A Weakly Dependent Bernstein–von Mises Theorem

Benjamin Connault*
Princeton University, Department of Economics

November 2014

Abstract

The Bernstein–von Mises theorem is the central result of classical asymptotic theory applied to Bayesian methods. I isolate a convenient set of assumptions and prove a new weakly dependent Bernstein–von Mises theorem along the lines of Le Cam (1986). This new theorem is valid under few assumptions beyond local asymptotic normality. An application in a microeconomic model of dynamic choice can be found in my job market paper.

The Bernstein–von Mises theorem is the central result of classical asymptotic theory applied to Bayesian methods. In sufficiently regular independent and identically distributed models, the Bayesian posterior will be asymptotically Gaussian, centered at the maximum likelihood, and with variance $1/T$ the asymptotic variance of the maximum likelihod. From a frequentist point of view, this implies that Bayesian methods can be used to obtain statistically efficient estimators and consistent confidence intervals. The limiting distribution does not depend on the Bayesian prior.

This paper proves a new Bernstein–von Mises theorem for weakly dependent models in time-series asymptotics. I give a set of weak, yet easy-to-check sufficient conditions for Bernstein–von Mises convergence, organized around local asymptotic normality. The assumptions and

*princeton.edu/~connault and connault@princeton.edu
I thank Andriy Norets for early encouragement and helpful comments. I also thank Sylvain Chassang, Bo Honoré, Ulrich Müller and participants to the Princeton Econometrics seminar for useful comments and discussion.

the proof strategy are along the lines of Le Cam (1986)'s theorem for smooth, independent and identically distributed models (see also van der Vaart (1998)). An example of application is in Connault (2014) in a microeconomic model of dynamic choice.

There are many Bernstein–von Mises theorems, including some for dependent processes. Le Cam (1953) formalized and generalized ideas already present with Laplace, Bernstein and von Mises. The results of Le Cam (1953) were extended in various directions (Walker (1969), Dawid (1970)), including for dependent processes (Heyde and Johnstone (1979), Chen (1985), Sweeting and Adekola (1987)). This family of results relies on relatively strong, functional assumptions about the likelihood and the density of the prior. In Le Cam (1986), Le Cam developed a proof strategy relying on much weaker assumptions. Uniformly consistent tests as in Schwartz (1965) play a significant role (see assumption (A3) below). The main lines of the approach are also described in Le Cam and Yang (2000), chapter 8. Although some of the ideas in Le Cam (1986) are developed for general sequences of experiments (chapter 12), for application purposes only the smooth, independent and identically distributed case is treated (chapter 17). van der Vaart (1998) improved and gave a streamlined version of the proof of Le Cam (1986) for smooth, independent and identically distributed models. As far as I know, Le Cam's approach has never been extended to the weakly-dependent case.

There are at least two reasons why a new weakly dependent Bernstein–von Mises theorem along the lines of Le Cam (1986) is of interest.

First, I obtain a Bernstein–von Mises theorem valid under very few assumptions besides local asymptotic normality. I comment further about the assumptions below, but essentially all the assumptions are close to necessity or expected to hold for a typical locally asymptotically normal model. Assumption (A6), a McDiarmid-type large deviation inequality for the score, might be the least easy to check in practice. One can usually obtain such an inequality from a weakly dependent concentration inequality granted one can prove the score has bounded differences.

Second, and related to first, the set of assumptions I consider is convenient when one is interested in a Bernstein–von Mises theorem as a part of a more comprehensive asymptotic analysis — which might also include for instance the asymptotic distribution of the maximum likelihood estimator. When this is the case, a natural first step is to first prove local asymptotic normality, and then "branch right and left" towards Bayesian and frequentist asymptotics. See for instance Connault (2014).

The exact statement of the Bernstein–von Mises theorem is as follows: the total variation distance between the scaled and centered posterior distribution and the Gaussian distribution centered at a random statistic $X \sim \mathcal{N}(0, I)$ and with variance-covariance matrix I goes to zero in probability, for some invertible matrix I. Note that the asymptotic Gaussianity of the Bayesian posterior is different in nature from the asymptotic normality of, e.g., the maximum likelihood. The posterior distribution is asymptotically Gaussian for a given path of observations, whereas the asymptotic normality of the maximum likelihood is a sampling property. There is sampling variation in the *centering* X of the posterior distribution.

Connault (2014) provides an example of application of the Bernstein–von Mises theorem developed in this paper. There, a uniformly consistent test (assumption (A3)) is obtained from a uniformly consistent estimator based on the empirical distribution of the marginal probability of a suitable number of consecutive observations. Assumptions specif to the weakly dependent setting are obtained from concentration bounds for hidden Markov models and a Taylor expansion for the score.

Section 1 describes the model, states and comments on the assumptions, and states the Bernstein–von Mises theorem. Section 2 gives the proof.

