Tests and model choice : asymptotics

J. Rousseau

CEREMADE, Université Paris-Dauphine & Oxford University (soon)

Greek Stochastics, Milos

Outline

Outline

1. Bayesian testing : the Bayes factor
   - Bayes factor
   - Consistency of the Bayes factor or of the model selection :
2. Non parametric tests
   - Parametric vs Nonparametric
   - Construction of nonparametric priors
   - A special case : Point null hypothesis
   - Parametric null hypothesis
3. General conditions for consistency of $B_{1/2}$ in non iid case for GOF types
4. Nonparametric/Nonparametric case
5. Conclusion

Bayes factor

- **Model** $X :\sim f(|\theta),\ \theta \in \Theta\,;\ \theta \sim \Pi$
- **Testing problem**

$$H_0 : \theta \in \Theta_0, \quad H_1 : \theta \in \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset \quad (\text{ or } \Pi[\Theta_0 \cap \Theta_1] = 0)$$

- **0-1 loss function** Bayes estimate :

$$\delta^\pi = 1 \quad \text{iff } \Pi[\Theta_1|X] > \Pi[\Theta_0|X]$$

$$\Downarrow \quad \text{Posterior odds ratio}$$

$$\frac{\Pi[\Theta_0|X]}{\Pi[\Theta_1|X]} = \underbrace{\frac{\int_{\Theta_0} f(X|\theta)d\Pi_0(\theta)}{\int_{\Theta_1} f(X|\theta)d\Pi_1(\theta)}}_{\frac{m_0(X)}{m_1(X)}} \times \frac{\pi[\Theta_0]}{\pi[\Theta_1]}$$

$$BF_{0/1} := \frac{m_0(X)}{m_1(X)}, \quad d\Pi_j(\theta) = \mathbb{1}_{\Theta_j} \frac{d\Pi(\theta)}{\Pi(\Theta_j)}$$

if $BF_{0/1}$ large $\rightarrow H_0$ else $\rightarrow H_1$

Model choice

- **Multiple models** $\mathfrak{M}_j : f_{j,\theta_j}, \theta_j \in \Theta_j$
- **Hierarchical Prior**

$$d\Pi(j, \theta_j) = \Pi(j) d\Pi_j(\theta_j)$$

- **Posterior**

$$d\Pi(j, \theta_j | X) \propto \Pi(j) d\Pi_j(\theta_j) f_{j,\theta_j}(X), \quad d\Pi(j|X) \propto \Pi(j) \underbrace{\int_{\Theta_j} f_{j,\theta_j}(X) d\Pi_j(\theta_j)}_{m_j(X)}$$

- **0-1 loss function** Bayes estimate

$$\delta^\pi(X) = \text{argmax}\{\Pi(\mathfrak{M}_j | X) \propto m_j(X) \times \Pi(\mathfrak{M}_j), j \in \mathcal{J}\}$$

Outline

Consistency of the Bayes factor $BF_{0/1} = \frac{m_0(X)}{m_1(X)}$

Definition 1 The Bayes factor is said *consistent* IFF

$$\begin{aligned} BF_{0/1} &\to +\infty, \quad \text{in proba when} \quad X \sim f_\theta \in \mathcal{H}_0 \\ BF_{0/1} &\to 0, \quad \text{in proba when} \quad X \sim f_\theta \in \mathcal{H}_1 \end{aligned}$$

Interpretation : As n goes to infinity you are bound to make the correct decision

Consistency of the model selection

- **true model** True parameter $\theta^* \in \cup_j \Theta_j$

$$\mathfrak{M}^* = \min\{\Theta_j; \theta^* \in \Theta_j\}, \quad \Theta_j \leqslant \Theta_{j'} \quad \Leftrightarrow \Theta_j \subset \Theta_{j'}$$

e.g. Mixtures, Variable selections, graph selection etc. . .

- **Consistency** δ^π is consistent iff

$$P_{\theta^*}\left(\delta^\pi \neq \mathfrak{M}^*\right) \rightarrow 0$$

Whether BF or model selection procedure : based on the asymptotic behaviour of the marginal likelihoods $m_j(X)$

BIC approximation of the marginal likelihood

- Bayes factor : marginal likelihood ratio : acts like a penalized likelihood ratio
  BIC approximation in **finite dim & regular** models
  $X = X^n = (X_1, ..., X_n)$

$$\log m_j(X^n) = \log f(X^n|\hat{\theta}_j) - \frac{d_j}{2}\log n + O_p(1)$$

$$d_j = dim(\Theta_j), \quad \hat{\theta}_j = \text{ MLE under } \Theta_j.$$

$\rightarrow$ Different penalization if **non regular or infinite dim**

How does BIC work ?

$$\ell_n(\theta; \mathfrak{M}) = \log f(X^n|\theta; \mathfrak{M}).$$

$$m_j(X^n) = e^{\ell_n(\hat{\theta}_j, \mathfrak{M}_j)} \int_{\Theta_j} e^{\ell_n(\theta; \mathfrak{M}_j) - \ell_n(\hat{\theta}_j; \mathfrak{M}_j)} \pi_j(\theta) d\theta, \quad \hat{\theta}_j = \text{mle in } \Theta_j$$

- **Local asymptotic normality+ consistency**

$$\ell_n(\theta) - \ell_n(\hat{\theta}_j) = \frac{-n(\theta - \hat{\theta}_j)^t I(\hat{\theta}_j)(\theta - \hat{\theta}_j)}{2}(1 + o_p(1))$$

$$\Pi_j(\|\theta - \hat{\theta}_j\| > \epsilon | X^n) = o_p(1)$$

- **positive, continuous and finite prior density**

$$0 < \inf_{\|\theta - \hat{\theta}_j\| < \epsilon} \pi_j(\theta) \leqslant \sup_{\|\theta - \hat{\theta}_j\| < \epsilon} \pi_j(\theta) < +\infty$$

Extension to irregular models : a more general approach

True parameter θ^*, $\Theta_j \subset \mathbb{R}^{d_j}$,

$$m_j(X^n) = e^{\ell_n(\theta_j^*)} \int_{\Theta_j} e^{\ell_n(\theta)-\ell_n(\theta_j^*)} \pi_j(\theta) d\theta, \quad \theta_j^* = \text{KL - projtn onto } \Theta_j$$

- **Deterministic approximation**

$$\ell_n(\theta) - \ell_n(\theta_j^*) \approx -nH(\theta_j^*, \theta), \quad H(\theta_j^*, \theta) = \frac{E_{\theta^*}\left(\ell_n(\theta_j^*) - \ell_n(\theta)\right)}{n}$$

$$\log m_j(X^n) = \underbrace{\ell_n(\theta_j^*)}_{\text{model bias}} + \underbrace{\log \Pi_j(H(\theta_j^*, \theta) \lesssim 1/n)}_{\text{prior penalty}} + o_p(\log n)$$

Understanding the prior penalty $\Pi_j(H(\theta_j^*, \theta) \lesssim 1/n)$

- **Regular models**

$$H(\theta_j^*, \theta) = \frac{(\theta - \theta_j^*)^t I(\theta_j^*)(\theta - \theta_j^*)}{2}(1 + o(1)) \quad \Rightarrow$$

$$\Pi_j(H(\theta_j^*, \theta) \lesssim 1/n) \approx \pi_j(\theta_j^*)(C/n)^{d_j/2} \quad \Rightarrow -\frac{d_j}{2}\log n + O_p(1)$$

- **Irregular models**

$$H(\theta_j^*, \theta) \neq \frac{(\theta - \theta_j^*)^t I(\theta_j^*)(\theta - \theta_j^*)}{2}(1 + o(1))$$

[We need to understand the geometry associated to $H(\theta_j^*, \theta)$.]

