On the frequentist validity of Bayesian limits

arXiv:1611.08444 [math.ST]

Bas Kleijn, KdV Institute for Mathematics
Part I
Introduction
## Bayesian and Frequentist statistics

<table>
<thead>
<tr>
<th>Sample spaces</th>
<th>((\mathcal{X}_n, \mathcal{B}_n))</th>
<th>Prob msr’s (M^1(\mathcal{X}_n))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td>(X^n = (X_1, \ldots, X_n) \in \mathcal{X}_n)</td>
<td>Sequential experiment</td>
</tr>
<tr>
<td><strong>Parameter space</strong></td>
<td>((\Theta, \mathcal{G}))</td>
<td>If i.i.d.: ((\mathcal{P}, \mathcal{G}))</td>
</tr>
<tr>
<td><strong>Parameter</strong></td>
<td>(\theta \in \Theta)</td>
<td>If i.i.d.: (P \in \mathcal{P})</td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>(\Theta \to M^1(\mathcal{X}<em>n): \theta \mapsto P</em>{\theta, n})</td>
<td>Not always i.i.d.</td>
</tr>
<tr>
<td><strong>Priors</strong></td>
<td>(\Pi_n: \mathcal{G} \to [0, 1])</td>
<td>Probability measure</td>
</tr>
<tr>
<td><strong>Posterior</strong></td>
<td>(\Pi(\cdot</td>
<td>X^n): \mathcal{G} \to [0, 1])</td>
</tr>
</tbody>
</table>

Frequentist: assume there is \(P_0\) \(X^n \sim P_0^n\)

Bayes: assume \(P \sim \Pi\) \(X^n | P \sim P^n\)
Definition of the posterior

Definition 4.1 Assume that all \( \theta \mapsto P_{\theta,n}(A) \) are \( \mathcal{G} \)-measurable. Fix \( n \geq 1 \). Given prior \( \Pi_n \), a posterior is any \( \Pi(\cdot|X^n = \cdot) : \mathcal{G} \times \mathcal{X}_n \to [0, 1] \)

(i) For any \( G \in \mathcal{G} \), \( x^n \mapsto \Pi(G|X^n = x^n) \) is \( \mathcal{B}_n \)-measurable

(ii) (Disintegration) For all \( A \in \mathcal{B}_n \) and \( G \in \mathcal{G} \)

\[
\int_A \Pi(G|X^n) \ dP^n \Pi = \int_G P_{\theta,n}(A) \ d\Pi_n(\theta)
\]

where \( P^n = \int P_{\theta,n} \ d\Pi_n(\theta) \) is the prior predictive distribution

Remark 4.2 For frequentists \( X^n \sim P_{0,n} \), so assume \( P_{0,n} \ll P^n \)
Asymptotic consistency of the posterior

**Definition 5.1** Given $\Theta$ *(Hausdorff completely regular)* and a Borel prior $\Pi$, the posterior is consistent at $\theta \in \Theta$ if for every nbd $U$ of $\theta$

$$\Pi(U|X^n) \xrightarrow{P} 1$$
The i.i.d. consistency theorems (I)

**Theorem 6.1** (Bayesian, Doob (1948))

Let $\mathcal{P}$ and $\mathcal{X}$ be Polish spaces and let $\Pi$ be a Borel prior. Assume that $P \mapsto P_n(A)$ is Borel measurable for all $n, A$. Then the posterior is consistent at $P$, for $\Pi$-almost-all $P \in \mathcal{P}$.

**Example 6.2** For some $Q \in \mathcal{P}$, take $\Pi = \delta_Q$. Then $\Pi(\cdot|X^n) = \delta_Q$ as well, $P_n^\Pi$-almost-surely. If $X_1, \ldots, X_n \sim P^\Pi_0$ (require $P^\Pi_0 \ll P_n^\Pi = Q^n$), the posterior is not frequentist consistent.