1 A general weakly dependent Bernstein-von Mises theorem

1.1 Notation

θ is a statistical parameter $\theta \in \Theta$, where Θ is an open subspace of a Euclidian space $\Theta \subset \mathbb{R}^{d_\theta}$. $y_{1:T}$ is a sequence of observations $y_t \in E$, where E is a Polish metric space. $\Theta \times E^T$ is endowed with its Borel σ-algebra. In accordance with the Bayesian paradigm, θ and $y_{1:T}$ have a joint distribution $\tilde{\mu}^T$. The Bayesian prior $\tilde{\pi}(d\theta)$, which is the marginal distribution of θ, is the same across different T's. We write $\tilde{\mu}^T(d\theta, dy)$ instead of $\tilde{\mu}^T(d\theta, dy_{1:T})$ when there is no ambiguity about the time-horizon.

The Bernstein-von Mises phenomenon is best studied under a local parametrization. $\theta^\star$ is a distinguished parameter value in the interior of Θ and $h = \sqrt{T}(\theta - \theta^\star)$ is the local parameter.

The usual Bayesian disintegration holds:

$$\underbrace{\pi^T(dh)}_{\text{prior}} \quad \underbrace{k^T(h, dy)}_{\text{conditional distribution}} = \underbrace{\mu^T(dh, dy)}_{\text{joint distribution}} = \underbrace{p^T(dy)}_{\text{marginal distribution}} \quad \underbrace{q^T(y, dh)}_{\text{posterior}}$$

Note that the prior π^T depends on T because of the local $\sqrt{T}$ scaling.

We will also consider the corresponding disintegration with the prior truncated and normalized to a subset $C \subset \mathbb{R}^{d_\theta}$:

$$\pi_C^T(dh) \quad k^T(h, dy) \quad = \quad \mu_C^T(dh, dy) \quad = \quad p_C^T(dy) \quad q_C^T(y, dh)$$

Sometimes we will abuse notation by writing p_h^T for $p_{\{h\}}^T = k^T(h, \cdot)$. p_h^T is simply the distribution of the data for a given value of the (local) parameter.

$\xrightarrow{P_{\theta^\star}}$ means converge in probability under $\theta^\star$ and ∇_{θ_1} means $\frac{\partial}{\partial\theta}\Big|_{\theta=\theta_1}$.

1.2 Assumptions

A domination condition is needed to apply Bayes formula:

Assumption (A1): Dominated family
For each T, the $k^T(h, dy)$'s are dominated by a common σ-finite measure. $\hat{k}^T(h, y)$ is the density with respect to the common dominating measure.

The prior puts mass around the distinguished parameter value $\theta^\star$:

Assumption (A2): Prior mass
$\tilde{\pi}$ is absolutely continuous with respect to the Lebesgue measure in a neighborhood of $\theta^\star$ with a continuous positive density at $\theta^\star$.

Assumption (A3): Uniformly consistent test
For every $\epsilon > 0$, there is a sequence of uniformly consistent tests ϕ_T, meaning tests such that:

$$\mathbb{E}_{\theta^\star}[\phi_T] \to 0 \qquad \textit{and} \qquad \sup_{\|\theta-\theta^\star\|\geq\epsilon} \mathbb{E}_\theta[1 - \phi_T] \to 0$$

Assumption (A4): Local asymptotic normality
There is an invertible matrix I and a sequence of random variables Δ_T converging weakly to $\mathcal{N}(0, I)$ under $\theta^\star$ such that, for every $h_T \to h$:

$$\log \frac{dp_{h_T}^T}{dp_0^T} = h'\Delta_T - \frac{1}{2}h'Ih + o_{p_0^T}(1)$$

Let $s_T = \frac{1}{\sqrt{T}}\Delta_T$. It is well-known that in the smooth, independent, identically distributed case, local asymptotic normality holds with:

$$s_T = \nabla_{\theta^\star} \frac{1}{T} \log f\left(Y_{1:T};\theta\right) = \frac{1}{T}\sum_{t=1}^{T} \frac{\nabla_{\theta^\star} f\left(Y_t;\theta\right)}{f\left(Y_t;\theta^\star\right)}$$

(Where f is simply the density parametrized by the global parameter, $f(y_{1:T};\theta) = \hat{k}^T(h, y)$.) Likewise, in a smooth, weakly dependent case, it is expected that local asymptotic normality holds with:

$$s_T = \nabla_{\theta^\star} \frac{1}{T} \log f\left(Y_{1:T};\theta\right) = \frac{1}{T} \frac{\nabla_{\theta^\star} f\left(Y_{1:T};\theta\right)}{f\left(Y_{1:T};\theta^\star\right)}$$

For this reason, we call s_T the *score* under local asymptotic normality.

The score must satisfy the following assumption, which can be checked using Taylor expansions in smooth models:

Assumption (A5): Local linear lower bound for the score
There is $T_0 \in \mathbb{N}$, $\delta < 1$ and $c > 0$ such that for any $\|\theta - \theta^\star\| \leq \delta$, any $T > T_0$:

$$\left\|\mathbb{E}_\theta\left[s_T\right] - \mathbb{E}_{\theta^\star}\left[s_T\right]\right\| \geq c\left\|\theta - \theta^\star\right\|$$

Assumption (A6): Large deviation inequality for the score
We assume a two-sided McDiarmid-type inequality for the score. There is $T_0 \in \mathbb{N}$ and $c < \infty$ such that, for any $\theta \in \Theta$, any $T > T_0$:

$$P_\theta\left(\left\|s_T - \mathbb{E}_\theta\left[s_T\right]\right\| > u\right) \leq 2\exp\left(-\frac{1}{2}\frac{u^2}{c/T}\right)$$