Example of mixture models : overspecified case

$$f_\theta(x) = \sum_{j=1}^{k} p_j g_{\gamma_j}(x), \quad \gamma_j \in \Gamma, \quad \sum_{j=1}^{k} p_j = 1$$

- **If true distribution**

$$f_{\theta^*}(x) = \sum_{j=1}^{k^*} p_j^* g_{\gamma_j^*}(x), \quad k^* < k$$

[Then the model is non identifyable]

$k = 2$ and $k^* = 1$

$$f_\theta(x) = p_1 g_{\gamma_1} + (1 - p_1) g_{\gamma_2}, \quad f^*(x) = g_{\gamma^*}$$

Then $f_\theta = f_{\theta^*}$ IFF

$$p_1 = 1, \quad \gamma_1 = \gamma^* \quad \textit{or} \quad \gamma_1 = \gamma_2 = \gamma^*$$

$\Theta^* = \{\theta : f_\theta = f_{\theta^*}\} \neq \{\theta^*\}$ even up to a permutation

Understanding $H(\theta_j^*, \theta)$ in mixtures $f_{\theta^*}(x) = \sum_{j=1}^{k^*} p_j^* g_{\gamma_j^*}(x)$, $\mathfrak{M}_k = \{f_\theta(x) = \sum_{j=1}^{k} p_j g_{\gamma_j}(x), \gamma_j \in \Gamma\}$

$$k^* < k \quad \& \quad \theta_k^* = \theta^* \quad H(\theta^*, \theta)?$$

- **varies with the configuration** $\mathrm{t} = (t_1, \cdots, t_{k^*}, t_{k^*+1})$

$\mathrm{t} = (t_j, j \leqslant k^* + 1)$ is the partition of $\{1, \cdots, k\}$

$$t_j = \{\ell; \|\gamma_\ell - \gamma_j^*\| \leqslant \epsilon\}, \quad j \leqslant k^*; \quad t_{k^*+1} = \{\ell; \min_{j \leqslant k^*} \|\theta_\ell - \theta_j^*\| > \epsilon\}$$

On t ; $p(j) = \sum_{\ell \in t_j} p_\ell$, $q_\ell = p_\ell / p(j)$ if $\ell \in t_j$

$$\sqrt{H(\theta^*, \theta)} \asymp \sum_{j=1}^{k^*} \left[\| \sum_{\ell \in t_j} q_\ell \gamma_\ell - \gamma_j^* \| + |p(j) - p_j^*| + \sum_{\ell \in t_j} \frac{p_\ell}{p(j)} \|\gamma_\ell - \gamma_j^*\|^2 \right] + \sum_{\ell \in t_{k^*+1}} p_\ell$$

$\Pi_k(H(\theta^*,\theta) \lesssim 1/n)$

- **prior**

$$(p_1,\cdots,p_k) \sim \mathcal{D}(\alpha,\cdots,\alpha), \quad \gamma_j \overset{iid}{\sim} \pi_\gamma$$

- **If** $\alpha < d/2$

$$\Pi_k(H(\theta^*,\theta) \lesssim 1/n) \approx n^{-(k^*d+k^*-1+(k-k^*)\alpha)/2}$$

[(prior) sparsest configuration : emptying the extra states]

- **If** $\alpha > d/2$

$$\Pi_k(H(\theta^*,\theta) \lesssim 1/n) \approx n^{-(k^*d+k^*-1+(k-k^*)d/2)/2}$$

[(prior) sparsest configuration : merging the extra states]

Prior penalty $D < dim(\Theta_k)$

Impact on consistency

$$BF_{k/k+1} = \frac{\int_{\Theta_k} f_\theta(X^n)\pi_k(\theta)d\theta}{\int_{\Theta_{k+1}} f_\theta(X^n)\pi_{k+1}(\theta)d\theta}$$

- **If** $\theta^* \in \Theta_{k+1} \setminus \Theta_k$ (Regular models)

$$\log m_k(X^n) = \ell_n(\theta_k^*) - \ell_n(\theta^*) + O_p(\log n) \approx -nKL(\theta^*, \theta_k^*) \to -\infty$$

- **If** $\theta^* \in \Theta_k \subset \Theta_{k+1}$ irregular model

$$D^* = \dim(\Theta_k) + \alpha, \quad \text{if } \alpha < d/2 \quad \& \quad D^* = \dim(\Theta_k) + d/2 \text{if } \alpha > d/2$$

Then

$$\log BF_{k/k+1} = -\frac{\dim(\Theta_k) - D^*}{2} \log n + o_p(\log n) \to +\infty$$

Nonparametric cases : At least one of the models is NP

- **Case 1 :** Parametric vs Non parametric :

$$\mathfrak{M}_0 : \theta_0 \in \Theta \subset \mathbb{R}^d, \quad \mathfrak{M}_1 : \theta_1 \in \Theta_1, \quad \dim(\Theta_1) = +\infty$$

e.g. Goodness of fit tests ; testing for linearity in regression etc . . .

- **Case 2 :** Nonparametric vs Nonparametric

$$\mathfrak{M}_0 : \theta_0 \in \Theta_0, \quad \mathfrak{M}_1 : \theta_1 \in \Theta_1, \quad \dim(\Theta_1), \dim(\Theta_0) = +\infty$$

e.g. Testing for monotonicity, independence, two sample tests etc. . .

- **Case 3 :** High dimensional model selection

$$\mathfrak{M}_k : \theta \in \Theta_k \subset \mathbb{R}^{d_k}, \quad k \in \mathcal{K}, \quad \text{card}(\mathcal{K})\text{large}$$

e.g. regression models (variable selection)

Outline

Parametric vs NP

$$X^n = (X_1, ..., X_n) \sim f_\theta^n,$$

- **Test problem** : $\Theta_0 \subset \mathbb{R}^d$, $\dim(\Theta_1) = +\infty$

$$H_0 : f_\theta^n \in \mathcal{F}_0 = \{f_\theta^n, \theta \in \Theta_0\}, \quad H_1 : f_\theta^n \in \mathcal{F}_1 = \{f_\theta^n, \theta \in \Theta_1\}$$

Examples :

- Goodness of fit in iid cases : $X_i \overset{iid}{\sim} f$,

$$\mathcal{F}_0 = \{f_\theta, \theta \in \Theta_0\}, \quad \mathcal{F}_1 = \mathcal{F}_0^c.$$

- Linear/partially linear

$$y_i = \alpha + \beta^t w_i + f(x_i) + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$H_0 : f = 0, \quad H_1 : f \neq 0$$

Parametric vs NP

$$X^n = (X_1, ..., X_n) \sim f_\theta^n,$$

- **Test problem** : $\Theta_0 \subset \mathbb{R}^d$, $\dim(\Theta_1) = +\infty$

$$H_0 : f_\theta^n \in \mathcal{F}_0 = \{f_\theta^n, \theta \in \Theta_0\}, \quad H_1 : f_\theta^n \in \mathcal{F}_1 = \{f_\theta^n, \theta \in \Theta_1\}$$

Examples :

- Goodness of fit in iid cases : $X_i \overset{iid}{\sim} f$,

$$\mathcal{F}_0 = \{f_\theta, \theta \in \Theta_0\}, \quad \mathcal{F}_1 = \mathcal{F}_0^c.$$

- Linear/partially linear

$$y_i = \alpha + \beta^t \boldsymbol{w}_i + f(x_i) + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$H_0 : f = 0, \quad H_1 : f \neq 0$$

Outline

Non parametric priors on $\mathcal{F}_1$

$$H_0 : f \in \mathcal{F}_0 = \{f_\theta, \theta \in \Theta\}, \quad H_1 : f \in \mathcal{F}_1 \quad \Theta \subset \mathbb{R}^d$$

- **Parameter space** $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$
- **Non parametric prior** π_1 **on** $\mathcal{F}_1$ Prior on H_1
  - Your favourite nonparametric prior *e.g. Gaussian processes, Levy processes, Mixtures*
  - Check that $\Pi_1[\mathcal{F}_0] = 0$, e.g.