Non-trivial counterexamples are due to Schwartz (1961) and Freedman (1963, 1965, ...)
The i.i.d. consistency theorems (II)

**Theorem 7.1** (*Frequentist, Schwartz (1965))

Let \( X_1, X_2, \ldots \) be i.i.d.-\( P_0 \) for some \( P_0 \in \mathcal{P} \). If,

(i) For every nbd \( U \) of \( P_0 \), there are \( \phi_n : X_n \to [0, 1] \), s.t.

\[
P_0^n \phi_n = o(1), \quad \sup_{Q \in U^c} Q^n (1 - \phi_n) = o(1),
\]

(ii) and \( \Pi \) is a Kullback-Leibler prior, i.e. for all \( \delta > 0 \),

\[
\Pi \left( P \in \mathcal{P} : -P_0 \log \frac{dP}{dP_0} < \delta \right) > 0,
\]

then \( \Pi(U|X^n) \xrightarrow{P_0\text{-a.s.}} 1. \)
The Dirichlet process

Definition 8.1 (Dirichlet distribution)  
A $p = (p_1, \ldots, p_k)$ 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, 1974)  
Let $\mathcal{X}$ be Polish and let $\alpha$ be a finite Borel measure on $(\mathcal{X}, \mathcal{B})$. The Dirichlet process $P \sim D_\alpha$ is defined by,

$$\left(P(A_1), \ldots, P(A_k)\right) \sim D_{(\alpha(A_1), \ldots, \alpha(A_k))}$$
The i.i.d. consistency theorems (III)

**Theorem 9.1** *(Frequentist, 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** *(Freedman (1963))*

Dirichlet priors are tailfree: if \( A' \) refines \( A \) and \( A'_i \cup \ldots \cup A'_{i_1} = A_i \), then \( (P(A'_{i_1}|A_i), \ldots, P(A'_{i_1}|A_i) : 1 \leq i \leq k) \) is independent of \( (P(A_1), \ldots, P(A_k)) \).

**Remark 9.3** \( X^n \mapsto \Pi(P(A)|X^n) \) is \( \sigma_n(A) \)-measurable where \( \sigma_n(A) \) is generated by products of the form \( \prod_{i=1}^n B_i \) with \( B_i = \{X_i \in A\} \) or \( B_i = \{X_i \notin A\} \).
Part II

Bayesian test sequences
Bayesian and Frequentist testability

For $B, V$ be two (disjoint) model subsets

**Definition 11.1** *Uniform* (or *minimax*) *testability*

$$\sup_{\theta \in B} P_{\theta,n} \phi_n \to 0, \quad \sup_{\theta \in V} P_{\theta,n} (1 - \phi_n) \to 0$$

**Definition 11.2** *Pointwise testability* for all $\theta \in B, \eta \in V$

$$\phi_n \xrightarrow{P_{\theta,n}} 0, \quad \phi_n \xrightarrow{P_{\eta,n}} 1$$

**Definition 11.3** *Bayesian testability* for $\Pi$-*almost-all* $\theta \in B, \eta \in V$

$$\phi_n \xrightarrow{P_{\theta,n}} 0, \quad \phi_n \xrightarrow{P_{\eta,n}} 1$$
Examples of uniform test sequences

**Lemma 12.1** *(Uniform Hellinger tests)* Let $B, V \subseteq \mathcal{P}$ 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 12.2** *(Uniform weak tests)* Let $n \geq 1$, $\epsilon > 0$, $P_0 \in \mathcal{P}$ and a msb $f : \mathbb{X}^n \to [0, 1]$ 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.

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

**Lemma 13.1** Let \((\mathcal{P}, \mathcal{G})\) be given. For any prior \(\Pi\), any test function \(\phi\) and any \(B, V \in \mathcal{G}\),

\[
\int_B P\Pi(V|X) \, d\Pi(P) \leq \int_B P\phi \, d\Pi(P) + \int_V Q(1 - \phi) \, d\Pi(Q)
\]

**Corollary 13.2** Consequently, for any sequences \((\Pi_n), (B_n), (V_n)\) such that \(B_n \cap V_n = \emptyset\) and \(\Pi_n(B_n) > 0\), we have,

\[
P^n_{\Pi_n|B_n}(V_n|X^n) := \int P_{\theta, n} \Pi(V_n|X^n) \, d\Pi_n(\theta|B_n)
\]

\[
\leq \frac{1}{\Pi_n(B_n)} \left( \int_{B_n} P_{\theta, n} \phi_n \, d\Pi_n(\theta) + \int_{V_n} P_{\theta, n}(1 - \phi_n) \, d\Pi_n(\theta) \right)
\]
Proposition 14.1 Let \((\Theta, \mathcal{G}, \Pi)\) be given. For any \(B, V \in \mathcal{G}\), the following are equivalent,

