Barron, Schervish, and Wasserman (1999) The consistency of posterior distributions in nonparametric problems

Flavio Argentieri

Bocconi University

November 15, 2021

Overview

1. General framework
2. Theoretical Core
3. Histograms Example
4. Main Proof
5. Appendix I: Histograms Assumption 1
6. Appendix II: Preliminary Lemmas

Motivating Question and Literature

- Does nonparametric Bayesian inference possess good **consistency** properties?

Motivating Question and Literature

- Does nonparametric Bayesian inference possess good **consistency** properties?
- Diaconis and Freedman (1986) showed that, even if the prior puts positive mass in weak neighborhoods of the true density, the posterior mass of every weak neighborhood of the true density does **not necessarily** goes to 1 almost surely with respect to the true underlying distribution.

Motivating Question and Literature

- Does nonparametric Bayesian inference possess good **consistency** properties?
- Diaconis and Freedman (1986) showed that, even if the prior puts positive mass in weak neighborhoods of the true density, the posterior mass of every weak neighborhood of the true density does **not necessarily** goes to 1 almost surely with respect to the true underlying distribution.
- Doob (1949) showed consistency of the posterior under very weak conditions, but **only almost surely w.r.t. the prior**.

Motivating Question and Literature

- Does nonparametric Bayesian inference possess good **consistency** properties?
- Diaconis and Freedman (1986) showed that, even if the prior puts positive mass in weak neighborhoods of the true density, the posterior mass of every weak neighborhood of the true density does **not necessarily** goes to 1 almost surely with respect to the true underlying distribution.
- Doob (1949) showed consistency of the posterior under very weak conditions, but **only almost surely w.r.t. the prior**.
- Schwartz (1965) showed that if the prior puts positive mass in each Kullback-Leibler neighborhood of the true density, then the posterior does accumulate in **weak neighborhoods** of f_0 a.s. w.r.t. the true density. However, weak neighborhoods may contain many distributions that do not resemble the true density.

Main Object of Discussion

- Barron, Schervish, and Wasserman (1999), then, show that, under two assumptions, the posterior accumulates in Hellinger neighborhoods of the true density a.s. $[P_o^\infty]$. Thus, **consistency in Hellinger distance holds**, which is equivalent to consistency in total variation.

General framework

Primitive elements

- $(\mathcal{X}, \mathcal{B})$ measurable space, $\mathcal{B}$ separable

Primitive elements

- $(\mathcal{X}, \mathcal{B})$ measurable space, $\mathcal{B}$ separable
- λ probability measure on $(\mathcal{X}, \mathcal{B})$

Primitive elements

- $(\mathcal{X}, \mathcal{B})$ measurable space, $\mathcal{B}$ separable
- λ probability measure on $(\mathcal{X}, \mathcal{B})$
- $\mathcal{Q} := \{\mu \text{ measure on } (\mathcal{X}, \mathcal{B}) : \mu(\mathcal{X}) < \infty,\ \mu \ll \lambda\}$ set of non-negative finite measures on $\mathcal{X}$ that are absolutely continuous w.r.t. λ

Primitive elements

- $(\mathcal{X}, \mathcal{B})$ measurable space, $\mathcal{B}$ separable
- λ probability measure on $(\mathcal{X}, \mathcal{B})$
- $\mathcal{Q} := \{\mu \text{ measure on } (\mathcal{X}, \mathcal{B}) : \mu(\mathcal{X}) < \infty,\ \mu \ll \lambda\}$ set of non-negative finite measures on $\mathcal{X}$ that are absolutely continuous w.r.t. λ
- $\mathcal{P} := \{\mu \in \mathcal{Q} : \mu(\mathcal{X}) = 1\}$ set of λ-a.c. probability measures on $\mathcal{X}$

Hellinger metric

- **Hellinger metric** on $\mathcal{Q}$:

$$d' : \mathcal{Q}^2 \to \mathbb{R}_+$$

$$(Q_1, Q_2) \mapsto d'(Q_1, Q_2) = \left\{ \int_{\chi} \left[\sqrt{f_1(x)} - \sqrt{f_2(x)} \right]^2 d\lambda(x) \right\}^{\frac{1}{2}}$$

where $f_i = \frac{dQ_i}{d\lambda_i}$ for $i = 1, 2$

Hellinger metric

- **Hellinger metric** on $\mathcal{Q}$:

$$d' : \mathcal{Q}^2 \to \mathbb{R}_+$$
$$(Q_1, Q_2) \mapsto d'(Q_1, Q_2) = \left\{ \int_{\chi} \left[\sqrt{f_1(x)} - \sqrt{f_2(x)} \right]^2 d\lambda(x) \right\}^{\frac{1}{2}}$$

where $f_i = \frac{dQ_i}{d\lambda_i}$ for $i = 1, 2$

- $\mathcal{D} := \mathcal{B}(\mathcal{Q})$ Borel σ-field of subsets of $\mathcal{Q}$ induced by open sets under the metric d'