$$\mathcal{F}_0 = \{\mathcal{N}(\mu, \sigma^2), (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+\}, \quad \Pi_1 : \sum_{j=1}^{k} p_j \mathcal{N}(\mu_j, \sigma_j^2), \quad k > 1$$

  - Should we care for more ? Yes : consistency

Non parametric priors on $\mathcal{F}_1$

$$H_0 : f \in \mathcal{F}_0 = \{f_\theta, \theta \in \Theta\}, \quad H_1 : f \in \mathcal{F}_1 \quad \Theta \subset \mathbb{R}^d$$

- **Parameter space** $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$
- **Non parametric prior** π_1 **on** $\mathcal{F}_1$ Prior on H_1
  - Your favourite nonparametric prior *e.g. Gaussian processes, Levy processes, Mixtures*
  - Check that $\Pi_1[\mathcal{F}_0] = 0$, e.g.

$$\mathcal{F}_0 = \{\mathcal{N}(\mu, \sigma^2), (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+\}, \quad \Pi_1 : \sum_{j=1}^{k} p_j \mathcal{N}(\mu_j, \sigma_j^2), \quad k > 1$$

  - Should we care for more ? Yes : consistency

Non parametric priors on $\mathcal{F}_1$

$$H_0 : f \in \mathcal{F}_0 = \{f_\theta, \theta \in \Theta\}, \quad H_1 : f \in \mathcal{F}_1 \quad \Theta \subset \mathbb{R}^d$$

- **Parameter space** $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$
- **Non parametric prior** π_1 **on** $\mathcal{F}_1$ Prior on H_1
  - Your favourite nonparametric prior *e.g. Gaussian processes, Levy processes, Mixtures*
  - Check that $\Pi_1[\mathcal{F}_0] = 0$, e.g.

$$\mathcal{F}_0 = \{\mathcal{N}(\mu, \sigma^2), (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+\}, \quad \Pi_1 : \sum_{j=1}^{k} p_j \mathcal{N}(\mu_j, \sigma_j^2), \quad k > 1$$

  - Should we care for more ? Yes : consistency

Outline

Case of point null hypothesis (case of iid observations)

$$H_0 : f = f_0, \quad H_1 : f \neq f_0$$

- **Prior on** $\mathcal{F}_1 = \mathcal{F} - \{f_0\}$
  - Corresponds to a prior on $\mathcal{F}$

$$\Pi(f) = p_0 \delta_{(f_0)} + (1 - p_0)\Pi_1(f), \quad \Pi_1 = \text{proba on } \mathcal{F} \quad \Pi_1(\{f_0\}) = 0$$

  - Dass & Lee :
    - $BF_{0/1}$ always consistent under f_0.
    - if $d\Pi_1(.|X^n)$ is consistent at $f \in \mathcal{F} - \{f_0\}$, then the $BF_{0,1}$ is consistent under f.

Case of point null hypothesis (case of iid observations)

$$H_0 : f = f_0, \quad H_1 : f \neq f_0$$

- **Prior on** $\mathcal{F}_1 = \mathcal{F} - \{f_0\}$
  - Corresponds to a prior on $\mathcal{F}$

    $$\Pi(f) = p_0 \delta_{(f_0)} + (1 - p_0)\Pi_1(f), \quad \Pi_1 = \text{proba on } \mathcal{F} \quad \Pi_1(\{f_0\}) = 0$$

  - Dass & Lee :
    • $BF_{0/1}$ always consistent under f_0.
    • if $d\Pi_1(.|X^n)$ is consistent at $f \in \mathcal{F} - \{f_0\}$, then the $BF_{0,1}$ is consistent under f.

Remark on Dass & Lee

- **Convergence under f_0**
Comes from Doob theorem because $\Pi[\{f_0\}] = p_0 > 0$

$$\Pi(f) = p_0 \delta_{(f_0)} + (1 - p_0)\Pi_1(f)$$

- **Harder to detect H_0 than H_1 :**
$f_0 \in \mathcal{F}, \rightarrow$ can be approximated under $\mathcal{F}_1$

$$\Downarrow \quad \text{typically}$$

$$H_1 : BF_{0/1} \lesssim e^{-nc(f)} \qquad H_0 : BF_{0/1} \lesssim n^{-q_0}$$

On posterior concentration

- **Posterior concentration** $\epsilon_n = o(1)$ Prove that

$$E_{\theta^*}\left[\Pi\left(d(\theta,\theta^*) > M\epsilon_n | Y^n\right)\right] = o(1), \quad \ell_n(\theta) = \log f_\theta(Y^n)$$

- **Conditions**

1. KL condition

$$\pi\left[KL_n(\theta^*,\theta) \leqslant n\epsilon_n^2, V_n(\theta^*,\theta) \leqslant \kappa n\epsilon_n^2\right] \gtrsim e^{-nc\epsilon_n^2}$$

$$KL_n(\theta^*,\theta) = E_{\theta^*}(\ell_n(\theta^*) - \ell_n(\theta)), \quad V_n(\theta^*,\theta) = V_{\theta^*}(\ell_n(\theta^*) - \ell_n(\theta))$$

2. Testing condition $\exists \Theta_n \subset \Theta : \pi(\Theta_n^c) \leqslant e^{-n(c+2)\epsilon_n^2}$ and $\phi_n \in [0,1]$

$$E_{\theta^*}(\phi_n) = o(1), \quad \sup_{d(\theta^*,\theta) > M\epsilon_n, \theta\in\Theta_n} E_\theta(1-\phi_n) \leqslant e^{-n(c+2)\epsilon_n^2}$$

On posterior concentration

- **Posterior concentration** $\epsilon_n = o(1)$ Prove that

$$E_{\theta^*}\left[\Pi\left(d(\theta,\theta^*) > M\epsilon_n | Y^n\right)\right] = o(1), \quad \ell_n(\theta) = \log f_\theta(Y^n)$$

- **Conditions**

1. KL condition

$$\pi\left[KL_n(\theta^*,\theta) \leqslant n\epsilon_n^2, V_n(\theta^*,\theta) \leqslant \kappa n\epsilon_n^2\right] \gtrsim e^{-nc\epsilon_n^2}$$

$$KL_n(\theta^*,\theta) = E_{\theta^*}(\ell_n(\theta^*) - \ell_n(\theta)), \quad V_n(\theta^*,\theta) = V_{\theta^*}(\ell_n(\theta^*) - \ell_n(\theta))$$

2. Testing condition $\exists \Theta_n \subset \Theta : \pi(\Theta_n^c) \leqslant e^{-n(c+2)\epsilon_n^2}$ and $\phi_n \in [0,1]$

$$E_{\theta^*}(\phi_n) = o(1), \quad \sup_{d(\theta^*,\theta) > M\epsilon_n, \theta \in \Theta_n} E_\theta(1-\phi_n) \leqslant e^{-n(c+2)\epsilon_n^2}$$