(i) There exist Bayesian tests \((\phi_n)\) for \(B\) versus \(V\);

(ii) There exist tests \((\phi_n)\) such that,

\[
\int_B P_{\theta,n} \phi_n \, d\Pi(\theta) + \int_V P_{\theta,n} (1 - \phi_n) \, d\Pi(\theta) \to 0,
\]

(iii) For \(\Pi\)-almost-all \(\theta \in B, \eta \in V\),

\[
\Pi(V|X^n) \xrightarrow{P_{\theta,n}} 0, \quad \Pi(B|X^n) \xrightarrow{P_{\eta,n}Q} 0
\]

Remark 14.2 Interpretation distinctions between model subsets are Bayesian testable, iff they are picked up by the posterior asymptotically, iff, the Bayes factor for \(B\) versus \(V\) is consistent.
Prior-almost-sure consistency

**Corollary 15.1** Let Hausdorff completely regular $\Theta$ with Borel prior $\Pi$ be given. Then the following are equivalent,

(i) for $\Pi$-almost-all $\theta \in \Theta$ and any nbd $U$ of $\theta$ there exist a msb $B \subset U$ with $\Pi(B) > 0$ and Bayesian tests $(\phi_n)$ for $B$ vs $V = \Theta \setminus U$,

(ii) the posterior is consistent at $\Pi$-almost-all $\theta \in \Theta$.

**Remark 15.2** Let $\mathcal{P}$ be a Polish space and assume that all $P \mapsto P^n(A)$ are Borel measurable. Then, for any prior $\Pi$, any Borel set $V \subset \mathcal{P}$ is Bayesian testable versus $\mathcal{P} \setminus V$.

**Corollary 15.3** (More than) Doob’s 1948 theorem
Part III
Remote contiguity
Le Cam’s inequality

**Definition 17.1** For $B \in \mathcal{G}$ such that $\prod_n(B) > 0$, the local prior predictive distribution is $P_n^{\prod|B} = \int P_{\theta,n} d\prod_n(\theta|B)$.

**Remark 17.2** (Le Cam, unpublished (197?) and (1986))

Rewrite the posterior concentration inequality

$$P_0^n \cap(V_n|X^n) \leq \left\| P_0^n - P_n^{\prod|B_n} \right\| + \int P^n \phi_n d\prod(P|B_n) + \frac{\prod(V_n)}{\prod(B_n)} \int Q^n(1 - \phi_n) d\prod(Q|V_n)$$

**Remark 17.3** Useful in parametric models (e.g. BvM) but “a considerable nuisance” [sic] (Le Cam (1986)) in non-parametric context.
Schwartz’s theorem revisited

Remark 18.1 Suppose that for all $\delta > 0$, there is a $B$ s.t. $\Pi(B) > 0$ and for $\Pi$-almost-all $\theta \in B$ and large enough $n$

$$P_0^n \Pi(V|X^n) \leq e^{n\delta} P_{\theta,n} \Pi(V|X^n)$$

then (by Fatou) for large enough $m$

$$\limsup_{n \to \infty} \left[ (P_0^n - e^{n\delta} P_{n|^B}) \Pi(V|X^n) \right] \leq 0$$

Theorem 18.2 Let $\mathcal{P}$ be a model with KL-prior $\Pi$; $P_0 \in \mathcal{P}$. Let $B, V \in \mathcal{G}$ be given and assume that $B$ contains a KL-neighbourhood of $P_0$. If there exist Bayesian tests for $B$ versus $V$ of exponential power then

$$\Pi(V|X^n) \xrightarrow{P_0-a.s.} 0$$

Corollary 18.3 (Schwartz’s theorem)
Remote contiguity

**Definition 19.1** Given \((P_n), (Q_n)\) of prob msr’s, \(Q_n\) is **contiguous** w.r.t. \(P_n\) \((Q_n \triangleleft P_n)\), if for any msb \(\psi_n : \mathcal{X}^n \rightarrow [0, 1]\)

\[ P_n\psi_n = o(1) \implies Q_n\psi_n = o(1) \]

