Notes for Probability Reading Seminar

Probability group of UW-Madison

January 24, 2016

Contents

1 Large deviations for Markov chains

Given by Elnur Emrah in September 2015 Madison

1.1 Large deviation principle

Let $(X_i)_{i\in\mathbb{N}}$ be i.i.d. random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathbf{E}\, X_i = 0$ and $\mathbf{Var} X_i = 1$. Let $S_n = n^{-1}\sum_{i=1}^n X_i$ and μ_n denote the distribution of S_n for $n \in \mathbb{N}$. For example, consider $X_i \sim \mathcal{N}(0,1)$. Then $S_n \sim \mathcal{N}(0, 1/n)$ and

$$\mathbf{P}(|S_n| \geqslant \ell) = \frac{2}{\sqrt{2\pi}} \int_{\ell\sqrt{n}}^{+\infty} e^{-y^2/2} dy = e^{-\frac{\ell^2}{2}n + o(n)}.$$

Similarly for $0 < \ell < \ell'$, we have

$$\mathbf{P}(|S_n| \in [\ell, \ell']) = e^{-\frac{\ell^2}{2}n + o(n)}.$$

It is natural to ask if X is a general random variable with measure μ, what should we put on the r.h.s. ? Motivation: for measure μ, what is the following I?

$$\mathbf{P}(|S_n| \in [\ell, \ell']) = e^{-I(\ell)n + o(n)}. \tag{1}$$

Definition 1.1. **We say that μ_n satisfies an LDP with a *rate function* I** *if $I : \mathbb{R} \to [0, +\infty]$ is lower semicontinuous and, for all Borel sets $B \subset \mathbb{R}$, we have*

$$-\inf_{x\in B^0} I(x) \leqslant \liminf_{n\to+\infty} \frac{\log \mu_n(B)}{n} \quad \textit{(lower bound)}$$
$$-\inf_{x\in \overline{B}} I(x) \geqslant \limsup_{n\to+\infty} \frac{\log \mu_n(B)}{n} \quad \textit{(upper bound).}$$

Here, B^0 and $\overline{B}$ denote the interior and the closure of B. Recall that I is lower-semicontinuous if the sublevel set $\{I \leqslant \alpha\}$ is closed for any $\alpha < +\infty$. This condition is equivalent to $\liminf_{y\to x} I(y) \geqslant I(x)$ for any $x \in \mathbb{R}$.

Remark: it may take a while to understand this form. Here is the equivalent expressions, which is very useful for me

$$e^{-(\inf_{x\in\bar{B}} I(x))\cdot n + o(n)} \geqslant \mu_n(B) \geqslant e^{-\left(\inf_{x\in B^0} I(x)\right)\cdot n + o(n)}$$

Note: you have to use $\bar{B}$ in l.h.s, and use B^0 in l.h.s

The definition of LDP can be given for sequences of measures on arbitrary topological spaces. I will refer to LDP for measures on Euclidean spaces below.

There are many basic properties of LDP in Prof. Varadhan, which some one can introduce to us in the future.

The answer to (1):

$$I(x) = \sup_{\lambda\in\mathbb{R}} \left\{\lambda x - \log \mathbf{E}^{\mu}[e^{\lambda X}]\right\} = \sup_{\lambda\in\mathbb{R}} \left\{\lambda x - \frac{1}{n} \log \mathbf{E}^{\mu_n}[e^{\lambda X}]\right\}.$$

where μ is measure of X_i and μ_n is measure of $\sum_i X_i$.

A useful tool to establish LDP is Gärtner-Ellis theorem. We consider the following setup. Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^d$. Let μ_n denote the distribution of Z_n. Consider the log-moment generating function $\Lambda_n(\lambda) = \log \mathbf{E}[e^{\lambda\cdot Z_n}]$ for $\lambda \in \mathbb{R}^d$. We assume that the following conditions hold:

1. The limit $\Lambda(\lambda) = \lim_{n\to+\infty} n^{-1}\Lambda_n(n\lambda) \in (-\infty, +\infty]$ exists.
2. $0 \in D_\Lambda^0$, where $D_\Lambda = \{\lambda \in \mathbb{R}^d : \Lambda(\lambda) < \infty\}$.
3. Λ is differentiable on D_Λ^0.
4. (Steepness condition) For any $x \in \partial D_\Lambda$, $\lim_{\substack{\lambda\to x \\ \lambda\in D_\Lambda^0}} |\nabla\Lambda(\lambda)| = +\infty$.

Theorem 1.1 (Gärtner-Ellis theorem). *Under assumptions (a)-(d), (μ_n) satisfy an LDP with convex, good (i.e. sublevel sets are compact) rate function Λ^*, the Legendre-Fenchel transform of Λ given by*

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}^d}\{\lambda \cdot x - \Lambda(\lambda)\}.$$

This theorem, in fact, a special case of the Gärtner-Ellis theorem; see, for example, [4], [12] for the full theorem and its proof.

Example: Sum of i.i.d. random variables.

1.2 Application to the Markov chains

We now present an application of this theorem to the Markov chains in discrete time with finite state space. We introduce some notation first. The state space is $[N] = \{1, \ldots, N\}$. Let $\Pi = [\pi(i,j)]_{i,j\in[N]}$ be a stochastic matrix, that is, $\pi(i,j) \geqslant 0$ and $\sum_j \pi(i,j) = 1$ for each $i \in [N]$. Let P_σ^π denote the Markov probability measure with transition matrix Π and initial state at $\sigma \in [N]$. Let Y_n denote the state the chain visits at time n. We have

$$P_\sigma^\pi(Y_1 = y_1, \ldots, Y_n = y_n) = \pi(\sigma, y_1)\pi(y_1, y_2)\ldots\pi(y_{n-1}, y_n)$$

for any path $(y_1, \ldots, y_n)$ in the state space. We assume that Π irreducible; this means that for each (i,j), there exists $m(i,j) \in \mathbb{N}$ such that $\Pi^{m(i,j)}(i,j) > 0$.

Our goal is to obtain an LDP for random variables $Z_n = n^{-1}\sum_{i=1}^n f(Y_i)$, where $f : [N] \to \mathbb{R}^d$ is a given function.

For the computation Λ, the limiting log-moment generating function, we will utilize the following result. For a vector u, we will write $u \gg 0$ if all components of u are positive.

Theorem 1.2 (Perron-Frobenius). *Let $B = [B(i,j)]_{i,j\in[N]}$ be an irreducible matrix with positive entries. Then B has a real eigenvalue ρ (called the Perron-Frobenius eigenvalue) with the following properties.*

(i) *$|\lambda| \leqslant \rho$ for any eigenvalue of B.*

(ii) *There exist a left eigenvector u and a right eigenvector v corresponding to ρ such that $u \gg 0$ and $v \gg 0$.*

(iii) *ρ has multiplicity 1.*

(iv) *For all $i \in [N]$ and $\varphi \gg 0$, we have*

$$\lim_{n\to+\infty}\frac{1}{n}\log\left[\sum_{j=1}^N B^n(i,j)\varphi_j\right] = \lim_{n\to+\infty}\frac{1}{n}\log\left[\sum_{i=1}^N B^n(i,j)\varphi_i\right] = \log\rho.$$

Proof of (iv). Let $c = \min_j \varphi_j / \min_j v_j$, where v is the right eigenvector corresponding to ρ. We have

$$\sum_{j=1}^N B^n(i,j)\varphi_j \geqslant \sum_{j=1}^N B^n(i,j)v_j c = c\rho^n v_i$$

Taking logarithms, dividing through by n and letting $n \to +\infty$ yields

$$\liminf_{n\to+\infty} \frac{1}{n} \log \left[\sum_{j=1}^{N} B^n(i,j)\varphi_j \right] \geqslant \log \rho.$$

We similarly obtain that the limsup is bounded by $\log \rho$. □

Theorem 1.3. *For Markov chain, random variables $Z_n = n^{-1} \sum_{i=1}^{n} f(Y_i)$ satisfy LDP with a rate function $I(x)$ with*

$$I(x) = \sup_{\lambda\in\mathbb{R}^d} \{\lambda \cdot x - \log \rho(\Pi_\lambda)\}$$

where $\Pi_\lambda = [\pi_\lambda(i,j)]_{i,j\in[N]}$ defined by $\pi_\lambda(i,j) = \pi(i,j)e^{\lambda\cdot f(j)}$

Proof: We now turn to LDP for the Markov chain (Y_n). We have

$$\begin{aligned}
\frac{\Lambda_n(n\lambda)}{n} &= \frac{1}{n} \log E_\sigma^\pi \left[\exp\left(\sum_{i=1}^{n} \lambda \cdot f(Y_i) \right) \right] \\
&= \frac{1}{n} \log \left[\sum_{(y_1,\ldots,y_n)\in[N]^n} \exp\left(\sum_i \lambda \cdot f(y_i) \right) \prod_i \pi(y_{i-1}, y_i) \right] \\
&= \frac{1}{n} \log \left[\sum_{(y_1,\ldots,y_n)\in[N]^n} \prod_i \pi(y_{i-1}, y_i) e^{\lambda\cdot f(y_i)} \right],
\end{aligned}$$

where $y_0 = \sigma$. We observe that the matrix $\Pi_\lambda = [\pi_\lambda(i,j)]_{i,j\in[N]}$ defined by $\pi_\lambda(i,j) = \pi(i,j)e^{\lambda\cdot f(j)}$ has positive entries and is irreducible because it is obtained from such a matrix Π by multiplying each entry with a positive number. Hence,

$$\frac{\Lambda_n(n\lambda)}{n} = \frac{1}{n} \log \left[\sum_{y_n=1}^{N} \Pi_\lambda^n(\sigma, y_n) \right] \to \log \rho(\Pi_\lambda)$$

as $n \to +\infty$, by the Perron-Frobenius theorem (applied with $\varphi = (1,\ldots,1)$). Since the Perron-Frobenius eigenvalue is positive, we have $\Lambda(\lambda) = \log \rho(\Pi_\lambda) \in (-\infty, +\infty)$ for all $\lambda \in \mathbb{R}^d$. Hence, (a), (b) hold and (d) is vacuously true. To check differentiability of Λ, we consider the characteristic equation

$$0 = \det[xI - \Pi_\lambda] = x^N + a_{N-1}(\lambda)x^{N-1} + \ldots + a_1(\lambda)X + a_0(\lambda),$$

where coefficients a_i are smooth functions of λ. Let $F(x,\lambda)$ denote the function of $(x,\lambda) \in \mathbb{R}^{d+1}$ on the far right-hand side. We have $F(\Lambda(\lambda),\lambda) = 0$ and, because the Perron-Frobenius eigenvalue has multiplicity 1, $\partial_x F(\Lambda(\lambda),\lambda) \neq 0$. Hence, it follows from the the implicit function theorem that Λ is a smooth function of λ.

Then, the conclusion from the Gärtner-Ellis theorem is that μ_n (the distribution of Z_n) satisfy an LDP with rate function $I(z) = \sup_{\lambda\in\mathbb{R}^d}\{\lambda \cdot z - \log \rho(\Pi_\lambda)\}$.

□

1.3 Key words:

1. Large deviation principle.
2. rate function
3. Gärtner-Ellis theorem
4. Perron-Frobenius theorem
5. Legendre-Fenchel transform
6. LDP of Markov chain

1.4 Exercise:

Exercise 1. Let μ_n be probability measures on $\mathbb{R}$ and $I : \mathbb{R} \to [0, +\infty]$ be a function (not necessarily lower semicontinuous). Define $\tilde{I}(x) = \min\{I(x), \liminf_{y \to x} I(y)\}$ for $x \in \mathbb{R}$.

(a) Show that $\tilde{I}$ is lower semicontinuous. (Hence, the assumption of lower semicontinuity is not restrictive. $\tilde{I}$ is called the lower semicontinuous regularization of I).

(b) Suppose that the lower and the upper bounds above hold for all Borel sets $B \subset \mathbb{R}$. Show that these bounds still hold if I is replaced with $\tilde{I}$, that is,

$$- \inf_{x \in B^0} \tilde{I}(x) \leqslant \liminf_{n \to +\infty} \frac{\log \mu_n(B)}{n}$$
$$- \inf_{x \in \overline{B}} \tilde{I}(x) \geqslant \limsup_{n \to +\infty} \frac{\log \mu_n(B)}{n}.$$

for all Borel sets $B \subset \mathbb{R}$. Moreover, $\tilde{I}$ is the unique lower semicontinuous function with range $[0, +\infty]$ that satisfy these bounds. (Hence, the rate function, if exists, is unique.)

