

Lecture 7: Asymptotic properties in parametric Bayesian models

Thibault Randrianarisoa

UTSC

March 11, 2026



Frequentist Perspective on Bayesian Posteriors

In this chapter, we explore how to study Bayesian posterior distributions from a **frequentist perspective**.

- We define the notions of **consistency** and **convergence** of these distributions in an asymptotic framework where the number of observations $n \rightarrow \infty$.
- We consider the limiting form of posterior distributions and state the **Bernstein-von Mises theorem**.
- We will examine important consequences, particularly for the construction of confidence regions.

Explicit Posterior Expressions

The following table presents classical models with prior distributions Π , explicit expressions for the posterior $\Pi[\cdot | \mathbf{X}]$, and the posterior mean $\mathbb{E}[\theta | \mathbf{X}]$.

Model \mathcal{P}	Prior Π	Posterior $\Pi[\cdot \mathbf{X}]$	$\mathbb{E}[\theta \mathbf{X}]$	MLE
$\mathcal{N}(\theta, 1)^{\otimes n}, \theta \in \mathbb{R}$	$\mathcal{N}(a, 1)$	$\mathcal{N}\left(\frac{a+n\bar{X}_n}{n+1}, \frac{1}{n+1}\right)$	$\frac{a+n\bar{X}_n}{n+1}$	\bar{X}_n
$\mathcal{B}(\theta)^{\otimes n}, \theta \in (0, 1)$	Beta(a, b)	Beta($a + n\bar{X}_n, b + n - n\bar{X}_n$)	$\frac{a+n\bar{X}_n}{a+b+n}$	\bar{X}_n
Poisson(θ) $^{\otimes n}, \theta > 0$	Gamma(a, b)	Gamma($a + n\bar{X}_n, n + b$)	$\frac{a+n\bar{X}_n}{n+b}$	\bar{X}_n
$\mathcal{E}(\theta)^{\otimes n}, \theta > 0$	Gamma(a, b)	Gamma($n + a, b + n\bar{X}_n$)	$\frac{n+a}{b+n\bar{X}_n}$	$\frac{1}{\bar{X}_n}$

Proximity Between Posterior Mean and MLE

Reading the last two columns of the table suggests a striking proximity between the posterior mean and the MLE as $n \rightarrow +\infty$.

- In this lecture, we use the Bayesian framework ($\mathbf{X} \mid \theta = \theta \sim P_\theta$ and $\theta \sim \Pi$) **only** to define the posterior distribution $\Pi[\cdot \mid \mathbf{X}]$.
- Once this distribution is defined, we study it in probability under $P_{\theta_0}^{\otimes n}$, with $\theta_0 \in \Theta$ fixed.
- This means assuming the **frequentist hypothesis**:

$$X_1, \dots, X_n \text{ i.i.d. } \sim P_{\theta_0}$$

Asymptotic Behavior and Concentration

What is the asymptotic behavior of $\Pi[\cdot | \mathbf{X}]$ when the data is generated under a fixed true parameter P_{θ_0} ?

In the previous examples, using an estimator like $\bar{X}_n \xrightarrow{\mathbb{P}} \theta_0$ under P_{θ_0} in the Gaussian model, we see that we also have:

$$\mathbb{E}[\theta | \mathbf{X}] \xrightarrow{\mathbb{P}} \theta_0$$

Furthermore, we can verify in each example that the posterior variance tends to 0 in probability (check it!). This signifies that, under P_{θ_0} , the **posterior mass concentrates around θ_0** .

Well-Defined Bayes Formula

Remark Letting $f(\mathbf{X}) = \int \prod_{i=1}^n p_{\theta}(X_i) d\Pi(\theta)$ be the marginal density of \mathbf{X} evaluated at \mathbf{X} :

$$\mathbb{P}(f(\mathbf{X}) = 0) = \mathbb{E}[\mathbf{1}_{f(\mathbf{X})=0}] = \int \mathbf{1}_{f(x)=0} f(x) dx = 0$$

- The denominator of Bayes' formula is non-zero almost surely under the **marginal distribution** of \mathbf{X} .
- However, it might be zero with non-zero probability under P_{θ_0} .
- For our frequentist study of the posterior $\Pi[\cdot | \mathbf{X}]$, we assume the denominator is non-zero P_{θ_0} -almost surely:

$$P_{\theta_0}(f(\mathbf{X}) = 0) = 0$$

- This makes Bayes' formula well-defined almost surely under the P_{θ_0} distribution of the data.

