Posterior consistency:

$D^n \sim p_n(d^n \mid \theta), \; \theta \in \Theta, \; \theta \sim \Pi_n$.

$$\Pi_n(B \mid D^n) = \frac{\int_B p_n(D^n \mid \theta)\, d\Pi_n(\theta)}{\int_\Theta p_n(D^n \mid \theta)\, d\Pi_n(\theta)}, \quad B \subset \Theta$$

$$p_n(d^n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

Expect: $\Pi_n(\cdot \mid D^n) \xrightarrow{\text{weakly}} \delta_{\theta_0}$.

$\rho_n(\theta) := d_W(\Pi_n(\cdot \mid D^n), \delta_{\theta_0})$

d_W = wasserstein metric

Def. 1: The posterior sequence is consistent at θ_0 if $\rho_n(\theta_0) \to 0$ a.s. or in prob. when $D^n \sim p_n(\cdot \mid \theta_0)$.

Lemma 1: If (Θ, d) is seperable, $\rho_n(\theta_0) \to 0$ a.s. or i.p. iff $\forall$ open neighborhood U of θ_0 $\Pi_n(U^c \mid D^n) \to 0$ a.s./i.p.

The Schwartz theorem:

$D^n = (X_1, \dots, X_n)$; $X_i \overset{iid}{\sim} f, \; f \in \Pi$

$$d_{KL}(p, q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx$$

Test fcn: $\Phi_n : \mathcal{X}^n \to [0,1]$.

$f = f_0$ is true density. P_0^∞ joint density of $(X_1, \dots, X_n)$. f_0 belongs to the KL support of Π if

$$\forall\ \varepsilon > 0,\ \Pi\left(f : d_{KL}(f_0, f) < \varepsilon\right) > 0.$$

Theorem: If f_0 belongs to the KL support of Π & $U_n \subset \mathcal{F}$ are neighborhoods of f_0 s.t. Φ_n with

$$\mathbb{E}_{f_0} \Phi_n \le B e^{-bn} \quad \text{Type I error},$$

$$\sup_{f \in U_n^c} \mathbb{E}_f (1 - \Phi_n) \le B e^{-bn}, \quad \text{Type II error}$$

for $b, B > 0$ then $\Pi(U_n^c \mid D_n) \to 0$ a.s. P_0^∞;

Technical extension (for infinite dim models)

A sequence of compact subsets $\mathcal{F}_n \subset \mathcal{F}$ are called sieves if:

The metric d on $\mathcal{F}$ is bounded from above by d_H. Fix $\varepsilon > 0$, $\mathcal{F}_n \subset \mathcal{F}$ and $\delta > 0$, $0 < \xi < \frac{1}{2}$, $C > 0$ s.t.

1) $\log(N(\delta, \mathcal{F}_n, d)) \leq \xi n \varepsilon^2$

2) $\Pi(\mathcal{F}_n^c) \leq e^{-Cn}$.

What about non-iid, dynamical systems?

There is a variational form of posterior consistency based on large deviation theory.

non-iid models:

Markov: $X_{t+1} = f(X_t; \theta)$

HMM: $X_{t+1} = f(X_t; \theta)$: state process

$Y_{t+1} = g(X_{t+1}; \theta)$: obs. process

Deterministic dynamics - $\theta \in \Theta$ we have $(X, \mathcal{X}, T_\theta, \mu_\theta)$

- X is complete seperable metric space with Borel σ-algebra $\mathcal{X}$
- $T_\theta : X \to X$ is a measurable map
- μ_θ: on $(X, \mathcal{X})$ is T_θ-invariant if $\mu_\theta(T_\theta^{-1} A) = \mu_\theta(A) \ \forall x \in \mathcal{X}$

- the system is ergodic: $T_\theta^{-1} A = A \Rightarrow \mu(A) = \{0, 1\}$

$$(X, \mathcal{X}, T_\theta, \mu_\theta)_{\theta \in \Theta} = (T_\theta, \mu_\theta)_{\theta \in \Theta}$$