Proof

$$\Pi\left(d(\theta,\theta^*) > \epsilon_n | Y^n\right) = \frac{\int_{B_n} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)}{\int_{\Theta} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)} := \frac{N_n}{D_n}$$

$$E_{\theta^*}\left[\Pi\left(d(\theta,\theta^*) > \epsilon_n | Y^n\right)\right] = E_{\theta^*}\left[\phi_n \times \frac{N_n}{D_n}\right] + E_{\theta^*}\left[(1-\phi_n) \times \frac{N_n}{D_n}\right]$$

$$\leqslant E_{\theta^*}[\phi_n] + P_{\theta^*}\left[D_n < e^{-n(c+3/2)\epsilon_n^2}\right] + e^{n(c+3/2)\epsilon_n^2} E_{\theta^*}\left(N_n(1-\phi_n)\right)$$

$$= E_{\theta^*}[\phi_n] + P_{\theta^*}\left[D_n < e^{-n(c+3/2)\epsilon_n^2}\right] + e^{n(c+3/2)\epsilon_n^2} \int_{B_n} E_{\theta}(1-\phi_n) d\pi(\theta$$

$$= o(1) + P_{\theta^*}\left[D_n < e^{-n(c+2)\epsilon_n^2}\right] + e^{n(c+3/2)\epsilon_n^2}\left(e^{-n(c+2)\epsilon_n^2} + \Pi(\Theta_n^c)\right)$$

$$= o(1) + P_{\theta^*}\left[D_n < e^{-n(c+2)\epsilon_n^2}\right]$$

Control of $P_{\theta^*}\left[D_n < e^{-n(c+2)\epsilon_n^2}\right]$: KL condition

$$
\begin{aligned}
D_n &= \int_\Theta e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta) \geqslant \int_{S_n} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta) \\
&\geqslant \int_{S_n} \mathbb{I}_{\ell_n(\theta)-\ell_n(\theta^*)>-3/2n\epsilon_n^2} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta) \\
&\geqslant e^{-3/2n\epsilon_n^2} \pi(S_n \cap \{\ell_n(\theta)-\ell_n(\theta^*) > -3/2n\epsilon_n^2\}) \\
&= e^{-3/2n\epsilon_n^2} \left(\pi(S_n) - \pi(S_n \cap \{\ell_n(\theta)-\ell_n(\theta^*) \leqslant -3/2n\epsilon_n^2\})\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&P_{\theta^*}\left(\pi(S_n \cap \{\ell_n(\theta)-\ell_n(\theta^*) \leqslant -3/2n\epsilon_n^2\}) > \pi(S_n)/2\right) \\
&\leqslant \frac{2}{\pi(S_n)} \int_{S_n} \underbrace{P_{\theta^*}\left(\ell_n(\theta)-\ell_n(\theta^*) \leqslant -3/2n\epsilon_n^2\right)}_{\text{KL condition:} \lesssim \frac{1}{n\epsilon_n^2}} d\pi(\theta) \lesssim \frac{1}{n\epsilon_n^2}
\end{aligned}
$$

What are these tests ?

- **The tests ϕ_n depend on the metric $d(\theta,\theta')$**
  - **Case of iid data with weak topology** $U = \{f; \left|\int \varphi_j f(x)dx - \int \varphi_j f_0(x)dx\right| \leqslant \epsilon, \quad j = 1, \cdots, J\}$:
    Tests always exist
  - **L_1 metric :** $d(f,f') = \int |f(x) - f'(x)|dx$

$$\phi_n = \max \phi_{n,f_\ell}, \quad \phi_{n,f_\ell} = \mathbb{I}_{\sum_{i=1}^n [\mathbb{I}_{y_i \in A(f_\ell,f_0)} - P_{f_0}(A(f_\ell,f_0))] > n\delta\|f_\ell - f_0\|_1}$$

with

$$A(f_\ell, f_0) = \{y;\ f_\ell(y) > f_0(y)\}$$

[existence of tests $\leftrightarrow$ Bound on the L_1 covering number $N(.) \leqslant e^{nc\epsilon_n^2}$]

What are these tests ?

- **The tests ϕ_n depend on the metric $d(\theta, \theta')$**
  - **Case of iid data with weak topology**
    $U = \{f; \left|\int \varphi_j f(x)dx - \int \varphi_j f_0(x)dx\right| \leqslant \epsilon, \quad j = 1, \cdots, J\}$:
    Tests always exist
  - **L_1 metric :** $d(f, f') = \int |f(x) - f'(x)|dx$

$$\phi_n = \max \phi_{n,f_\ell}, \quad \phi_{n,f_\ell} = \mathbb{I}_{\sum_{i=1}^n [\mathbb{I}_{y_i \in A(f_\ell, f_0)} - P_{f_0}(A(f_\ell, f_0))] > n\delta \|f_\ell - f_0\|_1}$$

with

$$A(f_\ell, f_0) = \{y; \ f_\ell(y) > f_0(y)\}$$

[existence of tests $\leftrightarrow$ Bound on the L_1 covering number $N(.) \leqslant e^{nc\epsilon_n^2}$]

- **Parametric**

$$\log m_j(Y^n) = \ell_n(\theta_j^*) - \underbrace{\frac{D_j}{2}\log n}_{\log \pi_j(d(\theta_j^*,\theta)\leqslant 1/n)} + o_p(\log n)$$

- Regular model $D_j = \dim(\Theta_j)$
- Irregular model $D_j = D_j(\theta^*) \neq \dim(\Theta_j)$
- Important for convergence : if $\theta^* \in \Theta_j \subset \Theta_{j+1}$

$$D_j(\theta^*) < D_{j+1}(\theta^*)$$

Summary of the first day : $BF_{0/1} = \frac{m_0(Y^n)}{m_1(Y^n)}$

- **Parametric**

$$\log m_j(Y^n) = \ell_n(\theta_j^*) - \underbrace{\frac{D_j}{2}\log n}_{\log \pi_j(d(\theta_j^*,\theta)\leqslant 1/n)} + o_p(\log n)$$

- Regular model $D_j = \dim(\Theta_j)$
- Irregular model $D_j = D_j(\theta^*) \neq \dim(\Theta_j)$
- Important for convergence : if $\theta^* \in \Theta_j \subset \Theta_{j+1}$

$$D_j(\theta^*) < D_{j+1}(\theta^*)$$

Summary of the first day : $BF_{0/1} = \frac{m_0(Y^n)}{m_1(Y^n)}$

- **Parametric**

$$\log m_j(Y^n) = \ell_n(\theta_j^*) - \underbrace{\frac{D_j}{2}\log n}_{\log \pi_j(d(\theta_j^*,\theta)\leqslant 1/n)} + o_p(\log n)$$

- Regular model $D_j = \dim(\Theta_j)$
- Irregular model $D_j = D_j(\theta^*) \neq \dim(\Theta_j)$
- Important for convergence : if $\theta^* \in \Theta_j \subset \Theta_{j+1}$

$$D_j(\theta^*) < D_{j+1}(\theta^*)$$

Summary 2 : towards nonparametric

- **Posterior concentration rates**

$$\pi(B_n|Y^n) = \frac{\int_{B_n} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)}{\int_{\Theta} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)} := \frac{N_n}{D_n}$$

- Lower bound on D_n : $P_{\theta^*}[D_n > e^{-n\epsilon_n^2(c+3/2)}] = 1 + o(1)$ If

$$\pi\left(\{\theta; KL_n(\theta^*,\theta) \leqslant n\epsilon_n^2,\ V_n(\theta^*,\theta) \lesssim n\epsilon_n^2\}\right) \geqslant e^{-nc\epsilon_n^2}$$

- Upper bound of $N_n = \int_{B_n} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)$

$$P_{\theta^*}[N_n = o(e^{-n\epsilon_n^2(c+3/2)}] = 1 + o(1)$$

- Either Existence of tests $\phi_n : H_0 : \theta = \theta^*$ vs $H_1 : \theta \in B_n$
- Or $\pi(B_n) = o(e^{-n(c+3/2)\epsilon_n^2})$

Summary 2 : towards nonparametric

- **Posterior concentration rates**

$$\pi(B_n|Y^n) = \frac{\int_{B_n} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)}{\int_\Theta e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)} := \frac{N_n}{D_n}$$

  - Lower bound on D_n : $P_{\theta^*}[D_n > e^{-n\epsilon_n^2(c+3/2)}] = 1 + o(1)$ If

$$\pi\left(\{\theta; KL_n(\theta^*, \theta) \leqslant n\epsilon_n^2,\ V_n(\theta^*, \theta) \lesssim n\epsilon_n^2\}\right) \geqslant e^{-nc\epsilon_n^2}$$

