Testability and consistency

Bas Kleijn, KdV Institute for Mathematics

Universiteit van Amsterdam
Part I

Introduction
Notation and conventions

<table>
<thead>
<tr>
<th>Sample space</th>
<th>$(X, \mathcal{B})$</th>
<th>Measurable space</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i.i.d.$ data</td>
<td>$X_n = (X_1, \ldots, X_n) \in \mathcal{X}^n$</td>
<td>Frequentist/Bayesian</td>
</tr>
<tr>
<td>Model</td>
<td>$(\mathcal{P}, \mathcal{G})$</td>
<td>Model subsets $B, V \in \mathcal{G}$</td>
</tr>
<tr>
<td>Prior</td>
<td>$\Pi : \mathcal{G} \rightarrow [0, 1]$</td>
<td>Probability measure</td>
</tr>
<tr>
<td>Posterior</td>
<td>$\Pi(\cdot</td>
<td>X_n) : \mathcal{G} \rightarrow [0, 1]$</td>
</tr>
<tr>
<td>Parameter</td>
<td>$\Theta \rightarrow \mathcal{P} : \theta \mapsto P_\theta$</td>
<td>Define prior on $\Theta$</td>
</tr>
<tr>
<td>Test function</td>
<td>$\phi_n : \mathcal{X}^n \rightarrow [0, 1]$</td>
<td>Statistical discrimination</td>
</tr>
</tbody>
</table>

Frequentism: assume $\underline{X}_n \sim P^0_n$
Posterior consistency for hypotheses

**Definition 4.1** Let $B, V$ be measurable model subsets. The posterior is consistent for testing $B$ versus $V$ if,

$$
\Pi(V|X_n) \xrightarrow{P-a.s.} 0, \quad \Pi(B|X_n) \xrightarrow{Q-a.s.} 0,
$$

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

**Remark 4.2** ... or in a Bayesian version:

for $\Pi$-almost-all $P \in B, Q \in V$. 

Posterior consistency

Definition 5.1 Given a model $\mathcal{P}$ with topology and a Borel prior $\Pi$, the posterior is consistent at $P \in \mathcal{P}$ if for every open nbd $U$ of $P$,

$$\Pi(U|X_n) \xrightarrow{P-a.s.} 1.$$
Doob’s prior-almost-sure consistency

**Theorem 6.1** *(Doob (1948))*

Let $\Theta$ and $\mathcal{X}$ be Polish spaces. Assume that the parametrization is one-to-one. Then for any prior $\Pi$ on $\Theta$ the posterior is consistent at $P$, for $\Pi$-almost-all $P \in \mathcal{P}$.

**Remark 6.2** *(Schwartz (1961), Freedman (1963))*

*Bayesian interpretation? Yes*

*Frequentist interpretation? No*

*The exceptional $\Pi$-nullset can be unexpectedly large*
Hellinger consistency with KL priors

**Theorem 7.1** *(Schwartz (1965))*

Let $X_1, X_2, \ldots$ be an i.i.d.-sample from $P_0 \in \mathcal{P}$. Let $\mathcal{P}$ be Hellinger totally bounded and let $\Pi$ be such that,

$$\Pi \left( P : -P_0 \log \frac{dP}{dP_0} < \epsilon \right) > 0,$$

for all $\epsilon > 0$. Then the posterior is consistent at $P_0$ in the Hellinger topology.