Exercise 2. Let $\mathcal{M}_1([N])$ denote the set of probability measures on the set $[N] = \{1, \ldots, N\}$. We can identify each $\mu \in \mathcal{M}_1([N])$ with the vector $(\mu_j, \ldots, \mu_N)$, where $\mu_j = \mu(\{j\})$ for $j \in [N]$. The relative entropy of $q \in \mathcal{M}_1([N])$ with respect to $\mu \in \mathcal{M}_1([N])$ is defined as

$$H(q|\mu) = \sum_j q_j \log\left(\frac{q_j}{\mu_j}\right),$$

where we interpret $0 \log 0$ and $0 \log(0/0)$ as 0. Suppose that $q_j > 0$ for all $j \in [N]$. Show that

$$H(q|\mu) = \sup_{\substack{u \in \mathcal{M}_1([N]) \\ u_i > 0}} \sum_j q_j \log\left(\frac{u_j}{\mu_j}\right).$$

2 Talagrand's Inequalities

Given by Dae Han Kang in September 2015 Madison

2.1 Concentration inequalities

In this talk, we give powerful concentration inequalities of Talagrand for product probability measures. This talk is primarily based on an article by Nicholas Cook [3].

Theorem 2.1. *Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$ and $\mathbb{P} = \mu_1 \times \cdots \times \mu_n$ a product probability measure on Ω. Then for all nonempty measurable subsets $A \subset \Omega$,*

$$\int_\Omega e^{d_C(x,A)^2/4}\, d\,\mathbb{P}(x) \leqslant 1/\mathbb{P}(A) \tag{2}$$

where $d_C(x, A)$ is the **convex distance** *(TBD) from x to A. As a consequence, by Chebyshev's inequality we have*

$$\mathbb{P}(A_t^c) \leqslant \frac{1}{\mathbb{P}(A)} e^{-t^2/4} \tag{3}$$

where $A_t = \{x \in \Omega : d_C(x, A) \leqslant t\}$.

One of the most useful corollary of Talagrand's inequality is the following concentration inequality for convex Lipschitz functions.

Corollary 2.2. *Let $X = (X_1, X_2, \ldots, X_n)$ be a random variable with independent components taking values in $[c, c + R]$ ($c \in \mathbb{R}, R > 0$). Let $F : \mathbb{R}^n \to \mathbb{R}$ be a convex L-Lipschitz function with respect to ℓ_2 norm. Let $MF(X)$ be a median for $F(X)$. Then for all $t \geqslant 0$*

$$\mathbb{P}(|F(X) - MF(X)| \geqslant t) \leqslant 4e^{-t^2/(4R^2L^2)}. \tag{4}$$

Note: Convex is very important here.

2.2 Idea of proof

We must define the convex distance. To motivate the definition and to understand Talagrand's inequality, first we consider the classical bounded difference inequality. A function $f : \mathcal{X}^n \to \mathbb{R}$ has the *bounded differences property* if for some nonnegative constants $c_1, \ldots, c_n$,

$$\sup_{\substack{x_1,\ldots,x_n \\ x_i' \in \mathcal{X}}} |f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leqslant c_i,\ 1 \leqslant i \leqslant n.$$

Theorem 2.3 (Bounded differences inequality)**.** *Assume that the function f satisfies the bounded differences assumption with constants $c_1, \ldots, c_n$ and denote $\nu = \frac{1}{4}\sum_{i=1}^n c_i^2$. Let $X = (X_1, \ldots, X_n)$ where X_i are independent random variables taking values in $\mathcal{X}$. Then*

$$P(f(X) - Ef(X) > t) \leqslant e^{-t^2/(2\nu)}.$$

For the proof, see Theorem 6.2 in [2]. For any $\alpha \in \mathbb{R}_+^n$, the **weighted Hamming distance** $d_\alpha(x, y)$ between the vectors $x, y \in \Omega_1 \times \cdots \times \Omega_n$ is defined as

$$d_\alpha(x, y) = \sum_{i=1}^n \alpha_i 1_{\{x_i \neq y_i\}}.$$

With this definition Theorem 2.3 implies that if $f : \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$ is Lipschitz with respect to d_α, then

$$P(f(X) - Ef(X) \geqslant t) \leqslant e^{-2t^2/\|\alpha\|^2}$$

where $\|\alpha\|$ is the euclidean norm of α.

On the other hand, if instead of 'bounded differences property', if we apply Theorem 2.3 for a convex Lipschitz function with respect to ℓ_2 norm as in Corollary 2.2, we have

$$P(f(X) - Ef(X) > t) \leqslant e^{-2t^2/(nR^2L^2)}.$$

The power of Talagrand's inequality is the dimension free subgaussian inequality.

Motivation for the definition of the convex distance

Now we give a motivation for the definition of the convex distance. Since the function $f(x) = d_\alpha(x, A)$ is Lipschitz with respect to d_α, by the bounded differences inequality

$$\mathbb{P}(\mathbb{E}\, d_\alpha(X, A) - d_\alpha(X, A) \geqslant t) \leqslant e^{-2t^2/\|\alpha\|^2}.$$

However, by taking $t = \mathbb{E}\, d_\alpha(X, A)$, the left-hand side becomes $\mathbb{P}(d_\alpha(X, A) \leqslant 0) = \mathbb{P}(A)$, so the above inequality implies

$$\mathbb{E}\, d_\alpha(X, A) \leqslant \sqrt{\frac{\|\alpha\|^2}{2} \log \frac{1}{P(A)}}.$$

Then, by using the bounded differences inequality again, we obtain

$$\mathbb{P}(d_\alpha(X, A) \geqslant t + \sqrt{\frac{\|\alpha\|^2}{2} \log \frac{1}{\mathbb{P}(A)}}) \leqslant e^{-2t^2/\|\alpha\|^2}.$$

Thus, for example, for all vectors α with unit norm $\|\alpha\| = 1$,

$$P(d_\alpha(X, A) \geqslant t + \sqrt{\frac{1}{2} \log \frac{1}{P(A)}}) \leqslant e^{-2t^2}.$$

Thus, denoting $u = \sqrt{\frac{1}{2} \log \frac{1}{P(A)}}$, for any $t \geqslant u$,

$$P(d_\alpha(X, A) \geqslant t) \leqslant e^{-2(t-u)^2}.$$

On the other hand, if $t \leqslant \sqrt{-2 \log P(A)}$, then $P(A) \leqslant e^{-t^2/2}$. On the other hand, since $(t-u)^2 \geqslant t^2/4$ for $t \geqslant 2u$, for any $t \geqslant \sqrt{2 \log \frac{1}{P(A)}}$ the inequality above implies $P(d_\alpha(X, A) \geqslant t) \leqslant e^{-t^2/2}$. Thus, for all $t > 0$, we have

$$\sup_{\alpha : \|\alpha\| = 1} P(A) \cdot P(d_\alpha(X, A) \geqslant t) \leqslant \sup_{\alpha : \|\alpha\| = 1} \min(P(A), P(d_\alpha(X, A) \geqslant t)) \leqslant e^{-t^2/2}.$$

The main message of Talagrand's inequality is that the above inequality remains true even if the supremum is taken within the probability. (See (3).)

Convex distance

The convex distance of x from the set A is defined by

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

There is an equivalent definition of the convex distance that is used for the proof. Let Ω as in the theorem, and let $A \subseteq \Omega$, $x \in \Omega$. We define $U_A(x) \subseteq \{0, 1\}^n$

$$U_A(x) = \{s \in \{0, 1\}^n : \exists y \in A \text{ with } y_i = x_i \text{ whenever } s_i = 0\}.$$

Now let $V_A(x) \subseteq \mathbb{R}^n$ be the convex hull of $U_A(x)$ in $\mathbb{R}^n$. If Ω is a vector space we say that a vector $s = (s_1, \ldots, s_n)$ in the binary cube *supports* a vector $z \in \Omega$ if $z_i \neq 0$ only when $s_i = 1$, and define

$U_A(x)$ to be the set of vectors in the binary cube that support some element of $A - x$. We claim that $d_C(x, A) = d_E(0, V_A(x))$ where d_E is the euclidean distance in $\mathbb{R}^n$ (Exercise).

Proof of the Corollary 2.2: We only consider the case $R = 1$, $c = 0$, and $L = 1$. Key in the passage from the theorem to the corollary is the observation that for the special case of A convex in $[0, 1]^n$, the convex distance controls the Euclidean distance.

Lemma 2.4. *Let A convex in $[0, 1]^n$ and $x \in [0, 1]^n$. Then $d_E(x, A) \leqslant d_C(x, A)$.*

Proof. Suppose $d_C(x, A) \leqslant t$. Then by the equivalent definition of convex distance, there exists a convex combination $w = \sum_{i=1}^m \lambda_i \vec{s_i}$ of vectors $\vec{s_i} \in U_A(x)$ $1 \leqslant i \leqslant m$ such that $\|w\| \leqslant t$. Now for each i, $\vec{s_i} \in U_A(x)$ means there exists $\vec{z_i} \in A - x$ supported by $\vec{s_i}$. Let $z = \sum_{i=1}^m \lambda_i \vec{z_i}$. Then $z \in A - x$ by convexity. Note that $\|z\| \leqslant \|w\|$ (Exercise). Thus $d_E(x, A) \leqslant \|z\| \leqslant t$, and the claim follows.

□

Now we return to prove corollary 2.2. Note that the lemma and the theorem imply

$$\mathbb{E}\, e^{d_E(X,A)^2/4} \leqslant \frac{1}{\mathbb{P}(X \in A)}$$

for any convex subset A of $[0, 1]^n$. Let $a \geqslant 0$ and take $A = \{F \leqslant a\}$. By the Lipschitz property, if $X \in \{F \geqslant a + t\}$ for some $t \geqslant 0$, then $d_E(X, A) \geqslant t$. Then by applying Chebyshev's inequality to the LHS, we have

$$\mathbb{P}(F(X) \geqslant a + t)e^{t^2/4} \leqslant 1/\mathbb{P}(F(X) \leqslant a).$$

Now taking $a = MF(X)$ we get the upper tail estimate

$$\mathbb{P}(F(X) - MF(X) \geqslant t) \leqslant 2e^{-t^2/4}$$

and taking $a = MF(X) - t$ we get the lower tail estimate

$$\mathbb{P}(F(X) - MF(X) \leqslant -t) \leqslant 2e^{-t^2/4}$$

where the definition of median has given us the prefactors 2.

□

Proof of the main Theorem 2.1: We only sketch the proof. For the details see [3]. One can prove the theorem by induction on n. For the case $n = 1$, we must show

$$e^{1/4}(1 - \mathbb{P}(A)) + \mathbb{P}(A) \leqslant 1/\mathbb{P}(A)$$

which follows easily from $e^{1/4}(1 - u) + u \leqslant 1/u$ for all $u \in [0, 1]$. For the inductive step we need a lemma.

Lemma 2.5. *For all u in $(0, 1]$ we have*

$$\inf_{\lambda \in [0,1]} e^{(1-\lambda)^2/4} u^{-\lambda} \leqslant 2 - u.$$

Assume the result holds for n. Let $\Omega' = \Omega_1 \times \cdots \times \Omega_n$ a product space with product measure P, and let Ω_{n+1} be another probability space with measure μ_{n+1}. Let $\Omega = \Omega' \times \Omega_{n+1}$. Let $A \subseteq \Omega$ and $x \in \Omega$. The proof of the result for $n + 1$ follows these steps:

1. Obtain an inequality for $d_C(x, A)$ from consideration of "slices" and the "projections" of A in Ω', and convexity. For a point $z \in \Omega$ we write $z = (z', w)$, $z' \in \Omega'$, $w \in \Omega_{n+1}$. Let $A(w) = \{z' \in \Omega' : (z', w) \in A\}$ be the w-slice of A, and $B = \cup_{w \in \Omega_{n+1}} A(w)$ be the projection of A to Ω'. The key observation that gets the proof is that we can bound the convex distance $d_C(x, A)$ in terms of the distances to the sections $A(w)$ and the projection B.

2. Apply Hölder's inequality and the induction hypothesis.

3. Optimize using the lemma.

4. Use Fubini.

□

2.3 Applications

Example 1. The largest eigenvalue
Let M be an $n \times n$ Hermitian matrix Then the largest eigenvalue $\lambda_1(M) = \|M\|_{op}$. Considering the operator nomr of M as a function of the n^2 components of the entries, we see that it is a convex and 1-Lipschitz function from $\mathbb{R}^{n^2}$ with euclidean distance to $\mathbb{R}_+$ (Exercise). Hence, if X is a random Hermitian matrix-where the diagonal entries and the real and imaginary parts of the strict upper-triangle entries are independent bounded scalar random variables, and we identify the space of Hermitian matrices with $\mathbb{R}^{n^2}$, then by Talagrand's inequality we have that the random variable $\lambda_1(X)$ is concentrated around its mean with sub-Gaussian tails independent of n.

