DEC statistics seminar, Milano, 12 April 2018

What is asymptotically testable and what is not?

Bas Kleijn

KdV Institute for Mathematics

Universiteit van Amsterdam

Part I

Introduction and Motivation

Asymptotic symmetric testing

Observe *i.i.d.* $X^n \sim P^n$, model $P \in \mathscr{P}$. For disjoint $B, V \subset \mathscr{P}$,

$$H_0 : P \in B, \quad \text{or} \quad H_1 : P \in V.$$

Tests $\phi_n : \mathscr{X}^n \to [0,1]$; asymptotically, require:

$$\begin{aligned} &(\text{type-I}) \quad P^n\phi_n \to 0 \text{ for } P \in B, \text{ and,} \\ &(\text{type-II}) \quad P^n(1-\phi_n) \to 0 \text{ for } P \in V. \end{aligned}$$

Equivalently, we want,

> A testing procedure that chooses for B or V based on X^n for every $n \geq 1$, has property (D) if it is wrong only a finite number of times with P^∞-probability one.

Property (D) is sometimes referred to as "discernibility".

Some examples and unexpected answers (I)

Consider non-parametric regression with $f : X \to \mathbb{R}$ and test for smoothness,

$$H_0 : f \in C^1(X \to \mathbb{R}), \quad H_1 : f \in C^2(X \to \mathbb{R}),$$

Consider a non-parametric density estimation with $p : \mathbb{R} \to [0, \infty)$ and test for square-integrability,

$$H_0 : \int x^2 \, p(x) \, dx < \infty, \quad H_1 : \int x^2 \, p(x) \, dx = \infty.$$

Practical problem we cannot use the data to determine with asymptotic certainty, if CLT applies with our data.

Some examples and unexpected answers (II)

Coin-flip $X^n \sim \text{Bernoulli}(p)^n$ with $p \in [0,1]$.

Consider Cover's rational mean problem: test for rationality:

$$H_0 : p \in [0,1] \cap \mathbb{Q}, \quad H_1 : p \in [0,1] \setminus \mathbb{Q}.$$

Consider also Dembo and Peres's irrational alternative:

$$H_0 : p \in [0,1] \cap \mathbb{Q}, \quad H_1 : p \in [0,1] \cap \sqrt{2} + \mathbb{Q},$$

Consider ultimately fractal hypotheses, *e.g.* with Cantor set C,

$$H_0 : p \in C, \quad H_1 : p \in [0,1] \setminus C.$$

The Le Cam-Schwartz theorem

Theorem 6.1 (Le Cam-Schwartz, 1960) *Let $\mathscr{P}$ be a model for i.i.d. data X^n with disjoint subsets B, V. The folllowing are equivalent:*

i. there exist (uniformly) consistent tests for B vs V,
ii. there is a sequence of $\mathscr{U}_\infty$-uniformly continuous $\psi_n : \mathscr{P} \to [0,1]$,

$$\psi_n(P) \to 1_V(P), \tag{1}$$

(uniformly) for all $P \in B \cup V$.

Topological context uniform space $(\mathscr{P}, \mathscr{U}_\infty)$.

The Dembo-Peres theorem

Theorem 7.1 (Dembo and Peres, 1995) *Let* $\mathscr{P}$ *be a model dominated by Lebesgue measure* μ *for* i.i.d. data X^n*. Model subsets* B, V *that are contained in disjoint countable unions of closed sets for* Prokhorov's weak topology *have tests with property (D). If there exists an* $\alpha > 1$ *such that* $\int (dP/d\mu)^{\alpha}\, d\mu < \infty$ *for all* $P \in \mathscr{P}$*, then the converse is also true.*

Topological context L^1-weakly compact, dominated model $\mathscr{P}$ with Prokhorov's weak topology.

Three forms of testability

Definition 8.1 (ϕ_n) *is a uniform test sequence for* B *vs* V*, if,*

$$\sup_{P\in B} P^n\phi_n \to 0, \qquad \sup_{Q\in V} Q^n(1-\phi_n) \to 0. \tag{2}$$

Definition 8.2 (ϕ_n) *is a pointwise test sequence for* B *vs* V*, if,*

$$\phi_n(X^n) \xrightarrow{P} 0, \qquad \phi_n(X^n) \xrightarrow{Q} 1, \tag{3}$$

for all $P \in B$ *and* $Q \in V$*.*

Definition 8.3 (ϕ_n) *is a Bayesian test sequence for* B *vs* V*, if,*

$$\phi_n(X^n) \xrightarrow{P} 0, \qquad \phi_n(X^n) \xrightarrow{Q} 1, \tag{4}$$

for Π*-almost-all* $P \in B$ *and* $Q \in V$*.*

Questions

Existence

Existence of uniform tests?

Existence of pointwise tests?

Existence of Bayesian tests?

Construction

How does one model-select? Are there constructive solutions?