Consistency of Posterior Distributions

Framework

We consider a model $\mathcal{P} = \{P_\theta^{\otimes n}, \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d, d \geq 1$. We equip Θ with a prior distribution Π , and to form the posterior $\Pi[\cdot | \mathbf{X}]$, we consider the Bayesian model:

$$\begin{aligned}\theta &\sim \Pi \\ \mathbf{X} &= (X_1, \dots, X_n) \mid \theta \sim P_\theta^{\otimes n}\end{aligned}$$

Once $\Pi[\cdot | \mathbf{X}]$ is formed, it is studied under the **frequentist hypothesis**:

$$\mathbf{X} = (X_1, \dots, X_n) \sim P_{\theta_0}^{\otimes n}$$

Definition of Consistency

Definition

We say that the posterior distribution $\Pi[\cdot | \mathbf{X}] = \Pi[\cdot | X_1, \dots, X_n]$ is **consistent** at the point $\theta_0 \in \Theta$ if, for all $\varepsilon > 0$, under P_{θ_0} :

$$\Pi(\{\theta \in \Theta : \|\theta - \theta_0\| > \varepsilon\} | \mathbf{X}) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$$

Remark For a random variable Z_n such that $0 \leq Z_n \leq 1$, we have:

$$Z_n \xrightarrow{\mathbb{P}} 0 \iff \mathbb{E}[Z_n] \rightarrow 0 \quad (n \rightarrow \infty)$$

In particular, to show that the posterior is consistent, it suffices to show that:

$$\mathbb{E}_{\theta_0} [\Pi(\{\theta \in \Theta : \|\theta - \theta_0\| > \varepsilon\} | \mathbf{X})] \xrightarrow[n \rightarrow \infty]{} 0$$

Example: A Non-Consistent Posterior

Here is an example of a non-consistent posterior distribution. Let $\mathcal{P} = \{\mathcal{N}(\theta, 1)^{\otimes n}, \theta \in \mathbb{R}\}$ and consider the prior $\Pi = \text{Unif}[0, 1]$.

The posterior density is proportional to:

$$\exp \left\{ - \sum_{i=1}^n (X_i - \theta)^2 / 2 \right\} \mathbb{1}_{[0,1]}(\theta).$$

In particular, the posterior density is zero outside of $[0, 1]$.

The posterior $\Pi[\cdot | \mathbf{X}]$ is therefore inconsistent outside of $[0, 1]$, for example at $\theta_0 = 2$, since:

$$\Pi[[3/2, 5/2] | \mathbf{X}] = 0.$$

Consistency in the Gaussian Model

Proposition

In the Gaussian model $\mathcal{P} = \{\mathcal{N}(\theta, 1)^{\otimes n}, \theta \in \mathbb{R}\}$ with a prior distribution $\Pi = \mathcal{N}(a, \sigma^2)$, the posterior distribution $\Pi[\cdot | \mathbf{X}]$ is consistent at every point $\theta_0 \in \mathbb{R}$.

Proof The posterior distribution is given by:

$$\Pi[\cdot | \mathbf{X}] = \mathcal{N}\left(\frac{a\sigma^{-2} + n\bar{X}_n}{n + \sigma^{-2}}, \frac{1}{n + \sigma^{-2}}\right).$$

Let $m_{\mathbf{X}} = \mathbb{E}[\theta | \mathbf{X}] = \frac{a\sigma^{-2} + n\bar{X}_n}{n + \sigma^{-2}}$.

For any real θ_0 and $\varepsilon > 0$, we have:

$$\begin{aligned}\mathbb{P}(|\theta - \theta_0| > \varepsilon \mid \mathbf{X}) &\leq \mathbb{P}(|\theta - m_{\mathbf{X}}| + |m_{\mathbf{X}} - \theta_0| > \varepsilon \mid \mathbf{X}) \\ &\leq \mathbb{P}\left(|\theta - m_{\mathbf{X}}| > \frac{\varepsilon}{2} \mid \mathbf{X}\right) + \mathbb{1}_{|m_{\mathbf{X}} - \theta_0| > \frac{\varepsilon}{2}},\end{aligned}$$

where we used the triangle inequality and then the fact that if $|\theta - m_{\mathbf{X}}| + |m_{\mathbf{X}} - \theta_0| > \varepsilon$, then at least one of the two terms is strictly greater than $\varepsilon/2$.

Since $\bar{X}_n \xrightarrow{\mathbb{P}} \theta_0$ under P_{θ_0} by the [law of large numbers](#), we have $m_{\mathbf{X}} \xrightarrow{\mathbb{P}} \theta_0$ under P_{θ_0} . Thus:

$$\mathbb{E}_{\theta_0} \mathbb{1}_{|m_{\mathbf{X}} - \theta_0| > \frac{\varepsilon}{2}} = \mathbb{P}_{\theta_0}\left(|m_{\mathbf{X}} - \theta_0| > \frac{\varepsilon}{2}\right) \xrightarrow[n \rightarrow \infty]{} 0.$$

On the other hand, according to the explicit expression of the posterior distribution:

$$\begin{aligned}\mathbb{P}\left(|\theta - m_{\mathbf{X}}| > \frac{\varepsilon}{2} \mid \mathbf{X}\right) &= \mathbb{P}\left(\left|\mathcal{N}\left(m_{\mathbf{X}}, \frac{1}{n + \sigma^{-2}}\right) - m_{\mathbf{X}}\right| > \frac{\varepsilon}{2} \mid \mathbf{X}\right) \\ &= \mathbb{P}\left(\left|\mathcal{N}\left(0, \frac{1}{n + \sigma^{-2}}\right)\right| > \frac{\varepsilon}{2}\right) \\ &= \mathbb{P}\left(|\mathcal{N}(0, 1)| > \frac{\varepsilon}{2} \sqrt{n + \sigma^{-2}}\right) \\ &\xrightarrow[n \rightarrow \infty]{} 0,\end{aligned}$$

since **Slutsky's lemma** gives $\frac{1}{\sqrt{n + \sigma^{-2}}} |\mathcal{N}(0, 1)| \xrightarrow{\mathbb{P}} 0$.