**Lemma 7.2** *(Existence of Hellinger tests)* Let $V \subset \mathcal{P}$ be convex with $H(P_0, V) > 0$. There exist a $D > 0$ and uniform test sequence $\phi_n : \mathcal{X}^n \rightarrow [0, 1]$ s.t.

$$P_0^n \phi_n \leq e^{-nD}, \quad \sup_{Q \in V} Q^n (1 - \phi_n) \leq e^{-nD}.$$
The Dirichlet process

**Definition 8.1** *(Dirichlet distribution)*

A random variable \( p = (p_1, \ldots, p_k) \) with \( p_l \geq 0 \) and \( \sum_l p_l = 1 \) is Dirichlet distributed with parameter \( \alpha = (\alpha_1, \ldots, \alpha_k) \), \( p \sim D_\alpha \), if it has density,

\[
f_\alpha(p) = C(\alpha) \prod_{l=1}^{k} p_l^{\alpha_l-1}.
\]

**Definition 8.2** *(Dirichlet process, Ferguson 1973-74)*

Let \( \alpha \) be a finite measure on \((\mathcal{X}, \mathcal{B})\). The Dirichlet process \( P \sim D_\alpha \) is defined by, (for all msb partitions \( \mathcal{X} = A_1 \cup \ldots \cup A_k \)),

\[
\left( P(A_1), \ldots, P(A_k) \right) \sim D(\alpha(A_1), \ldots, \alpha(A_k))
\]
Weak consistency with Dirichlet priors

**Theorem 9.1 (Dirichlet consistency)**
Let $X_1, X_2, \ldots$ be an i.i.d.-sample from $P_0$ If $\Pi$ is a Dirichlet prior $D_\alpha$ with finite $\alpha$ such that $\text{supp}(P_0) \subset \text{supp}(\alpha)$, the posterior is consistent at $P_0$ in the weak model topology.

**Remark 9.2** Dirichlet priors are not KL.

**Remark 9.3** Several proofs. Dirichlet distributions are tailfree.
Part II
Doob tests and prior-a.s.-consistency
Posterior concentration lemma

**Definition 11.1** Let $(\mathcal{P}, \mathcal{G}, \Pi)$ and $B, V \in \mathcal{G}$ be given. Tests $(\phi_n)$ form a sequence of Doob tests for $B$ versus $V$ (under $\Pi$), if,

\[
\int_B P^n \phi_n \, d\Pi(P) \to 0, \quad \int_V Q^n (1 - \phi_n) \, d\Pi(Q) \to 0.
\]

**Lemma 11.2** Let the model $\mathcal{P}$ and prior $\Pi$ be given. For any $B, V \in \mathcal{G}$ such that $\Pi(B) > 0$ and any test sequence $(\phi_n)$,

\[
\int P^n \Pi(V|X_n) \, d\Pi(P|B) \leq \frac{1}{\Pi(B)} \left(\int_B P^n \phi_n \, d\Pi(P) + \int_V Q^n (1 - \phi_n) \, d\Pi(Q)\right).
\]
Doob tests and the posterior

**Theorem 12.1** Let \((\mathcal{P}, \mathcal{G}, \Pi)\) be given. For any \(B, V \in \mathcal{G}\), the following are equivalent,

(i) There exist Doob tests \((\phi_n)\) for \(B\) versus \(V\) under \(\Pi\);

(ii) The posterior satisfies \(\Pi(V|X_n) \xrightarrow{P-a.s.} 0\) and \(\Pi(B|X_n) \xrightarrow{Q-a.s.} 0\), for \(\Pi\)-almost-all \(P \in B, Q \in V\).

**Remark 12.2** Loose interpretation: distinctions between model subsets are Doob testable under \(\Pi\), if and only if they are picked up by the posterior asymptotically.
Prior-almost-sure consistency

**Theorem 13.1** Let Hausdorff $\mathcal{P}$ with Borel prior $\Pi$ be given. Assume that for $\Pi$-almost-all $P \in \mathcal{P}$ and any open nbd $U$ of $P$, there exist a $B \subset U$ with $\Pi(B) > 0$ and Doob tests $(\phi_n)$ for $B$ versus $\mathcal{P} \setminus U$. Then the posterior is consistent at $\Pi$-almost-all $P \in \mathcal{P}$.