Examples

Select the correct directed, acyclical graph in a graphical model; select the right number of clusters in a clustering model.

Part II

Existence

Uniform testability has exponential power

Proposition 11.1 *Let $\mathscr{P}$ be a model for i.i.d. data with disjoint B and V. The following are equivalent:*

i. there exists a uniform test sequence (ϕ_n),

$$\sup_{P\in B} P^n\phi_n \to 0, \qquad \sup_{Q\in V} Q^n(1-\phi_n) \to 0,$$

ii. there is a exponentially powerful uniform test sequence (ψ_n), i.e. there is a $D > 0$ such that,

$$\sup_{P\in B} P^n\psi_n \le e^{-nD}, \qquad \sup_{Q\in V} Q^n(1-\psi_n) \le e^{-nD}.$$

The model as a uniform space (I)

Take $\mathscr{X}$ a separable metrizable space, with Borel σ-algebra $\mathscr{B}$.

The class $\mathscr{F}_n$ contains all bounded, $\mathscr{B}^n$-measurable $f : \mathscr{X}^n \to \mathbb{R}$.

For every $n \geq 1$ and $f \in \mathscr{F}_n$, define the entourage,

$$W_{n,f} = \{(P,Q) \in \mathscr{P} \times \mathscr{P} : |P^n f - Q^n f| < 1\}.$$

Defines uniformity $\mathscr{U}_n$ (with topology $\mathscr{T}_n$). Take $\mathscr{U}_\infty = \cup_{n\geq 1}\mathscr{U}_n$.

$$P \to Q \ \text{ in } \mathscr{T}_\infty \quad \Leftrightarrow \quad \int f\, dP^n \to \int f\, dQ^n,$$

for all $n \geq 1$ and all $f \in \mathscr{F}_n$. Note also,

$$\mathscr{U}_C \subset \mathscr{U}_1 \subset \cdots \subset \mathscr{U}_\infty \subset \mathscr{U}_{TV}.$$

The model as a uniform space (II)

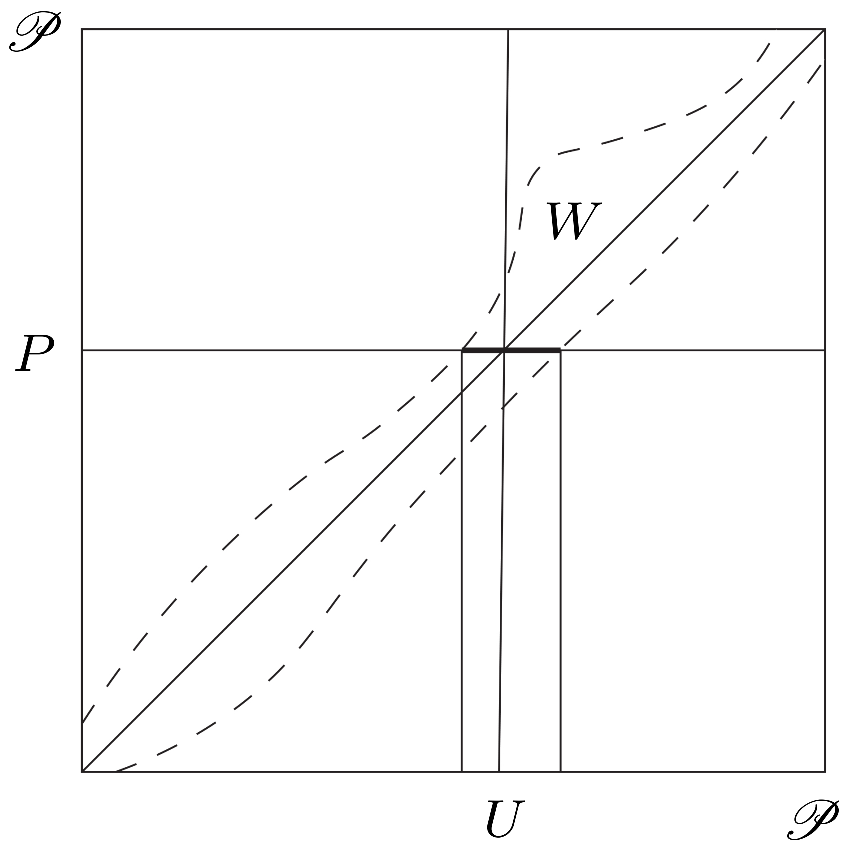

Fig 1. Let $P \in \mathscr{P}$ and entourage W be given. A neighbourhood U corresponds to $U = \{Q \in \mathscr{P} : (Q, P) \in W\}$

Uniform separation (I)