Observation process:

$$\int g_\theta(y|x)\, d\nu(y) = 1$$

$$g: \Theta \times X \times Y \to \mathbb{R}_+$$

$$p_\theta(y_0^n | x) = \prod_{k=0}^{n} g_\theta(y_k | T_\theta^k(x))$$

$$p_\theta(y_0^n) = \int p_\theta(y_0^n | x)\, d\mu_\theta(x)$$

$$X_0 \sim U[0,1]$$

$$x_{t+1} = \theta x_t (1 - x_t)$$

$$Y_{t+1} = N(x_{t+1}, \sigma^2)$$

Generalized Baye set-up

a) Observation system $(\mathcal{Y}, T, \nu)$,

$$T: \mathcal{Y} \to \mathcal{Y}$$

b) Tracking systems: $\mathcal{X} := X \times \Theta$,

$$S: \Theta \times X \to X, \quad S_\theta: X \to X$$

c) loss : $\mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$

$$\ell_n(x, y; \theta) := \ell_n(X_0^{n-1}, Y_0^{n-1})$$
$$= \sum_{k=0}^{n-1} \ell(x_k, y_k)$$

$$X_0^{n-1} = (x, S_\theta x, \dots, S_\theta^{n-1}(x))$$
$$Y_0^{n-1} = (y, Ty, \dots T^{n-1} y)$$

Gibbs posterior -

$$\Pi_n(A \mid y) = \frac{\int_A \exp(-\ell_n(x, y; \theta)\, d\pi(x)}{Z_n(y)}$$

$$Z_n(y) = \int_X \exp(-\ell_n(x, y; \theta)\, d\pi(x)$$

(1) is $\lim_{n \to \infty} \Pi_n(\cdot \mid y)$ unique

(2) does it concentrate around T

$$\Pi_n(A|y) = \frac{\int_A \exp(-\psi \, \ell_n(x, y; \theta)) \, d\sigma(x)}{Z_n(y)}$$

Sequence space model

A is a finite set ($|A| = N$), $\Sigma = A^{\mathbb{Z}}$

Left shift: $\sigma : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ + $(\sigma x)_i = x_{i+1}$

$$\Sigma_{\mathcal{F}} = \{ x \in A^{\mathbb{Z}} \mid x_{[i,j]} \neq u \;\; \forall i, j \in \mathbb{Z}, \; u \in \mathcal{F} \}$$

forbids a finite # of words $\mathcal{F}$

Shift of finite type (SFT)

Note: model system to study systems like axiom A

Can encode $\Sigma_{\mathcal{F}}$ by a matrix A:

$$\Sigma_A = \{ (a_i)_{i=-\infty}^{\infty} \in \Sigma_{\mathcal{F}}, \; A_{a_i, a_{i+1}} = 1 \;\; \forall i \in \mathbb{Z} \}$$

a topological Markov chain or 1-step SFT

Σ_A is mixing iff $\exists\, n \geq 1$ s.t. A^n consists of all positive entries.

Gibbs measure:

Let $f : \Sigma_f \to \mathbb{R}$ be continuous.

A measure μ on Σ_F has the Gibbs property for f is there exists $K > 1$ and $P \in \mathbb{R}$ s.t. $\forall$ $x \in A^{\mathbb{Z}}$ and $m \geq 1$

$$K^{-1} \leq \frac{\mu(x[0, m-1])}{\exp\left(-Pm + \sum_{k=0}^{m-1} f(\sigma^k(x))\right)} \leq K.$$

Theorem (Bowen):

If Σ_F is a mixing SFT and $f : \Sigma_F \to \mathbb{R}$ is Hölder continuous then there exists a unique Gibbs measure on Σ_F.

$f : \Sigma_F \to \mathbb{R}$ is called the potential

$P = P(f)$ is its pressure.

Model class:

Let Θ be a compact metric space.

