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 $t > 0$,

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

Cramé-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 $X_1, \ldots, X_n$ be independent random variables taking values in $\mathcal{X}$ and let $f : \mathcal{X}^n \to \mathbb{R}$ be given. Set $Z = f(X_1, \ldots, X_n)$ 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

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

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

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

and

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

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

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

and similarly

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

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,

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

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

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

- Ent(Y) *the* entropy *of* Y *defined by*

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

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

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

Then

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

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

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

Indeed

$$D(Q||P) = \sum_{x} f(x)p(x)\frac{f(x)p(x)}{p(x)} = \sum_{x} f(x)p(x) \log f(x) = \mathbb{E}ZLogZ = \text{Ent } Z$$

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*

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

where

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

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*

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

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 $Z = f(x)$,

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

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

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

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 Entropy method. The main idea is that one will try to use convexity of the entropy and the Herbst's argument in order to prove the desired inequalities. Important consequences of this method include the exponential Efron-Stein inequality and exponential inequality for self-bounded functions.

Let $Z = f(X_1, ..., X_n)$ where X_i's are independent RVs (not necessarily IID). Denote $\varphi(\lambda) = \log \mathbb{E}e^{\lambda(Z - \mathbb{E}Z)}$ and one could see that

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

The Herbst's argument lets you use an differential inequality to prove the following. If

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

then

$$\log \mathbb{E}e^{\lambda(Z - \mathbb{E}Z)} \leq C \,. \tag{1}$$

In this chapter, one theme that follows from the convexity of entropy is some sort of modified logarithmic Sobolev inequalities, which, more or less, has the following form

$$\lambda \mathbb{E}(Ze^{\lambda Z}) - \mathbb{E}(e^{\lambda Z}) \log \mathbb{E}(e^{\lambda Z}) \leq \sum_{i=1}^{n} \mathbb{E}(e^{\lambda Z}\phi(-\lambda(Z - Z_i))) \,, \tag{2}$$

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

The general philosophy is that, after translating the convexity into the Log Sobolev inequality (2), we use the result (1) of Herbst's argument to prove all sorts of exponential inequalities. We only name two here since they seem to be important.

THEOREM 6 (Exponential self-bounding inequality). *Assume Z satisfies the self-bounding property. Then for every $\lambda \in \mathbb{R}$,*

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

THEOREM 7 (Exponential Efron-Stein). *Denote $V^+ = \sum_{i=1}^{n} \mathbb{E}((Z - Z_i')_+^2 | X_1, ..., X_n)$, where $Z_i' = f(X_1, ..., X_i', ..., X_n)$ for $X_i \sim X_i'$. Let $\theta, \lambda > 0$ so that $\theta\lambda < 1$ and $\mathbb{E}e^{\lambda V^+/\theta} < \infty$. Then*

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

In conclusion, we have the following trick:

Entropy method. Entropy method = Herbst's argument (typically something like (1)) + convexity of entropy (typically translate to some sort of modified log Sobolev inequality (2)).

List of definitions

Conditional expectation ($\mathbb{E}^{(i)}$)
Conditional variance ($\mathbb{VAR}^{(i)}$)
Entropy (Ent) and conditional entropy ($\mathrm{Ent}^{(i)}$)

List of results

Hoeffding's inequality
Efron-Stein inequality
Sub-additivity of entropy
Logarithmic Sobolev inequality for the symmetric Bernoulli distribution
Gaussian logarithmic Sobolev inequality Exponential self-bounding inequality Exponential Efron-Stein

List of 'tricks'

Method of isoperimetric inequalities
Entropy method
Transport method
Cramér-Chernoff method
Fubini trick
Entropy as a Kullback-Leibler divergence (discrete case)
Herbst's argument
Entropy method

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

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.