mathbb{P} \in \mathcal{P}(P,Q)} \mathbb{E}_{\mathbb{P}} \left[\sum_{i=1}^{n} \left(\mathbb{P}^2(X_i \neq Y_i | X_i) + \mathbb{P}^2(X_i \neq Y_i | Y_i)\right)\right] \leq 2D(Q||P).$$

Transportation Lemma. The glue connecting optimal transport (including Theorem 10 and Theorem 11 above) to concentration inequalities is the following transportation lemma:

Lemma 12 (Transportation Lemma). *As before write $Z = f(X)$ for $f : \mathcal{X}^n \to \mathbb{R}$ and ψ_Z for the cumulant generating function, i.e. $\psi_Z(\lambda) = \log \mathbb{E}[e^{\lambda} Z]$ is the log moment generating function. If Z has mean zero then we say that Z is* sub-Gaussian *with* variance factor v *if*

$$\psi_Z(\lambda) \leq \frac{\lambda v^2}{2}$$

for every $\lambda \in \mathbb{R}$, in which cas we write $Z \sim SG(v)$. We have that $Z \sim SG(v)$ if and only if for every probability measure $Q << P$ such that $D(Q||P) < \infty$ we have that

$$\mathbb{E}_Q[f] - \mathbb{E}_P[f] \leq \sqrt{2vD(Q||P)}. \tag{2}$$

This allows us to prove sub-Gaussian concentration inequalities by proving the transportation inequality (2). We will typically assume that f satisfies the *1-sided Lipschitz condition* $f(y) - f(x) \leq \sum_{i=1}^{n} c_i(x) d(x_i, y_i)$ where c_i are some functions and d is a pseudometric. From this it follows that

$$\begin{aligned}\mathbb{E}_Q[f] - \mathbb{E}_P[f] \leq \left(\sum_{i=1}^{n} \mathbb{E}[c_i(X)]^2\right)^{1/2} &\left(\inf_{\mathbb{P} \in \mathcal{P}(P,Q)} \mathbb{E}_{\mathbb{P}}\left[\left(\sum_{i=1}^{n} \mathbb{E}_{\mathbb{P}}[d(X_i, Y_i)^2 | X_i]\right)\right]\right)^{1/2} \\ := \sqrt{v} \cdot \sqrt{\Theta(P,Q)}.&\end{aligned}$$

Thus it suffices to prove

$$\Theta(P, Q) \leq 2D(Q||P)$$

This is normally done in two steps (and is how Theorem 10 and Theorem 11 are proved):

(1) First one tries to *characterize* $\Theta(P, Q)$. That is to find its value in terms of some metric of P and Q. One may have to do this directly or could refer to the optimal transport literature for known results.
(2) Bounding the quantity found in step 1.

Taking $d(X_i, Y_i) = \mathbb{P}(X_i \neq Y_i)$ and $c_i(X) \equiv c_i$ together with Theorem 10 yields the bounded differences inequality, whereas that same distance and setting $c_i(x)$ to be the supremum over α in (1) together with Theorem 11 yields Talagrand's convex distance inequality.

For a new example take the Euclidian distance $d(x, y) = |x - y|$ and assume $f : \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz; this yields that $Z = f(X) \sim SG(L)$ if X is standard Gaussian by the following transportation theorem:

Theorem 13 (Talagrand's Gaussian Transportation Inequality). *Let P be the standard Gaussian measure on $\mathbb{R}^n$ and Q a probability measure with $Q << P$. Then*

$$\min_{\mathbb{P} \in \mathcal{P}(P,Q)} \sum_{i=1}^{n} \mathbb{E}_{\mathbb{P}}[(X_i - Y_i)^2] \leq 2D(Q||P)$$

which is proved in the standard two-step fashion explained above.

List of definitions

List of results

List of 'tricks'

Credits

Lecture 1	David Gutman
Lecture 2	Adrian Hagerty
Lecture 3	David Itkin
Lecture 4	Antoine Remond-Tiedrez
Lecture 5	Won Eui Hong
Lecture 6	Son Van
Lecture 7	Adrian Hagerty
Lecture 8	David Itkin
Editing	Antoine Remond-Tiedrez

References

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities.* Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.