**Theorem 13.2** Let $\ell$ be a loss function that is bounded, continuous and convex. If the posterior is consistent for $\Pi$-almost-all $P \in \mathcal{P}$, the posterior mean $(\hat{P}_n)$ is asymptotically Bayes admissible.
Part III

Existence of tests
Existence of Doob tests

**Definition 15.1 (Breiman, Le Cam, Schwartz (1964))**

Let \((\mathcal{P}, \mathcal{G}, \Pi)\) be given. An event \(B \in \mathcal{B}(\infty)\) is called an \(\Pi\)-zero-one set if \(P^\infty(B) = P^\infty(B)^2\) for \(\Pi\)-almost-all \(P \in \mathcal{P}\).

A model subset \(G \in \mathcal{G}\) is called a \(\Pi\)-one set if there exists a \(\Pi\)-zero-one set \(B\) such that \(G = \{P \in \mathcal{P} : P^\infty(B) = 1\}\).

**Theorem 15.2** Given \((\mathcal{P}, \mathcal{G}, \Pi)\), let \(V\) be a \(\Pi\)-one set. Then there exist Doob tests for \(V\) versus \(\mathcal{P}\setminus V\) under \(\Pi\) and \(\Pi(V\mid X_n) \xrightarrow{P\text{-a.s.}} 1_V(P)\) for \(\Pi\)-almost-all \(P \in \mathcal{P}\).
Polish spaces, tests and Doob’s theorem

Lemma 16.1 (Le Cam (1986))
If $\mathcal{P}$ is a Polish space, all Borel sets are $\Pi$-one sets.

Corollary 16.2 Let the model $\mathcal{P}$ be Polish with Borel prior $\Pi$. For any Borel set $V$ there exist Doob tests for $V$ versus $\mathcal{P} \setminus V$ under $\Pi$ and $\Pi(V|X_n) \xrightarrow{P-a.s.} 1_V(P)$ for $\Pi$-almost-all $P \in \mathcal{P}$.

Corollary 16.3 (Doob’s consistency theorem)

Remark 16.4 Schwartz’s models and the Dirichlet model are Polish.
Examples of uniform test sequences

**Lemma 17.1 (Minimax Hellinger tests)** Let $B, V \subset \mathcal{P}$ be convex with $H(B, V) > 0$. There exist a $D > 0$ and uniform test sequence $(\phi_n)$ s.t.

$$\sup_{P \in B} P^n \phi_n \leq e^{-nD}, \quad \sup_{Q \in V} Q^n (1 - \phi_n) \leq e^{-nD}.$$

**Lemma 17.2 (Weak-\ast tests)** Let $n \geq 1$, $\epsilon > 0$, $P_0 \in \mathcal{P}$ and a msb $f : \mathcal{X}^n \to [0, 1]$ be given. Define,

$$B = \left\{ P \in \mathcal{P} : |(P^n - P_0^n)f| < \epsilon \right\}, \quad V = \left\{ P \in \mathcal{P} : |(P^n - P_0^n)f| \geq 2\epsilon \right\}.$$

There exist a $D > 0$ and uniform test sequence $(\phi_n)$ s.t.

$$\sup_{P \in B} P^n \phi_n \leq e^{-nD}, \quad \sup_{Q \in V} Q^n (1 - \phi_n) \leq e^{-nD}.$$
Doob tests versus uniform tests, examples I

Example 18.1  (Cover’s rational mean problem (Cover, 1973))

Bernoulli($p$) i.i.d. coin tosses $X_1, X_2, \ldots$, $p \in [0, 1]$. Hypotheses

\[ H_0 : p \in \mathbb{Q}, \quad \text{versus} \quad H_1 : p \in \mathbb{R} \setminus \mathbb{Q}. \]

Prior mixes discrete $\Pi_0$ on $\mathbb{Q}$ and Lebesgue on $[0, 1]$. There are Doob tests but no uniform tests.