  - Upper bound of $N_n = \int_{B_n} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)$

$$P_{\theta^*}[N_n = o(e^{-n\epsilon_n^2(c+3/2)}] = 1 + o(1)$$

  - Either Existence of tests $\phi_n : H_0 : \theta = \theta^*$ vs $H_1 : \theta \in B_n$
  - Or $\pi(B_n) = o(e^{-n(c+3/2)\epsilon_n^2})$

Summary 2 : towards nonparametric

- **Posterior concentration rates**

$$\pi(B_n|Y^n) = \frac{\int_{B_n} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)}{\int_\Theta e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)} := \frac{N_n}{D_n}$$

- Lower bound on D_n : $P_{\theta^*}[D_n > e^{-n\epsilon_n^2(c+3/2)}] = 1 + o(1)$ If

$$\pi\left(\{\theta; KL_n(\theta^*,\theta) \leqslant n\epsilon_n^2,\ V_n(\theta^*,\theta) \lesssim n\epsilon_n^2\}\right) \geqslant e^{-nc\epsilon_n^2}$$

- Upper bound of $N_n = \int_{B_n} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)$

$$P_{\theta^*}[N_n = o(e^{-n\epsilon_n^2(c+3/2)}] = 1 + o(1)$$

  - Either Existence of tests ϕ_n : $H_0 : \theta = \theta^*$ vs $H_1 : \theta \in B_n$
  - Or $\pi(B_n) = o(e^{-n(c+3/2)\epsilon_n^2})$

Summary 2 : towards nonparametric

- **Posterior concentration rates**

$$\pi(B_n|Y^n) = \frac{\int_{B_n} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)}{\int_{\Theta} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)} := \frac{N_n}{D_n}$$

- Lower bound on D_n : $P_{\theta^*}[D_n > e^{-n\epsilon_n^2(c+3/2)}] = 1 + o(1)$ If

$$\pi\left(\{\theta; KL_n(\theta^*, \theta) \leqslant n\epsilon_n^2,\ V_n(\theta^*, \theta) \lesssim n\epsilon_n^2\}\right) \geqslant e^{-nc\epsilon_n^2}$$

- Upper bound of $N_n = \int_{B_n} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi(\theta)$

$$P_{\theta^*}[N_n = o(e^{-n\epsilon_n^2(c+3/2)}] = 1 + o(1)$$

  - Either Existence of tests ϕ_n : $H_0 : \theta = \theta^*$ vs $H_1 : \theta \in B_n$
  - Or $\pi(B_n) = o(e^{-n(c+3/2)\epsilon_n^2})$

Outline

General result under non i.i.d observations : sufficient conditions

▶ **Context**

• Under $\mathfrak{M}_0$: $Y \sim f^n_{0,\theta_0}$, $\theta_0 \in \Theta_0 \subset \mathbb{R}^d$
• Under $\mathfrak{M}_1$: $Y \sim f^n_{1,\theta_1}$, $\theta_1 \in \Theta_1$ (infinite dim)
$d_n(.,.)$ (semi) metric on the distributions (e.g. total variation, KL, etc.)

$$\liminf_n \inf_{\theta_0 \in \Theta_0} d_n(f_{1,\theta_1}, f_{0,\theta_0}) := d(\theta_1, \mathfrak{M}_0)$$

$$\liminf_n \inf_{\theta_1 \in \Theta_1} d_n(f_{0,\theta_0}, f_{1,\theta_1}) := d(\theta_0, \mathfrak{M}_1)$$

[Typically $\mathfrak{M}_0 \subset \mathfrak{M}_1$]

Conditions under $\mathfrak{M}_1$ (bigger model) : $\theta^* = \theta_1$ —- often easier

bound from above $m_0(Y^n)$ and from below $m_1(Y^n)$

$$m_j(Y^n) = \int_{\Theta_j} e^{\ell_n(j,\theta)-\ell_n(\theta^*)} d\pi_j(\theta)$$

$d(\theta^*, \mathfrak{M}_0) > 0$ (separated case)

- **If**

$$\forall \epsilon > 0, \quad P_{\theta^*}\left[m_1(Y^n) < e^{-n\epsilon^2}\right] = o(1)$$

[Consistency under $\mathfrak{M}_1$]

- **If** , $\forall \epsilon > 0$, $\exists a > 0$ and tests ϕ_n and $\Theta_{0,n} \subset \Theta_0$ s.t.

$$E_{\theta^*}[\phi_n] = o(1), \qquad \sup_{\theta_0 \in \Theta_{0,n}} E_{\theta_0}[1-\phi_n] \leqslant e^{-an}, \quad \pi_0(\Theta_{0,n}^c) < e^{-n\epsilon}$$

[Statistical separation between θ^* and $\mathfrak{M}_0$]

Then $B_{0/1} < e^{-an/2}$, with proba going to 1 under P_{θ^*}

Conditions under $\mathfrak{M}_0$ (smaller model) ; $\theta^* \in \mathfrak{M}_0 \subset \mathfrak{M}_1$

Case : $d(\theta^*, \mathfrak{M}_1) = 0$, & $\Theta_0 \subset \mathbb{R}^d$

- **(parametric) If** $\exists k_0 > 0$

$$n^{k_0/2}\pi_0\left[\{\theta \in \Theta_0 : KL(f^n_{0,\theta^*}, f^n_{0,\theta}) \leqslant 1, \quad V(f^n_{0,\theta^*}, f^n_{0,\theta}) \leqslant 1)\}\right] \geqslant C$$

for some positive constants C.

- **(NP)** $f_{\theta^*} = f_{0,\theta^*}$ If $\exists \epsilon_n \downarrow 0$ with

$A_{\epsilon_n}(\theta^*) = \{\theta \in \Theta_1 : d_n(f^n_{\theta^*}, f^n_{1,\theta}) < \epsilon_n\}$ such that

1. $$P^n_{\theta^*}\left[\pi_1\left[A^c_{\epsilon_n}(\theta^*)|Y\right]\right] = o(1)$$
2. $$n^{k_0/2}\pi_1\left[A_{\epsilon_n}(\theta^*)\right] = o(1).$$

Then $B_{0/1} \to +\infty$ under P_{θ^*}

Conditions under $\mathfrak{M}_0$ (smaller model) ; $\theta^* \in \mathfrak{M}_0 \subset \mathfrak{M}_1$

Case : $d(\theta^*, \mathfrak{M}_1) = 0$, & $\Theta_0 \subset \mathbb{R}^d$

- **(parametric) If** $\exists k_0 > 0$

$$n^{k_0/2}\pi_0 \left[\{\theta \in \Theta_0 : KL(f^n_{0,\theta^*}, f^n_{0,\theta}) \leqslant 1, \quad V(f^n_{0,\theta^*}, f^n_{0,\theta}) \leqslant 1)\}\right] \geqslant C$$

for some positive constants C.