Assumption (A7): Large deviation inequality for blocks
We assume a one-sided Hoeffding-type inequality for non-overlapping consecutive blocks of y's. Let $R \in \mathbb{N}$ and $\hat{y}_s$ be non-overlapping blocks of R consecutive y_t's, ie $\hat{y}_1 = y_{1:R}$, $\hat{y}_2 = y_{R+1:2R}$, etc. Write $g_s = g\left(\hat{y}_s\right)$. There is $T_0 \in \mathbb{N}$ and $c_R < \infty$ such that for any function $0 \leq g \leq 1$, any $\theta \in \Theta$, any $T > T_0$:

$$P_\theta\left(\frac{1}{S}\sum_{s=1}^{S} g_s < \mathbb{E}\left[\frac{1}{S}\sum_{s=1}^{S} g_s\right] - u\right) \leq \exp\left(-\frac{1}{2}\frac{u^2}{c_R/S}\right)$$

(In the independent, identically distributed case, this inequality holds with $c_R = 1/4$.)

Assumptions (A5) to (A7) are specific to the weakly-dependent setting. Assumptions (A5) and (A6) concern the score of the model, as defined in section 1.2. (A5) is a local linear bound for the score and (A6) a large deviation inequality. (A7) is a large deviation inequality for sums of blocks of data. Assumptions (A1) to (A4) are usual assumptions for independent and identically distributed Bernstein–von Mises theorems in the style of Le Cam (1986) (see for instance van der Vaart (1998)). (A1) asks that the model be dominated in order to apply Bayes formula. (A2) asks that the prior puts some mass around the true parameter value. A uniformly consistent test, as in Schwartz (1965), must exist according to (A3). Finally, (A4) is the local asymptotic normality assumption.

1.3 Statement of the theorem

Write $\mathcal{N}^T = \mathcal{N}(I^{-1}\Delta_T, I^{-1})$ and $\mathcal{N}^T_C$ for its truncation.

With this notation, the Bernstein-von Mises theorem can be concisely stated as:

Theorem: Weakly-dependent Bernstein-von Mises theorem
Under assumptions (A1) to (A7):

$$\left\| q^T - \mathcal{N}^T \right\|_{TV} \xrightarrow{P_{\theta^\star}} 0$$

2 Proof

The idea of the proof is to decompose $\left\| q^T - \mathcal{N}^T \right\|_{\text{TV}}$ in three terms : $\left\| q^T - \mathcal{N}^T \right\|_{\text{TV}} \leq \left\| q^T - q^T_{C_T} \right\|_{\text{TV}} + \left\| q^T_{C_T} - \mathcal{N}^T_{C_T} \right\|_{\text{TV}} + \left\| \mathcal{N}^T_{C_T} - \mathcal{N}^T \right\|_{\text{TV}}$. Local asymptotic normality can be used on the truncated term $\left\| q^T_{C_T} - \mathcal{N}^T_{C_T} \right\|_{\text{TV}}$. This is described in more detail in section 2.3 and carried out in the following sections. Before that, I derive two preliminary results in section 2.1 and 2.2.

2.1 From uniformly consistent to exponentially consistent tests

Assumption (A3) ensures the existence of *uniformly consistent test*, more precisely tests which are uniformly consistent against uniformly remote alternatives. This is not strong enough for the proof of the Bernstein-von Mises theorem (see section 2.4). We will need *exponentially consistent tests*, more precisely tests which are uniformly exponentially consistent against local alternatives: for $M_T \to \infty$, ψ_T is a sequence of exponentially consistent tests if there is a constant $c > 0$ and $T_0 \in \mathbb{N}$ such that:

$$\mathbb{E}_{\theta^\star}[\psi_T] \to 0$$

And for any $T \geq T_0$, for any θ such that $\|\theta - \theta^\star\| \geq \frac{M_T}{\sqrt{T}}$:

$$E_\theta[1 - \psi_T] \leq \exp\left(-cT\left(\|\theta - \theta^\star\|^2 \wedge 1\right)\right)$$

This section shows that exponentially uniformly consistent tests can be built from uniformly consistent tests.

The proof proceeds in two steps: we find some δ, and two sequences of tests ψ_T^1 and ψ_T^2 which satisfy the definitions of exponentially consistent tests on $\frac{M_T}{\sqrt{T}} \leq \|\theta - \theta^\star\| \leq \delta$ and $\delta \leq \|\theta - \theta^\star\|$ respectively. Then $\psi_T = \psi_T^1 \vee \psi_T^2$ is an exponentially consistent test on the whole region $\frac{M_T}{\sqrt{T}} \leq \|\theta - \theta^\star\|$.

Step 1.1: δ and ψ_T^1 on $\frac{M_T}{\sqrt{T}} \leq \|\theta - \theta^\star\| \leq \delta$

Define the score test:

$$\psi_T^1 = 1\left[\|s_T\| \geq \sqrt{\frac{M_T}{T}}\right]$$

First, because $\sqrt{T}s_T = \Delta_T \Rightarrow \mathcal{N}(0, I)$ under $h = 0$ (by assumption (A4)) and because $M_T \to \infty$:

$$P_{\theta^\star}\left(\left\|\sqrt{T}s_T\right\| > M_T\right) \to 0$$

Equivalently, $E_{\theta^\star}[\psi_T^1] \to 0$, as required.

