Concentration of measure

SIAM Working Group - Spring 2019

Contents

1. Introduction

Abstract. In this lecture, we introduced 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 P, and $Mf(X)$ is a median of $f(X)$. Then

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

where

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

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 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

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

In turn, this implies that

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

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

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

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

2. Basic inequalities

Abstract. We used the Cramér-Chernoff method to establish a family of inequalities which give quantitative estimates on concentration of measure in the case of 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 $X_1, \dots, X_n$ be independent random variables such that X_i takes its values in $[a_i, b_i]$ almost surely for all $i \leq n$. Let*

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

Then for every $t > 0$,

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

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*

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

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

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

where

$$\psi(\lambda) = \log \mathbf{E}\left(e^{\lambda X}\right)$$

$$\psi^*(t) = \sup_{\lambda>0}\left(t\lambda - \psi(\lambda)\right)$$

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

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 variance. 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 $X_1, \ldots, X_n$ be IRVs 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 $\mathbb{E}^{(i)}[Z] := \mathbb{E}[Z|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n]$ *and* $\mathbb{VAR}^i[Z] := \mathbb{E}^{(i)}[(Z-\mathbb{E}^{(i)}[Z])^2]$.

REMARK 4 (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

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

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

which gives us Efron-Stein.

Credits

Lecture 1	David Gutman
Lecture 2	Adrian Hagerty
Lecture 3	David Itkin

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.