- **(NP)** $f_{\theta^*} = f_{0,\theta^*}$ If $\exists \epsilon_n \downarrow 0$ with

$A_{\epsilon_n}(\theta^*) = \{\theta \in \Theta_1 : d_n(f^n_{\theta^*}, f^n_{1,\theta}) < \epsilon_n\}$ such that

1. $$P^n_{\theta^*}\left[\pi_1\left[A^c_{\epsilon_n}(\theta^*)|Y\right]\right] = o(1)$$

2. $$n^{k_0/2}\pi_1\left[A_{\epsilon_n}(\theta^*)\right] = o(1).$$

Then $B_{0/1} \to +\infty$ under P_{θ^*}

Application to Partial linear regressions

$$y_i = \alpha + \beta^t \boldsymbol{w}_i + f(x_i) + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$(\boldsymbol{w}_i, x_i) \sim G \quad i.i.d, \quad E[\boldsymbol{w}] = 0, \; E[\boldsymbol{w}\boldsymbol{w}^t] > 0$$

- $\mathfrak{M}_0 : \; (\alpha, \beta, f, \sigma)$ s.t.

$$H(f) := \inf_{(\alpha', \beta') \in \mathbb{R}^{d+1}} E\left[\left(\alpha' + (\beta')^t \boldsymbol{w} - f(x)\right)^2\right] = 0$$

$$\Leftrightarrow \quad y_i = \alpha_1 + \beta_1^t \boldsymbol{w}_i + \epsilon_i$$

- $\mathfrak{M}_1$ **:** $(\alpha, \beta, f, \sigma)$ s.t.

$$H(f) := \inf_{(\alpha', \beta') \in \mathbb{R}^{d+1}} E\left[\left(\alpha' + (\beta')^t \boldsymbol{w} - f(x)\right)^2\right] > 0$$

- **prior** hierarchical (adaptive) Gaussian process prior on f

$$f = \sum_{k=1}^{K} \tau_k Z_k \xi_k, \quad (\xi_k) = BON$$

$$K \sim P, \quad Z_k \sim \mathcal{N}(0,1)$$

$$(\alpha, \beta, \sigma) \sim \pi_0$$

$$h_j(x) = E[w_j | X = x], \quad \boldsymbol{w} = (w_1, \cdots, w_d)$$

- **If** $\exists j$ s.t. $< h_j, \xi_1 > \neq 0$ then $B_{0/1}$ is consistent under $\mathfrak{M}_0$ like n^ρ

Nonparametric/Nonparametric testing problem $B_{0/1} = m_0(Y)/m_1(Y)$

- **Both $\mathfrak{M}_0$ and $\mathfrak{M}_1$ are non parametric** & $\mathfrak{M}_0 \subset \mathfrak{M}_1$
- **Lower bound** Not too hard but sharp ?

$$m_j(Y) \geqslant e^{-(c_j+2)n\epsilon_{n,j}^2}, \quad \pi_j(B_{KL}(\theta^*, n\epsilon_{n,j}^2)) \gtrsim e^{-nc_j\epsilon_n^2}$$

- **upper bound** Harder
  - Tests control : ok – tools for it

$$\int_{d(\theta^*,\theta)>\epsilon_{n,j}} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi_j(\theta) \lesssim e^{-c_j' n\epsilon_{n,j}^2}$$

  - Control on $\int_{d(\theta^*,\theta)\leqslant\epsilon_{n,j}} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi_j(\theta)$: damn it

$$\text{If } d(\theta^*, \mathfrak{M}_j) = 0 \begin{cases} E_{\theta^*}[N_{n,j}] := E_{\theta^*}\left[\int_{d(\theta^*,\theta)\leqslant\epsilon_{n,j}} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi_j(\theta)\right] \\ \overset{\text{Fubini}}{=} \pi_j(d(\theta^*,\theta) \leqslant \epsilon_{n,j}) \overset{?}{\lesssim} e^{-c_j' n\epsilon_{n,j}^2} \end{cases}$$

Nonparametric/Nonparametric testing problem $B_{0/1} = m_0(Y)/m_1(Y)$

- **Both $\mathfrak{M}_0$ and $\mathfrak{M}_1$ are non parametric** & $\mathfrak{M}_0 \subset \mathfrak{M}_1$
- **Lower bound** Not too hard but sharp ?

$$m_j(Y) \geqslant e^{-(c_j+2)n\epsilon_{n,j}^2}, \quad \pi_j(B_{KL}(\theta^*, n\epsilon_{n,j}^2)) \gtrsim e^{-nc_j\epsilon_n^2}$$

- **upper bound** Harder
  - Tests control : ok – tools for it

$$\int_{d(\theta^*,\theta)>\epsilon_{n,j}} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi_j(\theta) \lesssim e^{-c_j' n\epsilon_{n,j}^2}$$

  - Control on $\int_{d(\theta^*,\theta)\leqslant\epsilon_{n,j}} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi_j(\theta)$: damn it

$$\text{If } d(\theta^*, \mathfrak{M}_j) = 0 \begin{cases} E_{\theta^*}[N_{n,j}] := E_{\theta^*}\left[\int_{d(\theta^*,\theta)\leqslant\epsilon_{n,j}} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi_j(\theta)\right] \\ \overset{\text{Fubini}}{=} \pi_j(d(\theta^*,\theta) \leqslant \epsilon_{n,j}) \overset{?}{\lesssim} e^{-c_j' n\epsilon_{n,j}^2} \end{cases}$$

Setups where this approach works Ghosal, Lembert, VdV & Rousseau, Szabo

- **Multiple models** Θ_λ, $\lambda \in \Lambda$ (possibly continuous) Aim

$$\hat{\lambda} = \text{argmax}\{\pi(\lambda|Y^n), \lambda \in \Lambda\} \quad \pi(\lambda|Y^n) \propto \pi(\lambda) m_\lambda(Y^n)$$

- **What is λ^* ?** Two contexts
  - *Almost finite case* : $\Lambda \subset \mathbb{N}$, $\Theta_\lambda \subset \Theta_{\lambda+1}$ and

$$\exists \lambda_0 \text{ s.t. } \theta^* \in \Theta_{\lambda_0}, \quad \lambda^* = \min\{\lambda; \theta^* \in \Theta_\lambda\}$$

[Sharp tools like BIC]

  - *Boundary case* $\theta^* \in \mathcal{C}\ell(\cup_\lambda \Theta_\lambda)$ but $\forall \lambda$, $\theta^* \notin \Theta_\lambda$

    Claim : λ^* = Oracle

$$\lambda^* = \min_\lambda\{\epsilon_{n,\lambda}, \lambda \in \Lambda\}$$

where

$$\epsilon_{n,\lambda} : \pi_\lambda(d(\theta^*, \theta) \leq \epsilon_{n,\lambda}) \asymp e^{-cn\epsilon_{n,\lambda}^2},$$

Setups where this approach works Ghosal, Lembert, VdV & Rousseau, Szabo

- **Multiple models** Θ_λ, $\lambda \in \Lambda$ (possibly continuous) Aim

$$\hat{\lambda} = \text{argmax}\{\pi(\lambda|Y^n), \lambda \in \Lambda\} \quad \pi(\lambda|Y^n) \propto \pi(\lambda) m_\lambda(Y^n)$$

- **What is λ^* ?** Two contexts
  - *Almost finite case* : $\Lambda \subset \mathbb{N}$, $\Theta_\lambda \subset \Theta_{\lambda+1}$ and

$$\exists \lambda_0 \text{ s.t. } \theta^* \in \Theta_{\lambda_0}, \quad \lambda^* = \min\{\lambda; \theta^* \in \Theta_\lambda\}$$

[Sharp tools like BIC]

  - *Boundary case* $\theta^* \in \mathcal{C}\ell(\cup_\lambda \Theta_\lambda)$ but $\forall \lambda$, $\theta^* \notin \Theta_\lambda$
    Claim : λ^* = Oracle

$$\lambda^* = \min_\lambda\{\epsilon_{n,\lambda}, \lambda \in \Lambda\}$$

where

$$\epsilon_{n,\lambda} : \pi_\lambda(d(\theta^*, \theta) \leqslant \epsilon_{n,\lambda}) \asymp e^{-cn\epsilon_{n,\lambda}^2},$$

Where does the oracle come from ?

$$m_\lambda(Y^n) = \int_{\Theta_\lambda} \frac{f_{\theta,\lambda}(Y^n)}{f_{\theta^*}(Y^n)} d\pi_\lambda(\theta) \geqslant \int_{d(\theta,\theta^*)\leqslant \epsilon_{n,\lambda}} e^{\ell_n(\theta)-\ell_n(\theta^*)} d\pi_\lambda(\theta)$$

$$\gtrsim e^{-c_1 n\epsilon_n^2} \pi_\lambda(d(\theta^*,\theta) \leqslant \epsilon_{n,})$$

$$\lesssim e^{-c_2 n\epsilon_n^2} \pi_\lambda(d(\theta^*,\theta) \leqslant \epsilon_n) + e^{-nc_2\epsilon_n^2}$$