**Definition 19.2** Given \((P_n), (Q_n)\) of prob msr’s and a \(a_n \downarrow 0\), \(Q_n\) is **\(a_n\)-remotely contiguous** w.r.t. \(P_n\) \((Q_n \triangleleft a_n^{-1}P_n)\), if for any msb \(\psi_n : \mathcal{X}^n \rightarrow [0, 1]\)

\[ P_n\psi_n = o(a_n) \implies Q_n\psi_n = o(1) \]

**Remark 19.3** Contiguity is stronger than remote contiguity
note that \(Q_n \triangleleft P_n\) iff \(Q_n \triangleleft a_n^{-1}P_n\) for all \(a_n \downarrow 0\).

**Definition 19.4** **Hellinger transform** \(\psi(P, Q; \alpha) = \int p^\alpha q^{1-\alpha} \, d\mu\)
Le Cam’s first lemma

Lemma 20.1 Given \((P_n), (Q_n)\) like above, \(Q_n \preceq P_n\) iff:

(i) If \(T_n \xrightarrow{P_n} 0\), then \(T_n \xrightarrow{Q_n} 0\)

(ii) Given \(\epsilon > 0\), there is a \(b > 0\) such that \(Q_n(dQ_n/dP_n > b) < \epsilon\)

(iii) Given \(\epsilon > 0\), there is a \(c > 0\) such that \(||Q_n - Q_n \wedge cP_n|| < \epsilon\)

(iv) If \(dP_n/dQ_n \xrightarrow{Q_n-w.} f\) along a subsequence, then \(P(f > 0) = 1\)

(v) If \(dQ_n/dP_n \xrightarrow{P_n-w.} g\) along a subsequence, then \(Eg = 1\)

(vi) \(\lim \inf_n \psi(P_n, Q_n; \alpha) \to 1\) as \(\alpha \uparrow 1\)
Criteria for remote contiguity

Lemma 21.1 Given \((P_n)\), \((Q_n)\), \(a_n \downarrow 0\), \(Q_n \prec a_n^{-1} P_n\) if any of the following holds:

(i) For any bnd msb \(T_n : \mathcal{X}^n \rightarrow \mathbb{R}\), \(a_n^{-1} T_n \xrightarrow{P_n} 0\), 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. \(\liminf_{n \rightarrow \infty} b a_n^{-1} P_n(dQ_n/dP_n > b a_n^{-1}) = 1\)

(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 \(Q_n\), \((a_n dQ_n/dP_n)\) are r.v.’s and every subseq has a weakly convergent subseq

(vi) \(\lim_{\alpha \uparrow 1} \liminf_n a_n^{-\alpha} \psi(P_n, Q_n; \alpha) > 0\)
Part IV

Frequentist consistency
Beyond Schwartz

**Theorem 23.1** Let \((\Theta, \mathcal{G}, \Pi)\) and \((X_1, \ldots, X_n) \sim P_{0,n}\) be given. Assume there are \(B, V \in \mathcal{G}\) with \(\Pi(B) > 0\) and \(a_n \downarrow 0\) s.t.

(i) There exist Bayesian tests for \(B\) versus \(V\) of power \(a_n\),

\[ \int_B P_{\theta,n} \phi_n \, d\Pi(\theta) + \int_V P_{\theta,n} (1 - \phi_n) \, d\Pi(\theta) = o(a_n) \]

(ii) The sequence \((P_{0,n})\) satisfies \(P_{0,n} \prec a_n^{-1} P_{n|B}^{\Pi|B}\)

Then \(\Pi(V_n|X^n) \xrightarrow{P_0} 0\)
Application to i.i.d. consistency (I)

Remark 24.1 (Schwartz (1965))

Take $P_0 \in \mathcal{P}$, and define

$$V_n = \{ P \in \mathcal{P} : H(P, P_0) \geq \epsilon \}$$

$$B_n = \{ P : -P_0 \log \frac{dP}{dP_0} < \frac{1}{2} \epsilon^2 \}$$

With $N(\epsilon, \mathcal{P}, H) < \infty$, and $a_n$ of form $\exp(-nD)$ the theorem proves Hellinger consistency with KL-priors.
Application to i.i.d. consistency (II)

**Remark 25.1** Dirichlet posteriors $X^n \mapsto \prod(P(A)|X^n)$ are msb $\sigma_n(A)$ where $\sigma_n(A)$ is generated by products of the form $\prod_{i=1}^{n} B_i$ with $B_i = \{X_i \in A\}$ or $B_i = \{X_i \notin A\}$.

