

Concentration of measure

SIAM Working Group - Spring 2019

CONTENTS

1. Introduction	1
2. Basic inequalities	2
3. Bounding the variance	3
4. Basic information inequalities	4
5. Logarithmic Sobolev inequalities	5
List of results	6
List of ‘tricks’	6
Credits	6
References	6

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

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

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 $\Phi : I \rightarrow \mathbb{R}$ 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

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

In turn, this implies that

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

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 (Hoeffding's Inequality). *Let X_1, \dots, 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 $t > 0$,

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

REMARK 2 (Cramér-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*

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

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

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

where

$$\begin{aligned} \psi(\lambda) &:= \log \mathbb{E}(e^{\lambda X}) \quad \text{and} \\ \psi^*(t) &:= \sup_{\lambda > 0} (t\lambda - \psi(\lambda)) \end{aligned}$$

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

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

where

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

and

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

REMARK 4 (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

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

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,

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

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

For every $1 \leq i \leq n$, denote by

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

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

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

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

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

and

- $\text{Ent}^{(i)}(Y)$ *the conditional entropy of Y given $X^{(i)}$ defined by*

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

Then

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

REMARK 6 (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 distribution of X , and let Q be defined via $q(x) = f(x)p(x)$. Then

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

Indeed

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

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

5. Logarithmic Sobolev inequalities

Abstract.

THEOREM 7 (Name of important result).

REMARK 8 (Name of the ‘trick’).

List of results

Hoeffding's inequality
Efron-Stein inequality
Sub-additivity of entropy

List of 'tricks'

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

Credits

Lecture 1 David Gutman
Lecture 2 Adrian Hagerty
Lecture 3 David Itkin
Lecture 4 Antoine Remond-Tiedrez
Lecture 5 Won Eui Hong

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.