Therefore, the posterior is consistent at θ_0 . ■

Consistency when Θ is Finite

Let $\Theta = \{1, \dots, k\}$. We consider the model:

$$\mathcal{P} = \{P_1, \dots, P_k\} = \{P_\theta, \theta \in \Theta\},$$

where P_j are probability measures on E .

Let Π be a prior distribution on Θ . This is defined by the values:

$$\Pi(\{j\}) = \pi_j, \quad j = 1, \dots, k.$$

Proposition

In the framework of the discrete model, assume the model is **identifiable** and let Π be a prior distribution on Θ such that $\pi_j > 0$ for all $j = 1, \dots, k$. Then the posterior distribution $\Pi[\cdot | \mathbf{X}]$ is consistent at every point $\theta_0 \in \{1, \dots, k\}$.

Proof Let $\theta_0 \in \{1, \dots, k\}$. It suffices to show that

$$\Pi[\{\theta_0\} | X_1, \dots, X_n] \xrightarrow{\mathbb{P}} 1$$

under P_{θ_0} .

Since the distribution Π is discrete, **Bayes' formula** is written here, for $A \subset \{1, \dots, k\}$,

$$\Pi[A | \mathbf{X}] = \frac{\sum_{j \in A} \pi_j \prod_{i=1}^n p_j(X_i)}{\sum_{j=1}^k \pi_j \prod_{i=1}^n p_j(X_i)}.$$

Let $l_j(\mathbf{X}) = \prod_{i=1}^n p_j(X_i)$, then

$$\Pi[\{\theta_0\} \mid \mathbf{X}] = \frac{\pi_{\theta_0} l_{\theta_0}(\mathbf{X})}{\sum_{j=1}^k \pi_j l_j(\mathbf{X})}.$$

For all $j \neq \theta_0$, we have $l_j(\mathbf{X}) \leq \max_{i \neq \theta_0} l_i(\mathbf{X})$. Since $\sum_{i \neq \theta_0} \pi_i = 1 - \pi_{\theta_0}$, we deduce:

$$\Pi[\{\theta_0\} \mid \mathbf{X}] \geq \frac{\pi_{\theta_0} l_{\theta_0}(\mathbf{X})}{\pi_{\theta_0} l_{\theta_0}(\mathbf{X}) + (1 - \pi_{\theta_0}) \max_{j \neq \theta_0} l_j(\mathbf{X})} = \frac{1}{1 + \frac{1 - \pi_{\theta_0}}{\pi_{\theta_0}} \frac{\max_{j \neq \theta_0} l_j(\mathbf{X})}{l_{\theta_0}(\mathbf{X})}}.$$

Let $\varepsilon > 0$. We have:

$$\mathbb{P}_{\theta_0} \left(\max_{j \neq \theta_0} l_j(\mathbf{X}) \geq \varepsilon l_{\theta_0}(\mathbf{X}) \right) \leq \sum_{j \neq \theta_0} \mathbb{P}_{\theta_0} (l_j(\mathbf{X}) \geq \varepsilon l_{\theta_0}(\mathbf{X})).$$

For $j \in [1, k] \setminus \{\theta_0\}$, **Markov's inequality** applied with the function $x \mapsto \sqrt{x}$ gives

$$\mathbb{P}_{\theta_0}(l_j(\mathbf{X}) \geq \varepsilon l_{\theta_0}(\mathbf{X})) \leq \frac{1}{\sqrt{\varepsilon}} \mathbb{E}_{\theta_0} \left[\sqrt{\frac{l_j(\mathbf{X})}{l_{\theta_0}(\mathbf{X})}} \right].$$

Now the expectation in this last expression is written:

$$\begin{aligned} \mathbb{E}_{\theta_0} \left[\sqrt{\frac{l_j(\mathbf{X})}{l_{\theta_0}(\mathbf{X})}} \right] &= \int \left[\frac{\prod_{i=1}^n p_j(x_i)}{\prod_{i=1}^n p_{\theta_0}(x_i)} \right]^{1/2} \prod_{i=1}^n p_{\theta_0}(x_i) d\mu(x_i) \\ &= \int \sqrt{\prod_{i=1}^n p_j(x_i) \prod_{i=1}^n p_{\theta_0}(x_i)} d\mu(x_i) \end{aligned}$$

Continuing with the expectation:

$$\begin{aligned}\mathbb{E}_{\theta_0} \left[\sqrt{\frac{\ell_j(\mathbf{X})}{\ell_{\theta_0}(\mathbf{X})}} \right] &= \rho(P_j^{\otimes n}, P_{\theta_0}^{\otimes n}) = \rho(P_j, P_{\theta_0})^n \\ &\leq \left[\max_{j \neq \theta_0} \rho(P_j, P_{\theta_0}) \right]^n := r^n,\end{aligned}$$

where we used the property of the **Hellinger affinity** ρ seen in Lecture ??, and where we define r as the maximum appearing in the last expression.

Since the model is identifiable, we have $\rho(P_j, P_{\theta_0}) < 1$ for all $j \neq \theta_0$ (otherwise the Hellinger distance between the measures would be zero and they would be equal), therefore $r < 1$.