Definition 14.1 *Subsets* $B, V \subset \mathscr{P}$ *are uniformly separated by* $\mathscr{U}_\infty$*, if there exists an entourage* $W \in \mathscr{U}_\infty$ *such that,*

$$(B \times V \cup V \times B) \cap W = \varnothing.$$

In other words, there are $J, m \geq 1$*,* $\epsilon > 0$ *and bounded, measurable functions* $f_1, \ldots, f_J : \mathscr{X}^m \to [0,1]$ *such that, for any* $P, Q \in B \cup V$*, if,*

$$\max_{1 \leq j \leq J} \left| P^m f_j - Q^m f_j \right| < \epsilon,$$

then either $P, Q \in B$*, or* $P, Q \in V$*. (If the model is* $\mathscr{T}_\infty$*-compact,* $m = 1$ *suffices).*

Uniform separation (II)

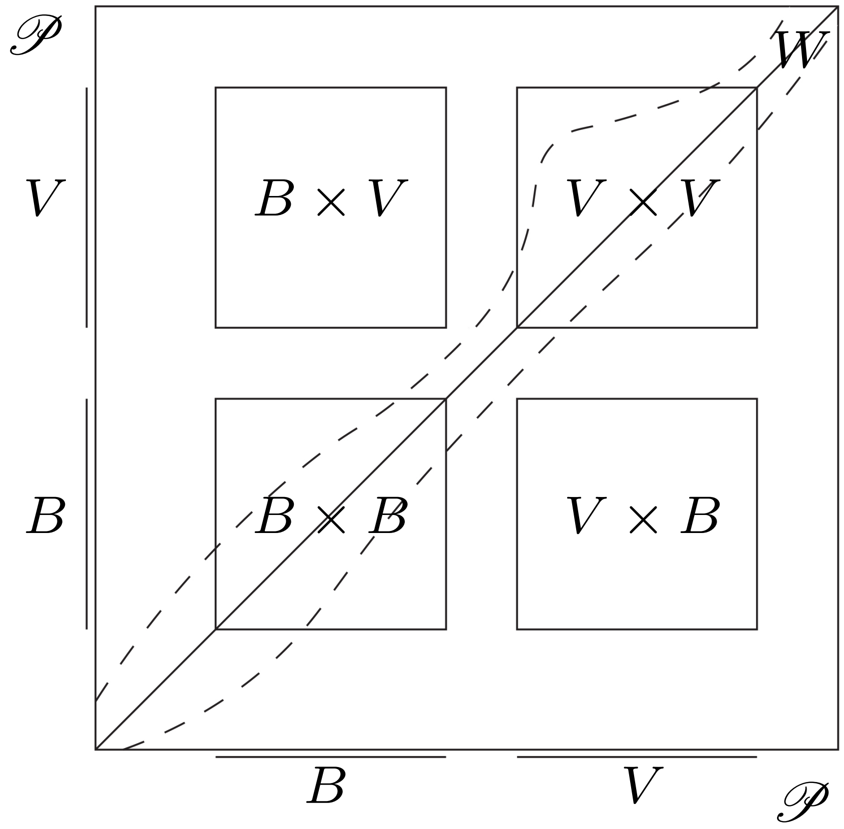

Fig 2. Let $B, V \subset \mathscr{P}$ and entourage W be given. W separates B and V if $B \times V$ and $V \times B$ do not meet W.

Characterisation of uniform testability

Theorem 16.1 *Let $\mathscr{P}$ be a model for i.i.d. data with disjoint B and V. The following are equivalent:*

(i.) there exist uniform tests ϕ_n for B versus V,
(ii.) the subsets B and V are uniformly separated by $\mathscr{U}_\infty$.

Corollary 16.2 (Parametrised models) *Suppose $\mathscr{P} = \{p_\theta : \theta \in \Theta\}$, with (Θ, d) compact, metric space and $\theta \to P_\theta$ identifiable and $\mathscr{T}_\infty$-continuous, (that is, for every $f \in \mathscr{F}_n$, $\theta \mapsto \int f \, dP_\theta^n$ is continuous). If $B_0, V_0 \subset \Theta$ with $d(B_0, V_0) > 0$, then the images $B = \{P_\theta : \theta \in B_0\}$, $V = \{P_\theta : \theta \in V_0\}$ are uniformly testable.*

Closures are important

Proposition 17.1 *Let $\mathscr{P}$ be a model for i.i.d. data and let B, V be disjoint model subsets with $\mathscr{T}_\infty$-closures $\overline{B}$ and $\overline{V}$. If B, V are uniformly separated by $\mathscr{U}_\infty$, then $\overline{B} \cap \overline{V} = \varnothing$. If $\mathscr{P}$ is relatively $\mathscr{T}_\infty$-compact, the converse is also true.*

Theorem 17.2 (Dunford-Pettis) *Assume $\mathscr{P}$ is dominated by a probability measure Q with densities in $\mathscr{P}_Q \subset L^1(Q)$; $\mathscr{P}_Q$ is relatively weakly compact, if and only if, for every $\epsilon > 0$ there is an $M > 0$ such that,*

$$\sup_{P \in \mathscr{P}} \int_{\{dP/dQ > M\}} \frac{dP}{dQ}\, dQ < \epsilon,$$

that is, $\mathscr{P}_Q$ is uniformly Q-integrable.