Example 18.2  (The $\sqrt{2}$-test (Dembo and Peres, 1994))

Same as Cover’s problem, but with hypotheses

\[ H_0 : p \in \mathbb{Q}, \quad \text{versus} \quad H_1 : p \in \sqrt{2} + \mathbb{Q}. \]

Prior mixes discrete $\Pi_0$ on $\mathbb{Q}$ and discrete $\Pi_1$ on $\sqrt{2} + \mathbb{Q}$. There are Doob tests but no uniform tests.
Doob tests versus uniform tests, examples II

**Example 19.1** (*Integrability with densities (Devroye, Lugosi, 1999))*

Sample $X_1, X_2, \ldots$ i.i.d. $P_0 \ll \mu$. For some measurable $h : \mathbb{R} \to \mathbb{R}$,

$$H_0 : \int |h| p_0 \, d\mu < \infty, \quad \text{versus} \quad H_1 : \int |h| p_0 \, d\mu = \infty.$$ 

With $\Pi(H_{0,1}) > 0$, there are *Doob tests but no uniform tests*.

**Example 19.2** (*Distinction between Sobolev classes in regression*)

Sample $(X_1, Y_1), (X_2, Y_2), \ldots$ i.i.d. $Y = f(X) + e$. ($\alpha > \beta$, $1 \leq p < \infty$)

$$H_0 : f \in W^{\alpha,p}([0, 1]), \quad \text{versus} \quad H_1 : f \in W^{\beta,p}([0, 1]) \setminus W^{\alpha,p}([0, 1]).$$

Prior mixes Gaussians on $W^{\alpha,p}([0, 1])$ and $W^{\beta,p}([0, 1])$. *Doob tests can be found but no uniform tests.*
Example 20.1  (*Distinctions between classes of graphs*) Θ is the set of all countable graphs, in which each vertex has exactly one parent vertex (possibly itself). Θ is the (Polish) space $\mathbb{N}^\mathbb{N}$. Let $\Theta \mapsto \mathcal{P}$ be a one-to-one model parametrization.

\[ B = \{ \theta \in \Theta : \theta \text{ finite forest} \}, \quad V = \{ \theta \in \Theta : \theta \text{ cyclic or infinite} \} \]

Then $\prod(V|X_n) \xrightarrow{P-a.s.} 0$, for $\prod$-almost-all $P \in B$ (and vice versa).
Part IV

Frequentism
Schwartz’s theorem revisited I

Remark 22.1 Suppose that for all $\delta > 0$, there is a $\mathcal{B}$ s.t. $\Pi(\mathcal{B}) > 0$ and for all $P \in \mathcal{B}$ and large enough $n$,

$$P^n \Pi(V|X_n) \geq P_0^n \Pi(V|X_n) e^{-n\delta},$$

then (by Fatou) for large enough $m$,

$$\sup_{n \geq m} \left[ P_0^n \Pi(V|X_n) - e^{n\delta} \int P^n \Pi(V|X_n) \ d\Pi(P|\mathcal{B}) \right] \leq 0,$$

Theorem 22.2 Let $\mathcal{P}$ be a model with KL-prior $\Pi$. Let $P_0 \in \mathcal{P}$ be given. Let $B, V \in \mathcal{G}$ be given and assume that $B$ contains a KL-neighbourhood of $P_0$. If there exist Doob tests for $B$ versus $V$ of exponential power then,

$$\Pi(V|X_n) \xrightarrow{P_0-a.s.} 0.$$
Schwartz’s theorem revisited II

**Theorem 23.1** Let $\mathcal{P}$ be Hausdorff with Borel prior $\Pi$ and let $P_0 \in \mathcal{P}$ be given. Assume that,

(i) For any $U$ open nbd of $P_0$, $U$ contains a KL-nbd of $P_0$;
(ii) For every open nbd $V$ of $P_0$, there is an open nbd $U \subset V$ of $P_0$ and Doob tests for $U$ versus $\mathcal{P} \setminus V$ of exponential power.

If $\Pi$ is a KL-prior, then the posterior is consistent at $P_0$.