Example 2. The longest increasing subsequence
Let $X = (x_1, \dots, x_n)$ be uniformly distributed in $\Omega = [0,1]^n$, $J(x)$ be the longest increasing subsequence of $(x_1, \dots, x_n)$, and let $F_n(x) = |J(x)|$ be its length. We will show that F_n concentrates tightly around its median MF_n.

Note that we cannot apply Corollary 2.2 as F_n is not convex. For example, with $n = 3$, taking $x = (0, 1, 0.6)$ and $y = (0.8, 0.0, 6)$ we have that $F_3(x) = 2 = F_3(y)$, but $F_3(\frac{x+y}{2}) = F_3((0.4, 0.5, 0.6)) = 3$. However, F_n is 1-Lipschitz with respect to the Hamming metric. While the full convex distance is not so easy to apply directly as euclidean distance, we will see that a weight function suggests itself. Let $a > 0$ and $A = \{F(y) \leqslant a\}$. For any x, $y \in \Omega$,

$$F_n(y) \geqslant F_n(x) - \sum_{i=1}^{n} 1_{\{x_i \in J(x), x_i \neq y_i\}}.$$

If we let $\alpha(x) = \frac{1}{\sqrt{|J(x)|}} 1_{J(x)} = \frac{1}{\sqrt{F_n(x)}} 1_{J(x)}$ we have

$$d_{\alpha(x)(x,y)} = \frac{1}{\sqrt{F_n(x)}} \sum_{i=1}^{n} 1_{\{x_i \in J(x)\}} 1_{\{x_i \neq y_i\}} \geqslant \frac{1}{\sqrt{F_n(x)}} (F_n(x) - F_n(y)).$$

For the convex distance from x to A we have

$$d_C(x, A) \geqslant \frac{F_n(x) - a}{\sqrt{F_n(x)}}.$$

which is the key step of applying Talagrand inequality in this example. For $t \geqslant a$ the function $g(t) = (t-a)/\sqrt{t}$ is monotone increasing. From this and Theorem 2.1 it follows that

$$\begin{aligned} P(F_n(x) \geqslant a + t) &\leqslant P(\frac{F_n(x) - a}{\sqrt{F_n(x)}} \geqslant \frac{t}{a+t}) \\ &\leqslant P(d_C(x, A) \geqslant \frac{t}{a+t}) \leqslant \frac{1}{P(A)} e^{-t^2/4(a+t)}. \end{aligned}$$

Taking $a = M_n := MF_n(x)$ we get the upper tail estimate

$$P(F_n(x) \geqslant M_n + t) \leqslant 2 \exp(-\frac{t^2}{4(M_n + t)})$$

and taking $a = M_n - t$ we get the lower tail estimate

$$P(F_n(x) \leqslant M_n - t) \leqslant 2 \exp(-\frac{t^2}{4M_n}).$$

It can be shown that $M_n = O(\sqrt{n})$, so the above concentration estimates are enough to prove $(F_n(x) - M_n)/\sqrt{n} \to 0$ a.s..

2.4 Key words:

1. Concentration inequality
2. Talagrand's inequality
3. Convex distance
4. Convex and Lipschitz function

2.5 Exercise:

1. Prove that $d_C(x, A) = d_E(0, V_A(x))$ as claimed.
2. Complete the proof of Lemma 2.4.
3. Prove that $\lambda_1(M)$ in Example 1 is a convex and 1-Lipschitz function on $\mathbb{R}^{n^2}$.

3 Concentration of Measure and Concentration Inequalities

Talks given by Jessica Lin October 2015 in Madison, WI

The purpose of these lectures is to discuss some examples of classical concentration inequalities used in probability theory, as well as to clarify the connection between probabilistic concentration inequalities and the analytic subject of *concentration of measure*.

The typical setting of concentration inequalities is to consider $X_1, X_2, \ldots, X_n$ independent random variables, and $f : \mathbb{R}^n \to \mathbb{R}$ a measurable function. We let $Z = f(X_1, X_2, \ldots, X_n)$ and we aim to identify a function $g : \mathbb{R} \to \mathbb{R}$, with $\lim_{\lambda \to 0} g(\lambda) = 0$ such that

$$\mathbb{P}\left[|Z - EZ| \geqslant \lambda\right] \leqslant g(\lambda).$$

The main things to consider will be:

- What hypotheses do we need to assume about f?
- What type of function is g?

.

As an introductory example, the classical Chebyshev inequality yields that for f measurable,

$$\mathbb{P}\left[|Z - EZ| \geqslant \lambda\right] \leqslant \mathbb{P}\left[(Z - EZ)^2 \geqslant \lambda^2\right] \leqslant \frac{\mathbf{Var}[Z]}{\lambda^2}.$$

This is an example of one of the most elementary concentration inequalities, and *so long as we can get control on the* $\mathbf{Var}[Z]$, then we have some type of concentration phenomena. The goal of these lectures is to see

- How can we control $\mathbf{Var}\, Z$?
- Can we get control of higher moments?

3.1 Concentration of Measure, Poincare Inequalities and the Efron-Stein Inequality

We begin by describing an analogous formulation of concentration inequalities, which is the subject of concentration of measure.

Definition 3.1. *Let (X, μ, d) denote a metric probability space. We say that μ satisfies concentration of measure with concentration rate $g : \mathbb{R} \to \mathbb{R}$ such that for all $A \subseteq X$ with $\mu(A) > \frac{1}{2}$, we have that for*

$$\mu(A_r^c) \leqslant g(r), \quad A_r := \{x \in X : d(x, A) \leqslant r\},$$

This looks very similar to one of the formulations of the Talagrand inequality we saw:

$$\mu(A)\mu(A_r^c) \leqslant e^{-cr^2}$$

By a simple rearrangement, this implies that

$$\mu(A_r) \geqslant 1 - \frac{1}{\mu(A)} e^{-cr^2} \geqslant 1 - Ce^{-cr^2},$$

where we used that $\mu(A) \geqslant \frac{1}{2}$.

We now state what is meant by a Poincare inequality, which we will see plays an important role in obtaining concentration phenomena:

Definition 3.2. *Consider a metric probability space (X, μ, d). The measure μ is said to satisfy a Poincare inequality with constant C if*

$$\mathbf{Var}[f] \leqslant C \int_X |\nabla f|^2 \, d\mu \quad \text{for all} \quad f : X \to \mathbb{R}\,. \tag{5}$$

The notion of ∇f is interpreted in the distributional sense, and thus the inequality makes sense for all functions $f \in H^1$.

An interpretation of a Poincare inequality is to say that given a value for variance, there is a smoothest function which has that variance (since any other function will have a *larger* L^2-norm of the gradient). We next show that indeed, having a Poincare inequality is a sufficient condition to have concentration of measure:

Theorem 3.1 ([1], Theorem 2, p. 15)**.** *Suppose (μ, X, d) satisfies a Poincare inequality, and μ is absolutely continuous with respect to the volume element. If $\mu(A) \geqslant \frac{1}{2}$, then for all $r > 0$,*

$$\mu(A_r^c) \leqslant e^{-\frac{r}{3\sqrt{c}}}.$$

Proof. Let A, B denote two subsets of X such that $d(A, B) = \varepsilon$, for ε to be chosen. (We should think of $B = A_\varepsilon^c$). Let $a := \mu(A)$, and $b = \mu(B)$. We then define

$$f(x) = \begin{cases} \frac{1}{a} & x \in A \\ \frac{1}{a} - \frac{1}{\varepsilon}\left(\frac{1}{a} + \frac{1}{b}\right) \min\{\varepsilon, d(x, A)\} & x \in X \backslash (A \cup B) \\ -\frac{1}{b} & x \in B \end{cases}$$

Note that f belongs to H^1. Thus, we are able to apply the Poincare inequality. We have that since f is constant on $A \cup B$,

$$\nabla f(x) = 0 \quad \text{for } x \in A \cup B.$$

Otherwise, we have that μ-almost surely,

$$|\nabla f(x)| \leqslant \frac{1}{\varepsilon}\left(\frac{1}{a} + \frac{1}{b}\right).$$

Therefore, we have that

$$\int |\nabla f(x)|^2 \, d\mu \leqslant \frac{1}{\varepsilon^2}\left(\frac{1}{a} + \frac{1}{b}\right)^2 (1 - a - b).$$

Moreover, we consider that if $\bar{f} = \int f \, d\mu$, then

$$\begin{aligned} \mathbf{Var}[f] &= \int (f - \bar{f})^2 \, d\mu \\ &\geqslant \int_A (f - \bar{f})^2 \, d\mu + \int_B (f - \bar{f})^2 \, d\mu \\ &\geqslant a\left(\frac{1}{a} - \bar{f}\right)^2 + b\left(-\frac{1}{b} - \bar{f}\right)^2. \end{aligned}$$

The right hand side is minimized when $\bar{f} = 0$, which implies in particular that

$$\mathbf{Var}[f] \geqslant \frac{1}{a} + \frac{1}{b}.$$

By the Poincare inequality, we have then that

$$\frac{1}{a} + \frac{1}{b} \leqslant \frac{C}{\varepsilon^2}\left(\frac{1}{a} + \frac{1}{b}\right)^2 (1 - a - b).$$

Rearranging this inequality, we have that

$$b \leqslant \frac{1-a}{1+\varepsilon^2/(2C)}$$

Now, we let $B = A_\varepsilon^c$ and $\frac{\varepsilon^2}{2C} = 1$, or $\varepsilon = \sqrt{2C}$. Recall $a := \mu(A)$, and $b = \mu(B)$. Thus,

$$\mu(A_\varepsilon^c) \leqslant \mu(A^c)/2.$$

Notice that

$$(A_\varepsilon)_\varepsilon = \{x : d(x, A_\varepsilon) \leqslant \varepsilon\} \subseteq \{x : d(x, A) \leqslant 2\varepsilon\}$$

which implies that

$$\mu(A_{2\varepsilon}^c) \leqslant \mu((A_\varepsilon)_\varepsilon^c).$$

Therefore, by iterating, this implies that

$$1 - \mu(A_{k\varepsilon}) \leqslant 2^{-k-1}.$$

Thus, for any $r > 0$, let k such that $k\varepsilon \leqslant r < (k+1)\varepsilon$. Since $\mu(A_r^c)$ is monotonically non-increasing in r, we have that by the choice of ε,

$$1 - \mu(A_r) \leqslant 2^{-k-1} \leqslant \exp\left(-\frac{\log 2}{\sqrt{2C}} r\right)$$

Since $\frac{\log 2}{\sqrt{2}} > \frac{1}{3}$, we get that

$$1 - \mu(A_r) \leqslant \exp\left(-\frac{r}{3\sqrt{C}}\right).$$

□

Thus, we see that the property of having a Poincare inequality leads to concentration of measure.

In the context of probability theory, we now would like to ask:

1. What does a Poincare inequality look like in the setting of random variables?
2. When does a Poincare inequality hold?

The main statement which encodes these two results is the *Efron-Stein* inequality:

Theorem 3.2 (Efron-Stein Inequality). *Let f be measurable, and $\{X_i\}_{i=1}^n$ independent random variables. Let $Z := f(X_1, X_2, \ldots, X_n)$. Then*

$$\mathbf{Var}[Z] \leqslant \sum_{i=1}^n \mathbb{E}[(Z - \mathbb{E}_i Z)^2] \tag{6}$$

where

$$\mathbb{E}_i Z := \mathbb{E}[Z \mid X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots X_n].$$

Equivalently, if we let $\left\{\tilde{X}_i\right\}_{i=1}^n$ denote an independent copy of random variables, then by defining

$$Z_i = f(X_1, X_2, \ldots, X_{i-1}, \tilde{X}_i, X_{i+1}, \ldots, X_n),$$

we have that

$$\mathbf{Var}[Z] \leqslant \frac{1}{2} \sum_{i=1}^n \mathbb{E}[(Z - Z_i)^2]. \tag{7}$$

This theorem tells us that so long as we are studying a measurable function of independent random variables, this is enough to guarantee a Poincare inequality. Instead of proving the statement, we will discuss why (6) is equivalent to (7): Consider that, in general, if X, Y are independent identically distributed random variables, then

$$\begin{aligned}\mathbb{E}[(X-Y)^2] &= \mathbb{E}[X^2 - 2XY + Y^2] \\ &= 2\,\mathbb{E}[X^2] - 2(\mathbb{E}[X])^2\end{aligned}$$

which implies that

$$\mathbf{Var}[X] = \frac{1}{2}\,\mathbb{E}[(X-Y)^2].$$

Thus, by definition of Z_i, Z_i is iid with Z. This implies that

$$\mathbb{E}_i\left[(Z - \mathbb{E}_i Z)^2\right] = \frac{1}{2}\,\mathbb{E}_i\left[(Z-Z_i)^2\right].$$

Taking expectation of both sides, we have that

$$\mathbb{E}[(Z - \mathbb{E}_i Z)^2] = \frac{1}{2}\,\mathbb{E}\left[(Z-Z_i)^2\right],$$

so by (6),

$$\mathbf{Var}[Z] \leqslant \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[(Z-Z_i)^2\right],$$

which is (7).

