Asymptotic Statistical Equivalence for Ergodic Diffusions

The Multi-Dimensional Case

Arnak Dalalyan and Markus Reiß

dalalyan@ccr.jussieu.fr

Université Paris 6, France

The Plan of my Talk

- Le Cam’s distance
- MD diffusions
- Continuous-time and descrete data
- Main result
- Historical remarks

Le Cam's distance 1

Notation: $\mathbb{E} = (E, \mathscr{F}, (\mathrm{P}_\theta)_{\theta\in\Theta})$ is a statistical experiment.
For two experiments $\mathbb{E}_1 = \left(E_1, \mathscr{F}_1, (\mathrm{P}_{1,\theta})_{\theta\in\Theta}\right)$ and $\mathbb{E}_2 = \left(E_2, \mathscr{F}, (\mathrm{P}_{2,\theta})_{\theta\in\Theta}\right)$ denote

$$\begin{array}{ll} X_1 \in E_1 & X_1 \sim \mathrm{P}_{1,\theta}, \\ X_2 \in E_2 & X_2 \sim \mathrm{P}_{2,\theta}. \end{array}$$

Le Cam's distance 2

Definition. The deficiency $\delta(\mathbb{E}_1, \mathbb{E}_2)$ of $\mathbb{E}_1$ with respect to $\mathbb{E}_2$ is the smallest number $\epsilon > 0$ such that

- for any loss function $\mathbf{W}_\theta(\mathbf{z})$ bounded by one,
- for any decision function $\rho_2(x_2, dz)$ in $\mathbb{E}_2$,

$\exists$ a decision function $\rho_1(x_1, dz)$ in $\mathbb{E}_1$ s. t. $\forall\, \theta \in \Theta$,

risk of $\rho_1 \leq$ risk of $\rho_2 + \varepsilon$.

Le Cam's distance between $\mathbb{E}_1$ and $\mathbb{E}_2$ is

$$\Delta(\mathbb{E}_1, \mathbb{E}_2) = \delta(\mathbb{E}_1, \mathbb{E}_2) \vee \delta(\mathbb{E}_2, \mathbb{E}_1).$$

Properties of Le Cam's distance

Let $\mathbb{E}_1$ and $\mathbb{E}_2$ be two statistical experiments.

1. Denote $L_{i,\theta}(x_i) = \frac{d\mathrm{P}_{i,\theta}}{d\mathrm{P}_i}(x_i), \quad i = 1, 2.$ If

   $$\mathscr{L}\left((L_{1,\theta}(X_1))_{\theta\in\Theta}\middle|\mathrm{P}_1\right) = \mathscr{L}\left((L_{2,\theta}(X_2))_{\theta\in\Theta}\middle|\mathrm{P}_2\right),$$

   $\Delta(\mathbb{E}_1, \mathbb{E}_2) = 0$.

2. If $(E_1, \mathscr{F}_1) = (E_2, \mathscr{F}_2)$, then
   $\Delta(\mathbb{E}_1, \mathbb{E}_2) \leq \frac{1}{2} \sup_{\theta\in\Theta} \|\mathrm{P}_{1,\theta} - \mathrm{P}_{2,\theta}\|_{TV}$.

Multi-dim. Diffusions

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space and W a d-dimensional BM, then

$$dX_t = b(X_t)\,dt + dW_t, \quad X_0, \quad t > 0,$$

where X_0 is independent of W, $b \in \Sigma$, Σ is the set of functions $f : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ s.t.

1. f is locally Lipschitz,
2. $|f(x)| \leq M(1 + |x|)$,
3. $(f(x) - f(y))^T (y - x) \geq K|x - y|^2$.

Properties of Diffusions

- If $b \in \Sigma$ then X has a unique invariant probability, let μ_b be its density.
- Let $P_{b,t}$ and $p_{b,t}$ be the transition operator and its density. Then $p_{b,t}(x,y) \leq C(t^{-d/2} + t^{2d}$.
- If $b \in \Sigma$ can be written as $b = -\nabla V$, then the spectral gap inequality holds: $\|P_{b,t}f - \mu_b(f)\|_{\mu_b} \leq e^{-t\rho}\|f\|_{\mu_b}$.
- In this case $\mu_b(x) = G_b^{-1}\exp(-2V(x))$.

