Probability concentration

Yuyuexi

September 23, 2020

Contents

1 Markov process

Definition (Markov process and Markov semigroup) *A (homogeneous) Markov process $(X_t)_{t\in\mathbb{R}_+}$ is a random process that satisfies the Markov property: for every bounded measurable function f and $s, t \in \mathbb{R}_+$, there is a bounded measurable function $P_s f$ such that*

$$\mathbf{E}\left[f\left(X_{t+s}\right) \mid \{X_r\}_{r\leq t}\right] = \left(P_s f\right)\left(X_t\right)$$

In particular, $\{P_t\}_{t\in\mathbb{R}_+}$ defines a semigroup of linear operators on $L^p(\mu)$, called Markov semigroup.

Definition (generator) *The generator $\mathscr{L}$ is defined as*

$$\mathscr{L} f := \lim_{t\to 0} \frac{P_t f - f}{t}$$

2 Poincaré inequality

Theorem (Poincareé inequality) *Let P_t be reversible ergodic Markov semigroup with stationary measure μ. The following are equivalent given $c \geq 0$:*

1. $\operatorname{Var}_\mu[f] \leq c\mathcal{E}(f, f)$ *for all* f

2. $\|P_t f - \mu f\|_{L^2(\mu)} \leq e^{-t/c}\|f - \mu f\|_{L^2(\mu)}$ *for all* f, t

3. $\mathcal{E}(P_t f, P_t f) \leq e^{-2t/c}\mathcal{E}(f, f)$ *for all* f, t

3 Subgaussian concentration

Theorem (Chernoff bound) *Define the log-moment generating function* ψ *of a random variable* X *as* $\psi(\lambda) := \log \mathbf{E}\left[e^{\lambda(X-\mathbf{E}X)}\right]$. *Then* $\mathbf{P}[X - \mathbf{E}X \geq t] \leq e^{-[\lambda t - \psi(\lambda)]}$ *for all* $t \geq 0$

Definition (subgaussian variable) *A random variable is called* σ^2 *-subgaussian if its log-moment generating function satisfies* $\psi(\lambda) \leq \lambda^2\sigma^2/2$ *for all* $\lambda \in \mathbb{R}$ *(and the constant* σ^2 *is called the variance proxy).*

4 Log-Sobolev inequality

Theorem (log-Soblev inequality) *Let* P_t *be a Markov semigroup with stationary measure* μ. *The following are equivalent:*

1. $\operatorname{Ent}_\mu[f] \leq c\varepsilon(\log f, f)$ *for all* f

2. $\operatorname{Ent}_\mu[P_t f] \leq e^{-t/c} \operatorname{Ent}_\mu[f]$ *for all* f, t

3. $\mathcal{E}(\log P_t f, P_t f) \leq e^{-t/c}\mathcal{E}(\log f, f)$ *for all* f, t *if* $\operatorname{Ent}_\mu[P_t f] \to 0$ *as* $t \to \infty$.

5 Lipschitz concentration

Definition (Wasserstein distance) *The Wasserstein distance between probability measures* $\mu, \nu \in \mathcal{P}_1(\mathbb{X}) := \left\{\rho : \int d(x, \cdot)\rho(dx) < \infty\right\}$ *is defined as*

$$W_1(\mu, \nu) := \sup_{f \in \operatorname{Lip}(\mathbb{X})} \left| \int f d\mu - \int f d\nu \right| \leq \inf_{\mathbf{M} \in \mathcal{C}(\mu,\nu)} \mathbf{E}_{\mathbf{M}}[d(X, Y)]$$

Definition (Relative entropy) *The relative entropy between probability measures* ν *and* μ *on any measurable space is defined as*

$$D(\nu\|\mu) := \begin{cases} \operatorname{Ent}_\mu\left[\frac{d\nu}{d\mu}\right] & \text{if } \nu \ll \mu \\ \infty & \text{otherwise} \end{cases}$$

Theorem (Bobkov-Götze) *Let* $\mu \in \mathcal{P}_1(\mathbb{X})$ *be a probability measure on a metric space* $(\mathbb{X}, d)$. *Then the following are equivalent for* $X \sim \mu$:

1. $f(X)$ *is* σ^2 *-subgaussian for every* $f \in \operatorname{Lip}(\mathbb{X})$

2. $W_1(\nu,\mu) \leq \sqrt{2\sigma^2 D(\nu\|\mu)}$ *for all* ν

Theorem (Marton) *Let* $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ *be a convex function, and let* $w_i : \mathbb{X}_i \times \mathbb{X}_i \to \mathbb{R}_+$ *be positive weight function. Suppose that for* $i = 1, \ldots, n$

$$\inf_{\mathbf{M}\in\mathcal{C}(\mu_i,\nu)} \varphi\left(\mathbf{E}_{\mathbf{M}}\left[w_i(X,Y)\right]\right) \leq 2\sigma^2 D\left(\nu\|\mu_i\right) \quad \textit{for all } \nu$$

Then we have

$$\inf_{\mathbf{M}\in\mathcal{C}(\mu_1\otimes\cdots\otimes\mu_n,\nu)} \sum_{i=1}^{n} \varphi\left(\mathbf{E}_{\mathbf{M}}\left[w_i\left(X_i,Y_i\right)\right]\right) \leq 2\sigma^2 D\left(\nu\|\mu_1\otimes\cdots\otimes\mu_n\right) \quad \textit{for all } \nu$$

6 Talagrand inequality

Theorem (T_1-inequality) *Suppose that the probability measures* μ_i *on* $(\mathbb{X}_i, d_i)$ *satisfy the transportation cost* (T_1) *inequality*

$$W_1\left(\mu_i,\nu\right) \leq \sqrt{2\sigma^2 D\left(\nu||\mu_i\right)} \quad \textit{for all } \nu$$

Then we have

$$W_1\left(\mu_1\otimes\cdots\otimes\mu_n,\nu\right) \leq \sqrt{2\sigma^2 D\left(\nu \mid \mu_1\otimes\cdots\otimes\mu_n\right)} \quad \textit{for all } \nu$$

on $(\mathbb{X}_1\times\cdots\times\mathbb{X}_n, d_c)$, *where* $d_c = \sum_i^n c_i d_i$ *and* $\sum_i^n c_i^2 = 1$

Definition (Quadratic Wasserstein metric) *The quadratic Wasserstein metric for probability measures* μ,ν *on a metric space* $(\mathbb{X}, d)$ *is*

$$W_2(\mu,\nu) := \inf_{\mathbf{M}\in\mathcal{C}(\mu,\nu)} \sqrt{\mathbf{E}\left[d(X,Y)^2\right]}$$

Theorem (T_2-inequality) *Suppose that the probability measures* μ_i *on* $(\mathbb{X}_i, d_i)$ *satisfy the quadratic transportation cost* (T_2) *inequality*

$$W_2\left(\mu_i,\nu\right) \leq \sqrt{2\sigma^2 D\left(\nu||\mu_i\right)} \quad \textit{for all } \nu$$

Then we have

$$W_2\left(\mu_1\otimes\cdots\otimes\mu_n,\nu\right) \leq \sqrt{2\sigma^2 D\left(\nu \mid \mu_1\otimes\cdots\otimes\mu_n\right)} \quad \textit{for all } \nu$$

on $\left(\mathbb{X}_1\times\cdots\times\mathbb{X}_n, \left[\sum_{i=1}^n d_i^2\right]^{1/2}\right)$