Second, by the triangle inequality:

$$\|s_T\| = \|s_T - \mathbb{E}_{\theta^\star}[s_T]\| \geq \|\mathbb{E}_\theta[s_T] - \mathbb{E}_{\theta^\star}[s_T]\| - \|s_T - \mathbb{E}_\theta[s_T]\|$$

From assumption (A5), there is $\delta < 1$ and c such that for any $\|\theta - \theta^\star\| \leq \delta$:

$$\|\mathbb{E}_\theta[s_T] - \mathbb{E}_{\theta^\star}[s_T]\| \geq c\|\theta - \theta^\star\|$$

Now consider any θ such that $\frac{M_T}{\sqrt{T}} \leq \|\theta - \theta^\star\| \leq \delta$. In particular, for T big enough, $\frac{2}{c\sqrt{M_T}} \leq 1$

and $\sqrt{\frac{M_T}{T}} \leq \frac{2}{c\sqrt{M_T}} \frac{c}{2} \|\theta - \theta^{\star}\| \leq \frac{c}{2} \|\theta - \theta^{\star}\|$. We have:

$$\begin{aligned}
\mathbb{E}_\theta \left[1 - \psi_T^1\right] &= P_\theta \left(\|s_T\| \leq \sqrt{\frac{M_T}{T}} \right) \\
&\leq P_\theta \left(\|\mathbb{E}_\theta [s_T] - \mathbb{E}_{\theta^\star} [s_T]\| - \|s_T - \mathbb{E}_\theta [s_T]\| \leq \sqrt{\frac{M_T}{T}} \right) \\
&\leq P_\theta \left(c \|\theta - \theta^\star\| - \|s_T - \mathbb{E}_\theta [s_T]\| \leq \sqrt{\frac{M_T}{T}} \right) \\
&\leq P_\theta \left(\|s_T - \mathbb{E}_\theta [s_T]\| \geq \frac{1}{2} c \|\theta - \theta^\star\| \right)
\end{aligned}$$

We can apply the large deviation inequality for the score (A6) and as required:

$$E_\theta[1 - \psi_T^1] \leq \exp\left(-c_2 T \|\theta - \theta^\star\|^2\right)$$

Step 1.2: ψ_T^2 on $\delta \leq \|\theta - \theta^\star\|$

Consider uniformy consistent tests ϕ_T for δ (by assumption (A3)):

$$\mathbb{E}_{\theta^\star}[\phi_T] \to 0 \qquad \text{and} \qquad \sup_{\|\theta - \theta^\star\| \geq \delta} \mathbb{E}_\theta[1 - \phi_T] \to 0$$

Take R such that:

$$\mathbb{E}_{\theta^\star}[\phi_R] < \frac{1}{4} \qquad \text{and} \qquad \sup_{\|\theta - \theta^\star\| > \delta} \mathbb{E}_\theta[1 - \phi_R] < \frac{1}{4}$$

Define blocks $\hat{y}_s$ to be non-overlapping blocks of R consecutive y_t's, and $\hat{\phi}_s = \phi_R(\hat{y}_s)$. Note that by definition of R, $\mathbb{E}_\theta \left[\hat{\phi}_s\right] > \frac{3}{4}$. Finally, define:

$$S = \lfloor T/R \rfloor \qquad \text{and} \qquad \psi_T^2 = 1\left[\frac{1}{S}\sum_{s=1}^{S} \hat{\phi}_s \geq \frac{1}{2}\right]$$

Then:

$$\begin{aligned}
\mathbb{E}_\theta[1 - \psi_T^2] &= P\left(\frac{1}{S}\sum_{s=1}^{S} \hat{\phi}_s < \frac{1}{2}\right) \\
&= P\left(\frac{1}{S}\sum_{s=1}^{S} \hat{\phi}_s < \mathbb{E}\left[\frac{1}{S}\sum_{s=1}^{S} \hat{\phi}_s\right] - \left(\mathbb{E}\left[\frac{1}{S}\sum_{s=1}^{S} \hat{\phi}_s\right] - \frac{1}{2}\right)\right)
\end{aligned}$$

Apply the large deviation inequality for blocks of data (A7):

$$\begin{aligned}\mathbb{E}_\theta[1-\psi_T^2] &\leq \exp\left(-\frac{1}{2}\frac{\left(\mathbb{E}\left[\frac{1}{S}\sum_{s=1}^{S}\hat{\phi}_s\right]-\frac{1}{2}\right)^2}{\frac{1}{S}c_R}\right)\\ &\leq \exp\left(-\frac{1}{32c_R}S\right)\end{aligned}$$

$S \sim T$, so this concludes the proof for $\mathbb{E}_\theta[1-\psi_T^2]$.

In a similar fashion, $\mathbb{E}_{\theta^\star}[\psi_T^2] \to 0$.

Step 2: $\psi_T = \psi_T^1 \vee \psi_T^2$.

