Statistical Foundations Workshop, Cambridge, 22 June 2017

On the frequentist validity of Bayesian limits

arXiv:1611.08444 [math.ST]

Bas Kleijn, KdV Institute for Mathematics

Universiteit van Amsterdam

Part I

Introduction

Bayesian and Frequentist statistics

sample spaces	$(\mathscr{X}_n, \mathscr{B}_n)$	prob msr's $M^1(\mathscr{X}_n)$
data	$X^n = (X_1, \ldots, X_n) \in \mathscr{X}_n$	sequential experiment
parameter space	$(\Theta, \mathscr{G})$	if *i.i.d.*: $(\mathscr{P}, \mathscr{G})$
parameter	$\theta \in \Theta$	if *i.i.d.*: $P \in \mathscr{P}$
model	$\Theta \to M^1(\mathscr{X}_n) : \theta \mapsto P_{\theta,n}$	not always *i.i.d.*
priors	$\Pi_n : \mathscr{G} \to [0,1]$	probability measure
posterior	$\Pi(\,\cdot\,\vert X^n) : \mathscr{G} \to [0,1]$	Bayes's rule, inference

Frequentist	assume there is P_0	$X^n \sim P_0^n$
Bayes	assume $P \sim \Pi$	$X^n \,\vert\, P \sim P^n$

Definition of the posterior

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

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

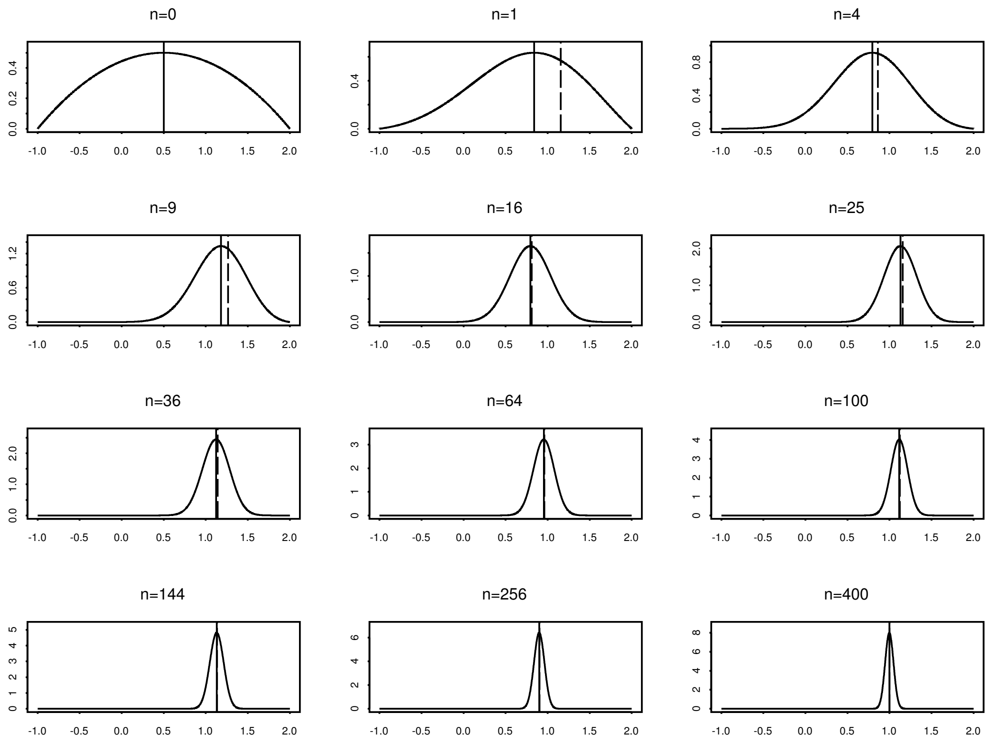
(ii) (Disintegration) For all $A \in \mathscr{B}_n$ and $G \in \mathscr{G}$

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

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

Remark 4.2 *For frequentists $X^n \sim P_{0,n}$, so assume $P_{0,n} \ll P_n^{\Pi}$*

Asymptotic consistency of the posterior



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

$$\Pi(U|X^n) \xrightarrow{P} 1$$

The i.i.d. consistency theorems (I)

