NOTES ON CONCENTRATION OF MEASURE, PART 1

HAN CHENG LIE

Abstract. Notes for a 90-minute presentation on Concentration of Measure for the Uncertainty Quantification seminar at Freie Universität Berlin, held on Tuesday, the 3rd of May, 2016. Based on resources from Wikipedia, unpublished notes by A. M. Stuart and T. J. Sullivan [2] on uncertainty quantification, and M. Ledoux's book on concentration of measure [1].

1. Motivation: uncertainty quantification (from [2])

Let $(\Theta, \mathcal{A}, \mathbb{P})$ be a probability space, $D \subset \mathbb{R}^d$ be a bounded, open set, and $u : D \times \Theta \to \mathbb{R}$ be a random field, so that $u(x, \cdot) : \Theta \to \mathbb{R}$ is a random variable and $u(\cdot, \theta) : D \to \mathbb{R}$ is a function for $\mathbb{P}$-almost all θ. Define the Hilbert space

$$H := L^2(D; \mathbb{R}) = \left\{ u : D \to \mathbb{R} \mid \int_D |u(x)|^2 dx < \infty \right\}$$

In UQ applications, dim $H = +\infty$, and we seek to approximate a random field u (e.g. the solution to some PDE with random coefficients) in some orthonormal basis $(\psi_n)_{n\in\mathbb{N}}$, where the $\psi_n : D \to \mathbb{R}$ are deterministic.

Key question: convergence of sequence of partial sums

$$u^N(x, \theta) := m(x) + \sum_{n=1}^{N} \gamma_n \theta_n \psi_n(x)$$

to a limit u.

Assumption 1.1 (Assumption 2.4.2 in [2]). *The sequence $\gamma = (\gamma_j)_{j\in\mathbb{N}} \in \ell^2(\mathbb{N}; \mathbb{R})$ and the sequence $\theta = (\theta_j)_{j\in\mathbb{N}}$ is a sequence of i.i.d. $N(0, 1)$ random variables.*

Theorem 1.2 (Construction of Gaussian prior (Theorem 2.4.3 in [2]). *Under Assumption 1.1, the sequence of partial sums $(u^N)_{N\in\mathbb{N}}$ is Cauchy in the Hilbert space $\mathcal{H} := L^2_{\mathbb{P}}(\Theta; H)$, and the limit u is a H-valued random variable with law given by a probability measure μ.*

High-dimensional probability:

Let $\theta = (\theta_j)_{j\in\mathbb{N}}$ be a sequence of independent, $N(0, 1)$ random variables. Then $\theta^N = (\theta_j)_{j=1}^N$ is a $\mathbb{R}^N$-valued random variable, distributed according to $N(0, I_N)$ for the $N \times N$ identity matrix I_N.

Question: Where is θ^N most likely to be found?

2. First example (excerpt from §1.1 of [1])

Notation:

- (X, d) - metric space X endowed with metric $d : X \times X \to [0, \infty)$
- $\mathcal{B}(X)$ - Borel sigma-algebra of X
- $\mathcal{M}_1(X)$ - space of probability measures on X (with Borel sigma-algebra)
- $B^{n+1}_r(x) = \left\{ y \in \mathbb{R}^{n+1} \mid |y - x| < r \right\}$ - open ball of radius r, centre x in $\mathbb{R}^{n+1}$
- λ^{n+1} - Lebesgue measure in $\mathbb{R}^{n+1}$
- $\omega(n + 1) = \lambda^{n+1}(B^{n+1}_1(0))$ - volume w.r.t Lebesgue measure of $B^{n+1}_1(0)$
- $S^n = \left\{ x \in \mathbb{R}^{n+1} \mid \|x\| = 1 \right\}$ - sphere of radius 1 in $\mathbb{R}^{n+1}$ $(= \partial B^{n+1}_1(0))$

Given $A \in \mathcal{B}(X)$ and $r > 0$, we shall write the open r-neighbourhood of A as

$$A_r := \{x \in X \mid d(x, A) < r\}. \tag{2.1}$$

Date: May 3, 2016.