Notice that (7) truly looks like a Poincare inequality since

$$Z - Z_i \sim \nabla_{X_i} Z$$

since it measures the changes in Z with respect to changes in X_i.

Remark 3.3. *We point out that equality in* (7) *is achieved in the case when* $Z = \sum_{i=1}^n X_i$*. This implies that sums of random variables are the* ***least*** *concentrated of all measurable functions of* $(X_1, X_2, \ldots, X_n)$*.*

Remark 3.4. *The primary purpose of the Efron-Stein inequality is to provide a way of computing* $\mathbf{Var}[Z]$*. In most applications, we will couple the Efron-Stein inequality with Chebyshev's inequality. However, we can take this further by not just considering* f *which is given, but we can apply Efron-Stein to any monotone function* $h(f)$ *to compute the variance. This is why one is able to obtain something with* exponential *decay, but not* Gaussian*. For reference, we also state the "concentration version" with exponential bounds, which we refer to as the Gromov-Milman Theorem:*

Theorem 3.5 (Gromov-Milman, [10], p.34)**.** *Let* (X, μ, d) *denote a metric probability space, and suppose it satisfies a Poincare inequality with constant* C*. Let* f *be a 1-Lipschitz function. Then for every* $t > 0$*,*

$$\mathbb{P}[|f - \mathbb{E}\,f| > t] \leqslant 240 e^{-\sqrt{\frac{2}{C}}t}.$$

Next, we provide some examples of interesting applications of the Efron-Stein inequality in the context of probability:

Example: Bounded Differences. The Efron-Stein will prove to be particularly useful in the context of random variables which have *bounded differences*. Let f be measurable, and $\{X_i\}_{i=1}^n$ independent random variables. Let $Z := f(X_1, X_2, \ldots, X_n)$, and suppose f has the property that for each i, there exists c_i so that

$$\Big| f(X_1, X_2, \ldots, X_i, \ldots, X_n) - f((X_1, X_2, \ldots, \tilde{X}_i, \ldots, X_n) \Big| \leqslant c_i.$$

Then by the Efron-Stein inequality, we have

$$\mathbf{Var}[Z] \leqslant \frac{1}{2}\sum_{i=1}^{n} c_i^2.$$

Checking the property of bounded differences is usually quite straightforward, and thus the Efron-Stein inequality an easy way of obtaining concentration from this. We next show an application to bin-packing:

Suppose $\{X_1, X_2, \ldots, X_n\} \subseteq [0,1]$ are independent random variables. What is the minimal number of bins into which $\{X_i\}$ can be packed such that the sum of X_i in each bin does not exceed 1?

Let

$$Z = f(X_1, \ldots, X_n) = \text{minimum number of bins to pack satisfying the rule above.}$$

Notice that if we adjust any X_i,

$$|f(X_1, \ldots, X_i, \ldots) - f(X_1, \ldots, \tilde{X}_i, \ldots)| \leqslant 1.$$

Thus, this choice of f satisfies bounded differences, which implies by the Efron-Stein inequality that

$$\mathbf{Var}[Z] \leqslant \frac{n}{2}.$$

Therefore, with Chebyshev's inequality, we have that

$$\mathbb{P}\left[|Z - \mathbb{E}\, Z| \geqslant n^{\frac{1+\varepsilon}{2}}\right] \leqslant Cn^{-\varepsilon}.$$

3.2 Logarithmic-Sobolev Inequalities

We next consider the following. The Efron-Stein gives us excellent control over the concentration, with estimates that are even exponential in nature. However, how can we obtain *Gaussian* bounds? It turns out to obtain Gaussian bounds, we need something which is **stronger** than a Poincare inequality. The right tool we need is a Log-Sobolev inequality (LSI):

Definition 3.3. *Let (X, μ, d) denote a metric probability space. μ satisfies a LSI with constant C if*

$$\int f^2 \log f^2 \, d\mu - \int f^2 \log\left(\int f^2 d\mu\right) d\mu \leqslant 2C \int |\nabla f|^2 \, d\mu$$
$$\mathbb{E}[f^2 \log f^2] - \mathbb{E}[f^2 \log \mathbb{E}[f^2]] \leqslant 2C\, \mathbb{E}[|\nabla f|^2]$$

For specialists in the field, the left-hand side can be identified as the entropy of the function f. This property is referred to as a Log-Sobolev inequality in light of its connection with classical Sobolev inequalities:

Consider that the classical Sobolev inequality states that $W^{1,p}(\mathbb{R}^d) \subseteq L^{p*}(\mathbb{R}^d)$ for

$$\frac{1}{p*} = \frac{1}{p} - \frac{1}{d}.$$

Thus, we see integrability of $|\nabla f|$ implies higher integrability of the function itself, depending on the dimension. The LSI can be seen as a similar estimate *which does not depend on dimension*! This is why it is referred to as the LSI, since it gives us similar improvement of integrability with logarithmic weights.

Next, we show that a Log-Sobolev inequality is indeed stronger than a Poincare inequality:

Proposition 3.6 ([1], Proposition 3, p.28)**.** *Let (X, μ, d) denote a metric probability space. If μ satisfies LSI with constant C, then μ satisfies a Poincare inequality with constant C.*

Proof. The proof follows by a clever, yet standard trick in Taylor expansion. We study the Taylor expansion of the LSI applied to $1 + \varepsilon f$, where f is any bounded function with 0 mean.

On the right-hand side of LSI, we have that

$$\int |\nabla(1 + \varepsilon f)|^2 \, d\mu = \varepsilon^2 \int |\nabla f|^2 \, d\mu.$$

Next, we investigate the asymptotics of the left-hand side. We consider that

$$E\left[(1 + \varepsilon f)^2 \log((1 + \varepsilon f)^2)\right] = 2E\left[(1 + \varepsilon f)^2 \log(1 + \varepsilon f)\right].$$

Next, we recall that according to the Taylor expansion of $\log(1 + x)$,

$$\log(1 + \varepsilon f) = \varepsilon f - \frac{\varepsilon^2 f^2}{2} + o(\varepsilon^2).$$

Thus,

$$\begin{aligned}
\mathbb{E}\left[(1 + \varepsilon f)^2 \log((1 + \varepsilon f)^2)\right] &= 2\,\mathbb{E}\left[(1 + \varepsilon f)^2 \left(\varepsilon f - \frac{\varepsilon^2 f^2}{2}\right)\right] + o(\varepsilon^2) \\
&= 2\varepsilon\,\mathbb{E}[f] + 4\varepsilon^2\,\mathbb{E}[f^2] - \varepsilon^2\,\mathbb{E}[f^2] + o(\varepsilon^2) \\
&= 3\varepsilon^2\,\mathbb{E}[f^2] + o(\varepsilon^2)
\end{aligned}$$

where in the last line, we used that f has mean 0. Next, we check the second term,

$$\begin{aligned}
\mathbb{E}[(1 + \varepsilon f)^2] \log \mathbb{E}[(1 + \varepsilon f)^2] &= (1 + \varepsilon^2\,\mathbb{E}[f^2] \log(1 + \varepsilon^2 E[f^2]) \\
&= (1 + \varepsilon^2\,\mathbb{E}[f^2])(\varepsilon^2\,\mathbb{E}[f^2]) + o(\varepsilon^2) \\
&= \varepsilon^2\,\mathbb{E}[f^2] + o(\varepsilon^2).
\end{aligned}$$

Therefore, combining these, we have that according to the LSI,

$$2\varepsilon^2\,\mathbb{E}[f^2] + o(\varepsilon^2) \leqslant 2C\varepsilon^2 \int |\nabla f|^2 \, d\mu$$

which implies that

$$\mathbf{Var}[f] \leqslant C \int |\nabla f|^2 \, d\mu,$$

which is indeed the Poincare inequality. Moreover, the inequality is unchanged by adding constants to f, and for any smooth f with compact support. By density, this implies that it holds for any $f \in H^1(\mathbb{R}^d)$ as desired.

□

Finally, we show that (X, μ, d) satisfying LSI implies that we have Gaussian concentration bounds:

Theorem 3.7 (Herbst)**.** *[[10], p. 35, Theorem 21] Let (X, μ, d) denote a metric probability space, satisfying a Log-Sobolev inequality. Then for every $f : X \to \mathbb{R}$ Lipschitz with constant 1, for every $\lambda \in \mathbb{R}$,*

$$\mathbb{E}[e^{\lambda(f - \mathbb{E}[f])}] \leqslant e^{C\lambda^2/4}$$

and

$$\mathbb{P}[|f - \mathbb{E}\, f| > t] \leqslant 2e^{-t^2/C}$$

Proof. Let $\lambda > 0$. The case $\lambda < 0$ can be proved similarly. We apply the LSI to the function $e^{\lambda f/2}$. This implies that

$$\begin{aligned}\int e^{\lambda f}\lambda f\, d\mu - \int e^{\lambda f}\log\left(\int e^{\lambda f} d\mu\right) d\mu &= \lambda f E[e^{\lambda f}] - E[e^{\lambda f}\log(E[e^{\lambda f}]) \\ &\leqslant C\int \left|\nabla e^{\lambda f/2}\right|^2 d\mu \\ &\leqslant \frac{C}{4}\int \lambda^2 |\nabla f|^2 e^{\lambda f}\, d\mu \\ &\leqslant \frac{C}{4}\lambda^2 E[e^{\lambda f}]\end{aligned}$$

using that f is 1-Lipschitz.

Next, we define $h(\lambda) := E[e^{\lambda f}]$, so that $h'(\lambda) = E[fe^{\lambda f}]$. We may rewrite the above inequality as

$$\lambda h'(\lambda) - h(\lambda)\log h(\lambda) \leqslant \frac{C}{4}\lambda^2 h(\lambda) \tag{8}$$

so that

$$\left(\frac{1}{\lambda}\log h(\lambda)\right)' \leqslant \frac{C}{4}.$$

Also, we have that

$$\lim_{\lambda\to 0}\frac{\log(h(\lambda))}{\lambda} = \lim_{\lambda\to 0} h'(\lambda)h(\lambda) = \lim_{\lambda\to 0}\frac{E[e^{\lambda f}]}{E[e^{\lambda f}]} = E[f]$$

Combining these two pieces of information, we have

$$\frac{1}{\lambda}\log h(\lambda) \leqslant E[f] + \frac{C}{4}\lambda$$

Thus,

$$E[e^{\lambda f}] = h(\lambda) \leqslant e^{\lambda E f} e^{\frac{C}{4}\lambda^2}$$

and this implies that

$$E[e^{\lambda(f-Ef)}] \leqslant e^{\frac{C}{4}\lambda^2}$$

Notice that the same argument holds for $\lambda < 0$, since we multiply/divide by λ twice.

Finally, by Markhov's inequality, we have that

$$\mathbb{P}[|f-Ef| \geqslant \lambda] \leqslant \mathbb{P}[f-Ef \geqslant \lambda] + \mathbb{P}[f-Ef \geqslant -\lambda] \leqslant 2\frac{E[e^{\lambda(f-Ef)}]}{e^{\lambda^2}} \leqslant 2e^{-C\lambda^2}.$$

□

Next, as before, we state the probabilitistic version of the LSI:

Proposition 3.8. *Let $Z = f(X_1, X_2, \ldots, X_n)$, where $\{X_i\}_{i=1}^n$ are independent, and f is measurable. Then we have for every $s \in \mathbb{R}$,*

$$s\,\mathbb{E}\left[Ze^{sZ}\right] - \mathbb{E}[e^{sZ}]\log\mathbb{E}[e^{sZ}] \leqslant \sum_{i=1}^n \mathbb{E}[e^{sZ}\psi(-s(Z-Z_i'))] \tag{9}$$

where $\psi(x) = e^x - x - 1$ and $Z_i' := f(X_1, \ldots, X_i', \ldots X_n)$.