Pointwise testability: equivalent formulations

Proposition 18.1 *Let* $\mathscr{P}$ *be a model for i.i.d. data and let* B, V *be disjoint model subsets. The following are equivalent:*

i. there are tests (ϕ_n) *such that, for all* $P \in B$ *and* $Q \in V$*,*

$$P^n \phi_n \to 0, \quad Q^n(1 - \phi_n) \to 0,$$

ii. there are tests (ϕ_n) *such that, for all* $P \in B$ *and* $Q \in V$*,*

$$\phi_n(X^n) \xrightarrow{P} 0, \quad (1 - \phi_n(X^n)) \xrightarrow{Q} 0,$$

iii. there are tests (ϕ_n) *such that, for all* $P \in B$ *and* $Q \in V$*,*

$$\phi_n(X^n) \xrightarrow{P\text{-a.s.}} 0, \quad (1 - \phi_n(X^n)) \xrightarrow{Q\text{-a.s.}} 0.$$

Pointwise testability from consistent estimators

Consistent estimators $\hat{P}_n : \mathscr{X}^n \to \mathscr{P}$: for all P and nbd U of P,

$$P^n\Big(\hat{P}_n(X^n) \in U\Big) \to 1, \text{ as } n \to \infty.$$

For open $B, V \subset \mathscr{P}$, define $\phi_n(X^n) = 1\{\hat{P}_n \in V\}$. For any $P \in B$, B is a neighbourhood of P so $P^n\phi_n = P^n(\hat{P}_n \in V) \leq P^n(\hat{P}_n \notin B) \to 0$. For any $Q \in V$, $Q^n(1-\phi_n) \to 0$. So (ϕ_n) is a pointwise test sequence for B vs V.

Restrict to $\mathscr{P}' = B \cup V$, then B and V are *clopen* sets.

Proposition 19.1 *If $P \in \mathscr{P}$ can be estimated consistently and B is* clopen, *there exist pointwise tests for B vs its complement.*

Necessary conditions: pointwise non-testability

Suppose that there exist pointwise tests (ϕ_n) for B, V. Define,

$$g_n : \mathscr{P} \to [0,1] : P \mapsto P^n \phi_n,$$

which are all $\mathscr{U}_\infty$-uniformly continuous.

Proposition 20.1 *If there is a pointwise test (ϕ_n) for B vs V, then B, V are both G_δ- and F_σ-sets with respect to $\mathscr{T}_\infty$ (in the subspace $B \cup V$).*

Corollary 20.2 *Suppose $\mathscr{P} = B \cup V$ is Polish in the $\mathscr{T}_\infty$-topology. Pairs B, V that are pointwise testable, are both Polish spaces.*

Corollary 20.3 *If there exists a Baire subspace D of $\mathscr{P}$ in which both $D \cap B$ and $D \cap V$ are dense, then B is not testable versus V.*

Pointwise non-testability: examples (I)

Example 21.1 *Is Cover's rational means problem testable?*

Dunford-Pettis theorem shows that $\mathscr{P}$ is $\mathscr{T}_\infty$-compact and $[0,1] \to \mathscr{P} : p \mapsto P_p$ is a $\mathscr{T}_\infty$-homeomorphism. Since $[0,1]$ is a complete metric space, $\mathscr{P}$ is a Baire space for the $\mathscr{T}_\infty$-topology. Because both $[0,1] \cap \mathbb{Q}$ and $[0,1] \setminus \mathbb{Q}$ are dense in $[0,1]$, the images $\mathscr{P}_0 := \{P_p : p \in [0,1] \cap \mathbb{Q}\}$ and $\mathscr{P}_1 := \{P_p : p \in [0,1] \setminus \mathbb{Q}\}$ are $\mathscr{T}_\infty$-dense in $\mathscr{P}$: there is no pointwise test for $p \in [0,1] \cap \mathbb{Q}$ versus $p \in [0,1] \setminus \mathbb{Q}$.

Pointwise non-testability: examples (II)

Example 22.1 *Is Dembo and Peres's irrational alternative testable?*

Any countable $\mathscr{P}$ is Polish in the discrete topology. Any subset B of $\mathscr{P}$ is a countable union of closed sets $(B = \cup_{b \in B}\{b\})$, so it remains possible that there exists a pointwise test for Dembo and Peres's problem.

Example 22.2 *Is Cantor's fractal alternative testable?*

The interval $[0, 1]$ is Polish and $\mathscr{P}$ is homeomorphic. The Cantor set C is closed and its complement is open. Open sets in metrizable spaces are F_σ-sets. So it remains possible there exists a pointwise test for Cantor's fractal alternative.