2.1. **Paul Lévy's example (from Wikipedia).** We recall the 'wedge' construction of the uniform measure σ^n on S^n. Equip S^n with the geodesic distance on the sphere. Given $A \in \mathcal{B}(S^n)$, the corresponding *wedge* is a subset of the closed unit ball $\overline{B_1^{n+1}(0)}$, given by

$$wed(A) := \{tx \mid x \in A,\ t \in [0,1]\},$$

i.e. the volume obtained by scaling the set A using a scaling parameter $t \in [0,1]$. Then the uniform measure σ^n defined on S^n is given by

$$\sigma^n(A) := \frac{1}{\omega(n+1)} \lambda^{n+1}(wed(A)), \quad \forall A \in \mathcal{B}(S^n) \tag{2.2}$$

Theorem 2.1 (Conc. of unif. meas. on S^n for $n \geq 2$)**.** *Given the uniform measure σ^n constructed in (2.2), suppose that $A \in \mathcal{B}(S^n)$ satisfies $\sigma^n(A) \geq 0.5$. Then*

$$\sigma^n(A_r) \geq 1 - \exp\left(-(n-1)\frac{r^2}{2}\right) \quad \forall 0 < r \leq \pi. \tag{2.3}$$

Remark 1. The result (2.3) may be interpreted (see [1]) as the statement that almost all points on S^n (for $n \geq 2$) are within geodesic distance $\frac{1}{\sqrt{n}}$ from A.

Note that the lower bound is less useful for 'large' values of r: since if $r = \pi$, then the lower bound given by (2.3) is rather bad, since

$$1 - \exp\left(-(n-1)\frac{\pi^2}{2}\right) < 1 \quad \forall n \in \mathbb{N}$$

whereas it must hold that $\sigma^n(A_\pi) = 1$, since the geodesic distance between any two distinct points on S^n is at most π, by definition of geodesic distance. Since we are interested in high-dimensional spheres, an appropriate way of interpreting (2.3) is to fix $0 < r \ll \pi$, and to observe that the σ^n- measure of A_r increases rapidly in the dimension n. Alternatively, if we let n grow large, r must decrease as $O(n^{-1/2})$ in order for the measure of A_r to remain $O(1)$. A nice feature of the concentration result is that it does not matter what the set A looks like, provided that $\sigma^n(A) \geq 1/2$.

3. §1.2 from [1]: Concentration functions

Definition 3.1 (Concentration function)**.** Let (X, d) be a metric space. The concentration function associated to $\mu \in \mathcal{M}_1(X)$ is given by

$$\alpha_{(X,d,\mu)}(r) := \sup\left\{1 - \mu(A_r) \mid A \in \mathcal{B}(X),\ \mu(A) \geq \frac{1}{2}\right\}, \quad \forall r > 0. \tag{3.1}$$

Note that by definition, a concentration function is always bounded from above by $\frac{1}{2}$.

If (X, d) is a bounded metric space, the r-enlargement A_r defined in (2.1) is sensible for $0 < r$ bounded above by

$$\mathrm{Diam}(X, d) := \sup\{d(x, y) \mid x, y \in X\}. \tag{3.2}$$

We have the properties that

$$\alpha_{(X,d,\mu)}(r) = 0 \ \forall r > \mathrm{Diam}(X, d) \tag{3.3a}$$

$$\alpha_{(X,d,\mu)}(r) \searrow 0 \text{ as } r \to +\infty. \tag{3.3b}$$

The basic idea of concentration of measure can be described as follows:

$$\alpha_{(X,d,\mu)}(r) \searrow 0 \text{ rapidly with } r \text{ or } \dim(X). \tag{3.4}$$

3.1. **Concentration for Gaussian measure.** Let γ (γ^k) denote the canonical Gaussian probability measure on $\mathbb{R}$ ($\mathbb{R}^k$) equipped with the Euclidean metric. We will abuse notation and also use γ as a density, viz., for the cumulative distribution function (CDF) of the Gaussian probability measure on $\mathbb{R}$ and any $a \in \mathbb{R}$, we shall write $\Phi(a) := \int_{-\infty}^{a} \gamma(x)dx$.

Let $A \in \mathcal{B}(\mathbb{R}^k)$ such that $\gamma^k(A) = \Phi(a)$ for some $-\infty \leq a \leq +\infty$. Then $\forall r > 0$,

$$\gamma^k(A_r) \geq \Phi(a + r). \tag{3.5}$$

Define the *concentration function* for γ by

$$\alpha_\gamma(r) := \sup\left\{1 - \gamma^k(A_r) \mid A \in \mathcal{B}(\mathbb{R}^k),\ \gamma^k(A) \geq \frac{1}{2}\right\} \quad \forall r > 0. \tag{3.6}$$

Since $\Phi(0) = \frac{1}{2}$ and $1 - \Phi(r) \leq \exp\left(-\frac{r^2}{2}\right)$ for $r > 0$, the concentration function α_γ defined in (3.6) satisfies

$$\alpha_{\gamma^k}(r) \leq \exp\left(-\frac{r^2}{2}\right), \quad \forall r > 0. \tag{3.7}$$

The interpretation of (3.7) is that for any $A \in \mathcal{B}(\mathbb{R}^k)$ with $\gamma^k(A) \geq \frac{1}{2}$, γ^k-almost all points in $\mathbb{R}^k$ are within distance 10 of A. In this case, "γ-almost all" has the quantitative meaning

$$\exp(-\frac{10^2}{2}),$$

which is very close to zero.