Putting steps 1.1 and 1.2 together:

$$\mathbb{E}_{\theta^\star}[\psi_T^1 \vee \psi_T^2] \leq \mathbb{E}_{\theta^\star}[\psi_T^1] + \mathbb{E}_{\theta^\star}[\psi_T^2] \to 0$$

$$\begin{aligned}\mathbb{E}_\theta[1-\psi_T^1 \vee \psi_T^2] &\leq \mathbb{E}_\theta[1-\psi_T^1] \wedge \mathbb{E}_\theta[1-\psi_T^2]\\ &\begin{cases}\leq \exp\left(-c_1 T\left\|\theta-\theta^\star\right\|^2\right) & \text{on } \frac{M_T}{\sqrt{T}} \leq \|\theta-\theta^\star\| \leq \delta\\ \leq \exp\left(-c_2 T\right) & \text{on } \delta \leq \|\theta-\theta^\star\|\end{cases}\\ &\begin{cases}\leq \exp\left(-(c_1\wedge c_2)T\left(\left\|\theta-\theta^\star\right\|^2\wedge 1\right)\right) & \text{on } \frac{M_T}{\sqrt{T}} \leq \|\theta-\theta^\star\| \leq \delta\\ \leq \exp\left(-(c_1\wedge c_2)T\left(\left\|\theta-\theta^\star\right\|^2\wedge 1\right)\right) & \text{on } \delta \leq \|\theta-\theta^\star\|\end{cases}\\ &\leq \exp\left(-(c_1\wedge c_2)T\left(\left\|\theta-\theta^\star\right\|^2\wedge 1\right)\right) \qquad \text{on } \frac{M_T}{\sqrt{T}} \leq \|\theta-\theta^\star\|\end{aligned}$$

This shows $\psi_T = \psi_T^1 \vee \psi_T^2$ is an exponentially uniformly consistent test.

2.2 Contiguity

Under local asymptotic normality (assumption (A4)), it is well known that, for any bounded sequence h_T:

$$p_{h_T}^T \lhd\rhd p_0^T$$

We will need the stronger contiguity property: for any U closed ball around 0:

$$p_U^T \lhd\rhd p_0^T$$

Let A_T be such that $p_0^T(A_T) \to 0$. By LAN and contiguity, for any $h \in U$, $p_h^T(A_T) \to 0$. By dominated convergence, $p_U^T(A_T) = \int_U \pi_U^T(dh) p_h^T(A_T) \to 0$.

Let A_T be such that $p_U^T(A_T) \to 0$. Let $h_T \in U$ be such that $p_{h_T}^T(A_T) \leq p_U^T(A_T)$. Then $p_{h_T}^T(A_T) \to 0$. By LAN and contiguity, $p_{h_T}^T \lhd\rhd p_0^T$, so that $p_0^T(A_T) \to 0$.

2.3 Decomposition in three terms

We turn to the core of the proof. We want to show:

$$\left\| q^T - \mathcal{N}^T \right\|_{\mathrm{TV}} \xrightarrow{P_{\theta^\star}} 0$$

Or expressed in local parameters:

$$\left\| q^T - \mathcal{N}^T \right\|_{\mathrm{TV}} \xrightarrow{p_0^T} 0$$

We will use the following decomposition, for some sequence of balls C_T of diameters $M_T \to \infty$:

$$\left\| q^T - \mathcal{N}^T \right\|_{\mathrm{TV}} \leq \left\| q^T - q_{C_T}^T \right\|_{\mathrm{TV}} + \left\| q_{C_T}^T - \mathcal{N}_{C_T}^T \right\|_{\mathrm{TV}} + \left\| \mathcal{N}_{C_T}^T - \mathcal{N}^T \right\|_{\mathrm{TV}}$$

More precisely, we will show that for any sequence C_T, $M_T \to \infty$:

$$\left\| q^T - q_{C_T}^T \right\|_{\mathrm{TV}} \xrightarrow{p_0^T} 0 \quad \text{(step 1)} \qquad \text{and} \qquad \left\| \mathcal{N}_{C_T}^T - \mathcal{N}^T \right\|_{\mathrm{TV}} \xrightarrow{p_0^T} 0 \quad \text{(step 3)}$$

And that for some particular sequence C_T, $M_T \to \infty$:

$$\left\| q_{C_T}^T - \mathcal{N}_{C_T}^T \right\|_{\mathrm{TV}} \xrightarrow{p_0^T} 0 \qquad \text{(step 2)}$$

2.4 Step 1: posterior and truncated posterior

For any B and some particular C:

$$\begin{aligned}
q^T(B) - q^T_C(B) &= q^T(B) - q^T(B|C) \\
&= q^T(B \cap C) + q^T(B \cap C^c) - q^T(B|C) \\
&= q^T(B \cap C^c) - q^T(B|C)(1 - q^T(C)) \\
&= q^T(B \cap C^c) - q^T(B|C) q^T(C^c) \\
|q^T(B) - q^T_C(B)| &\leq q^T(B \cap C^c) + q^T(B|C) q^T(C^c) \\
&\leq 2q^T(C^c)
\end{aligned}$$