Pointwise non-testability: examples (III)

Example 23.1 *Is integrability of a real-valued X, $P|X| < \infty$, testable?*

Model $\mathscr{P} = \{$all probability distributions on $\mathbb{R}\}$. $\mathscr{P}$ is Baire space for $\mathscr{T}_{TV}$. Define,

$$B = \{P \in \mathscr{P} : P|X| < \infty\}, \quad V = \{P \in \mathscr{P} : P|X| = \infty\}.$$

B cannot be tested versus V.

Namely Let $P \in B$ and $Q \in V$ be given. For any $0 < \epsilon < 1$, $P' = (1-\epsilon)P + \epsilon Q$ satisfies $\|P' - P\| = \epsilon\|(P+Q)\| \leq 2\epsilon$, but $P' \in V$. Conclude that V lies $\mathscr{T}_{TV}$-dense in $\mathscr{P}$.

Conversely, Q is tight, so for every $\epsilon > 0$, there exists an $M > 0$ such that $|Q(A) - Q(A||X| \leq M)| < \epsilon$ for all measurable $A \subset \mathbb{R}$. Since $Q(\cdot||X| \leq M) \in B$, we also see that B lies $\mathscr{T}_{TV}$-dense in $\mathscr{P}$.

Pointwise testability in dominated models

Definition 24.1 *The testing problem has a (uniform) representation on* X*, if there exists a* $\mathscr{T}_\infty$*-(uniformly-)continuous, surjective map* $f : B \cup V \to X$ *such that* $f(B) \cap f(V) = \varnothing$.

Definition 24.2 *The model is parametrised by* Θ*, if there exists a* $\mathscr{T}_\infty$*-continuous bijection* $P. : \Theta \to \mathscr{P}$ *(i.e. for every* $m \geq 1$ *and measurable* $f : \mathscr{X}^m \to [0,1]$*, the map* $\theta \mapsto \int f \, dP_\theta^m$ *is continuous).*

If Θ is compact, any parametrization is a homeomorphism, so the inverse gives rise to representations of testing problems in $\mathscr{P}$.

Characterisation of pointwise testability

Theorem 25.1 *Let $\mathscr{P}$ be a dominated model for i.i.d. data with disjoint B,V. The following are equivalent,*

i. there exists a pointwise test for B vs V,

ii. the problem has a representation $f : B \cup V \to X$ on a normal space X and there exist disjoint F_σ-sets $B', V' \subset X$ such that $f(B) \subset B', f(V) \subset V'$,

iii. the problem has a uniform representation $\psi : B \cup V \to X$ on a separable, metrizable space X with $\psi(B), \psi(V)$ both F_σ- and G_δ-sets.

Pointwise testability: corollaries (I)

Corollary 26.1 *Suppose that $\mathscr{P}$ is dominated and there exist disjoint F_σ-sets B', V' in the completion $\widehat{\mathscr{P}}$ (for $\mathscr{U}_\infty$) with $B \subset B'$, $V \subset V'$. Then B is pointwise testable versus V.*

Corollary 26.2 *Suppose that $\mathscr{P}$ is dominated and complete (for $\mathscr{U}_\infty$) with disjoint subsets B, V. Then B is pointwise testable versus V, if and only if, there exist disjoint F_σ-sets $B', V' \subset \mathscr{P}$ with $B \subset B'$, $V \subset V'$.*

Pointwise testability: corollaries (II)

Corollary 27.1 *Suppose that $\mathscr{P}$ is dominated and TV-totally-bounded. Then disjoint $B, V \subset \mathscr{P}$ are pointwise testable, if and only if, B, V are both F_σ- and G_δ-sets in $B \cup V$ (for $\mathscr{T}_{TV}$).*

Corollary 27.2 *Suppose that $\mathscr{P}$ is dominated by a probability measure, with a uniformly integrable family of densities. Then disjoint $B, V \subset \mathscr{P}$ are pointwise testable, if and only if, B, V are both F_σ- and G_δ-sets in $B \cup V$ (for $\mathscr{T}_C$).*

Pointwise testability: examples (I)

Example 28.1 *Is independence of two events A and B testable?*

Let $A, B \in \mathscr{B}$ be msb subsets. Consider,

$$H_0 : P(A \cap B) = P(A)P(B), \quad H_1 : P(A \cap B) \neq P(A)P(B).$$

Define $\mathscr{U}_1$-continuous $f_i : \mathscr{P} \to [0,1]$, $(i = 1, 2, 3)$,

$$f_1(P) = P(A \cap B), \quad f_2(P) = P(A), \quad f_3(P) = P(B),$$

and continuous $g : [0,1]^3 \to [1,-1]$, $g(x_1, x_2, x_3) = x_1 - x_2 x_3$. Now,

$$h : \mathscr{P} \to [0,1] : P \mapsto |g \circ (f_1, f_2, f_3)(P)|,$$

is $\mathscr{U}_1$-continuous. Then $B = h^{-1}(\{0\})$ is closed (for $\mathscr{T}_\infty$) and (since the complement V' is open in $[0,1]$, it is F_σ, so) $V = h^{-1}(V')$ is F_σ (for $\mathscr{T}_\infty$). So independence of events A and B is asymptotically testable.