**Remark 25.2** (Freedman (1965), Ferguson (1973), ...)
Take $P_0 \in \mathcal{P}$, and define

$$V_n = V := \{P \in \mathcal{P} : |P_0(A) - P(A)| \geq 2\epsilon\}$$
$$B_n = B := \{P : |P_0(A) - P(A)| < \epsilon\}$$

for some measurable $A$. **Impose remote contiguity only for $\psi_n$ that are $\sigma_n(A)$-measurable!** Take $a_n$ of form $\exp(-nD)$. The theorem then proves weak consistency with a Dirichlet prior $D_\alpha$, if $\text{supp}(P_0) \subset \text{supp}(\alpha)$. 


Consistency with $n$-dependence

**Theorem 26.1** Let $(\mathcal{P}, \mathcal{G})$ with priors $(\Pi_n)$ and $(X_1, \ldots, X_n) \sim P_{0,n}$ be given. Assume there are $B_n, V_n \in \mathcal{G}$ and $a_n, b_n \geq 0, a_n = o(b_n)$ s.t.

(i) There exist **Bayesian tests** for $B_n$ versus $V_n$ of power $a_n$,

$$\int_{B_n} P_{\theta,n} \phi_n d\Pi_n(\theta) + \int_{V_n} P_{\theta,n}(1 - \phi_n) d\Pi_n(\theta) = o(a_n)$$

(ii) The prior mass of $B_n$ is lower-bounded by $b_n$, $\Pi_n(B_n) \geq b_n$

(iii) The sequence $(P_{0,n})$ satisfies $P_{0}^n \prec b_n a_n^{-1} P_{n|B_n}^{\Pi_n}$

Then $\Pi_n(V_n|X^n) \overset{P_0}{\longrightarrow} 0$

Take $P_0 \in \mathcal{P}$, and define

$$V_n = \{P \in \mathcal{P} : H(P, P_0) \geq \epsilon_n\}$$

$$B_n = \{P : -P_0 \log dP/dP_0 < \frac{1}{2} \epsilon_n^2, P_0 \log^2 dP/dP_0 < \frac{1}{2} \epsilon_n^2\}$$

With $\log N(\epsilon_n, \mathcal{P}, H) \leq n \epsilon_n^2$, and $a_n$ and $b_n$ of form $\exp(-Kn \epsilon_n^2)$ the theorem proves Hellinger consistency at rate $\epsilon_n$

Remark 27.2 Larger $B_n$ are possible, under conditions on the model (see Kleijn and Zhao (201x))
Consistent Bayes factors

**Theorem 28.1** Let the model \((\mathcal{P}, \mathcal{G})\) with priors \((\prod_n)\) be given. Given \(B, V \in \mathcal{G}\) with \(\prod(B), \prod(V) > 0\) s.t.

(i) There are Bayesian tests for \(B\) versus \(V\) of power \(a_n \downarrow 0\),

\[
\int_B P_{\theta,n} \phi_n d\prod_n(\theta) + \int_V P_{\theta,n}(1 - \phi_n) d\prod_n(\theta) = o(a_n)
\]

(ii) For every \(\theta \in B\), \(P_{\theta,n} \prec a_n^{-1} P_n^{\prod_n|B}\)

(iii) For every \(\eta \in V\), \(P_{\eta,n} \prec a_n^{-1} P_n^{\prod_n|V}\)

Then or Bayes factors (or posterior odds),

\[
B_n = \frac{\prod(B|X^n) \prod(V)}{\prod(V|X^n) \prod(B)}
\]

for \(B\) versus \(V\) are consistent.
Random-walk goodness-of-fit testing (I)

Given \((S, \mathcal{L})\) state space for a discrete-time, stationary Markov process with transition kernel \(P(\cdot|\cdot) : \mathcal{L} \times S \rightarrow [0,1]\), the data consists of random walks \(X^n\).