Theorem 6.1 *(Bayesian, Doob (1948))*
Let $\mathscr{P}$ and $\mathscr{X}$ be Polish spaces and let Π 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 Π-almost-all $P \in \mathscr{P}$

Example 6.2 *For some $Q \in \mathscr{P}$, take $\Pi = \delta_Q$. Then $\Pi(\cdot|X^n) = \delta_Q$ as well, P_n^{Π}-almost-surely. If $X_1, \ldots, X_n \sim P_0^n$ (require $P_0^n \ll P_n^{\Pi} = Q^n$), 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 $P_0 \in \mathscr{P}$. If,

(i) For every nbd U of P_0, there are $\phi_n : \mathscr{X}_n \to [0,1]$, s.t.

$$P_0^n \phi_n = o(1), \qquad \sup_{Q \in U^c} Q^n(1 - \phi_n) = o(1), \tag{1}$$

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

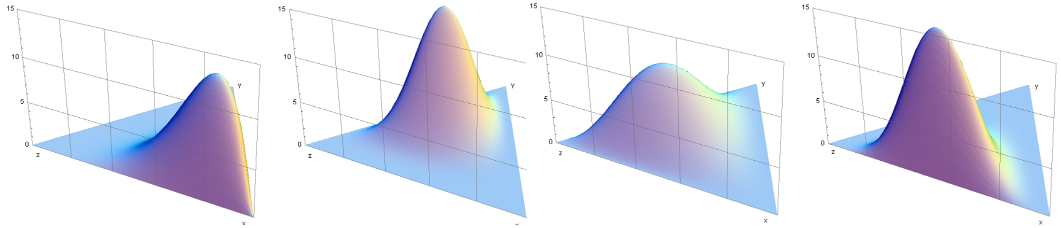
$$\Pi\Big(P \in \mathscr{P} \,:\, -P_0 \log \frac{dP}{dP_0} < \delta \Big) > 0, \tag{2}$$

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

The Dirichlet process

Definition 8.1 *(Dirichlet distribution)*
A $p = (p_1, \ldots, p_k)$ $p_l \geq 0$ *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 $\mathscr{X}$ *be Polish and let* α *be a finite Borel msr on* $(\mathscr{X}, \mathscr{B})$. *The Dirichlet process* $P \sim D_\alpha$ *is defined by,*

$$\big(P(A_1), \ldots, P(A_k)\big) \sim D_{(\alpha(A_1), \ldots, \alpha(A_k))}$$

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 Π is a Dirichlet prior D_α with finite α such that $\mathrm{supp}(P_0) \subset \mathrm{supp}(\alpha)$, 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 $A'_{i1} \cup \ldots \cup A'_{il_i} = A_i$, then $(P(A'_{i1}|A_i), \ldots, P(A'_{il_i}|A_i) : 1 \leq i \leq k)$ 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* Π*-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 $B, V \subset \mathscr{P}$ be convex with $H(B,V) > 0$. There exist a $D > 0$ and uniform test sequence (ϕ_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 $n \geq 1$, $\epsilon > 0$, $P_0 \in \mathscr{P}$ and a msb $f : \mathscr{X}^n \to [0,1]$ be given. Define*

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

There exist a $D > 0$ and uniform test sequence (ϕ_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* $(\mathscr{P},\mathscr{G})$ *be given. For any prior* Π*, any test function* ϕ *and any* $B, V \in \mathscr{G}$*,*

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

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

$$\begin{aligned} P_n^{\Pi|B_n}\Pi(V_n|X^n) &:= \int P_{\theta,n}\Pi(V_n|X^n)\,d\Pi_n(\theta|B_n) \\ &\leq \frac{1}{\Pi_n(B_n)}\Big(\int_{B_n} P_{\theta,n}\phi_n\,d\Pi_n(\theta) + \int_{V_n} P_{\theta,n}(1-\phi_n)\,d\Pi_n(\theta)\Big) \end{aligned}$$

Martingale convergence

Proposition 14.1 *Let* $(\Theta, \mathscr{G}, \Pi)$ *be given. For any* $B, V \in \mathscr{G}$*, the following are equivalent,*

(i) There exist Bayesian tests (ϕ_n) *for* B *versus* V*;*

(ii) There exist tests (ϕ_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 Π*-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 Θ with Borel prior Π be given. Then the following are equivalent,*