3.2. **Concentration for uniform measure on S^n.** Note that we can define the concentration function for the uniform measure σ^n on S^n similarly, as

$$\alpha_{\sigma^n}(r) := \sup\left\{1 - \sigma^n(A_r) \mid A \subset \mathbb{R}^k,\ \sigma^n(A) \geq \frac{1}{2}\right\}$$

and by (2.3) we have, analogous to (3.7),

$$\alpha_{\sigma^n}(r) \leq \exp\left(-(n-1)\frac{r^2}{2}\right). \tag{3.8}$$

Remark 2. Note that, unlike the statement for the product measure γ in (3.7), the bound on the concentration function in (3.8) depends on the dimension n.

3.3. **Formal definitions of concentration.**

Definition 3.2. A probability measure $\mu \in \mathcal{M}_1(X)$ exhibits normal (exponential) concentration on (X, d) if $\exists C, c > 0$ not depending on r such that

$$\alpha_{(X,d,\mu)}(r) \leq C \exp\left(-cr^p\right), \quad \forall r > 0, \tag{3.9}$$

for $p = 2$ ($p = 1$).

Examples of normal concentration: $(X, d, \mu) = (S^n, d, \sigma^n)$ or $(\mathbb{R}^k, |\cdot|, \gamma^k)$.

Note: the bound $\mu(A) \geq 1/2$ may be replaced by $\mu(A) \geq \varepsilon > 0$, which yields corresponding statements on concentration functions.

4. §1.3 Deviation inequalities

4.1. **Concentration around medians for continuous functions.** Let $\mu \in \mathcal{M}_1(X)$, $F : X \to \mathbb{R}$ measurable.

Definition 4.1 (Median of F)**.** A median of F for μ is a number m_F such that

$$\mu\left(\{F \leq m_F\}\right) \geq \frac{1}{2} \text{ and } \mu\left(\{F \geq m_F\}\right) \geq \frac{1}{2} \tag{4.1}$$

Definition 4.2 (Modulus of continuity)**.** For $F \in C(X; \mathbb{R})$, the modulus of continuity of F is a function $\omega_F : \mathbb{R}_{++} \to \mathbb{R}_+ \cup \{+\infty\}$ defined by

$$\omega_F(\eta) = \sup\left\{|F(x) - F(y)| \mid d(x, y) < \eta\right\}, \ \eta > 0. \tag{4.2}$$

Interpretation: Functions that exhibit large, highly localised oscillations have large values of $\omega_F(\eta)$ for small η. Functions that exhibit very small local oscillations, i.e. are almost constant, have very small values of $\omega_F(\eta)$ for large η.

Observe that if m_F is a median and $A := \{F \leq m_F\}$, then for all $x \in X$ such that $d(x, y) < \eta$ for some $y \in A$,

$$F(x) \leq F(y) + \omega_F(\eta) \leq m_F + \omega_F(\eta),$$

where the first inequality follows from (4.2) and the second inequality follows from the fact that $y \in A$. From the above inequality we may deduce

$$\mu\left(\{F > m_F + \omega_F(\eta)\}\right) \leq \alpha_\mu(\eta). \tag{4.3}$$

Why is this important? It gives us a quantity which we can control using the concentration function. Analogously, if $A = \{F \geq m_F\}$, then

$$\mu\left(\{F < m_F + \omega_F(\eta)\}\right) \leq \alpha_\mu(\eta),$$

which with (4.3) yields the two-sided inequality

$$\mu\left(\{|F-m_F|>\omega_F(\eta)\}\right)\leq 2\alpha_\mu(\eta). \tag{4.4}$$

The left-hand side of (4.4) gives the measure of the set of points at which the function deviates from a given median value. Recall that the concentration function α_{S^n,d,σ^n} decays exponentially with the dimension n.

Interpretation: Equation (4.4) expresses that continuous functions defined on X with small local oscillations are almost constant on μ-almost all of X as n increases.

Note: for $X=S^n$ and d the geodesic distance on S^n, equations (4.3) and (4.4) are called "L'evy's inequalities"'.