Pointwise testability: examples (II)

Example 29.1 *Is independence of real-valued X and Y testable?*

Let $A_k \in \sigma_X, B_l \in \sigma_Y$ be generators. Consider,

$H_0 : \forall_{k,l}\, P(A_k \cap B_l) = P(A_k)P(B_l),\ H_1 : \exists_{k,l}\, P(A_k \cap B_l) \neq P(A_k)P(B_l).$

Define $\mathscr{U}_1$-continuous $f_{kl,i} : \mathscr{P} \to [0,1]$, $(i = 1,2,3)$,

$$f_{kl,1}(P) = P(A_k \cap B_l), \quad f_{k,2}(P) = P(A_k), \quad f_{l,3}(P) = P(B_l),$$

and continuous $g : [0,1]^3 \to [1,-1]$, $g(x_1,x_2,x_3) = x_1 - x_2x_3$. Now,

$$h : \mathscr{P} \to [0,1]^{\mathbb{N}} : P \mapsto (|g \circ (f_{kl,1}, f_{k,2}, f_{l,3})(P)| : k,l \geq 1),$$

is $\mathscr{U}_1$-continuous. Then $B = h^{-1}(\{0\})$ is closed (for $\mathscr{T}_\infty$) and (since the complement V' is open in $[0,1]^{\mathbb{N}}$, it is F_σ, so) $V = h^{-1}(V')$ is F_σ (for $\mathscr{T}_\infty$). So independence of X and Y is asymptotically testable.

Bayesian testability: equivalent formulations

Theorem 30.1 *Let a model* $(\mathscr{P}, \mathscr{G}, \Pi)$ *with* $B, V \in \mathscr{G}$ *be given, with* $\Pi(B) > 0, \Pi(V) > 0$. *The following are equivalent,*

i. there exist Bayesian tests for B *vs* V*,*

ii. there are tests ϕ_n *such that for* Π*-almost-all* $P \in B, Q \in V$*,*

$$P^n \phi_n \to 0, \quad Q^n(1 - \phi_n) \to 0,$$

iii. there are tests $\phi_n : \mathscr{X}^n \to [0, 1]$ *such that,*

$$\int_B P^n \phi_n \, d\Pi(P) + \int_V Q^n(1 - \phi_n) \, d\Pi(Q) \to 0,$$

iv. for Π*-almost-all* $P \in B$, $Q \in V$*,*

$$\Pi(V|X^n) \xrightarrow{P} 0, \quad \Pi(B|X^n) \xrightarrow{Q} 0.$$

Characterisation of Bayesian testability

Definition 31.1 *Given model* $(\mathscr{P}, \mathscr{G}, \Pi)$*. An event* $B \in \mathscr{B}^\infty$ *is called a* Π*-zero-one set, if* $P^\infty(B) \in \{0, 1\}$*, for* Π*-almost-all* $P \in \mathscr{P}$*. A model subset* $G \in \mathscr{G}$ *is called a* Π*-one set if there is a* Π*-zero-one set* B *such that* $G = \{P \in \mathscr{P} : P^\infty(B) = 1\}$*.*

Proposition 31.2 (Martingale convergence) *Let* $(\mathscr{P}, \mathscr{G}, \Pi)$ *be given. Let* V *be a* Π*-one set. Then, for* Π*-almost-all* $P \in \mathscr{P}$*,*

$$\Pi(V|X^n) \xrightarrow{P\text{-a.s.}} 1_V(P). \tag{5}$$

Theorem 31.3 *Let* $(\mathscr{P}, \mathscr{G})$ *be a measurable model with a prior* Π *that is a Radon measure and hypotheses* B, V*. There is a Bayesian test sequence for* B *vs* V*, if and only if,* B, V *are* $\mathscr{G}$*-measurable.*

Part III

Constructive results

Bayesian testing power

Denote the density for the local prior predictive distribution $P_n^{\Pi|B}$ with respect to $\mu_n = P_n^{\Pi|B} + P_n^{\Pi|V}$ by $p_{B,n}$, and similar for $P_n^{\Pi|V}$.