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

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

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

Corollary 15.3 *(More than) Doob's 1948 theorem*

Part III

Remote contiguity

Le Cam's inequality

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

Remark 17.2 *(Le Cam, unpublished (197?) and (1986))*
Rewrite the posterior concentration inequality

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

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 Π-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

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

Theorem 18.2 *Let $\mathscr{P}$ be a model with KL-prior Π; $P_0 \in \mathscr{P}$. Let $B, V \in \mathscr{G}$ 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 ($Q_n \lhd P_n$), if for any msb $\psi_n : \mathscr{X}^n \to [0,1]$*

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

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 ($Q_n \lhd a_n^{-1} P_n$), if for any msb $\psi_n : \mathscr{X}^n \to [0,1]$*

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

Remark 19.3 *Contiguity is stronger than remote contiguity*
note that $Q_n \lhd P_n$ iff $Q_n \lhd a_n^{-1} P_n$ 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,* $Q_n \lhd P_n$ *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* $\|Q_n - Q_n \wedge c\, P_n\| < \epsilon$

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

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

(vi) $\liminf_n \psi(P_n, Q_n; \alpha) \to 1$ *as* $\alpha \uparrow 1$

Criteria for remote contiguity

Lemma 21.1 *Given* (P_n), (Q_n), $a_n \downarrow 0$, $Q_n \lhd a_n^{-1} P_n$ *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, $(a_n\, dQ_n/dP_n)$ *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* $(\Theta, \mathscr{G}, \Pi)$ *and* $(X_1, \ldots, X_n) \sim P_{0,n}$ *be given. Assume there are* $B, V \in \mathscr{G}$ *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* $P_{0,n} \lhd a_n^{-1} \, P_n^{\Pi|B}$

Then $\Pi(V_n|X^n) \xrightarrow{P_0} 0$

Application to i.i.d. consistency (I)

Remark 24.1 *(Schwartz (1965))*
Take $P_0 \in \mathscr{P}$, and define

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

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

With $N(\epsilon, \mathscr{P}, H) < \infty$, 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* $X^n \mapsto \Pi(P(A)|X^n)$ *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 $P_0 \in \mathscr{P}$, *and define*

$$V_n = V := \{P \in \mathscr{P} : |P_0(A) - P(A)| \geq 2\epsilon\}$$

$$B_n = B := \{P : |P_0(A) - P(A)| < \epsilon\}$$

for some measurable A. *Impose remote contiguity only for* ψ_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_α, *if* $\text{supp}(P_0) \subset \text{supp}(\alpha)$.

Consistency with n-dependence

Theorem 26.1 *Let* $(\mathscr{P},\mathscr{G})$ *with priors* (Π_n) *and* $(X_1,\ldots,X_n) \sim P_{0,n}$ *be given. Assume there are* $B_n, V_n \in \mathscr{G}$ *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* $P_0^n \lhd b_n a_n^{-1}\, P_n^{\Pi_n|B_n}$

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

Application to i.i.d. consistency (III)

Remark 27.1 *(Barron-Schervish-Wasserman (1999), Ghosal-Ghosh-vdVaart (2000), Shen-Wasserman (2001))*
Take $P_0 \in \mathscr{P}$, and define

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

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

With $\log N(\epsilon_n, \mathscr{P}, H) \leq n\epsilon_n^2$, and a_n and b_n of form $\exp(-Kn\epsilon_n^2)$ the theorem proves Hellinger consistency at rate ϵ_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* $(\mathscr{P}, \mathscr{G})$ *with priors* (Π_n) *be given. Given* $B, V \in \mathscr{G}$ *with* $\Pi(B), \Pi(V) > 0$ *s.t.*

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

$$\int_B P_{\theta,n}\phi_n \, d\Pi_n(\theta) + \int_V P_{\theta,n}(1-\phi_n)\, d\Pi_n(\theta) = o(a_n)$$

(ii) For every $\theta \in B$*,* $P_{\theta,n} \lhd a_n^{-1}\, P_n^{\Pi_n|B}$

(iii) For every $\eta \in V$*,* $P_{\eta,n} \lhd a_n^{-1}\, P_n^{\Pi_n|V}$

Then or Bayes factors (or posterior odds),

$$B_n = \frac{\Pi(B|X^n)}{\Pi(V|X^n)} \frac{\Pi(V)}{\Pi(B)}$$

for B *versus* V *are consistent.*

Random-walk goodness-of-fit testing (I)

Given $(S, \mathscr{S})$ state space for a discrete-time, stationary Markov process with transition kernel $P(\cdot|\cdot) : \mathscr{S} \times S \to [0,1]$, 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_α. Z^n is stationary Markov chain on S_α with transition probabilities

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

We assume that p_α is ergodic with equilibrium distribution π_α.