Thus, for all $\varepsilon > 0$,

$$\mathbb{P}_{\theta_0} \left(\max_{j \neq \theta_0} \ell_j(\mathbf{X}) \geq \varepsilon \ell_{\theta_0}(\mathbf{X}) \right) \leq \sum_{j \neq \theta_0} \frac{r^n}{\sqrt{\varepsilon}} = \frac{(k-1)r^n}{\sqrt{\varepsilon}} \xrightarrow{n \rightarrow \infty} 0.$$

In other words, under P_{θ_0} ,

$$\frac{\max_{j \neq \theta_0} \ell_j(\mathbf{X})}{\ell_{\theta_0}(\mathbf{X})} \xrightarrow{\mathbb{P}} 0.$$

Since $\Pi[\{\theta_0\} | \mathbf{X}] \leq 1$, we indeed obtain $\Pi[\{\theta_0\} | \mathbf{X}] \xrightarrow{\mathbb{P}} 1$ under P_{θ_0} . ■

Rates of Convergence

We can naturally extend the notion of consistency by allowing ε to vary, and typically to tend to 0 with n .

Definition

We say that the posterior distribution $\Pi[\cdot | \mathbf{X}] = \Pi[\cdot | X_1, \dots, X_n]$ converges at rate (at least) ε_n at the point $\theta_0 \in \Theta$ if, under P_{θ_0} ,

$$\Pi[\{\theta : \|\theta - \theta_0\| \leq \varepsilon_n\} | \mathbf{X}] \xrightarrow{\mathbb{P}} 1.$$

In the framework of parametric models, we will typically be able to show convergence at a rate M_n/\sqrt{n} , for any sequence $M_n \rightarrow +\infty$, *as slowly as desired*.

We say in this case that we have **convergence at a rate of the order of $1/\sqrt{n}$** .

Lower Bounds and the Gaussian Model

Remark The fact that a given rate cannot be improved in order of magnitude can sometimes be obtained by deriving a 'lower bound'. We say that ζ_n is a **lower bound for the rate of convergence** of $\Pi[\cdot | \mathbf{X}]$ at point $\theta_0 \in \Theta$ if, under P_{θ_0} ,

$$\Pi[\{\theta : \|\theta - \theta_0\| \leq \zeta_n\} | \mathbf{X}] \xrightarrow{\mathbb{P}} 0.$$

Proposition

In the Gaussian model $\mathcal{P} = \{\mathcal{N}(\theta, 1)^{\otimes n}, \theta \in \mathbb{R}\}$ with a prior $\Pi = \mathcal{N}(a, 1)$ on θ , the posterior distribution $\Pi[\cdot | \mathbf{X}]$ converges at every point $\theta_0 \in \mathbb{R}$, at a rate of the order of $1/\sqrt{n}$.

More precisely, for all $\theta_0 \in \mathbb{R}$ and m_n, M_n two sequences such that $m_n \rightarrow 0$ and $M_n \rightarrow +\infty$, under P_{θ_0} ,

$$\Pi\left[\left\{\theta : \frac{m_n}{\sqrt{n}} \leq \|\theta - \theta_0\| \leq \frac{M_n}{\sqrt{n}}\right\} | \mathbf{X}\right] \xrightarrow{\mathbb{P}} 1.$$

Limiting Form and Bernstein-von Mises Theorem

In parametric models, the rate of convergence will often be $1/\sqrt{n}$. This results from the **Bernstein-von Mises theorem**.

We will state a limiting form result for the posterior distribution. This result can be seen as a sort of **central limit theorem**, for much more general objects than an empirical mean.

Asymptotically, posterior distributions typically resemble Gaussian distributions, centered at an "optimal" estimator, and with a variance equal to a constant divided by n .

Total Variation Distance

To show such a result, it is first useful to recall a notion of proximity for two distributions, already seen in Lecture ??: the **total variation distance**.

Let P, Q be two probability measures with densities p and q . The total variation distance between P and Q satisfies:

$$d_{\text{TV}}(P, Q) = \frac{1}{2} \int |p(x) - q(x)| d\mu(x).$$

Example Let $P_n = \text{Unif}[0, 1 + \frac{1}{n}]$ and $P = \text{Unif}[0, 1]$. We calculate:

$$2d_{\text{TV}}(P, P_n) = \int_0^1 \left| \frac{1}{1 + \frac{1}{n}} - 1 \right| du + \int_1^{1 + \frac{1}{n}} \frac{1}{1 + \frac{1}{n}} du = \frac{2}{n+1} = o(1).$$