Equivalently, we have that

$$s\,\mathbb{E}[Ze^{sZ}] - \mathbb{E}[e^{sZ}]\log\mathbb{E}[e^{sZ}] \leqslant s^2\,\mathbb{E}\left[\sum_{i=1}^n (Z-Z_i')^2 e^{sZ}\,\mathbf{1}_{\{Z>Z_i'\}}\right] \tag{10}$$

Next we discuss some applications and examples:

Bounded Differences. Again, we study the situation where we have bounded differences:

$$|Z - Z'_i| \leqslant c_i.$$

By (10), we have

$$s\,\mathbb{E}[Ze^{sZ}] - \mathbb{E}[e^{sZ}]\log \mathbb{E}[e^{sZ}] \leqslant s^2 \sum_{i=1}^{n} c_i^2\,\mathbb{E}[e^{sZ}].$$

Notice it has the same form of (8). Then with similar proof as the one for the Herbst Theorem, we have

$$\mathbb{P}[|Z - \mathbb{E}\,Z| > t] \leqslant e^{-t^2/\sum_{i=1}^n c_i^2} \tag{11}$$

Finally, we provide an application to Random Matrices, found in [7] . Let A_{ij} denote a symmetric, real matrix with entries $X_{i,j}$ for $1 \leqslant i \leqslant j \leqslant n$, which are independent random variables, and suppose that $|X_{i,j}| \leqslant 1$.

Let $Z = \lambda_1$ denote the largest eigenvalue of A_{ij}. We check to see if Z satisfies the bounded differences property. We have that

$$\lambda_1 = v^T Av = \sup_{\|u\|=1} u^T Au.$$

Let $A'_{i,j}$ denote the matrix $A_{i,j}$ but replacing $X_{i,j}$ by an independent copy, called $X'_{i,j}$. Then we have that

$$\begin{aligned} |Z - Z'_{i,j}| &\leqslant |v^T(A_{ij} - A'_{ij})v| \\ &\leqslant (v_i v_j (X_{i,j} - X'_{i,j})) \\ &\leqslant 2|v_i v_j|, \end{aligned}$$

using that $|X_{i,j}| \leqslant 1$. Therefore, we have that

$$\sum_{1\leqslant i\leqslant j\leqslant n} (Z - Z_{i,j})^2 \leqslant 4 \sum_{1\leqslant i\leqslant j\leqslant n} |v_i v_j|^2 \leqslant 4\left(\sum_{i=1}^{n} v_i^2\right) = 4$$

Therefore, by (11), we have that

$$\mathbb{P}[|Z - EZ| > t] \leqslant e^{-t^2/16}$$

3.3 Key Words:

1. Concentration of Measure
2. Poincare Inequality
3. Efron-Stein inequality
4. Bounded Differences
5. Log-Sobolev Inequality

3.4 Exercises

Exercise 1: Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be two sequences of coin flips. Show that with large probability (larger than 1/2), the length of the longest common sequence is within $O(\sqrt{n})$ of the mean.

Exercise 2: Show that (9) implies (10)

4 Comparison methods and applications.

Talks given by HaoKai Xi, November 2015 in Madison, WI

Let $\{X_1, ..., X_n, ...\}$ be a sequence of random variables. For each $n = 1, 2, ...$ let F_n be a function of n variables. If $F_n(X_1, ..., X_n)$ has a limiting distribution, how to find it? Assume there is another sequence of variables $\{Y_1, ..., Y_n, ...\}$, Y_n is "close" to X_n for each n, and we know the limiting distribution of $F_n(Y_1, ..., Y_n)$. Then we might verify that $F_n(X_1, ..., X_n)$ has the same limiting distribution by showing that $\mathbb{E}\, g\,(F_n(X_1, ..., X_n)) - \mathbb{E}\, g\,(F_n(Y_1, ..., Y_n))$ converges to 0 for all bounded continuous function g through a interpolation or some replacement trick. In this paper, I will focus on three methods:

1. Lindeberg replacement trick;
2. Comparison through $X_n^s = \sqrt{s}X_n^1 + \sqrt{1-s}X_n^0$, where $X_n^0 = X_n,\ X_n^1 = Y_n$.
3. Comparison through $X_n^s = \mathcal{X}_n^s X_n^1 + (1 - \mathcal{X}_n^s)X_n^0$ where X_n^s is a Bernoulli random variable with $\mathbb{P}(\mathcal{X}_n^s = 1) = s$ and $\mathbb{P}(\mathcal{X}_n^s = 0) = 1 - s$.

4.1 Lindeberg replacement trick

The main idea is replacing X_k with Y_k iteratively for $k = 1, 2, ..., n$ in $F_n(X_1, ..., X_n)$ so it will become $F_n(Y_1, ..., Y_n)$ and also produces some error in distribution. Next we illustrate how to use this method to prove the Central Limit Theorem. It suffices to show the following version

Theorem 4.1 (Central Limit Theorem). *Assume $X_1, ..., X_n, ...$ is a sequence of bounded i.i.d. random variables with $\mathbb{E}\, X_1 = 0$ and $\mathbb{E}\, X_1^2 = 1$, then we have*

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Proof. Let $Y_1, ..., Y_n, ...$ be a sequence of i.i.d. Gaussian random variables with mean 0 and variance 1 that are independent of $X_1, ..., X_n, ...$. For $X = X_1,\ Y_1$ or 0, denote

$$Z_{i,n}(X) := (Y_1 + \cdots + Y_{i-1} + X + X_{i+1} + \cdots + X_n)/\sqrt{n}$$

where $1 \leqslant i \leqslant n$ and $n = 1, 2, ...$ Clearly we have $Z_{n,n}(Y_n) \overset{d}{=} \mathcal{N}(0, 1)$ for all n. Therefore it suffices to show that for any $g \in C^3(\mathbb{R})$ with bounded derivatives up to 3rd order,

$$\mathbb{E}[g(Z_{n,n}(Y_n)) - g(Z_{1,n}(X_1))] \to 0$$

Actually

$$\mathbb{E}[g(Z_{n,n}(Y_n)) - g(Z_{1,n}(X_1))] = \sum_{i=1}^{n} \mathbb{E}[g(Z_{i,n}(Y_i)) - g(Z_{i,n}(X_i))] \tag{12}$$

By Taylor expansion, we have

$$\begin{aligned} &\mathbb{E}\, g(Z_{i,n}(X)) \\ &\quad = \mathbb{E}\, g(Z_{i,n}(0)) + \frac{1}{\sqrt{n}}\,\mathbb{E}\, g'(Z_{i,n}(0))X + \frac{1}{2n}\,\mathbb{E}\, g''(Z_{i,n}(0))X^2 + \frac{1}{6n^{3/2}}\,\mathbb{E}\, g'''(Z_{i,n}(\tilde{X}))X^3 \end{aligned} \tag{13}$$

where $\tilde{X}$ is between 0 and X. Notice that $Z_{i,n}(0)$ is independent of X_i and Y_i, and X_i and Y_i have the same first and second moment, hence we have

$$|\mathbb{E}[g(Z_{i,n}(Y_i)) - g(Z_{i,n}(X_i))]| \leqslant Cn^{-3/2} \tag{14}$$

where $C = \mathbb{E}\, X_1^3 \cdot \sup_{x\in\mathbb{R}} g'''(x)$. Thus

$$|\,\mathbb{E}[g(Z_{n,n}(Y_n)) - g(Z_{1,n}(X_1))]| \leqslant Cn^{-1/2} \tag{15}$$

Let $n \to \infty$ and we are done. $\square$

4.2 $X^s = \sqrt{s}X^1 + \sqrt{1-s}X^0$ type interpolation

We use this interpolation to prove Theorem 4.1:

Proof. Denote

$$Z_n^s = \frac{X_1^s + \cdots + X_n^s}{\sqrt{n}}$$

where $X_k^s = \sqrt{s}X_k^1 + \sqrt{1-s}X_k^0$ for all k and $s \in [0,1]$. For any $g \in C^3(\mathbb{R})$ with bounded derivatives up to 3rd order, we want to show that

$$\mathbb{E}[g(Z_n^1) - g(Z_n^0)] \to 0.$$

By foundamental theorem of calculus, it suffices to show that

$$\sup_{0<s<1} \mathbb{E}\,\frac{dg(Z_n^s)}{ds} = O(n^{-1/2})$$

Actually,

$$\begin{aligned}\mathbb{E}\,\frac{dg(Z_n^s)}{ds} &= \mathbb{E}\,g'(Z_n^s)\sum_{i=1}^n \frac{\partial Z_n^s}{\partial X_i^s}\frac{dX_i^s}{ds} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}\,g'(Z_n^s)(s^{-1/2}X_i^1 - (1-s)^{-1/2}X_i^0) \end{aligned} \tag{16}$$

For each $1 \leqslant i \leqslant n$, denote $Z_n^{s(i)} = Z_n^s - X_i^s/\sqrt{n}$, we have

$$g'(Z_n^s) = g'(Z_n^{s(i)}) + \frac{1}{\sqrt{n}}g''(Z_n^{s(i)})X_i^s + \frac{1}{2n}g'''(\widetilde{Z}_n^{s(i)})(X_i^s)^2 \tag{17}$$

for some $\widetilde{Z}_n^{s(i)}$ between Z_n^s and $Z_n^{s(i)}$. Now plug (19) into (16), since $Z_n^{(i)s}$ is independent of X_i^0 and X_i^1, X_i^0 and X_i^1 have vanishing first moment and the same second moment, we get

$$\left|\mathbb{E}\,\frac{dg(Z_n^s)}{ds}\right| \leqslant \sum_{i=1}^n \frac{1}{2n^{3/2}} \sup_{x\in\mathbb{R}} |g'''(x)| \cdot |\,\mathbb{E}\,X_1|^3 = Cn^{-1/2}.$$

□

Next let's look at an example from random matrix, which I find in [?]. Denote by X_n an $n \times n$ Wigner matrix, in which the upper triangular entries $(X_n)_{ij}, i < j$ are real random variables with mean 0 and variance $1/n$ and the diagonal entries $(X_n)_{ii}$ are real random variables with mean 0 and variance $2/n$. Denote by Y_n an $n \times n$ GOE, in which the upper triangular entries has distribution $\mathcal{N}(0, 1/n)$ and the diagonal entries has distribution $\mathcal{N}(0, 2/n)$. Assume we have that the asymptotic eigenvalue density of GOE is the semi-circular law, we show by a interpolation trick that the asymptotic eigenvalue density of Wigner matrices is also the semi-circular law.

Lemma 4.2. *Let $\{X_n\}$ be the Wigner matrices that satisfy $\mathbb{E}(\sqrt{n}(X_n)_{ij})^p \leqslant C_p$ for all n, p, i, j where C_p is independent of n, i, j. Let Y_n be GOE, then for any z with* $\mathrm{Im}z > 0$

$$\mathbb{E}\left(\frac{1}{n}TrG(X_n, z)\right) - \mathbb{E}\left(\frac{1}{n}TrG(Y_n, z)\right) \to 0.$$

where $G(X, z) = (X - z)^{-1}$, the resolvent of X.