**Corollary 23.2** *(Schwartz's theorem)*

**Corollary 23.3** Let $\mathcal{P}$ with Borel KL prior $\Pi$ be given. Then the posterior is weak-$*$-consistent.
Weak consistency

Let $\mathcal{A}$ denote the set of all partitions $A = \{A_1, \ldots, A_N\}$ of $\mathcal{X} = [0, 1]$ into equal-sized intervals and $N(A) = N$.

**Theorem 24.1** Let $\mathcal{P}$ be the full model on $([0, 1], \mathcal{B})$ (with the weak topology and a Borel prior $\Pi$). Let $U$ be an open nbd of $P_0$. Assume,

(i) For every $n \geq 1$, $x_n \mapsto \Pi(U|X_n = x_n)$ is continuous,

(ii) For every $\epsilon > 0$,

$$\inf_{A \in \mathcal{A}} \Pi \left( P \in U : \sum_{k=1}^{N(A)} P_0(A_k) \log \frac{P_0(A_k)}{P(A_k)} < \epsilon \right) > 0.$$ 

Then $\Pi(U|X_n) \xrightarrow{P_0\text{-a.s.}} 1$. 

24
Conjectures on the Dirichlet prior

**Conjecture 25.1** If the prior $\Pi$ is Dirichlet $D_\alpha$ with finite $\alpha$ such that $\text{supp}(P_0) \subset \text{supp}(\alpha)$, then condition (ii) is satisfied.

**Conjecture 25.2** If the prior $\Pi$ is Dirichlet $D_\alpha$, condition (i) holds.

**Conjecture 25.3** Tailfree-ness of the posterior can replace condition (i).
Open questions

Remark 26.1 What can we show weak consistency for Polya-tree, Pitman-Yor (or Gibbs-type) and other families of priors?

Remark 26.2 Different constructions, or variations on thm 24.1?

Remark 26.3 Which hypotheses $B,V$ can the posterior distinguish in $P_0$-probability? (Think of the Dembo-Peres $\sqrt{2}$-test)

Lemma 26.4 Let $a_n \downarrow 0$; let $(\phi_n)$ be Doob tests for $B$ versus $V$ of power $(a_n)$. Let $\Pi$ be such that, for every (continuous, tailfree?) $f_n : \mathcal{X}^n \to [0,1]$, there exists a $(b_n)$, $b_n \downarrow 0$ such that,

$$a_n^{-1} \Pi(\{P \in B : P^n f_n \geq P_0^n f_n - b_n\}) \to \infty.$$

Then $\Pi(V|X_n) \xrightarrow{P_0} 0$. 

The weak-* topology $\mathcal{T}_\infty$

Uniformity $\mathcal{U}_n$: basis is finite intersections of $W_f$'s,

$$W_f = \{(P, Q) : |(P^n - Q^n)f| < \epsilon\}, \quad \text{(bnd msb } f : X^n \to [0, 1]).$$

and $\mathcal{U}_\infty = \bigcup_n \mathcal{U}_n \subset \mathcal{U}_H$. Weak $\mathcal{U}_C \subset \mathcal{U}_1$ for (bnd cont $f : X \to [0, 1])$.

$(\mathcal{P}, \mathcal{U}_\infty)$ is tot bounded. $\mathcal{T}_C \subset \mathcal{T}_1 \subset \mathcal{T}_n \subset \mathcal{T}_\infty \subset \mathcal{T}_H$ are $T_{3.5}$.

$(\mathcal{P}, \mathcal{T}_1)$ sep $\iff (\mathcal{P}, \mathcal{T}_\infty)$ sep $\iff (\mathcal{P}, \mathcal{T}_H)$ sep $\iff \mathcal{P}$ is dominated

**Theorem 27.1 (Dunford-Pettis)**

$(\mathcal{P}, \mathcal{T}_1)$ is relatively compact iff there exists a $P$ such that $\mathcal{P} \ll P$ and $\{dQ/dP : Q \in \mathcal{P}\}$ is UI($\mathcal{P}$). In that case $\mathcal{U}_1 = \mathcal{U}_\infty$. 

27