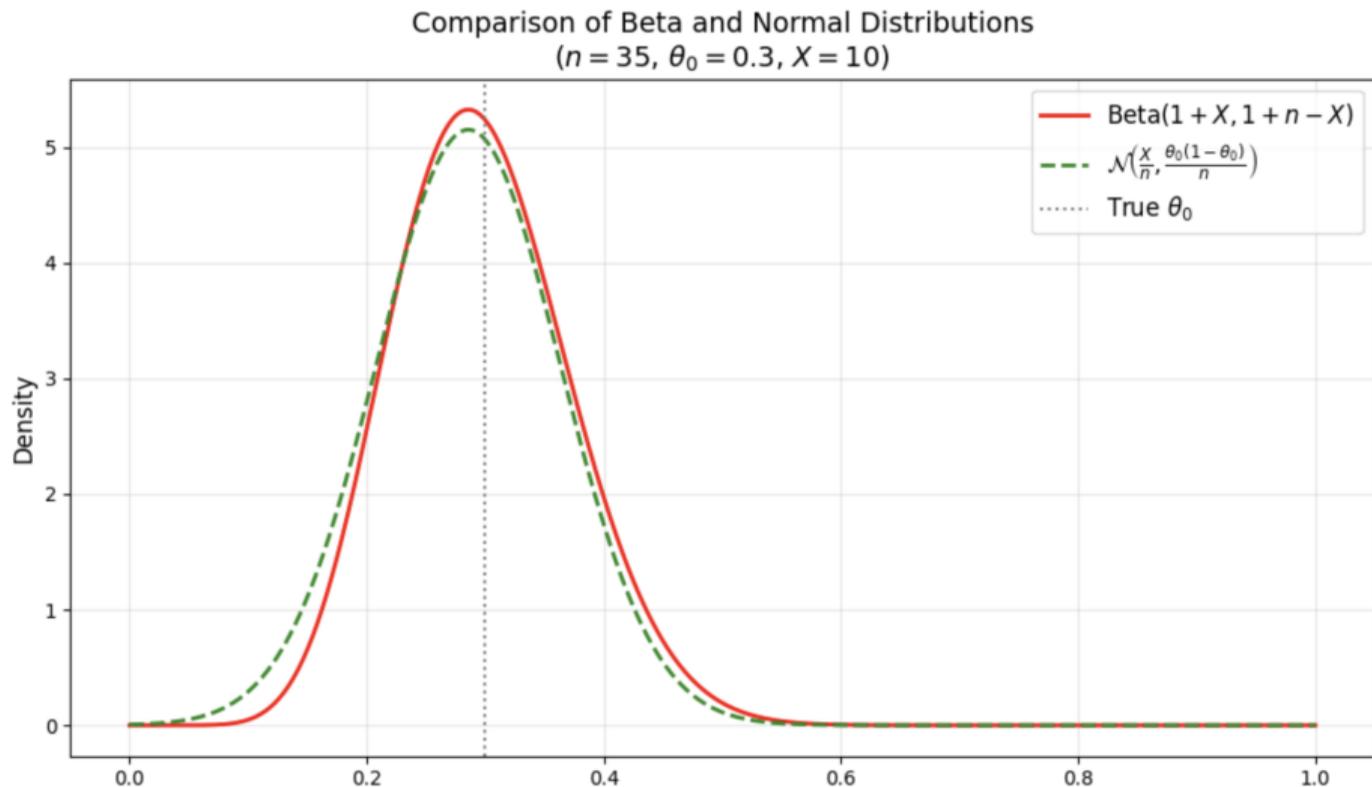
Historical Context: Laplace and the Binomial Model

Laplace, in the early 1800s, noticed and demonstrated that in the binomial model $\{\mathcal{B}(n, \theta), \theta \in (0, 1)\}$, with a uniform prior distribution on θ (i.e., the model considered by T. Bayes):

- The posterior distribution is a $\text{Beta}(1 + \mathbf{X}, 1 + n - \mathbf{X})$ distribution.
- This distribution strangely resembles a $\mathcal{N}\left(\frac{\mathbf{X}}{n}, \frac{\theta_0(1-\theta_0)}{n}\right)$ distribution if \mathbf{X} actually follows a $\mathcal{B}(n, \theta_0)$ distribution.

Note that \mathbf{X}/n happens to be the **maximum likelihood estimator** in this model. Since then, many statisticians have been interested in this phenomenon, including Bernstein, von Mises, and Le Cam.

Historical Context: Laplace and the Binomial Model



Consistent Tests

Let $\mathcal{P} = \{P_\theta^{\otimes n}, \theta \in \Theta\}$, with $\Theta \subset \mathbb{R}$ open.

We will assume that we can always find a **consistent test** to test a point against the complement of a ball: for all $\theta_0 \in \Theta$ and $\varepsilon > 0$, there exists a sequence of tests $\phi_n = \phi_n(\mathbf{X})$ with, as $n \rightarrow \infty$,

$$\mathbb{E}_{\theta_0} \phi_n \rightarrow 0, \quad \sup_{|\theta - \theta_0| \geq \varepsilon} \mathbb{E}_\theta (1 - \phi_n) \rightarrow 0,$$

where we recall that \mathbb{E}_θ denotes the expectation under the distribution $P_\theta^{\otimes n}$.

This condition is generally quite easily verified. For example, in the model $\mathcal{P} = \{\mathcal{N}(\theta, 1)^{\otimes n}, \theta \in \mathbb{R}\}$ it suffices to take $\phi_n = \mathbb{1}_{|\bar{\mathbf{x}} - \theta_0| \geq \varepsilon/2}$ (check it!).