Proposition 33.1 *Let* $(\mathscr{P}, \mathscr{G}, \Pi)$ *be a model with measurable* B, V*. There are tests* ϕ_n *such that,*

$$\int_B P^n \phi_n \, d\Pi(P) + \int_V Q^n (1 - \phi_n) \, d\Pi(Q) \leq \int \Bigl(\Pi(B) \, p_{B,n}(x) \Bigr)^{\alpha} \Bigl(\Pi(V) \, p_{V,n}(x) \Bigr)^{1-\alpha} d\mu_n(x), \qquad (6)$$

for every $n \geq 1$ *and any* $0 \leq \alpha \leq 1$*.*

Proposition 33.2 *For every* $n \geq 1$*, the test,*

$$\phi_n(X^n) = 1\{X^n : \Pi(V|X^n) \geq \Pi(B|X^n)\},$$

based on posterior odds has optimal Bayesian testing power.

Posterior odds model selection for frequentists

Theorem 34.1 *For all $n \geq 1$, let the model be a probability space $(\mathscr{P}, \mathscr{G}, \Pi_n)$. Consider disjoint, measurable $B, V \subset \Theta$ with $\Pi_n(B), \Pi_n(V) > 0$ such that,*

i. There are Bayesian tests for B vs V of power $a_n \downarrow 0$,

$$\int_B P^n \phi_n \, d\Pi_n(P) + \int_V Q^n (1 - \phi_n) \, d\Pi_n(Q) = o(a_n),$$

ii. for all $P \in B$, $P^n \lhd a_n^{-1} P_n^{\Pi_n|B}$; for all $Q \in V$, $Q^n \lhd a_n^{-1} P_n^{\Pi_n|V}$.

Then the indicators $\phi_n(X^n) = 1\{X^n : \Pi(V|X^n) \geq \Pi(B|X^n)\}$ for posterior odds form a pointwise test sequence for B vs V.

See arXiv:1611.08444 [math.ST]

Remote contiguity

Definition 35.1 *Given* (P_n), (Q_n) *and a* $a_n \downarrow 0$, Q_n *is* a_n*-remotely contiguous w.r.t.* P_n ($Q_n \lhd a_n^{-1} P_n$), *if for any msb* $\psi_n : \mathscr{X}^n \to [0,1]$

$$P_n \psi_n = o(a_n) \quad \Rightarrow \quad Q_n \psi_n = o(1)$$

Lemma 35.2 $Q_n \lhd a_n^{-1} P_n$ *if any of the following holds:*

(i) *For any bnd msb* $T_n : \mathscr{X}^n \to \mathbb{R}$, $a_n^{-1} T_n \xrightarrow{P_n} 0$, *implies* $T_n \xrightarrow{Q_n} 0$

(ii) *Given* $\epsilon > 0$, *there is a* $\delta > 0$ *s.t.* $Q_n(dP_n/dQ_n < \delta\, a_n) < \epsilon$ *f.l.e.n.*

(iii) *There is a* $b > 0$ *s.t.* $\liminf_{n\to\infty} b\, a_n^{-1}\, P_n(dQ_n/dP_n > b\, a_n^{-1}) = 1$

(iv) *Given* $\epsilon > 0$, *there is a* $c > 0$ *such that* $\|Q_n - Q_n \wedge c\, a_n^{-1} P_n\| < \epsilon$

(v) *Under* Q_n, *every subsequence of* $(a_n(dP_n/dQ_n)^{-1})$ *has a further subsequence that converges in* $\mathscr{T}_C$.

Example: KL-neighbourhoods

Example 36.1 *Let* $\mathscr{P}$ *be a model for i.i.d. data* X^n. *Let* P_0, P *and* $\epsilon > 0$ *be such that* $-P_0 \log(dP/dP_0) < \epsilon^2$. *Then, for large enough* n,

$$\frac{dP^n}{dP_0^n}(X^n) \geq e^{-\frac{n}{2}\epsilon^2}, \tag{7}$$

with P_0^n*-probability one. So for any tests* ψ_n,

$$P^n \psi_n \geq e^{-\frac{1}{2}n\epsilon^2} P_0^n \psi_n. \tag{8}$$

So if $P^n\phi_n = o(\exp(-\frac{1}{2}n\epsilon^2))$ *then* $P_0^n\phi_n = o(1)$*:* $P_0^n \lhd a_n^{-1} P^n$ *with* $a_n = \exp(-\frac{1}{2}n\epsilon^2)$.

Consistent model selection

Let $\mathscr{P}$ be a model for *i.i.d.* data $X^n \sim P^n$, $(n \geq 1)$, and suppose that $(\mathscr{P}, \mathscr{G}, \Pi)$ has finite, measurable partition,

$$P \in \mathscr{P} = \mathscr{P}_1 \cup \ldots \cup \mathscr{P}_M.$$

Model-selection Which $1 \leq i \leq M$? (such that $P \in \mathscr{P}_i$)

Theorem 37.1 *Assume that for all $1 \leq i < j \leq M$,*

$\mathscr{P}_i$ and $\mathscr{P}_j$ are $\mathscr{U}_\infty$-uniformly separated.