Continuous time and Descrete Data

Assume that $b \in \Sigma \subset \mathcal{H}(\beta, L)$ (the set of Hölder continuous functions).

Proposition 1 (Milstein, Nußbaum). *If* $Th_T^{\beta} \longrightarrow 0$ *as* $T \longrightarrow +\infty$*, then the experiments defined by observing*

$$(X_0, X_h, \ldots, X_{nh}) \quad \text{and} \quad X^T = (X_t,\ t \in [0, T])$$

with $n = [T/h]$ *are asymp. equivalent as* $T \longrightarrow \infty$.

From now on, we deal with cont.-time observations X^T.

Main result

Define $\Sigma_0(\epsilon, \eta) \subset \Sigma \cap \mathcal{H}(\beta, L)$ such that

$$|b(x) - b_0(x)| \le \varepsilon \mathbb{1}(|x| \le a), \; |\mu_b - \mu_{b_0}| \le \eta$$

Theorem 1. *If* $(T\varepsilon^2)^{2\beta+d-2}\varepsilon^{-d} \longrightarrow 0$ *and* $T\varepsilon^2\eta \longrightarrow 0$ *as* $T \longrightarrow +\infty$, *then the experiment of ergodic diffusion is asymp. equivalent to the Gaussian shift model*

$$dZ_i(x) = b_i(x)\sqrt{\mu_{b_0}(x)}\, dx + T^{-1/2}\, dB_i(x),$$

where $i = 1, \dots, d$ *and* $B_1, \dots, B_d$ *are independent standard Wiener processes defined on* $\mathbb{R}^d$.

Main steps of the proof

- Discetization in space variable: $\boldsymbol{b} \mapsto \bar{\boldsymbol{b}}$ where $\bar{\boldsymbol{b}}$ is piecewise constant.
- Design modification : the time of visits of a set $\boldsymbol{A}$ should be approximately $\boldsymbol{T\mu_{b_0}(A)}$. Thus,
  - if $\boldsymbol{A}$ is not sufficiently frequently visited we add observations,
  - if $\boldsymbol{A}$ is too frequently visited and $\boldsymbol{X_t \in A}$, we replace $\bar{\boldsymbol{b}}(\boldsymbol{X_t})$ by $\bar{\boldsymbol{b}}_{\mathbf{0}}(\boldsymbol{X_t})$.
- Going back to continuous space variable: $\bar{\boldsymbol{b}} \mapsto \boldsymbol{b}$.

Remarks

1. If $\beta > \frac{d-2}{4} + \frac{1}{2}\sqrt{d^2 + (\frac{d}{2} - 1)^2}$ then the local neighbourhood Σ_0 can be reached by an estimator. Therefore the equivalence result can be globalised.
2. If $b \in \Sigma_0$, then $b(x) = b_0(x)$ for any $|x| > a$. But we show that there is a Markov kernel from diffusion experiment to GWN experiment, which does not depend on $(b_0(x) : |x| > a)$ and realises the asymp. equivalence.
3. We have only one-way Markov kernel : $\text{Diffusion} \mapsto \text{GWN}$.

Historical remarks

- Brown and Low (1996): $X_i = f(i/n) + \xi_i$ as. equiv. $dX_t = f(t)\,dt + \frac{1}{\sqrt{n}}\,dW_t$.
- Nußbaum (1996): i.i.d. with density f as. equiv. $dX_t = \sqrt{f(t)}\,dt + \frac{1}{2\sqrt{n}}\,dW_t$.
- Grama and Nußbaum (1998,2002): GLM is asymptotically equivalent to GWN.
- Brown and Zhang (1998): the asymptotic equivalence fails for smoothness $\beta \leq 1/2$.

Historical remarks 2

- Genon-Catalot, Larédo and Nußbaum (2000): drift est. from hitting times of $dX_t = f(X_t)\,dt + \epsilon\,dW_t$ is as. equiv. to a Poisson experiment with intensity f/ϵ^2.
- Delattre and Hoffmann (2002): Null recurrent diffusion is as. equivalent to a mixed GWN.
- Dalalyan and Reiss (2004): Scalar ergodic diffusion is as. equivalent to a GWN.
- Brown and Zhao (2003): Regression with random design (with unknown density) is locally as. equivalent to a GWN.