Regular Models

We also assume that \mathcal{P} is a regular model that satisfies the following conditions, denoting p_θ the density of P_θ :

- (a) for almost every x , the function $\theta \mapsto p_\theta(x)$ is absolutely continuous on Θ .
- (b) for all $\theta_0 \in \Theta$, for almost all x , the function $\theta \mapsto p'_\theta(x)$ is continuous at θ_0 ;
- (c) for all $\theta_0 \in \Theta$, the Fischer information $\mathbf{I}(\theta_0)$ exists and the function $\theta \mapsto \mathbf{I}(\theta)$ is continuous on Θ .
- (d) $\theta \mapsto \sqrt{p_\theta(x)}$ is of class C^1 for all x .
- (e) The MLE is **consistent** (at point θ_0).
- (f) There exists a measurable function $\lambda(\cdot)$ with $\mathbb{E}_{\theta_0}[\lambda^2(X_1)] < \infty$ such that, for all θ_1, θ_2 in a neighborhood of θ_0 and all x ,

$$|\log p_{\theta_1}(x) - \log p_{\theta_2}(x)| \leq \lambda(x)|\theta_1 - \theta_2|.$$

Bernstein-von Mises (BvM)

Theorem (Bernstein-von Mises Theorem)

Let $\mathcal{P} = \{P_\theta^{\otimes n}, \theta \in \Theta\}$, with $\Theta \subset \mathbb{R}$ open, be a regular model. Let $\theta_0 \in \Theta$. Assume there exists ϕ_n as above and that the prior distribution Π on Θ satisfies:

- Π has a density π with respect to the Lebesgue measure on \mathbb{R} .
- $\pi(\theta_0) > 0$ and $\pi(\cdot)$ is continuous at point θ_0 .

Assume that the Fisher information $\mathbf{I}(\theta_0)$ at point θ_0 is strictly positive. Let $\hat{\theta}_n(\mathbf{X})$ be the maximum likelihood estimator in this model. Then as $n \rightarrow \infty$,

$$d_{\text{TV}} \left(\Pi[\cdot | \mathbf{X}], \mathcal{N} \left(\hat{\theta}_n(\mathbf{X}), \frac{\mathbf{I}(\theta_0)^{-1}}{n} \right) \right) \xrightarrow{\mathbb{P}} 0,$$

under P_{θ_0} .

Implications of the Bernstein-von Mises Theorem

This result implies a remarkable proximity between frequentist limit distributions and Bayesian limit distributions.

Indeed, the BvM theorem gives:

$$\mathcal{L}(\theta - \hat{\theta}_n(\mathbf{X}) \mid \mathbf{X}) \approx \mathcal{N}\left(0, \frac{\mathbf{I}(\theta_0)^{-1}}{n}\right).$$

Moreover, one of the fundamental results on maximum likelihood in regular models is that:

$$\mathcal{L}(\hat{\theta}_n(\mathbf{X}) - \theta_0) \approx \mathcal{N}\left(0, \frac{\mathbf{I}(\theta_0)^{-1}}{n}\right).$$

We note that this is the **same limit distribution!**

Proof (BvM)

Proof in the Gaussian model for a Gaussian prior. We set $\mathcal{P} = \{\mathcal{N}(\theta, 1)^{\otimes n}, \theta \in \mathbb{R}\}$ and $\Pi = \mathcal{N}(a, 1)$ for a fixed $a \in \mathbb{R}$.

Given the explicit expression of the posterior distribution and the MLE, and the fact that $\mathbf{I}(\theta) = 1$ for all θ in the Gaussian model, we need to show that under P_{θ_0} ,

$$d_{\text{TV}} \left(\mathcal{N} \left(\frac{a + n\bar{X}_n}{n+1}, \frac{1}{n+1} \right), \mathcal{N} \left(\bar{X}_n, \frac{1}{n} \right) \right) \xrightarrow{\mathbb{P}} 0.$$

There are several possible proofs, here we use a comparison of distances and an explicit calculation.

Proof (BvM)

By the translation invariance of the total variation distance and the **triangle inequality**, we have:

$$\begin{aligned} d_{\text{TV}} \left(\mathcal{N} \left(\frac{a + n\bar{X}_n}{n+1}, \frac{1}{n+1} \right), \mathcal{N} \left(\bar{X}_n, \frac{1}{n} \right) \right) &= d_{\text{TV}} \left(\mathcal{N} \left(\frac{a - \bar{X}_n}{n+1}, \frac{1}{n+1} \right), \mathcal{N} \left(0, \frac{1}{n} \right) \right) \\ &\leq \underbrace{d_{\text{TV}} \left(\mathcal{N} \left(\frac{a - \bar{X}_n}{n+1}, \frac{1}{n+1} \right), \mathcal{N} \left(0, \frac{1}{n+1} \right) \right)}_{\text{(I)}} \\ &\quad + \underbrace{d_{\text{TV}} \left(\mathcal{N} \left(0, \frac{1}{n+1} \right), \mathcal{N} \left(0, \frac{1}{n} \right) \right)}_{\text{(II)}} \end{aligned}$$

Proof (BvM)

We now combine the distances with the square root of the corresponding **Kullback-Leibler divergence** (using Pinsker's inequality below). For term **(I)**:

$$\begin{aligned} \text{(I)} &\leq \sqrt{\text{KL} \left(\mathcal{N} \left(\frac{a - \bar{X}_n}{n+1}, \frac{1}{n+1} \right), \mathcal{N} \left(0, \frac{1}{n+1} \right) \right)} \\ &= \frac{|\bar{X}_n - a|}{\sqrt{2(n+1)}}. \end{aligned}$$

Since under P_{θ_0} , $\bar{X}_n \xrightarrow{\mathbb{P}} \theta_0$, we conclude by **Slutsky's lemma** that under P_{θ_0} ,

$$\frac{|\bar{X}_n - a|}{\sqrt{2(n+1)}} \xrightarrow{\mathbb{P}} 0.$$

Lemma (Pinsker's Inequality)

Let P, Q be two probability measures. Then $d_{\text{TV}}(P, Q) \leq \sqrt{\text{KL}(P, Q)}$.