4.2. **Concentration around medians of Lipschitz functions.** Now consider the subset of Lipschitz functions in $C(X;\mathbb{R})$.

Definition 4.3 (Lipschitz functions)**.** $F\in C(X;\mathbb{R})$ is Lipschitz if

$$\|F\|_{Lip}:=\sup_{x\neq y}\frac{|F(x)-F(y)|}{d(x,y)}<\infty, \tag{4.5}$$

and F is "1-Lipschitz" if $\|F\|_{Lip}\leq 1$.

Note that by (4.2) and (4.5) we have

$$\omega_F(\eta)\leq\eta\|F\|_{Lip},\ \forall\eta>0. \tag{4.6}$$

In particular, if F is Lipschitz, then its modulus of continuity is finite for all finite η. Lipschitz continuity is a stronger property than continuity.

Our first "deviation inequality" is of the form

$$\mu\left(\{F\geq m_F+r\}\right)\leq\alpha_\mu\left(\frac{r}{\|F\|_{Lip}}\right),\ \forall r>0, \tag{4.7}$$

and from it we can derive the "concentration inequality" of F around the median m_F with rate α_μ,

$$\mu\left(\{|F-m_F|\geq r\}\right)\leq 2\alpha_\mu\left(\frac{r}{\|F\|_{Lip}}\right),\ \forall r>0. \tag{4.8}$$

We interpret (4.8) as saying that F concentrates around m_F on a portion of its domain that has large μ-measure.

Remark 3. No constraints have been placed on m_F and $\|F\|_{Lip}$, especially their relative sizes. The concentration phenomenon does not yield any information on $\|F\|_{Lip}$ or m_F.

We can replace the median m_F in (4.7) with the mean $\int Fd\mu$:

Proposition 4.1. *For $\mu\in\mathcal{M}_1(X)$, $\alpha:\mathbb{R}_+\to\mathbb{R}_+$ and F 1-Lipschitz such that*

$$\mu\left(\left\{F\geq\int Fd\mu+r\right\}\right)\leq\alpha(r),\ \forall r>0, \tag{4.9}$$

it holds that

$$\forall A\in\mathcal{B}(X)\ \text{such that}\ \mu(A)>0,\ 1-\mu(A_r)\leq\alpha(\mu(A)r)\ \forall r>0,$$

and in particular $\alpha_{(X,d,\mu)}(r)\leq\alpha(r/2)$.

4.3. **Some consequences of normal concentration.** Next proposition shows that normal concentration implies strong integrability properties for Lipschitz functions.

Proposition 4.2. *Let $F:X\to\mathbb{R}$ be measurable with respect to $(X,\mathcal{A},\mu)$ such that $\exists a_F\in\mathbb{R}$, $C,c>0$ constants such that*

$$\mu\left(\{|F-a_F|\geq r\}\right)\leq C\exp(-cr^2),\ \forall r>0.$$

Then

$$\begin{aligned}&\int\exp(\rho F^2)d\mu<\infty\ \forall\rho<c,\\&\left|\int Fd\mu-a_F\right|\leq\frac{C}{2}\sqrt{\frac{\pi}{c}},\\&Var_\mu(F)\leq\frac{C}{c}.\end{aligned}$$

Prop. 4.2 is typical of CoM: we don't know anything about the size of F - only about the fluctuations of F about its mean.

We end with the following equivalent characterisation for normal concentration.

Proposition 4.3. *Let $\mu \in \mathcal{M}_1(X, d)$. Then normal concentration holds if and only if $\exists K > 0$ (depending on the constant $C > 0$ in normal concentration statement) such that for any 1-Lipschitz function $F : X \to \mathbb{R}$,*

$$\|F - \int F d\mu\|_q \leq K\sqrt{\frac{q}{c}} \ \forall q \geq 1.$$

The significance of the above result is that if normal concentration holds, then

$$\|F\|_q \leq \|F\|_1 + K\sqrt{q}\|F\|_{Lip} \tag{4.10}$$

If the concentration functions are of type $C\exp(-cr^p)$ for $p > 0$, then the growth rate in $q \geq 1$ is $q^{1/p}$ in Proposition 4.3.

References

1. Michel Ledoux, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs, American Mathematical Society, Providence (R.I.).
2. Stuart, Andrew M. and Sullivan, Tim J., *Lectures on Uncertainty Quantification*, Unpublished notes.

Arnimallee 9, Institut für Mathematik, Freie Universität Berlin, 14195 Berlin
Current address: Arnimallee 9, Institut für Mathematik, Freie Universität Berlin, 14195 Berlin
E-mail address: hlie@math.fu-berlin.de