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' $p_\alpha(k,l) = p_\alpha(k|l)\pi_\alpha(l)$,

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

Define, for $\delta_n \downarrow 0$,

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

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

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

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

Random-walk goodness-of-fit testing (III)

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

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

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 Δ denote a collection of measurable subsets of Θ

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

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

P^{Π}-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,*

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

for all $\theta_0 \in \Theta$.

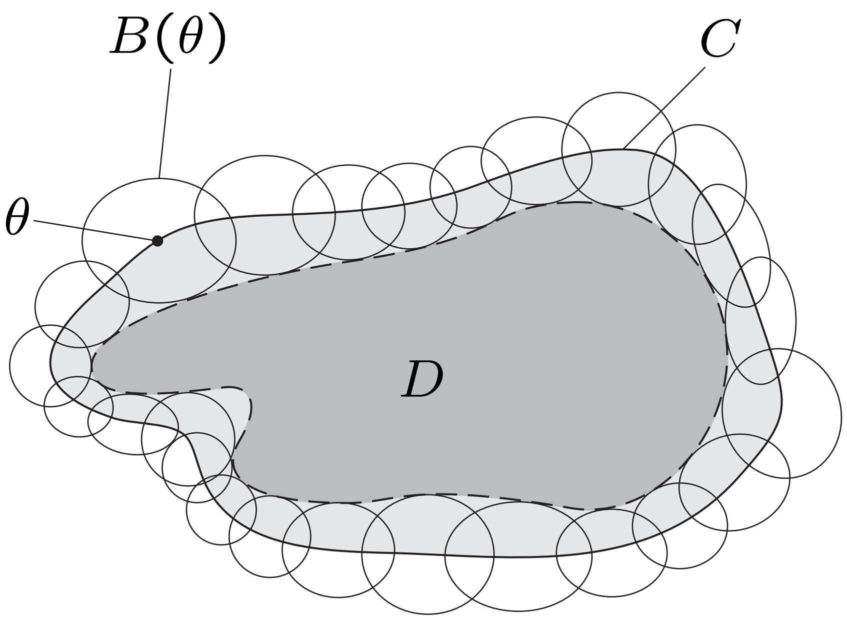
Enlargement of credible sets (I)

Definition 34.1 *Let D be a credible set in Θ 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 $\theta \in \Theta \setminus C$,*

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

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 $B(\theta) \cap D \neq \varnothing$.

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) $\Pi_n(B_n(\theta_0)) \geq b_n$,

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

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

$$P_{\theta_0,n}\Big(\theta_0 \in C_n(X^n) \Big) \to 1.$$

Methodology: confidence sets from posteriors (I)

Corollary 37.1 *Given $(\Theta, \mathscr{G})$, (Π_n) and (B_n) with $\Pi_n(B_n) \geq b_n$ and $P_{\theta,n} \lhd P_n^{\Pi_n|B_n}$, 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 $(X_1, X_2, \ldots, X_n) \in \mathscr{X}^n \sim P_0^n$ for some $P_0 \in \mathscr{P}$.

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

$$\Pi_n\Big(P \in \mathscr{P} \,:\, -P_0 \log \frac{dP}{dP_0} < \epsilon_n^2,\, P_0\Big(\log \frac{dP}{dP_0}\Big)^2 < \epsilon_n^2 \Big) \geq e^{-Cn\epsilon_n^2}.$$

Given credible sets D_n of level $1 - \exp(-C'n\epsilon_n^2)$, for some $C' > C$. Then radius-ϵ_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 ϵ_n and credible sets $D_n(X^n)$ have diameters $\leq \epsilon_n$.

If ϵ_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]