Convention of notations: In the rest of the paper I ignore the parameter n for matrices, so without mentioning specifically, all matrices are n by n.
Denote by $X^{s,\lambda}_{(ij)}$ the matrix that satisfies

$$(X^{s,\lambda}_{(ij)})_{ab} = \begin{cases} X^s_{ab} & \text{if } \delta_{ai}\delta_{bj} = 0 \text{ and } \delta_{aj}\delta_{bi} = 0 \\ \lambda & \text{if } \delta_{ai}\delta_{bj} = 1 \\ \overline{\lambda} & \text{if } \delta_{aj}\delta_{bi} = 1 \end{cases}$$

Denote by $\Delta^s_{(ij)}$ the matrix that satisfies

$$(\Delta^s_{(ij)})_{ab} = \begin{cases} 0 & \text{if } \delta_{ai}\delta_{bj} = 0 \text{ and } \delta_{aj}\delta_{bi} = 0 \\ X^s_{ij} & \text{if } \delta_{ai}\delta_{bj} = 1 \\ \overline{X^s_{ji}} & \text{if } \delta_{aj}\delta_{bi} = 1 \end{cases}$$

Also,

$$G^s := G(X^s, z),$$

$$G^{s,\lambda}_{(ij)} := G(X^{s,\lambda}_{(ij)}, z)$$

Proof. It is easy to check that if X is Hermitian then $\|G(X,z)\| \leqslant \frac{1}{\mathrm{Im}z}$.
We choose the interpolation $X^s = \sqrt{s}X^1 + \sqrt{1-s}X^0$ with $X^0 = X$ and $X^1 = Y$. We have

$$\frac{1}{n}\mathbb{E}\frac{d}{ds}TrG^s = -\frac{1}{n}\mathbb{E}Tr\{(G^s)(s^{-1/2}X^1 - (1-s)^{-1/2}X^0)(G^s)\} \tag{18}$$

Expanding G^s at X^s_{ij} we get

$$G^s = G^s_{(ij)} - G^s_{(ij)}\Delta^s_{(ij)}G^s_{(ij)} + G^s\Delta^s_{(ij)}G^s_{(ij)}\Delta^s_{(ij)}G^s_{(ij)} \tag{19}$$

Plug (19) into (18) and multiply out $s^{-1/2}X^1$ we get

$$\begin{aligned}\mathbb{E}Tr(G^s)(s^{-1/2}X^1)(G^s) =& s^{-1/2}\mathbb{E}Tr(G^s)(\sum_{1\leqslant i\leqslant j\leqslant n}\Delta^1_{(ij)})(G^s) \\ =& \sum_{1\leqslant i\leqslant j\leqslant n}\Big\{Tr[-2\mathbb{E}(G^s_{(ij)})^3\mathbb{E}(\Delta^1_{(ij)})^2] + O(n^{-3/2})\Big\} \\ =& -2\sum_{1\leqslant i\leqslant j\leqslant n}\Big\{\mathbb{E}(G^s_{(ij)})^3_{ii}\mathbb{E}(\Delta^1_{(ij)})^2)_{ii} + (1-\delta_{ij})\mathbb{E}(G^s_{(ij)})^3_{jj}\mathbb{E}(\Delta^1_{(ij)})^2)_{jj} + O(n^{-3/2})\Big\}\end{aligned}$$

We get the same first two terms in the result of the computation of $\mathbb{E}Tr(G^s)((1-s)^{-1/2}X^0)(G^s)$.
Therefore in (18) these terms cancel out and it remains $O(n^{-1/2})$. □

4.3 $X^s = \mathcal{X}^s X^1 + (1-\mathcal{X}^s)X^0$ type interpolation

This method is first developed in [6].
Just like the previous methods, we start with computing $\partial_s\mathbb{E}F(X^s)$ for some function $F:\mathbb{R}^n \to \mathbb{C}$. The advantage of such interpolation is that this derivative is very clean, as shown in the following lemma.

Lemma 4.3. *For $F:\mathbb{R}^n \to \mathbb{C}$ we have*

$$\partial_s\mathbb{E}F(X^s) = \sum_{1\leqslant i\leqslant n}\mathbb{E}F(X^{s,X^1_i}_{(i)}) - \mathbb{E}F(X^{s,X^0_i}_{(i)})$$

provided all the expectations exist. $X^{s,\lambda}_{(i)}$ is X^s with the ith component replaced with λ.

Proof. Let μ_i^s be the distribution of X_i^s so we have

$$\mathrm{d}\mu_i^s = s\mathrm{d}\mu_i^1 + (1-s)\mathrm{d}\mu_i^0.$$

For simplicity of notations, $\prod_k^{(i)}$ means $\prod_{1\leqslant k\leqslant n, k\neq i}$ and $\sum_k$ means $\sum_{1\leqslant k\leqslant n}$ in the equations below.

$$\begin{aligned}
\partial_s \mathbb{E}\, F(X^s) =& \partial_s \int F(X) \prod_k d\mu_k^s \\
=& \sum_i \int \partial_s \left[sF(X)d\mu_i^1 + (1-s)F(X)d\mu_i^0 \right] \prod_k^{(i)} d\mu_k^s \\
=& \sum_i \int F(X)d\mu_i^1 \prod_k^{(i)} d\mu_k^s - \int F(X)d\mu_i^0 \prod_k^{(i)} d\mu_k^s \\
=& \sum_i \mathbb{E}\, F(X_{(i)}^{s,X_i^1}) - \mathbb{E}\, F(X_{(i)}^{s,X_i^0})x
\end{aligned}$$

□

Exercise: Prove the CLT using this interpolation.

We can exploit $\partial_s \mathbb{E}\, F(X^s)$ further. Consider the following question, suppose we know that $\mathbb{E}\, F(X^0) \leqslant \Psi$ for some small Ψ, and we want to show that $\mathbb{E}\, F(X^1) \leqslant c\Psi$ for some constant c. We can use Gronwall's inequality, with which it suffices to show

$$\partial_s \mathbb{E}\, F(X^s) \leqslant \Psi + \mathbb{E}\, F(X^s).$$

This requires us to have a self-consistent estimate for $\partial_s \mathbb{E}\, F(X^s)$, i.e. we want to write it in term of X^s.

Lemma 4.4. *Assume F is analytic and X^s has finite moments for all s,*

$$\partial_s \mathbb{E}\, F(X^s) = \sum_{k\geqslant 1} \sum_i K_{k,i}^s \,\mathbb{E} \left(\frac{\partial}{\partial X_i^s} \right)^k F(X^s)$$

where $K_{k,i}^s$ is the coefficient of the kth order term of the formal power series of

$$\frac{\mathbb{E}\, e^{tX_i^1} - \mathbb{E}\, e^{tX_0^1}}{\mathbb{E}\, e^{tX_i^s}}$$

Proof. We fix i and abbreviate $f(x) := F(X_{(i)}^{s,x})$, $\xi := X_i^s$, $\zeta := X_i^0$, $\zeta' := X_i^1$. By Taylor expansion,

$$\mathbb{E}\, f(\zeta) = \mathbb{E}\, f(0) + \sum_{k\geqslant 1} \mathbb{E}\, f^{(k)}(0)\,\mathbb{E}\,\zeta^k/k! \tag{20}$$

We use

$$\mathbb{E}\, f^{(l)}(0) = \mathbb{E}\, f^{(l)}(\xi) - \sum_{k\geqslant 1} \mathbb{E}\, f^{(l+k)}(0)\,\mathbb{E}\,\xi^l/l!$$

repeatedly on (20) to get

$$\mathbb{E}(f(\zeta)-f(0)) = \sum_{q\geqslant 0}(-1)^q \sum_{k,k_1,\dots,k_q\geqslant 1} \mathbb{E}\, f^{(k+k_1+\cdots+k_q)}(\xi)\,\mathbb{E}\,\zeta^k/k! \prod_{j=1}^q \mathbb{E}\,\xi^{k_j}/k_j! = \sum_{m\geqslant 1} K_m(\zeta,\xi)\,\mathbb{E}\, f^{(m)}(\xi)$$

where

$$K_m(\zeta,\xi) = \sum_{q\geqslant 0}(-1)^q \sum_{k,k_1,\dots,k_q = m} \mathbb{E}\,\zeta^k/k! \prod_{j=1}^q \mathbb{E}\,\xi^{k_j}/k_j! = \frac{1}{m!}\left(\frac{d}{dt}\right)^m \Bigg|_{t=0} \frac{\mathbb{E}\, e^{t\zeta}-1}{\mathbb{E}\, e^{t\xi}}$$

Now by Lemma 4.3, summing over i we get the desired result.

□

4.4 Key words

- Comparison method
- Lindeberg replacement trick
- Interpolation
- Gronwall's inequality

4.5 Exercise

Prove the CLT using the second interpolation method. (The 3rd method in this talk.).

5 An introduction to initial enlargement of filtration.

Talks given by Christoper Janjigian, November-December 2015 in Madison, WI

These incomplete notes are a work in progress. They are intended for a two-part talk in the graduate probability seminar at the University of Wisconsin - Madison. The goal is to introduce listeners to the techniques of initial enlargement of filtration and the Doob h transform, which offer perspectives on what it means to condition a stochastic process on a potentially measure zero event. We will be particularly interested in cases when this conditioning preserves some regularity property, such as being a semi-martingale, being a diffusion, or being Markov. The first part of this talk is a reworked and corrected version of a talk that I gave in this seminar three years ago.

5.1 Introduction - discrete random walk bridge and Brownian bridges

5.1.1 Discrete random walk bridge

In our first example, we consider a random walk in discrete time and we want to condition the random walk to arrive at a specified position at a certain time. Our goal is to describe the law of this conditioned process in a nice way.

Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables with $E[X_1] = 0$ and denote by $S_n = \sum_{i=0}^{n} X_i$ with the convention that $X_0 = 0$. Fix some integer $N > 1$. Our first goal is going to be to understand what happens to the distribution of the process S_n for $0 \leqslant n \leqslant N$ if we condition on the random variable S_N. To do this, we first note that S_n is a martingale in the filtration $\mathcal{F}_n = \sigma(X_i : 0 \leqslant i \leqslant n)$. We will attempt to write the semi-martingale decomposition of S_n in the filtration $\mathcal{G}_n = \mathcal{F}_n \vee \sigma(S_N)$.

To do this, we will chase the proof of the existence of the semi-martingale decomposition. We want

- $S_n = M_n + A_n$,
- $E[M_n | \mathcal{G}_{n-1}] = M_{n-1}$,
- $E[A_n | \mathcal{G}_{n-1}] = A_n$,

which implies a recursion that yields

$$M_n = \sum_{k=1}^{n} S_k - E\left[S_k \middle| \mathcal{G}_{k-1}\right],$$
$$A_n = \sum_{k=1}^{n} E\left[S_k \middle| \mathcal{G}_{k-1}\right] - S_{k-1}.$$

We take the convention that the empty sum is zero. Notice that since S_N and S_k are both $\mathcal{G}_k$ measurable for each k, so $S_N = E[S_N | \mathcal{G}_{k-1}]$. On the other hand, by symmetry for each $n, m \geqslant k$ we see that $E[X_n | \mathcal{G}_{k-1}] = E[X_m | \mathcal{G}_{k-1}]$. It follows that

$$\begin{aligned} S_N = E[S_N | \mathcal{G}_{k-1}] &= E[S_{k-1} + \sum_{n=k}^{N} X_n | \mathcal{G}_{k-1}] \\ &= S_{k-1} + (N - k + 1) E[X_k | \mathcal{G}_{k-1}]. \end{aligned}$$

Rearranging, we see that for $k \leqslant n \leqslant N$,

$$E[X_n | \mathcal{G}_{k-1}] = \frac{S_N - S_{k-1}}{N - (k-1)}.$$

Consequently, for $n \geqslant 1$,

$$M_n = \sum_{k=1}^{n} S_k - E[S_k | \mathcal{G}_{k-1}] = \sum_{k=1}^{n} X_k - E[X_k | \mathcal{G}_{k-1}]$$

$$= \sum_{k=1}^{n} X_k - \frac{S_N - S_{k-1}}{N-(k-1)} = S_n - \sum_{k=1}^{n} \frac{S_N - S_{k-1}}{N-(k-1)}$$

and

$$A_n = \sum_{k=1}^{n} E\left[S_k \middle| \mathcal{G}_{k-1}\right] - S_{k-1} = \sum_{k=1}^{n} E\left[X_k \middle| \mathcal{G}_{k-1}\right] = \sum_{k=1}^{n} \frac{S_N - S_{k-1}}{N-(k-1)}.$$

The key here is that we can still identify the martingale part of S_n in its natural filtration *even if we condition on S_N.*

5.1.2 Brownian bridge

Typically, Brownian bridge is introduced as a 'Brownian motion conditioned on $B_1 = 0$.' One should be wary of this definition at first glance: $P(B_1 = 0) = 0$, so it is not entirely trivial to say what this means. Questions of this type are going to be the main focus of this talk and we will loosely structure the exposition around the Brownian bridge. Other, more involved, examples will be interspersed along the way.

Weak limit using the Gaussian structure

One natural way to condition on the event $\{B_1 = 0\}$ would be to take a limit in some sense of what we get if we condition $\{B_t\}_{t\geqslant 0}$ on $\{|B_1| < \epsilon\}$ as $\epsilon \downarrow 0$.

Write $B_t = B_t - tB_1 + tB_1$. Since Brownian motion is a Gaussian process, linear combinations of coordinate projections are jointly normal so we may compute covariances to see that for $b(t) = B_t - tB_1$,

$$\{b(t)\}_{0\leqslant t\leqslant 1} \qquad \text{and} \qquad B_1$$

are independent.

Exercise 5.1. *Verify that the finite dimensional distributions of $\{b(t)\}_{0\leqslant t\leqslant 1}$ and B_1 are independent (i.e. show that the vector $(b(t_1), \ldots, b(t_n))$ is independent of B_1 for $0 \leqslant t_1 < \cdots < t_n \leqslant 1$).*

Once one knows that $b(t)$ and B_1 are independent, we can now take weak limits without too much difficulty.

Exercise 5.2. *Let $F \in C_b(C([0,1]))$. Show that as $\epsilon \to 0$*

$$E\left[F(B_\cdot) \middle| |B_1| \leqslant \epsilon\right] = E\left[F(b(\cdot) + \cdot B_1) \middle| |B_1| \leqslant \epsilon\right] \to E\left[F(b(\cdot))\right].$$

This shows that defining $b(t) = B_t - tB_1$ gives a sensible definition of Brownian bridge. Are there others?