So that:

$$\left\| q^T - q^T_C \right\|_{\mathrm{TV}} \leq 2q^T(C^c)$$

Now taking C_T to be any sequence of balls of diameter $M_T \to \infty$, we show:

$$q^T(C^c_T) \xrightarrow{p^T_0} 0$$

Taking ψ_T an exponentially uniformly consistent test (see section 2.1), write:

$$q^T(C^c_T) = \underbrace{q^T(C^c_T)(1 - \psi_T)}_{\text{step 1.1}} + \underbrace{q^T(C^c_T)\psi_T}_{\text{step 1.2}}$$

Step 1.1 Take U a fixed ball around 0. Then:

$$\begin{aligned}
\mathbb{E}_{p^T_U}[q^T(C^c)(1 - \psi_T)] &= \int_y \underbrace{p^T_U(dy)}_{= \int_{h_1 \in U} \pi^T_U(dh_1) k^T(h_1, dy)} (1 - \psi_T(y)) \int_{h_2 \in C^c} q^T(y, dh_2) \\
&= \int_y \int_{h_1 \in U} \frac{1}{\pi^T(U)} \underbrace{\pi^T(dh_1) k^T(h_1, dy)}_{= p^T(dy) q^T(y, dh_1)} (1 - \psi_T(y)) \int_{h_2 \in C^c} q^T(y, dh_2) \\
&= \frac{1}{\pi^T(U)} \int_y p(dy)(1 - \psi_T(y)) \int_{h_1 \in U} \int_{h_2 \in C^c} q^T(y, dh_1) q^T(y, dh_2) \\
&= \frac{\pi^T(C^c)}{\pi^T(U)} \mathbb{E}_{p^T_{C^c}}[q^T(U)(1 - \psi_T)] \qquad \text{by Fubini}
\end{aligned}$$

And:

$$\begin{aligned}\frac{\pi^T(C_T^c)}{\pi^T(U)}\mathbb{E}_{p^T_{C_T^c}}[q^T(U)(1-\psi_T)] &= \frac{\pi^T(C_T^c)}{\pi^T(U)}\int_y \underbrace{p^T_{C_T^c}(dy)}_{=\int_h \pi^T_{C_T^c}(dh)k^T(h,dy)} \underbrace{q^T(y,U)}_{\leq 1}(1-\psi_T(y))\\ &\leq \frac{\pi^T(C_T^c)}{\pi^T(U)}\int_y\int_{h\in C_T^c}\frac{\pi^T(dh)}{\pi^T(C_T^c)}k^T(h,dy)(1-\psi_T(y))\\ &= \frac{1}{\pi^T(U)}\int_{h\in C_T^c}\pi^T(dh)\mathbb{E}_h[1-\psi_T]\end{aligned}$$

Now note that $\pi^T(U) > c_1\frac{1}{\sqrt{T}^{d_\theta}}$ and that, according to assumption (A2), there is $D \leq 1$ small enough and c_2 such that π has a density with respect to the Lebesgue measure which is bounded uniformely by c_2 on $\|\theta-\theta^\star\| \leq D$. Then using exponential consistency (section 2.1):

$$\begin{aligned}\frac{1}{\pi^T(U)}\int_{h\in C_T^c}\pi^T(dh)\mathbb{E}_h[1-\psi_T] &\leq c_1\sqrt{T}^{d_\theta}\int_{M_T\leq\|h\|}\pi^T(dh)e^{-c_3(\|h^2\|\wedge T)}\\ &\leq c_1\sqrt{T}^{d_\theta}\left(\int_{M_T\leq\|h\|\leq D\sqrt{T}}\underbrace{\pi^T(dh)}_{\leq\frac{c_4}{\sqrt{T}^{d_\theta}}c_2 dh}\underbrace{e^{-c_3(\|h^2\|\wedge T)}}_{\leq e^{-c_3\|h^2\|}}+\int_{D\sqrt{T}\leq\|h\|}\pi^T(dh)\underbrace{e^{-c_3(\|h^2\|\wedge T)}}_{\leq e^{-c_3 D^2 T}}\right)\\ &\leq c_5\int_{M_T\leq\|h\|\leq D\sqrt{T}}dh \quad e^{-c_3(\|h^2\|)}+c_1\sqrt{T}^{d_\theta}e^{-c_3D^2T}\\ &\leq c_5\int_{M_T\leq\|h\|}dh \quad e^{-c_3(\|h^2\|)}+c_1\sqrt{T}^{d_\theta}e^{-c_3D^2T}\end{aligned}$$

So that, if $M_T\to\infty$:

$$\frac{\pi^T(C_T^c)}{\pi^T(U)}\mathbb{E}_{p^T_{C_T^c}}[q^T(U)(1-\psi_T)]\to 0$$

And finally:

$$\begin{aligned}\mathbb{E}_{p_U^T}[q^T(C_T^c)(1-\psi_T)]\to 0 &\Longleftrightarrow q^T(C_T^c)(1-\psi_T)\to 0 \qquad \text{in } p_U^T\text{-mean}\\ &\Longrightarrow q^T(C_T^c)(1-\psi_T)\xrightarrow{p_U^T} 0\\ &\Longrightarrow q^T(C_T^c)(1-\psi_T)\xrightarrow{p_0^T} 0 \qquad \text{by LAN and contiguity}\end{aligned}$$