Let $\{f_\theta : \theta \in \Theta\}$ be a continuous parametrized family of Hölder continuous potential fcn's.

Let $\{\mu_\theta : \theta \in \Theta\}$ be the corresponding family of Gibbs measures.

Markov chains are an example.

Observation process: Let λ be a Borel measure on $\mathcal{Y}$

Let $g : \Theta \times \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ be a measurable fcn. s.t. $\forall\ \theta \in \Theta$ and $x \in \mathcal{X}$

$$\int g(\theta, x, y)\, \lambda(dy) = 1$$

$$g_\theta(\cdot \mid x) = g(\theta, x, \cdot)$$

Hidden Gibbs process:

$$p_\theta(y_0^{n-1}) = \int \prod_{k=0}^{n-1} g_\theta(y_k \mid \sigma^k(x))\, \mu_\theta(dx)$$

$$X_0 \sim \mu_\theta, \quad X_{n+1} = \sigma(X_n), \quad Y_n \sim g_\theta(y \mid X_n)\, \lambda(dy)$$

$\mathbb{P}_\theta^Y$ denote the distribution of $\{Y_n\}_{n\geq 0}$ under θ.

For $\theta \in \Theta$, let $[\theta] = \{\theta' \in \Theta : \mathbb{P}_\theta^Y = \mathbb{P}_{\theta'}^Y\}$

Theorem: Suppose π is fully supported on Θ, let $\theta_0 \in \Theta$. Then for any neighborhood U of $[\theta_0]$

$$\Pi_n(\Theta \setminus U \mid Y_0^{n-1}) \to 0, \quad P_{\theta_0}^Y\text{-a.s.}$$

$$\Pi_n(A \mid y_0^{n-1}) = \frac{\int_A \exp(-\ell(\theta, x; y_0^{n-1}))\, P_0(d\theta, dx)}{Z_n(y_0^{n-1})}$$

Questions:

1. What is
$$\lim_n \frac{1}{n} \log Z_n(Y_0^{n-1})$$

2. Convergence of $\{\Pi_n\}$

Coupling: A coupling of two r.v.'s X, X' taking values in $(E, \mathcal{E})$ is any pair of r.v.'s (Y, Y') in $(E \times E, \mathcal{E} \times \mathcal{E})$ with $X \overset{D}{=} Y$, $X' \overset{D}{=} Y'$

Joining: Let (X, A, μ, T) and (Y, B, ν, S) be two dynamical systems. A joining of S, T is a prob. measure λ on $X \times Y$ with marginals μ, ν, and invariant to the product map $T \times S$.

A stationary $\mathcal{X}$-valued process $\{X_n\}_{n \geq 0}$ is in $\mathcal{P}(\mathcal{X}, \sigma)$ if

$$X_{n+1} = \sigma(X_n), \quad \forall n \text{ w.p. } 1.$$

A joining of $(\mathcal{X}, \sigma)$ with $\{Y_n\}_{n \geq 0}$ is a stationary bi-variate process $(U, V) = \{(U_n, V_n)\}_{n \geq 0}$ on $\mathcal{X} \times \mathcal{Y}$ s.t.

$U = \{U_n\}_{n \geq 0}$ is in $\mathcal{P}(\mathcal{X}, \sigma)$

$V = \{U_n\}_{n \geq 0} \overset{d}{=} \{Y_n\}_{n \geq 0}$

We denote the set of joinings of $(\mathcal{X}, \sigma)$ with $\{Y_n\}$ as $\mathcal{J}$

Theorem. There exists a lower semicontin fcn. $\phi : \Theta \to \mathbb{R}$ s.t. w.p. 1

$$\lim_n -\frac{1}{n} \log Z_n(y) = \inf_{\theta \in \Theta} \phi(\theta)$$

The r.h.s. is the rate fcn. in the large deviation sense.