Under *stringent* conditions

$$\pi(\lambda : \epsilon_{n,\lambda} \leqslant M\epsilon_{n,\lambda^*} | Y^n) \to 1$$

$$\hat{\lambda} \to \{\lambda : \epsilon_{n,\lambda} \leqslant M\epsilon_{n,\lambda^*}\}$$

[Only useful if $\lambda \neq \lambda^* \Rightarrow \epsilon_{n,\lambda} >> \epsilon_{n,\lambda^*}$]

NP priors $\mathfrak{M}_0$ and $\mathfrak{M}_1$

- **Testing for monotonicity**

$$Y = (y_1, \cdots, y_n) \overset{iid}{\sim} f, \quad \text{or} \quad y_i = f(x_i) + \epsilon_i,$$

- Model $\mathfrak{M}_0$: $f =$ decreasing
- Model $\mathfrak{M}_1$: f has no shape constraints.

- **1 rst idea**
  - Prior on $\mathfrak{M}_0$: Mixtures of $\mathcal{U}(0, \theta)$

$$f_P(y) = \int_0^\infty \frac{\mathbb{I}_{y \leqslant \theta}}{\theta} dP(\theta), \quad P \sim DP(AG)$$

  - Prior on smooth densities : e.g. Mixtures of Betas

$$f_P(y) = \int_0^\infty g_{z,z/\epsilon}(y) dP(\epsilon), \quad (z, P) \sim \pi_z \times DP(AG)$$

[Bad idea]

NP priors $\mathfrak{M}_0$ and $\mathfrak{M}_1$

- **Testing for monotonicity**

$$Y = (y_1, \cdots, y_n) \overset{iid}{\sim} f, \quad \text{or} \quad y_i = f(x_i) + \epsilon_i,$$

• Model $\mathfrak{M}_0$: f = decreasing
• Model $\mathfrak{M}_1$: f has no shape constraints.

- **1 rst idea**
  - Prior on $\mathfrak{M}_0$: Mixtures of $\mathcal{U}(0, \theta)$

$$f_P(y) = \int_0^\infty \frac{\mathbb{I}_{y \leqslant \theta}}{\theta} dP(\theta), \quad P \sim DP(AG)$$

  - Prior on smooth densities : e.g. Mixtures of Betas

$$f_P(y) = \int_0^\infty g_{z,z/\epsilon}(y) dP(\epsilon), \quad (z, P) \sim \pi_z \times DP(AG)$$

[Bad idea]

1rst solution

$y_i \in [0, 1]$

$$f(x) = f_{\omega,k} := \sum_{j=1}^{k} k\omega_j \mathbf{1}_{[(j-1)/k, j/k[}(x), \quad \sum_{j=1}^{k} \omega_j = 1$$

- **Prior**

$$k \sim P; \quad (\omega_1, \cdots, \omega_k) \sim \pi_k, \quad \text{e.g. } \mathcal{D}(\alpha_1, \cdots, \alpha_k)$$

- **Posterior probabilities**

$$\pi[\mathfrak{M}_0 | Y] = \pi\left[\max_j(\omega_j - \omega_{j-1}) \leqslant 0 \middle| Y\right]$$

- **Problem** : if $f \in \mathfrak{M}_0$ but has flat parts

Result

- **Question :** ? ? $\pi[\mathfrak{M}_0|Y] = 1 + o_{p_f}(1)$ if $f \in \mathfrak{M}_0$.
- **Answers : Not always**

$$g(\delta) := \mathcal{L}eb\left(\left\{x; \inf_{y>x} \frac{f(x)-f(y)}{y-x} \leqslant \delta\right\}\right)$$

• For all f_0 s.t. $g(\delta) = 0$, if

$$P^{\pi}\left[k \lesssim (n/\log n)^{1/5} \Big| Y\right] = 1 + o_p(1)$$

$$P^{\pi}[\mathfrak{M}_0|Y] = 1 + o_p(1) \rightarrow \text{Consistent test}$$

e.g. if $k^3 \sim \mathcal{P}(\lambda)$.
• For all f_0 s.t. $g(\delta) = \sqrt{\delta}$, idem with $k^7 \sim \mathcal{P}(\lambda)$.

Inconsistencies

If f_0 *is constant on* $[a, b] \subset (0, 1)$ *and if*

$$\pi(3/(b-a) < k < \sqrt{n/\log n}|Y^n) \to 1$$

then with probability greater than $1/2$ *(asymptotically)*

$$\pi(\mathfrak{M}_0|Y^n) < 1/2$$

[Inconsistency of the testing procedure]

Comments

- **Not satisfying** : Cannot be consistent over whole set of decreasing functions.
Increasing the set of functions by lowering k $\Rightarrow$ leads to poor power.
- **Change of loss function** $\rightarrow$ new test based on

$$P^{\pi}\left[d(f, \mathfrak{M}_1) \leqslant \epsilon | Y\right]$$

Choice of ϵ ?

Moving away (slightly) from the 0-1 loss function for unseparated hypotheses [JB Salomond]

- **New loss function**

$$L_\tau(\theta, \delta) = \begin{cases} 0 & \text{if} \quad d(\theta, \mathfrak{M}_0) \leqslant \tau \\ 1 & \text{if} \quad d(\theta, \mathfrak{M}_0) > \tau \end{cases}$$

Equivalent to testing

$$H_0 : d(\theta, \mathfrak{M}_0) \leqslant \tau, \quad H_1 : d(\theta, \mathfrak{M}_0) > \tau$$

- **Bayes estimator**

$$\delta^\pi = \begin{cases} 0 & \text{if} \quad \pi\left(d(\theta, \mathfrak{M}_0) \leqslant \tau | Y^n\right) > 1/2 \\ 1 & \text{if} \quad \text{otherwise} \end{cases}$$

Choice of τ

- If prior information on tolerance level τ chosen a priori
- Calibration by τ_n smallest s.t. Type I error

$$\sup_{\theta \in \mathfrak{M}_0} P_\theta[\pi(d(\theta, \mathfrak{M}_0) \leqslant \tau_n | Y^n) \leqslant 1/2] = o(1)$$

- Study of ρ_n s.t. (separation rate)

$$\sup_{\theta : d(\theta, \mathfrak{M}_0) > \rho_n} P_\theta[\pi(d(\theta, \mathfrak{M}_0) \leqslant \tau_n | Y^n) < 1/2] = o(1)$$

[Control of type II error over ρ_n separated hypotheses]

- In quite a few examples : ρ_n minimax separation rate

Choice of τ

- If prior information on tolerance level τ chosen a priori
- Calibration by τ_n smallest s.t. Type I error

$$\sup_{\theta \in \mathfrak{M}_0} P_\theta[\pi(d(\theta, \mathfrak{M}_0) \leqslant \tau_n | Y^n) \leqslant 1/2] = o(1)$$

- Study of ρ_n s.t. (separation rate)

$$\sup_{\theta : d(\theta, \mathfrak{M}_0) > \rho_n} P_\theta[\pi(d(\theta, \mathfrak{M}_0) \leqslant \tau_n | Y^n) < 1/2] = o(1)$$

[Control of type II error over ρ_n separated hypotheses]

- In quite a few examples : ρ_n minimax separation rate

Choice of τ

- If prior information on tolerance level τ chosen a priori
- Calibration by τ_n smallest s.t. Type I error

$$\sup_{\theta \in \mathfrak{M}_0} P_\theta[\pi(d(\theta, \mathfrak{M}_0) \leqslant \tau_n | Y^n) \leqslant 1/2] = o(1)$$