Step 1.2

$$q^T(C_T^c)\psi_T\xrightarrow{p_0^T} 0 \qquad \text{by definition of } \psi_T$$

2.5 Step 2: truncated posterior and truncated Gaussian

Thanks to the domination assumption (assumption (A1)), we can apply Bayes formula:

$$q_C^T(x, dh) = \pi_C^T(dh)\frac{\hat{k}^T(h,x)}{\int_{g\in C}\pi_C^T(dg)\hat{k}^T(g,x)} = \pi^T(dh)\frac{\hat{k}^T(h,x)}{\int_{g\in C}\pi^T(dg)\hat{k}^T(g,x)}$$

Let us start by fixing C and showing that $\left\|\mathcal{N}_C^T - q_C^T\right\|_{TV} \xrightarrow{p_0^T} 0$. For any two measure P and Q, $\|P-Q\|_{\mathrm{TV}} = 2\int\left(1-\frac{p}{q}\right)^+ dQ$ where p and q are the densities of P and Q with respect to any common dominating measures. For T big enough, by assumption (A2), π^T is dominated by the Lebesgue measure λ on $\mathbb{R}^{d_\theta}$. Let $\hat{\pi}^T$ and $\hat{\mathcal{N}}_C^T$ be the respective densities with respect to λ. Suppressing y's from notation for this display:

$$\begin{aligned}
\frac{1}{2}\left\|\mathcal{N}_C^T - q_C^T\right\|_{TV} &= \int_{h\in C} q_C^T(dh)\left(1 - \frac{\hat{\mathcal{N}}_C^T(h)}{\frac{\hat{k}^T(h)}{\int_{g\in C}\pi^T(dg)\hat{k}^T(g)}\hat{\pi}^T(h)}\right)^+ \\
&= \int_{h\in C} q_C^T(dh)\left(1 - \int_{g\in C}\pi^T(dg)\frac{\hat{\mathcal{N}}_C^T(h)}{\hat{\pi}^T(h)}\frac{\hat{k}^T(g)}{\hat{k}^T(h)}\right)^+ \\
&= \int_{h\in C} q_C^T(dh)\left(1 - \int_{g\in C}\mathcal{N}_C^T(dg)\frac{\hat{\pi}^T(g)}{\hat{\mathcal{N}}_C^T(g)}\frac{\hat{\mathcal{N}}_C^T(h)}{\hat{\pi}^T(h)}\frac{\hat{k}^T(g)}{\hat{k}^T(h)}\right)^+ \\
&\leq \int_{h\in C} q_C^T(dh)\int_{g\in C}\mathcal{N}_C^T(dg)\left(1 - \frac{\hat{k}^T(g)}{\hat{k}^T(h)}\frac{\hat{\mathcal{N}}_C^T(h)}{\hat{\pi}^T(h)}\frac{\hat{\pi}^T(g)}{\hat{\mathcal{N}}_C^T(g)}\right)^+
\end{aligned}$$

(In the last line we used the fact that for any random variable ξ, $(1-\mathbb{E}[\xi])^+ \leq \mathbb{E}[(1-\xi)^+]$.)

Now call:

$$a(h,g,y) = \left(1 - \frac{\hat{k}^T(g,y)}{\hat{k}^T(h,y)}\frac{\hat{\mathcal{N}}_C^T(y,h)}{\hat{\pi}^T(h)}\frac{\hat{\pi}^T(g)}{\hat{\mathcal{N}}_C^T(g)}\right)^+$$

Integrating the double integral above with respect to the "truncated" marginal $p_C^T(dy)$ and applying $\pi_C^T(dh)k^T(h,dy) = p_C^T(dy)q_C^T(y,dh)$:

$$\int p_C^T(dy)q_C^T(y,dh)\mathcal{N}_C^T(dg)a(h,g,y) = \int \pi_C^T(dh)k^T(h,dy)\mathcal{N}_C^T(dg)a(h,g,y)$$

Now, because $k^T(h,dy) \triangleleft\triangleright k^T(0,dy)$ (for any h) and $\pi_C^T(dh)$ and $\mathcal{N}_C^T(dg)$ are contiguous with the uniform distribution on C (the boundedness of C is crucial here, and the reason why we made this three-term decomposition in the first place):

$$\pi_C^T(dh)k^T(h,dy)\mathcal{N}_C^T(dg) \lhd\rhd \lambda_C(dh)k^T(0,dy)\lambda_C(dg)$$

Using:

$$h'\Delta_T - \frac{1}{2}h'Ih = -\frac{1}{2}(h - I^{-1}\Delta_T)'I(h - I^{-1}\Delta_T) + constant$$

the LAN property implies that:

$$\frac{\hat{k}^T(g,y)}{\hat{k}^T(h,y)}\frac{\mathcal{N}_C^T(y,h)}{\mathcal{N}_C^T(y,g)} = \frac{\frac{\hat{k}^T(g,y)/\hat{k}^T(0,y)}{\mathcal{N}^T(y,g)}}{\frac{\hat{k}^T(h,y)/\hat{k}^T(0,y)}{\mathcal{N}^T(y,h)}} = \frac{\exp\left(o^{(1)}_{p_0^T}\right)}{\exp\left(o^{(2)}_{p_0^T}\right)} \xrightarrow{p_0^T} 1$$