Enlargement of filtration

Let $\tilde{B}_t$ be standard Brownian motion and consider the process

$$X_t = (1-t)\int_0^t \frac{1}{1-s} d\tilde{B}_s.$$

One can check that $\{X_t\}_{0\leqslant t\leqslant 1}$ and $\{B_t - tB_1\}_{0\leqslant t\leqslant 1}$ have the same distribution.

Exercise 5.3. *(If you have taken stochastic calculus.) Verify that $\{X_t\}_{0\leqslant t\leqslant 1}$ has the same distribution as $\{B_t - tB_1\}_{0\leqslant t\leqslant 1}$. Hint: Show that $\{X_t\}$ and $\{B_t - tB_1\}$ are Gaussian processes with the same mean and covariance structure.*

Where did this formula come from? If we want to condition $\{B_t\}$ on $\{B_1 = 0\}$, we might hope to understand what happens to the distribution of the process $\{B_t\}$ if we condition on the random variable B_1 directly and then 'set $B_1 = 0$'. Recall that B_t comes equipped with the right continuous completion of the filtration $\mathcal{F}_t = \sigma(B_s : s \leqslant t)$. One way to understand what happens to B_t if we condition on B_1 would be to try to write a stochastic differential equation for B_t in the right continuous completion of the filtration $\mathcal{F}_t^{(B_1)} = \sigma(B_s : s \leqslant t) \vee \sigma(B_1)$. In both cases, I am going to abuse notation and refer to both these filtrations and their completions by $\mathcal{F}_t$ and $\mathcal{F}_t^{(B_1)}$. Define

$$\tilde{B}_t = B_t - \int_0^t \frac{B_1 - B_s}{1-s} ds.$$

I claim that $\tilde{B}_t$ is $\mathcal{F}_t^{(B_1)}$ Brownian motion. Note that based on Donsker's theorem, we could have guessed that this would be Brownian motion based on the semi-martingale decomposition of S_n in $\sigma(X_i : i \leqslant n) \vee \sigma(S_N)$ above.

If true, then the semi-martingale decomposition of B_t in this filtration is then given by

$$B_t = \tilde{B}_t + \int_0^t \frac{B_1 - B_s}{1-s} ds.$$

A natural interpretation of Brownian bridge would then be a solution to

$$dX = -\frac{X_t}{1-t} dt + d\tilde{B}$$

which one can compute explicitly to be equal to X_t defined above. As a comment, it follows from the Markov property and the fact that $\tilde{B}_0 = 0$ that $\tilde{B}_t$ is independent of B_1 for all t.

Last time I gave this part of the talk, I went through the computation showing that $\tilde{B}_t$ is Brownian motion. It is a bit tedious and I want to cover other things, so I will give a sketch and leave the details as an exercise. Observe that for $t \in [0, 1)$

$$E[B_t - B_s | \mathcal{F}_s \vee \sigma(B_1)] = E[B_t - B_s | \mathcal{F}_s \vee \sigma(B_1 - B_s)] = E[B_t - B_s | \sigma(B_1 - B_s)].$$

$B_t - B_s$ and $B_1 - B_s$ are jointly normal with mean and covariance matrix given by

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} t-s & t-s \\ t-s & 1-s \end{pmatrix}$$

so direct computation shows that

$$E[B_t - B_s | \sigma(B_1 - B_s)] = \frac{t-s}{1-s}(B_1 - B_s).$$

Using the previous comments and applying application the conditional Fubini's lemma, we have

$$E\left[\tilde{B}_t - \tilde{B}_s | \mathcal{F}_t^{(B_1)}\right] = E\left[B_t - B_s + \int_s^t \frac{B_1 - B_u}{1-u} du | \sigma(B_1 - B_s)\right] = 0.$$

Exercise 5.4. *Fill in the details of this argument showing that $\tilde{B}_t$ is a $\mathcal{F}_t^{(B_1)}$ martingale. Showing that this is actually Brownian motion is part of a later exercise.*

5.2 Initial enlargement of filtration

The previous example is a bit unfulfilling because it is not clear where the formula for β_t came from. It turns out that this is an example of a general phenomenon.

Let X be a random variable and let $\mathcal{F}_t$ be the right continuous completion of the filtration generated by a Brownian motion B_t. Let $\lambda_t(f)$ be a continuous version of the process $E[f(X)| \mathcal{F}_t]$.

Since $\lambda_t(f)$ is a continuous martingale in a Brownian filtration, it is a theorem that there exists a stochastic process $\hat{\lambda}_t(f)$ such that

$$\lambda_t(f) = E[f(X)] + \int_0^t \hat{\lambda}_s(f) dB_s$$

A bit of work shows that there exists a predictable family of measures $\lambda_t(dx)$ with the property that

$$\lambda_t(f) = \int f(x)\lambda_t(dx).$$

We will **assume** that there exists a predictable family of measures $\hat{\lambda}_t(dx)$ satisfying

$$\hat{\lambda}_t(dx) = \rho(t,x)\lambda_t(dx), \qquad \hat{\lambda}_t(f) = \int f(x)\hat{\lambda}_t(dx) \tag{21}$$

The meaning of this condition will become clear in the examples that follow. With this notation, we have the following theorem. Note: It is the main result of this note.

Theorem 5.5. *[Yor] Suppose that $M_t = \int_0^t m_s dB_s$ is a continuous martingale in the filtration $\mathcal{F}_t$ and that X satisfies (21). Then under appropriate integrability conditions, there exists an $\mathcal{F}_t^{(X)}$ local martingale $\tilde{M}_t$ with the property that*

$$M_t = \tilde{M}_t + \int_0^t \rho(X,s) m_s ds$$

Proof. Let $A \in \mathcal{F}_s$ and let f be a sufficiently regular test function. Then

$$\begin{aligned} E\left[1_A f(X)(M_t - M_s)\right] &= E\left[1_A \left(\lambda_t(f)M_t - \lambda_s(f)M_s\right)\right] \\ &= E\left[1_A \left([\lambda(f), M]_t - [\lambda(f), M]_s\right)\right] \end{aligned}$$

where $[X,Y]$. denotes the quadratic variation of X and Y. This follows from an application of stochastic integration by parts

$$\lambda_t(f)M_t - \lambda_s(f)M_s = \int_s^t \lambda_u(f) dM_u + \int_s^t M_u d\lambda_u(f) + [\lambda(f), M]_t - [\lambda(f), M]_s$$

where we require enough integrability that the first two terms on the right hand side, which are $\mathcal{F}_t$ local martingales, are $\mathcal{F}_t$ martingales. Now, we note that

$$\begin{aligned} E\left[1_A \left([\lambda(f), M]_t - [\lambda(f), M]_s\right)\right] &= E\left[1_A \int_s^t \hat{\lambda}_u(f) m_u du\right] \\ &= E\left[1_A \int_s^t \int \rho(X,u)\lambda_u(dx) m_u du\right]. \end{aligned}$$

The result follows from the observation that

$$E\left[f(X)\rho(X,u) | \mathcal{F}_u\right] = \int f(x)\rho(x,u)\lambda_u(dx).$$

$\square$

Exercise 5.6. *In the setting of Theorem 5.5, show that if M_t is $\mathcal{F}_t$ Brownian Motion, then $\tilde{M}_t$ is $\mathcal{F}_t^{(X)}$ Brownian motion and is independent of X. Hint: Apply Lévy's critereon and note that $X \in \mathcal{F}_0^{(X)}$.*

5.2.1 Examples

Brownian bridge

In the setting that we started with, let $X = B_1$, we can use the Markov property to compute

$$\begin{aligned}\lambda_t(f) = E[f(B_1)|\,\mathcal{F}_t] &= E[f(B_1)|B_t] \\ &= \int_{\mathbb{R}} f(x)\frac{1}{\sqrt{2\pi(1-t)}}e^{-\frac{(B_t-x)^2}{2(1-t)}}dx\end{aligned}$$

In order to compute $\hat{\lambda}_t(f)$, we can compute

$$d\frac{1}{\sqrt{2\pi(1-t)}}e^{-\frac{(B_t-x)^2}{2(1-t)}} = \frac{x-B_t}{1-t}\frac{1}{\sqrt{2\pi(1-t)}}e^{-\frac{(B_t-x)^2}{2(1-t)}}dB_t$$

For this measure, we then see that $\rho(x,t) = \frac{x-B_t}{1-t}$. We conclude that the semi-martingale decomposition of B_t in the filtration $\mathcal{F}_t^{(B_1)}$ is

$$B_t = \tilde{B}_t + \int_0^t \frac{B_1 - B_s}{1-s}ds$$

where $\tilde{B}_t$ is $\mathcal{F}_t^{(B_1)}$ Brownian Motion.

Aside: why is this a natural way to condition?

One might wonder at this point how this method of conditioning fits into the usual framework of conditioning random variables. We will use the Brownian Bridge as an example.

Take $g \in C[0,1]$ fixed and let $x \in \mathbb{R}$. We consider the ordinary differential equation

$$f(t) = g(t) + \int_0^t \frac{x - f(s)}{1-s}ds.$$

Showing that this ODE has a unique solution and that the solution map is nice is left as an exercise.

Exercise 5.7. *Check that if $g \in C[0,1]$ and $x \in \mathbb{R}$ then the ordinary integral equation*

$$f(t) = g(t) + \int_0^t \frac{x - f(s)}{1-s}ds,$$

has at most one solution $f \in C[0,1]$. Show that if we define $f(t)$ by

$$f(t) = (1-t)g(0) + xt + (1-t)\int_0^t \frac{1}{1-s}dg(s),$$

then f lies in $C[0,1]$, satisfies the ODE, and has $f(1) = x$. Define the map

$$F(x,g) = \left\{(1-t)g(0) + xt + (1-t)\int_0^t \frac{1}{1-s}dg(s)\right\}_{0\leqslant t\leqslant 1}.$$

Find $C > 0$ so that $\|F(x,g) - F(y,h)\|_{C[0,1]} \leqslant C\left(|x-y| + \|g-h\|_{C[0,1]}\right)$.
Hints: To show uniqueness, note that if f_1 and f_2 are solutions then for $t \leqslant T < 1$,

$$\sup_{0\leqslant s\leqslant t} |f_1(s) - f_2(t)| \leqslant \frac{1}{1-T}\int_0^t \sup_{0\leqslant r\leqslant s}|f_1(r) - f_2(r)|ds.$$

To show that f solves the ODE, it may help to find a differential equation satisfied by $H(t) = \int_0^t \frac{f(s)}{1-s}ds$. Note that for $t < 1$, $\int_0^t \frac{1}{1-s}dg(s)$ is a Riemann-Stieltjes integral satisfying

$$\int_0^t \frac{1}{1-s}dg(s) = g(t)\frac{1}{1-t} - g(0) - \int_0^t \frac{g(s)}{(1-s)^2}ds.$$

We saw above that there is a $\sigma(B_s : s \leqslant t) \vee \sigma(B_1)$ Brownian motion $\tilde{B}$ so that

$$B_t = \tilde{B}_t + \int_0^t \frac{B_1 - B_s}{1-s} ds.$$

The previous exercise shows that there is a Borel measurable (Lipschitz continuous) function $F : C[0,1] \times \mathbb{R} \to C[0,1]$ so that $B = F(\tilde{B}, B_1)$. It follows that for $G \in \mathcal{B}_b(C[0,1])$ and $A \in \sigma(B_1)$,

$$E\left[G(B) 1_A\right] = E\left[\tilde{E}\left[G(F(\tilde{B}, B_1))\right] 1_A\right]$$

where $\tilde{E}$ is the expectation with respect to the Brownian motion $\tilde{B}$, which is independent of B_1. More generally, this argument will work when we have strong solutions to our stochastic differential equations, even if the integrals are not Riemann-Stieltjes. Put another way, we have identified the conditional distribution of the process B given B_1 in the usual sense of a conditional expectation.

Stochastic integral with a deterministic integrand

Exercise 5.8. *[5] Show that the semimartingale decomposition of B_t in the filtration $\mathcal{F}_t^{(\int_0^\infty e^{-s} dB_s)}$ is*

$$B_t = \tilde{B}_t + \int_0^t 2e^{-s} \int_s^\infty e^{-r} dB_r ds$$

where $\tilde{B}_t$ is $\mathcal{F}_t^{(\int_0^\infty e^{-s} dB_s)}$ Brownian motion. Hint: $\int_t^\infty e^{-s} dB_s$ is normally distributed and independent of $\mathcal{F}_t$.