Hellinger metric

- **Hellinger metric** on $\mathcal{Q}$:

$$d' : \mathcal{Q}^2 \to \mathbb{R}_+$$

$$(Q_1, Q_2) \mapsto d'(Q_1, Q_2) = \left\{ \int_{\chi} \left[\sqrt{f_1(x)} - \sqrt{f_2(x)} \right]^2 d\lambda(x) \right\}^{\frac{1}{2}}$$

where $f_i = \frac{dQ_i}{d\lambda_i}$ for $i = 1, 2$

- $\mathcal{D} := \mathcal{B}(\mathcal{Q})$ Borel σ-field of subsets of $\mathcal{Q}$ induced by open sets under the metric d'

- $\mathcal{C} := \{A \cap \mathcal{P} : A \in \mathcal{D}\}$ restriction of $\mathcal{D}$ on $\mathcal{P}$

Hellinger metric

- **Hellinger metric** on $\mathcal{Q}$:

$$d' : \mathcal{Q}^2 \to \mathbb{R}_+$$

$$(Q_1, Q_2) \mapsto d'(Q_1, Q_2) = \left\{ \int_{\chi} \left[\sqrt{f_1(x)} - \sqrt{f_2(x)} \right]^2 d\lambda(x) \right\}^{\frac{1}{2}}$$

where $f_i = \frac{dQ_i}{d\lambda_i}$ for $i = 1, 2$

- $\mathcal{D} := \mathcal{B}(\mathcal{Q})$ Borel σ-field of subsets of $\mathcal{Q}$ induced by open sets under the metric d'
- $\mathcal{C} := \{A \cap \mathcal{P} : A \in \mathcal{D}\}$ restriction of $\mathcal{D}$ on $\mathcal{P}$
- $A_\epsilon = \{P \in \mathcal{P} : d'(P_0, P) \leq \epsilon\}$

Kullback-Leibler Information

- $\forall P, Q \in \mathcal{P}$ the **Kullback-Leibler information** is defined as:

$$\mathcal{I}(P, Q) = \int_{\chi} log\left(\frac{f_P(x)}{f_Q(x)}\right) f_P(x) d\lambda(x)$$

Kullback-Leibler Information

- $\forall P, Q \in \mathcal{P}$ the **Kullback-Leibler information** is defined as:

$$\mathcal{I}(P, Q) = \int_{\chi} log\left(\frac{f_P(x)}{f_Q(x)}\right) f_P(x) d\lambda(x)$$

- The integrand is intended to be 0 whenever $f_P(x) = 0$

Kullback-Leibler Information

- $\forall P, Q \in \mathcal{P}$ the **Kullback-Leibler information** is defined as:

$$\mathcal{I}(P, Q) = \int_{\mathcal{X}} log\left(\frac{f_P(x)}{f_Q(x)}\right) f_P(x) d\lambda(x)$$

- The integrand is intended to be 0 whenever $f_P(x) = 0$
- $N_\epsilon = \{P \in \mathcal{P} : \mathcal{I}(P_0, P) \leq \epsilon\}$ KL neighborhood of P_0

More Primitive Elements

- $(X_n)_{n\geq 1}$ i.i.d. random variables on $(\mathcal{X}, \mathcal{B})$ each with distribution P_0 λ-a.c. and Radon Nikodym derivative $f_0 = \frac{dP_0}{d\lambda}$

More Primitive Elements

- $(X_n)_{n\geq 1}$ i.i.d. random variables on $(\mathcal{X}, \mathcal{B})$ each with distribution P_0 λ-a.c. and Radon Nikodym derivative $f_0 = \frac{dP_0}{d\lambda}$
- $p_n(x^n) = \prod_{i=1}^{n} f_0(x_i)$ density of the true n-fold product measure of P_0

More Primitive Elements

- $(X_n)_{n\geq 1}$ i.i.d. random variables on $(\mathcal{X}, \mathcal{B})$ each with distribution P_0 λ-a.c. and Radon Nikodym derivative $f_0 = \frac{dP_0}{d\lambda}$
- $p_n(x^n) = \prod_{i=1}^{n} f_0(x_i)$ density of the true n-fold product measure of P_0
- $D_n(x^n, P) = \frac{1}{n} log\left(\frac{p_n(x^n)}{\prod_{i=1}^{n} f_P(x_i)}\right)$ sample KL information for $P \in \mathcal{P}$