Choose a finite partition \(\alpha = \{A_1, \ldots, A_N\}\) of \(S\) and `bin the data`: \(Z^n\) in finite state space \(S_\alpha\). \(Z^n\) is stationary Markov chain on \(S_\alpha\) with transition probabilities

\[ p_\alpha(k|l) = P(X_i \in A_k | X_{i-1} \in A_l), \]

We assume that \(p_\alpha\) is ergodic with equilibrium distribution \(\pi_\alpha\).

We are interested in Bayes factors for goodness-of-fit testing of transition probabilities.
Random-walk goodness-of-fit testing (II)

Fix $P_0, \epsilon > 0$ and hypothesize on ‘bin probabilities’ $p_\alpha(k, l) = p_\alpha(k|l)\pi_\alpha(l)$,

$$H_0 : \max_{k,l} \left| p_\alpha(k, l) - p_0(k, l) \right| < \epsilon, \quad H_1 : \max_{k,l} \left| p_\alpha(k, l) - p_0(k, l) \right| \geq \epsilon,$$

Define, for $\delta_n \downarrow 0$,

$$B_n = \{p_\alpha \in \Theta : \max_{k,l} \left| p_\alpha(k, l) - p_0(k, l) \right| < \epsilon - \delta_n \}$$

$$V_{k,l} = \{p_\alpha \in \Theta : \left| p_\alpha(k, l) - p_0(k, l) \right| \geq \epsilon \},$$

$$V_{+,k,l,n} = \{p_\alpha \in \Theta : p_\alpha(k, l) - p_0(k, l) \geq \epsilon + \delta_n \},$$

$$V_{-,k,l,n} = \{p_\alpha \in \Theta : p_\alpha(k, l) - p_0(k, l) \leq -\epsilon - \delta_n \}.$$
Random-walk goodness-of-fit testing (III)

Choquet $p_\alpha(k|l) = \sum_{E \in \mathcal{E}} \lambda_E E(k|l)$ where the $N^N$ transition kernels $E$ are deterministic. Define,

$$S_n = \left\{ \lambda_\mathcal{E} \in S_{N^N} : \lambda_E \geq \lambda_n/N^{N-1}, \text{for all } E \in \mathcal{E} \right\},$$

for $\lambda_n \downarrow 0$.

**Theorem 31.1** Choose a prior $\Pi \ll \mu$ on $S_{N^N}$ with continuous density that is everywhere strictly positive. Assume that,

(i) $n\lambda_n^2\delta_n^2/\log(n) \to \infty$,

(ii) $\Pi(B \setminus B_n), \Pi(\Theta \setminus S_n) = o(n^{-(N^N/2)})$,

(iii) $\Pi(V_{k,l} \setminus (V_{+,k,l,n} \cup V_{-,k,l,n})) = o(n^{-(N^N/2)})$, for all $1 \leq k, l \leq N$.

Then the Bayes factors $F_n$ for $H_0$ versus $H_1$ are consistent.
Part V

Uncertainty quantification
Credible sets and confidence sets

Let $\Delta$ denote a collection of measurable subsets of $\Theta$

Definition 33.1 Let $(\Theta, \mathcal{G})$ with prior $\Pi$ be given, denote the posterior by $\Pi(\cdot | \cdot) : \mathcal{G} \times \mathcal{X} \rightarrow [0, 1]$. For $0 \leq \alpha \leq 1$, a credible set $D$ of credible level $1 - \alpha$ is a set-valued map $D : \mathcal{X} \rightarrow \Delta$ such that:

$$\Pi(\Theta \setminus D_n(X^n) | X^n) = o(a_n)$$

$P^{\Pi}$-almost-surely.