A perpetuity, Dufresne identities, and the O'Connell-Yor polymer

This section is based in part on [8, Example 1.8]. Define a functional by

$$A_t = \int_0^t e^{2B_s - s} ds$$

and note that A_∞ exists almost surely.

Exercise 5.9. *Prove that* $\lim_{t\to\infty} A_t$ *exists almost surely. Hint: You can use the law of the iterated logarithm for B_t.*

Let f be a smooth test function compactly supported in $(0,\infty)$:

$$\begin{aligned}
\lambda_t(f) &= E\left[f(A_\infty) | \mathcal{F}_t\right] \\
&= E\left[f\left(A_t + e^{2B_t - t}\int_t^\infty e^{2B_s - 2B_t - (s-t)} ds\right) | \mathcal{F}_t\right] \\
&= E\left[f\left(A_t + e^{2B_t - t}\hat{A}_\infty\right) | \mathcal{F}_t\right]
\end{aligned}$$

where $\hat{A}_\infty$(which depends on t) is independent of $\mathcal{F}_t$. We then see that

$$\begin{aligned}
\lambda_t(f) &= \hat{E}\left[f\left(A_t + e^{2B_t - t}\hat{A}_\infty\right)\right] \\
\hat{\lambda}_t(f) &= \hat{E}\left[2e^{B_t - t}\hat{A}_\infty f'\left(A_t + e^{2B_t - t}\hat{A}_\infty\right)\right]
\end{aligned}$$

where the expectation $\hat{E}$ is only with respect to $\hat{A}_\infty$. We would like to identify the distribution of $\hat{A}_\infty$. This will be presented in two ways.

Hints of Lamperti's relation: Recall that $G_t = e^{B_t - \frac{1}{2}t}$ solves

$$dG_t = G_t dB_t$$

It follows from Dubins-Schwarz that G_t admits a representation as

$$G_t = \beta_{\int_0^t e^{2B_s - s} ds} = \beta_{A_t}$$

where β is Brownian motion. From this coupling, we see that $\beta_0 = 1$ and A_∞ has the same distribution as $\inf\{t > 0 : \beta_t = 0\}$. If you happen to know the distribution of this already, feel free to jump to the semi-martingale decomposition directly.

Dufresne identity. For each fixed t we have

$$\begin{aligned} Z_t = e^{2B_t - t} \int_0^t e^{-2B_s + s} ds &= \int_0^t e^{2(B_t - B_s) + s - t} ds \\ &= \int_0^t e^{2(B_t - B_{t-u}) - u} du \\ &\overset{d}{=} \int_0^t e^{2B_u - u} du = A_t \end{aligned}$$

Z_t solves the SDE

$$dZ_t = (1 + Z_t)\, dt + 2Z_t dB_t$$

In particular, Z_t is Markov. The limit $A_t \to A_\infty$ holds almost surely and therefore in distribution, so Z_t converges in distribution as $t \to \infty$ to A_∞. It will be helpful to know that this SDE is ergodic on $\mathbb{R}_+$. The next exercise provides a proof of this.

Exercise 5.10. *Check that the solution to the SDE*

$$\begin{aligned} dZ_t &= (1 + Z_t)\, dt + 2Z_t dB_t \\ Z_0 &= x \end{aligned}$$

is given by

$$Z_t = xe^{2B_t - t} + e^{2B_t - t} \int_0^t e^{-2B_s + s} ds$$

Conclude that Z_t has exactly one stationary distribution, given by the distribution of A_∞. Note that this distribution is supported on $\mathbb{R}_+$. (Hint: $xe^{2B_t - t} \to 0$ a.s. and we have shown that $e^{2B_t - t} \int_0^t e^{-2B_s + s} ds$ converges in distribution)

To compute the stationary distribution for Z_t, we recall that its generator $\mathcal{L}$ is given by

$$\mathcal{L} = 2x^2 \frac{d^2}{dx^2} + (1 + x)\frac{d}{dx}$$

Exercise 5.11. *Verify that $\rho(x) = x^{-\frac{3}{2}} e^{-\frac{1}{2x}} 1_{\{x>0\}}$ is a non-negative and integrable solution to*

$$2\frac{d^2}{dx^2}\left[x^2 \rho(x)\right] - \frac{d}{dx}\left[x\rho(x) + \rho(x)\right] = 0.$$

Show that for all $\varphi \in C_c^\infty(\mathbb{R}_+)$,

$$\int_{\mathbb{R}_+} \rho(x)\, \mathcal{L}\, \varphi(x) dx = 0$$

and conclude that the distribution of A_∞ is given by $Cx^{-\frac{3}{2}} e^{-\frac{1}{2x}} 1_{\{x>0\}}$ for some C.

In other words, A_∞ has an inverse gamma distribution with parameters $(\frac{1}{2},\frac{1}{2})$. Put yet another way, if Y has density proportional to $y^{-\frac{1}{2}}e^{-y}1_{\{y>0\}}$ (that is, a gamma distribution with parameter $\frac{1}{2}$), then A_∞ has the same distribution as $\frac{1}{2Y}$.

Semi-martingale decomposition. Now that we have this distribution, we can compute the semi-martingale decomposition of B_t in the filtration $\mathcal{F}_t^{A_\infty}$

$$\begin{aligned}
\lambda(f) &= C\int_0^\infty f(A_t+e^{2B_t-t}x)x^{-\frac{3}{2}}e^{\frac{1}{2x}}\,dx\\
\hat{\lambda}(f) &= C\int_0^\infty 2e^{2B_t-t}xf'(A_t+e^{2B_t-t}x)x^{-\frac{3}{2}}e^{-\frac{1}{2x}}\,dx\\
&= 2Ce^{2B_t-t}\int_0^\infty f'(A_t+e^{2B_t-t}x)x^{\frac{-1}{2}}e^{-\frac{1}{2x}}\,dx\\
&= C\int_0^\infty f(A_t+e^{2B_t-t}x)\left(1-\frac{1}{x}\right)x^{-\frac{3}{2}}e^{\frac{-1}{2x}}\,dx.
\end{aligned}$$

Changing variables in both expressions, $y=A_t+e^{2B_t-t}x$, $x=e^{t-2B_t}(y-A_t)$ and $e^{t-2B_t}dy=dx$ gives

$$\rho(x,t)=1-\frac{e^{2B_t-t}}{x-A_t}.$$

It follows that

$$B_t=\tilde{B}_t+t-\int_0^t\frac{e^{2B_s-s}}{A_\infty-A_s}ds \tag{22}$$

where $\tilde{B}_t$ is $\mathcal{F}_t^{A_\infty}$ Brownian motion *and therefore is independent of* A_∞. It is convenient to rewrite this as

$$B_t-\frac{t}{2}=\tilde{B}_t+\frac{t}{2}-\int_0^t\frac{e^{2B_s-s}}{A_\infty-A_s}ds$$

A process level identity

This section is based on [9, 11]. Note that pointwise (22) is of the form

$$f(t)=g(t)+\int_0^t e^{\alpha f(s)}\varphi\left(\int_0^s e^{\alpha f(s)}\right)ds,$$

where $f(t)=B_t-\frac{t}{2}$, $g(t)=\tilde{B}_t+\frac{t}{2}$, $\alpha=2$, and $\varphi(x)=\frac{1}{x-A_\infty}$. As above, for nice f,g,φ this is a solvable ordinary differential equation with a unique solution [9, Appendix]. We find that

$$B_t-\frac{t}{2}=\tilde{B}_t+\frac{t}{2}-\log\left(1+A_\infty^{-1}\int_0^t e^{2\tilde{B}_s+s}ds\right).$$

This result is particularly interesting read backwards. Given a random variable γ with inverse Gamma $(\frac{1}{2},\frac{1}{2})$ distribution which is independent of a Brownian motion B_t,

$$B_t+t+\log(\gamma)-\log\left(\gamma+\int_0^t e^{2B_s+s}ds\right)$$

is standard Brownian motion. Now, suppose that we have a two-sided Brownian motion B_t and recall that $\sigma(B_s:s\leqslant 0)$ is independent of $\sigma(B_s:s\geqslant 0)$. We now see that

$$B_t+t+\log\int_{-\infty}^0 e^{2B_s+s}ds-\log\int_{-\infty}^t e^{2B_s+s}ds$$

is Brownian motion and independent of $\int_{-\infty}^{0} e^{2B_s+s}ds$. Hence if we define $\hat{B}$ by

$$\begin{aligned}\hat{B}_t &= -B_t - t + \log\int_{-\infty}^{t} e^{2B_s+s}ds - \log\int_{-\infty}^{0} e^{2B_s+s}ds \\ &= B_t + \log\int_{-\infty}^{0} e^{2B_s+s}ds + \log\int_{-\infty}^{t} e^{2(B_s-B_t)+s-t}ds\end{aligned}$$

then $\hat{B}$ is Brownian motion. It will be convenient to have this identity in the form

$$\hat{B}_t + \frac{t}{2} = -B_t - \frac{t}{2} + \log\int_{-\infty}^{t} e^{2B_s+s}ds - \log\int_{-\infty}^{0} e^{2B_s+s}ds$$

Define a functional by

$$\alpha_t = \log\int_{-\infty}^{t} e^{2(B_t-B_s)-(t-s)}ds = -2B_t - t + \log\left(e^{\alpha_0} + \int_0^t e^{-2B_s+s}ds\right)$$

With this definition, we see that

$$\hat{B}_t = B_t + \alpha_t - \alpha_0$$

Perhaps surprisingly, we can also show that $\{\hat{B}_s : 0 \leqslant s \leqslant t\}$ is independent of $\{\alpha_s : s \geqslant t\}$. This follows from the fact that

$$\alpha_t = \log\int_t^{\infty} e^{2(\hat{B}_t-\hat{B}_s)+t-s}ds$$

This can be proven with a little calculus. To see this, set

$$Q_t = \int_{-\infty}^{t} e^{2B_s+s}ds$$

and notice that

$$\begin{aligned}\alpha_t &= \log\int_{-\infty}^{t} e^{2(B_s-B_t)-t+s}ds = -2B_t - t + \log Q_t \\ 2\hat{B}_t + t &= -2B_t - t + 2\log Q_t - 2\log Q_0\end{aligned}$$

Recall that

$$\begin{aligned}\int_{s\geqslant t} e^{2(\hat{B}_t-\hat{B}_s)+t-s}ds &= e^{2\hat{B}_t+t}\int_{s\geqslant t} e^{-2\hat{B}_s-s}ds \\ &= e^{-2B_t-t}Q_t^2\int_{s\geqslant t} Q_s^{-2}e^{2B_s+s}ds.\end{aligned}$$

Notice that

$$\frac{d}{dt}Q_t^{-1} = -Q_t^{-2}e^{2B_s-s}.$$

Consequently

$$\begin{aligned}\int_{s\geqslant t} e^{2(\hat{B}_t-\hat{B}_s)+t-s}ds &= e^{-2B_t-t}Q_t^2Q_t^{-1} \\ &= e^{-2B_t-t+\log Q_t} = e^{\alpha_t}.\end{aligned}$$

5.3 Key words

1. Initial enlargement of filtration
2. Brownian bridge
3. Dufresne identities

References

[1] N. Berestycki and R. Nickl. Concentration of measure, December 2009. http://www.statslab.cam.ac.uk/∼beresty/teach/cm10.pdf.

[2] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. OUP Oxford, 2013.

[3] N. Cook. Notes on talagrand's inequalities. lecture notes available at `http://www.math.ucla.edu/ nickcook/talagrand.pdf`.

[4] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, second edition, 1998.

[5] M. Jeanblanc. *Enlargements of Filtrations.*

[6] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *arXiv:1410.3516*, 2015.

[7] G. Lugosi. Concentration-of-measure inequalities, February 2006. https://web.math.princeton.edu/∼ naor/homepage%20files/Concentration%20of%20Measure.pdf.

[8] R. Mansuy and M. Yor. *Random times and enlargements of filtrations in a Brownian setting*, volume 1873 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006.

[9] H. Matsumoto and M. Yor. A relationship between Brownian motions with opposite drifts via certain enlargements of the Brownian filtration. *Osaka J. Math.*, 38(2):383–398, 2001.

[10] A. Naor. Concentration of measure, December 2008. https://web.math.princeton.edu/∼ naor/homepage%20files/Concentration%20of%20Measure.pdf.

[11] N. O'Connell and M. Yor. Brownian analogues of Burke's theorem. *Stochastic Process. Appl.*, 96(2):285–304, 2001.

[12] T. Seppäläinen. Large deviations for increasing sequences on the plane. *Probab. Theory Relat. Fields*, 112(2):221–244, 1998.