Limiting average cost:

$$\lim_{n \to \infty} \frac{1}{n} \int_{\mathcal{X}} \ell_n(x, y) \, d\lambda_y(x) = \int \ell \, d\lambda$$

Entropy: measures π, μ on $\mathcal{X}$, finite measurable partition ξ

$$\mu \ll_\xi \pi \Rightarrow \pi(C) = 0 \Rightarrow \mu(C) = 0 \text{ for } C \in \xi$$

$$L(\mu \| \pi, \xi) = \begin{cases} \sum_{C \in \xi} \mu(C) \log \pi(C) \\ -\infty \end{cases}$$

$$F(\mu, \pi) = \sup_\xi L(\mu \| \pi, \xi)$$

Theorem: Given a Gibbs prior

$$\lim_{n\to\infty} -\frac{1}{n}\log Z_n(y) = \inf_{\lambda\in\mathcal{J}} \left\{ \ell\, d\lambda + F[\lambda, \mu_0] \right\}$$

recall

$$\Pi^n(\theta|x) = \arg\min_{\mu} \left\{ \int_\Theta \ell(\theta, x)\, d\mu(\theta) + d_{KL}(\mu, \pi) \right\}$$

Prop.: Suppose a Gibbs prior and pressure

$$P = \inf_{\lambda\in\mathcal{J}} \left\{ \int \ell\, d\lambda + F(\lambda, \mu_0) \right\}$$

$$\Theta_* = \arg\min_{\theta\in\Theta} P$$

For all $\varepsilon > 0$

$$P(d(s_\theta, T) < \varepsilon) \to 1 \text{ a.s. } n\to\infty$$

Large deviations approach:

Two conditions need to hold for post. consistency

1) Conditional large deviations for one process on $\mathcal{X} \times \mathcal{Y}$

2) Prove exponential continuity over the map $\theta \mapsto \mu_\theta$; allows us to extend the LDP for one family of all of Θ.

Given a Polish space Z and a lower semicontinuous fcn. $I : Z \to [0, \infty]$.
A family $(\eta_t)_{t \in T}$ of prob measures satisfies the large deviations principle with rate fcn I if for every closed set $E \in Z$

$$\limsup_{t \to \infty} \frac{1}{t} \log \eta_t(E) \leq \inf_{z \in E} I(z)$$

and open set $U \subset Z$

$$\liminf_{t \to \infty} \frac{1}{t} \log \eta_t(V) \geq -\inf_{z \in U} I(z).$$

Step 4: For a fixed ϑ show

$$\lim_{t\to\infty} \frac{1}{t} \log Z_t^\vartheta(y) = \inf_{\lambda \in J(S;\nu)} \left\{ \int \ell \, d\lambda + F(\lambda, \pi) \right\}$$

$$= -V(\vartheta).$$

Exponential continuity:

$\{\mu_\vartheta\}_\vartheta$ is exponentially continuous w.r.t. ℓ if the following holds.

For all $\vartheta \in \Theta$ it holds for ν-a.e. $y \in \mathcal{Y}$ the following limit exists

$$\lim_{t\to\infty} \frac{1}{t} \log \int_X \exp[-\ell_\vartheta^t(x,y)] \, d\mu_\vartheta(x)$$

$$=: -V(\vartheta)$$

and if $(\theta_t)_{t\in T}$ is a family of parameters s.t. $\vartheta_t \to \vartheta$ in Θ then

$$\lim_{t\to\infty} \log \int_\chi \exp(-\ell_\vartheta^t(x,y)) \, d\mu_{\theta_t}(x) = -V(\vartheta)$$

Prop. $\{\mu_\theta\}_{\theta\in\Theta}$ is an exponentially continuous family w.r.t. ℓ and π is a Borel measure on Θ. For ν-almost every $y \in \mathcal{Y}$

$$\lim_{t\to\infty} -\frac{1}{t}\log Z_t^{\pi}(y) = \inf_{\theta\in\operatorname{supp}(\pi)} V(\theta).$$