On the other hand, for term **(II)**:

$$\begin{aligned} \text{(II)} &\leq \sqrt{\text{KL} \left(\mathcal{N} \left(0, \frac{1}{n+1} \right), \mathcal{N} \left(0, \frac{1}{n} \right) \right)} \\ &= \sqrt{\frac{1}{2} \left(\log \left(1 + \frac{1}{n} \right) - \frac{1}{n+1} \right)} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Thus, under P_{θ_0} , the sum of the two distances converges in probability to 0, which is what we needed to demonstrate. ■

Asymptotic Confidence of Credible Regions

Again, we place ourselves in dimension $d = 1$, meaning $\Theta \subset \mathbb{R}$.

We assume the posterior cumulative distribution function (CDF) $F_{\theta|\mathbf{X}}$ is strictly continuous and increasing, and we consider the **credible region** $[a_n(\mathbf{X}), b_n(\mathbf{X})]$ of level $1 - \alpha$ formed by the quantiles of the posterior distribution:

$$\begin{aligned}\Pi(] - \infty, a_n(\mathbf{X})] | \mathbf{X}) &= \frac{\alpha}{2}, \\ \Pi([b_n(\mathbf{X}), +\infty[| \mathbf{X}) &= \frac{\alpha}{2}.\end{aligned}$$

Asymptotic Expansion of Bounds

Below, $o_P(1)$ denotes a quantity that tends to 0 in probability under $P_{\theta_0}^{\otimes n}$.

Theorem

Let $0 < \alpha < 1$ and z_α be the quantile of level $1 - \frac{\alpha}{2}$ of a standard normal distribution. Assume the BvM theorem holds. Then, for $a_n(\mathbf{X}), b_n(\mathbf{X})$ as before, and $\hat{\theta}_n$ the MLE,

$$[a_n(\mathbf{X}), b_n(\mathbf{X})] = \left[\hat{\theta}_n(\mathbf{X}) - \frac{z_\alpha}{\sqrt{n\mathbf{l}(\theta_0)}}(1 + o_P(1)), \hat{\theta}_n(\mathbf{X}) + \frac{z_\alpha}{\sqrt{n\mathbf{l}(\theta_0)}}(1 + o_P(1)) \right].$$

Let A and B be the measurable sets defined by:

$$A =] - \infty, a_n(\mathbf{X})], \quad B = [b_n(\mathbf{X}), +\infty[.$$

By definition of $a_n(\mathbf{X})$ and $b_n(\mathbf{X})$, we have:

$$\Pi[A | \mathbf{X}] = \Pi[B | \mathbf{X}] = \frac{\alpha}{2}.$$

The BvM theorem is satisfied by hypothesis and according to the definition of the total variation distance, we thus have:

$$\sup_{\Lambda} \left| \Pi[\Lambda | \mathbf{X}] - \mathcal{N} \left(\hat{\theta}_n(\mathbf{X}), \frac{\mathbf{I}(\theta_0)^{-1}}{n} \right) (\Lambda) \right| = o_P(1).$$

In particular, applying this to $\Lambda = A$ (or B), we deduce that:

$$\mathcal{N}\left(\hat{\theta}_n(\mathbf{X}), \frac{\mathbf{I}(\theta_0)^{-1}}{n}\right)(A) = \frac{\alpha}{2} + o_P(1).$$

Denoting by Φ the CDF of a $\mathcal{N}(0, 1)$ distribution, this can be rewritten as:

$$\Phi\left(\sqrt{n\mathbf{I}(\theta_0)}(a_n(\mathbf{X}) - \hat{\theta}_n(\mathbf{X}))\right) = \frac{\alpha}{2} + o_P(1),$$

or equivalently:

$$\sqrt{n\mathbf{I}(\theta_0)}(a_n(\mathbf{X}) - \hat{\theta}_n(\mathbf{X})) = \Phi^{-1}\left(\frac{\alpha}{2} + o_P(1)\right).$$

Now, Φ^{-1} is continuous, so by the **continuous mapping theorem**, we deduce that the preceding expression converges in probability to $\Phi^{-1}(\alpha/2) = -z_\alpha$.

We obtain:

$$a_n(\mathbf{X}) = \hat{\theta}_n(\mathbf{X}) - \frac{z_\alpha}{\sqrt{n\mathbf{I}(\theta_0)}}(1 + o_{\mathbb{P}}(1)),$$

and the result for $b_n(\mathbf{X})$ is obtained in the same way. ■

Connection to Frequentist Confidence Intervals

This result gives a first-order asymptotic expansion of the bounds of the credible interval $[a_n(\mathbf{X}), b_n(\mathbf{X})]$ defined from the posterior quantiles.