Also, for T big enough, if $\hat{\tilde{\pi}}(\theta)$ is the density of the prior $\tilde{\pi}(d\theta)$ with respect to the Lebesgue measure around $\theta^\star$ (by assumption (A2)):

$$\hat{\pi}^T(h) = \frac{1}{\sqrt{T}^{d_\theta}}\hat{\tilde{\pi}}\left(\theta^\star + \frac{h}{\sqrt{T}}\right) \qquad \text{and thus} \qquad \frac{\hat{\pi}^T(g)}{\hat{\pi}^T(h)} \to 1$$

Finally:

$$\begin{aligned}
a(h,g,y) \xrightarrow{p_0^T} (1 - 1\cdot 1)^+ = 0 &\quad \text{for each } h \text{ and } g \\
\implies a(h,g,y) \to 0 &\quad \text{in } p_0^T \text{ mean, for each } h \text{ and } g\text{, by uniform integrability} \\
\implies a(h,g,y) \to 0 &\quad \text{in } \lambda_C(dh)\lambda_C(dg)k^T(0,dy) \text{ mean, by dominated convergence} \\
\iff a(h,g,y) \to 0 &\quad \text{in } \lambda_C(dh)k^T(0,dy)\lambda_C(dg) \text{ mean} \\
\implies a(h,g,y) \to 0 &\quad \text{in } \lambda_C(dh)k^T(0,dy)\lambda_C(dg) \text{ probability} \\
\implies a(h,g,y) \to 0 &\quad \text{in } \pi_C^T(dh)k^T(h,dy)\mathcal{N}_C^T(dg) \text{ probability, by contiguity} \\
\implies a(h,g,y) \to 0 &\quad \text{in } \pi_C^T(dh)k^T(h,dy)\mathcal{N}_C^T(dg) \text{ mean, by uniform integrability} \\
\implies \left\|\mathcal{N}_C^T - q_C^T\right\|_{\mathrm{TV}} \to 0 &\quad \text{in } p_C^T \text{ mean} \\
\implies \left\|\mathcal{N}_C^T - q_C^T\right\|_{\mathrm{TV}} \to 0 &\quad \text{in } p_C^T \text{ probability} \\
\implies \left\|\mathcal{N}_C^T - q_C^T\right\|_{\mathrm{TV}} \to 0 &\quad \text{in } p_0^T \text{ probability, by contiguity}
\end{aligned}$$

Thus we have shown that for any fixed C, $\left\|\mathcal{N}_C^T - q_C^T\right\|_{TV} \xrightarrow{p_0^T} 0$.

To conclude the proof of this step, note that in general if u_{mn} is a doubly-indexed sequence of random variables such that for every m, $u_{mn} \xrightarrow[n\to\infty]{} 0$ in μ_n-probability (μ_n any sequence of probability measures), then there is an extraction $m_n \to \infty$ such that $u_{m_T n} \xrightarrow[n\to\infty]{} 0$ in

μ_n-probability.

This implies that there is some sequence C_T with $M_T \to \infty$ such that:

$$\left\| \mathcal{N}_{C_T}^T - q_{C_T}^T \right\|_{\mathrm{TV}} \xrightarrow{p_0^T} 0$$

2.6 Step 3: truncated Gaussian and Gaussian

Similarly to section 2.4, for any ball of diameters $M_T \to \infty$:

$$\left\| \mathcal{N}_{C_T}^T - \mathcal{N}^T \right\|_{\mathrm{TV}} \xrightarrow{p_0^T} 0$$

This concludes the proof.

References

Chen, C.-F. (1985): "On asymptotic Normality of Limiting Density Functions with Bayesian Implications," *Journal of the Royal Statistical Society. Series B (Methodological)*, 540–546.

Connault, B. (2014): "Hidden Rust Models," *Job Market Paper*, http://www.princeton.edu/~connault/.

Dawid, A. (1970): "On the Limiting Normality of Posterior Distributions," in *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge Univ Press, vol. 67, 625–633.

Heyde, C. and I. Johnstone (1979): "On Asymptotic Posterior Normality for Stochastic Processes," *Journal of the Royal Statistical Society. Series B (Methodological)*, 184–189.

Le Cam, L. (1953): "On some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes' Estimates," *Univ. California Publ. Statist.*, 1, 277–329.

——— (1986): *Asymptotic Methods in Statistical Decision Theory*, Springer.

Le Cam, L. and G. Yang (2000): *Asymptotics in Statistics: Some Basic Concepts*, Springer.

Schwartz, L. (1965): "On Bayes Procedures," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 4, 10–26.

Sweeting, T. and A. Adekola (1987): "Asymptotic Posterior Normality for Stochastic Processes Revisited," *Journal of the Royal Statistical Society. Series B (Methodological)*, 215–222.

van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge University Press.

Walker, A. (1969): "On the Asymptotic Behaviour of Posterior Distributions," *Journal of the Royal Statistical Society. Series B (Methodological)*, 80–88.