Let $1 \leq i \leq M$ be such that $P \in \mathscr{P}_i$. If Π is a KL-prior, then indicators for posterior odds,

$$\phi_n(X^n) = 1\Big\{X^n : \Pi(\mathscr{P}_i|X^n) \geq \sum_{j \neq i} \Pi(\mathscr{P}_j|X^n)\Big\},$$

are a pointwise test for $\mathscr{P}_i$ vs $\cup_{j\neq i}\mathscr{P}_j$.

Example: select the DAG (I)

Observe an *i.i.d.* X^n of vectors of discrete random variables $X_i = (X_{1,i}, \ldots, X_{k,i}) \in \mathbb{Z}^k$. We assume that $X \sim P$ follows a graphical model,

$$P_{\mathscr{A},\theta}(X_1 \in B_1, \ldots, X_k \in B_k) = \prod_{i=1}^{k} P_{\theta_i}(X_i \in B_i | \mathscr{A}_i)$$

where $\mathscr{A}_i \subset \{1, \ldots, k\}$ denotes the *parents* of X_k. Together, the $\mathscr{A}_i$ describe a directed, a-cyclical graph.

Family $\mathscr{F}$ of kernels $p_\theta(\cdot|\cdot) : \mathbb{Z} \times \mathbb{Z}^l \to [0,1]$, for $\theta \in \Theta$, $1 \leq l \leq k$. Assume that Θ is compact and,

$$\theta \mapsto \sum_{x \in \mathbb{Z}} f(x) P_\theta(x | z_1, \ldots, z_l)$$

is continuous, for every bounded $f : \mathbb{Z} \to \mathbb{R}$ and all $z_1, \ldots, z_l \in \mathbb{Z}$.

Example: select the DAG (II)

The DAG $\mathscr{A} = (\mathscr{A}_i : 1 \leq i \leq k)$ represents a number of conditional independence statements concerning the components $X_1, \ldots, X_k$: for all $1 \leq i < j \leq k$, given $X_l = z$ for all $l \in \mathscr{A}_i \cup \mathscr{A}_j$, X_i is independent of X_j.

Define the submodels $\mathscr{P}_{\mathscr{A}} = \{P_{\mathscr{A},\theta} : \theta \in \Theta^k\}$, for all $\mathscr{A}$. Given a conditional independence relation for $\mathscr{A}$, we require that, for all θ, all $z \in \mathbb{Z}$, all $A, B \subset \mathbb{Z}$, any $\mathscr{A}' \neq \mathscr{A}$,

$$\Big| P_{\mathscr{A}',\theta}(X_i \in A, X_j \in B | X_l = z) \\ - P_{\mathscr{A}',\theta}(X_i \in A | X_l = z)\, P_{\mathscr{A}',\theta}(X_j \in B | X_l = z) \Big| > \epsilon,$$

for some $\epsilon > 0$ that depends only on $\mathscr{A}$ and $\mathscr{A}'$.

With a KL-prior posterior odds for $\mathscr{P}_{\mathscr{A}}$ select the correct DAG $\mathscr{A}$.

Example: how many clusters? (I)

Observe *i.i.d.* $X^n \sim P^n$, where P dominated with density p.

Clusters Family $\mathscr{F}$ of kernels $\varphi_\theta : \mathbb{R} \to [0, \infty)$, with parameter $\theta \in \Theta$. Assume Θ compact and,

$$\theta \mapsto \int f(x) \varphi_\theta(x)\, dx,$$

is continuous, for every bounded, measurable $f : \mathbb{R} \to \mathbb{R}$. Define $\Theta'_M = \Theta^M / \sim$.

Model Assume that there is an $M > 0$ such that p can be written as,

$$p_{\lambda,\theta}(x) = \sum_{m=1}^{M} \lambda_m p_{\theta_m}(x),$$

for some $M \geq 1$, with $\lambda \in S_M = \{\lambda \in [0,1]^M : \sum_m \lambda_m = 1\}$, $\theta \in \Theta'_M$.

Example: how many clusters? (II)

Assume M less than some known M'. Choose prior $\Pi_{\lambda,M}$ for $\lambda \in S_M$ such that, for some $\epsilon > 0$,

$$\Pi_{\lambda,M}\Big(\lambda \in S_M : \epsilon < \min\{\lambda_m\}, \max\{\lambda_m\} < 1 - \epsilon\Big) = 1.$$

For $\theta \in \Theta'_M$ also choose a prior $\Pi_{\theta,M}$ that 'stays away from the edges'. Define,

$$\Pi = \sum_{M=1}^{M'} \mu_M \, \Pi_{\lambda,M} \times \Pi_{\theta,M}.$$

(for $\sum_M \mu_M = 1$).

If Π is a KL-prior, posterior odds select the correct number of clusters M. If there are no M' and ϵ known, there are sequences $M'_n \to \infty$ and $\epsilon_n \downarrow 0$ with priors Π_n that finds the correct number of clusters.