Definition 33.2 A sequence of maps $x \mapsto C_n(x) \subset \Theta$ forms an asymptotically consistent sequence of confidence sets, if,

$$P_{\theta_0,n}(\theta_0 \in C_n(X^n)) \rightarrow 1$$

for all $\theta_0 \in \Theta$. 
Enlargement of credible sets (I)

**Definition 34.1** Let $D$ be a credible set in $\Theta$ and let $B$ denote a set function $\theta \mapsto B(\theta) \subset \Theta$. A model subsets $C$ is said to be a confidence set associated with $D$ under $B$, if for all $\theta \in \Theta \setminus C$,

$$B(\theta) \cap D = \emptyset$$

**Definition 34.2** The intersection $C_0$ of all $C$ like above is a confidence set associated with $D$ under $B$, called the minimal confidence set associated with $D$ under $B$. 
Enlargement of credible sets (II)

A credible set $D$ and its associated confidence set $C$ under $B$ in terms of Venn diagrams: additional points $\theta \in C \setminus D$ are characterized by non-empty intersection $B(\theta) \cap D \neq \emptyset$. 
Enlarged credible sets are confidence sets

**Theorem 36.1** Let $0 \leq a_n \leq 1$, $a_n \downarrow 0$ and $b_n > 0$ such that $a_n = o(b_n)$ be given and let $D_n$ denote level-$(1 - a_n)$ credible sets. Furthermore, for all $\theta \in \Theta$, let $B_n$ be set functions such that,

(i) $\prod_n(B_n(\theta_0)) \geq b_n$,

(ii) $P_{\theta_0,n} \triangleleft b_n a_n^{-1} P_n \prod_n|B_n(\theta_0)$.

Then any confidence sets $C_n$ associated with the credible sets $D_n$ under $B_n$ are asymptotically consistent, that is,

$$P_{\theta_0,n}(\theta_0 \in C_n(X^n)) \to 1.$$
Methodology: confidence sets from posteriors (I)

**Corollary 37.1** Given $(\Theta, \mathcal{G})$, $(\Pi_n)$ and $(B_n)$ with $\Pi_n(B_n) \geq b_n$ and $P_{\theta,n} < P_{n|B_n}$, any credible sets $D_n$ of level $1 - a_n$ with $a_n = o(b_n)$ have associated confidence sets under $B_n$ that are asymptotically consistent.

Next, assume that $(X_1, X_2, \ldots, X_n) \in \mathcal{X}^n \sim P_0^n$ for some $P_0 \in \mathcal{P}$.

**Corollary 37.2** Let $\Pi_n$ denote Borel priors on $\mathcal{P}$, with constant $C > 0$ and rate sequence $\epsilon_n \downarrow 0$ such that:

$$\Pi_n\left( P \in \mathcal{P} : -P_0 \log \frac{dP}{dP_0} < \epsilon_n^2, P_0\left( \log \frac{dP}{dP_0} \right)^2 < \epsilon_n^2 \right) \geq e^{-C\epsilon_n^2}.$$ 

Given credible sets $D_n$ of level $1 - \exp(-C'n\epsilon_n^2)$, for some $C' > C$. Then radius-$\epsilon_n$ Hellinger-enlargements $C_n$ are asymptotically consistent confidence sets.
Methodology: confidence sets from posteriors (II)

Note the relation between diameters,

$$\text{diam}_H(C_n(X^n)) = \text{diam}_H(D_n(X^n)) + 2\epsilon_n.$$ 

If, in addition, tests satisfying

$$\int_{B_n} P_{\theta,n} \phi_n(X^n) \, d\Pi_n(\theta) + \int_{V_n} P_{\theta,n}(1 - \phi_n(X^n)) \, d\Pi_n(\theta) = o(a_n),$$

with $a_n = \exp(-C'n\epsilon_n^2)$ exist, the posterior is Hellinger consistent at rate $\epsilon_n$ and credible sets $D_n(X^n)$ have diameters $\leq \epsilon_n$.

If $\epsilon_n$ is the minimax rate of convergence for the problem, the confidence sets $C_n(X^n)$ are rate-optimal.

**Remark 38.1** Rate-adaptivity is not possible like this because a definite choice for the sets in $B_n$ is required.
Conclusions

(i) There is a systematic way of taking Bayesian limits into frequentist limits based on generalization of Schwartz’s prior condition

(ii) Bayesian tests are natural: place low prior weight where testing is difficult, and high weight where testing is easy, ideally.

(iii) Development of new Bayesian methods benefits from a simple, insightful, fully general perspective to guide the search for suitable priors

(iv) Methodology: use priors that induce remote contiguity to enable conversion of credible sets to confidence sets

Thank you for your attention

arXiv:1611.08444 [math.ST]