Note that this interval asymptotically coincides with the "ideal" frequentist asymptotic confidence interval that we would like to be able to construct from the MLE $\hat{\theta}_n(\mathbf{X})$.

Indeed, if we assume the conditions are met to obtain

$$\sqrt{n}(\hat{\theta}_n(\mathbf{X}) - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{I}(\theta_0)^{-1}),$$

then the interval

$$I^*(\mathbf{X}) = \left[\hat{\theta}_n(\mathbf{X}) \pm \frac{z_\alpha}{\sqrt{n\mathbf{I}(\theta_0)}} \right]$$

has an asymptotic confidence level of $1 - \alpha$, since:

$$\mathbb{P}_{\theta_0} \left(\sqrt{n\mathbf{I}(\theta_0)} \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| \leq z_\alpha \right) \xrightarrow{n \rightarrow \infty} \mathbb{P}(|\mathcal{N}(0, 1)| \leq z_\alpha) = 1 - \alpha.$$

Unknown Fisher Information & Bayesian Advantage

⚠ In general, however, the Fisher information $\mathbf{I}(\theta_0)$ is unknown since it depends on θ_0 .

A standard solution consists of replacing $\mathbf{I}(\theta_0)$ with an **estimator**, for example $\mathbf{I}(\hat{\theta}_n(\mathbf{X}))$.

Example: Under usual regularity conditions, $\theta \mapsto \mathbf{I}(\theta)$ is continuous, so the convergence in probability of $\hat{\theta}_n(\mathbf{X})$ to θ_0 implies that of $\mathbf{I}(\hat{\theta}_n(\mathbf{X}))$ to $\mathbf{I}(\theta_0)$, and we can apply Slutsky's lemma.

One of the interests of the Bayesian approach is that obtaining the credible region is **"automatic"** (⚠ provided we know how to calculate the posterior quantiles, which is not always obvious).

Asymptotic Confidence

Asymptotic confidence of credible regions

Assume the BvM theorem holds. Then the credible interval $I(\mathbf{X}) = [a_n(\mathbf{X}), b_n(\mathbf{X})]$ is an asymptotic confidence interval at level $1 - \alpha$, that is:

$$\mathbb{P}_{\theta_0}(\theta_0 \in [a_n(\mathbf{X}), b_n(\mathbf{X})]) \xrightarrow{n \rightarrow \infty} 1 - \alpha.$$

Proof. It suffices to show that $\mathbb{P}_{\theta_0}(\theta_0 < a_n(\mathbf{X})) \rightarrow \alpha/2$ and that $\mathbb{P}_{\theta_0}(\theta_0 > b_n(\mathbf{X})) \rightarrow \alpha/2$. For this, we use the asymptotic expansions obtained in the previous theorem:

$$\begin{aligned} \mathbb{P}_{\theta_0}(\theta_0 < a_n(\mathbf{X})) &= \mathbb{P}_{\theta_0} \left(\theta_0 < \hat{\theta}_n(\mathbf{X}) - \frac{z_\alpha}{\sqrt{nl(\theta_0)}} (1 + o_{\mathbb{P}}(1)) \right) \\ &= \mathbb{P}_{\theta_0} \left(\sqrt{nl(\theta_0)}(\hat{\theta}_n(\mathbf{X}) - \theta_0) - o_{\mathbb{P}}(1) > z_\alpha \right). \end{aligned}$$

The quantity on the left of the ' $>$ ' sign converges in law to a $\mathcal{N}(0, 1)$ variable, so the expression converges to $\alpha/2$. We do the same for $\mathbb{P}_{\theta_0}(\theta_0 > b_n(\mathbf{X}))$, which concludes the proof. ■

Remarks on Posterior Analysis

From the perspective of mathematical analysis of posterior distributions:

- **Decision Theory Framework:** Particularly useful for suggesting estimators adapted to given loss functions (e.g., posterior mean for quadratic loss, Bayes classifiers for classification loss).
- **Frequentist Analysis:** To have 'absolute' results (an optimality notion independent of the prior choice), we analyze under \mathbb{P}_{θ_0} of $\Pi[\cdot | \mathbf{X}]$.
- **BvM Guarantees:** For regular parametric models and a prior with 'some' mass around all true potential θ_0 , BvM guarantees a $1/\sqrt{n}$ convergence rate, and a remarkable **duality with limit theory for the MLE** (recovering the same optimal limit variance).

Empirical Bayes procedures

Can we propose a convergence theory for empirical Bayes posteriors $\Pi_{\hat{\alpha}(\mathbf{X})}[\cdot \mid \mathbf{X}]$?

- The BvM theorem cannot be applied directly if the prior depends on the data.
- In regular models, empirical Bayes posteriors can have the same behavior as a hierarchical posterior: we say they "merge".
- Generally, this relies on showing the estimator $\hat{\alpha}(\mathbf{X})$ concentrates with high probability.
- However, this is delicate; merging does not always happen, especially in **high-dimensional models** (next week) where the choice of prior details matters even more.