- Study of ρ_n s.t. (separation rate)

$$\sup_{\theta : d(\theta, \mathfrak{M}_0) > \rho_n} P_\theta[\pi(d(\theta, \mathfrak{M}_0) \leqslant \tau_n | Y^n) < 1/2] = o(1)$$

[Control of type II error over ρ_n separated hypotheses]

- In quite a few examples : ρ_n minimax separation rate

Choice of τ

- If prior information on tolerance level τ chosen a priori
- Calibration by τ_n smallest s.t. Type I error

$$\sup_{\theta \in \mathfrak{M}_0} P_\theta[\pi(d(\theta, \mathfrak{M}_0) \leqslant \tau_n | Y^n) \leqslant 1/2] = o(1)$$

- Study of ρ_n s.t. (separation rate)

$$\sup_{\theta : d(\theta, \mathfrak{M}_0) > \rho_n} P_\theta[\pi(d(\theta, \mathfrak{M}_0) \leqslant \tau_n | Y^n) < 1/2] = o(1)$$

[Control of type II error over ρ_n separated hypotheses]

- In quite a few examples : ρ_n minimax separation rate

Difficulties : τ_n chosen asymptotically

- Difficult analysis in some cases : Not necessarily to determine τ_n . e.g. in the case of testing for monotonicity – only for simple families of priors
- optimal $\tau_n = \tau_0 v_n$ where v_n known τ_0 arbitrary. Calibration of τ_0 for finite n ?
- Can we trust purely asymptotic results to calibrate a procedure ?

Difficulties : τ_n chosen asymptotically

- Difficult analysis in some cases : Not necessarily to determine τ_n . e.g. in the case of testing for monotonicity – only for simple families of priors
- optimal $\tau_n = \tau_0 v_n$ where v_n known τ_0 arbitrary. Calibration of τ_0 for finite n ?
- Can we trust purely asymptotic results to calibrate a procedure ?

Difficulties : τ_n chosen asymptotically

- Difficult analysis in some cases : Not necessarily to determine τ_n . e.g. in the case of testing for monotonicity – only for simple families of priors
- optimal $\tau_n = \tau_0 v_n$ where v_n known τ_0 arbitrary. Calibration of τ_0 for finite n ?
- Can we trust purely asymptotic results to calibrate a procedure ?

Semi-parametric : $\theta = (\psi, \eta)$

$$\mathfrak{M}_1 : \psi = \psi_0, \quad \mathfrak{M}_2 : \psi \neq \psi_0$$

- Cox hazard :

$$h_\theta(x|z) = e^{\psi^t z}\eta(x), \quad \mathfrak{M}_1 : \psi = 0$$

- Long memory parameter : $\mathfrak{M}_1 : h = 0$

$$Y \sim \mathcal{N}(0, T_n(f_\theta)), \quad f_\theta(x) = (1 - \cos x)^h \eta(x), \quad \eta > 0 \in \mathcal{C}_b$$

▶ **Saved by BvM**

$$\pi(v_n(\psi - \hat{\psi}) \in A|Y) \approx Pr(\mathcal{N}_d(0,1) \in A) \quad \Rightarrow B_{1/2} \gtrsim v_n^{-d/2} \to +\infty$$

Semi-parametric : $\theta = (\psi, \eta)$

$$\mathfrak{M}_1 : \psi = \psi_0, \quad \mathfrak{M}_2 : \psi \neq \psi_0$$

- Cox hazard :

$$h_\theta(x|z) = e^{\psi^t z}\eta(x), \quad \mathfrak{M}_1 : \psi = 0$$

- Long memory parameter : $\mathfrak{M}_1 : h = 0$

$$Y \sim \mathcal{N}(0, T_n(f_\theta)), \quad f_\theta(x) = (1 - \cos x)^h \eta(x), \quad \eta > 0 \in \mathcal{C}_b$$

- **Saved by BvM**

$$\pi(v_n(\psi - \hat{\psi}) \in A|Y) \approx Pr(\mathcal{N}_d(0,1) \in A) \quad \Rightarrow B_{1/2} \gtrsim v_n^{-d/2} \to +\infty$$

The 2 - samples test

$$X^n = (X_1, \cdots, X_n) \overset{iid}{\sim} F_X, \quad Y^n = (Y_1, \cdots, Y_n) \overset{iid}{\sim} F_Y$$

$$H_0 : F_X = F_Y = F \quad H_1 : F_X \neq F_Y$$

- **Prior** $H_0 : F \sim \pi$ and $H_1 : F_X, F_Y \sim \pi$ ind.

$$B_{1/2} = \frac{m_{2n}(X^n, Y^n)}{m_n(X^n) m_n(Y^n)}, \quad \text{consistency ?}$$

- **Hardly any literature (Bayesian)** on 2 - sample test

• Holmes et al. [PolyaT]
• Dunson & Lock [$\approx$ parametric]
• Chen & Hansen [censored-PolyaT]

Other ways of getting away from 0-1 loss functions

- **distance penalized losses** Robert & Rousseau, Rousseau Goodness of fit test : $Y^n = (y_1, \cdots, y_n)$, $y_i \overset{iid}{\sim} f$

$$H_0 : f \in \mathcal{F}_0 = \{f_\theta, \theta \in \Theta \subset \mathbb{R}^d\}, \quad H_1 : f \notin \mathcal{F}_0$$

$$L(\theta, \delta) = \begin{cases} (d(\theta, \mathfrak{M}_0) - \epsilon)\,\mathbb{I}_{d(\theta,\mathfrak{M}_0) > \epsilon} & \text{when} \quad \delta = 0 \\ (\epsilon - d(\theta, \mathfrak{M}_0))\,\mathbb{I}_{d(\theta,\mathfrak{M}_0) \leqslant \epsilon} & \text{when} \quad \delta = 1 \end{cases}$$

- **Bayes test**

$$\delta^\pi = 1 \quad \Leftrightarrow \quad E^\pi[d(\theta, \mathfrak{M}_0) | Y^n] > \epsilon$$

- **Calibration of** ϵ using a conditional predictive p-value

$$p(y^n) = \int_\Theta P_\theta(T(Y^n) > T(y^n) | \hat{\theta}, y^n) \pi_0(\theta) d\theta$$

Other ways

There are many other ways . . .

- Point null hypothesis : Consistency of BF $\leftrightarrow$ consistency of the posterior under H_1
- Parametric null : more involved . Π_1 can be more favorable to f_{θ_0} than Π_0.
- Nonparametric Null . e.g. $\mathcal{F}_0 = \{f \downarrow\} \rightarrow$ need to be even more carefull
- Are BIC or BIC type formula better than BF ?

- Point null hypothesis : Consistency of BF $\leftrightarrow$ consistency of the posterior under H_1
- Parametric null : more involved . Π_1 can be more favorable to f_{θ_0} than Π_0.
- Nonparametric Null . e.g. $\mathcal{F}_0 = \{f \downarrow\} \rightarrow$ need to be even more carefull
- Are BIC or BIC type formula better than BF ?

- Point null hypothesis : Consistency of BF $\leftrightarrow$ consistency of the posterior under H_1
- Parametric null : more involved . Π_1 can be more favorable to f_{θ_0} than Π_0.
- Nonparametric Null . e.g. $\mathcal{F}_0 = \{f \downarrow\} \rightarrow$ need to be even more carefull
- Are BIC or BIC type formula better than BF ?

- Point null hypothesis : Consistency of BF $\leftrightarrow$ consistency of the posterior under H_1
- Parametric null : more involved . Π_1 can be more favorable to f_{θ_0} than Π_0.
- Nonparametric Null . e.g. $\mathcal{F}_0 = \{f \downarrow\} \rightarrow$ need to be even more carefull
- Are BIC or BIC type formula better than BF ?