More Primitive Elements

- $(X_n)_{n\geq 1}$ i.i.d. random variables on $(\mathcal{X}, \mathcal{B})$ each with distribution P_0 λ-a.c. and Radon Nikodym derivative $f_0 = \frac{dP_0}{d\lambda}$
- $p_n(x^n) = \prod_{i=1}^n f_0(x_i)$ density of the true n-fold product measure of P_0
- $D_n(x^n, P) = \frac{1}{n} log\big(\frac{p_n(x^n)}{\prod_{i=1}^n f_P(x_i)}\big)$ sample KL information for $P \in \mathcal{P}$
- For $\delta > 0,\ C \in \mathcal{C},$ we define the δ-**upper metric entropy** as:

$$\mathcal{K}(\delta, C) := inf\left\{k : \exists f_1^U, \dots, f_k^U : \quad i)\ \int_{\mathcal{X}} f_i^U(x) d\lambda(x) \leq 1 + \delta \right.$$

$$\left. ii)\ \forall P \in C,\ \exists i : f_P \leq f_i^U\ a.e.[\lambda] \right\}$$

Bayesian Updating

- π probability measure on $(\mathcal{P}, \mathcal{C})$

Bayesian Updating

- π probability measure on $(\mathcal{P}, \mathcal{C})$
- $m_n(x^n) = \int_{\mathcal{P}} \prod_{i=1}^n f_P(x_i) d\pi(P)$ **predictive density** of X^n under π

Bayesian Updating

- π probability measure on $(\mathcal{P}, \mathcal{C})$
- $m_n(x^n) = \int_{\mathcal{P}} \prod_{i=1}^n f_P(x_i) d\pi(P)$ **predictive density** of X^n under π
- The **posterior probability** of $B \in \mathcal{C}$ given $X^n = x^n$ is computed as:

$$\pi(B|x^n) = \frac{\int_B \prod_{i=1}^n f_P(x_i) d\pi(P)}{m_n(x^n)}$$

Theoretical Core

Assumptions

Assumption 1

$$\forall \epsilon > 0,\ \pi(N_\epsilon) > 0$$

Assumptions

Assumption 1

$$\forall \epsilon > 0,\ \pi(N_\epsilon) > 0$$

Assumption 2

$\forall \epsilon > 0,\ \exists\ (\mathcal{F}_n)_{n \geq 1} \in \mathcal{C}^{\mathbb{N}},\ \exists\ c, c_1, c_2, \delta \in \mathbb{R}_{++} : \delta < \frac{\epsilon^2}{4}\ ;\ c < \frac{(\epsilon - \sqrt{\delta})^2 - \delta}{2}$
and, for all but finitely many n,

$i)\quad \pi(\mathcal{F}_n^C) \leq c_1 exp(nc_2)$

$ii)\quad \mathcal{K}(\mathcal{F}_n, \delta) \leq nc$

Main Result

Theorem 1

Let A_ϵ be an Hellinger neighborhood of the true density. Under assumptions 1 and 2, $\forall \epsilon > 0$:

$$\lim_{n\to\infty} \pi(A_\epsilon | x^n) = 1 \quad a.s.\ [P_0^\infty]$$

Important Implication

Corollary 1

Let $\hat{f}_n(\cdot) = \int f_P(\cdot) d\pi(P|x^n)$. Under assumptions 1 and 2:

$$\lim_{n \to \infty} d(f_0, \hat{f}_n) = 0 \quad a.s.\ [P_0^\infty]$$

Where d is the Hellinger pseudo-metric on the set of non-negative functions that are integrable w.r.t. λ.

Conditions to Check

Lemma 8

Let $(\mathcal{T}_n)_{n\geq 1}$ be a sequence of finite measurable partitions of $\mathcal{X}$ and let $N_n = |\mathcal{T}_n|$. $\forall n$, let $a_n > 0$ and suppose $\lambda(A) = \frac{1}{N_n}\ \forall A \in \mathcal{T}_n$. Define $\mathcal{F}_n := \{P \in \mathcal{P} : \forall A \in \mathcal{T}_n, \forall x, y \in A, |f_P(x) - f_P(y)| \leq a_n\}$. Then

$$\mathcal{K}(\mathcal{F}_n, 2a_n) \leq N_n \left[1 + log\left(1 + \frac{1}{2N_n a_n}\right)\right]$$

Conditions to Check

Lemma 8

Let $(\mathcal{T}_n)_{n\geq 1}$ be a sequence of finite measurable partitions of $\mathcal{X}$ and let $N_n = |\mathcal{T}_n|$. $\forall n$, let $a_n > 0$ and suppose $\lambda(A) = \frac{1}{N_n}$ $\forall A \in \mathcal{T}_n$. Define $\mathcal{F}_n := \{P \in \mathcal{P} : \forall A \in \mathcal{T}_n, \forall x, y \in A, |f_P(x) - f_P(y)| \leq a_n\}$. Then

$$\mathcal{K}(\mathcal{F}_n, 2a_n) \leq N_n \left[1 + log\left(1 + \frac{1}{2N_n a_n}\right)\right]$$

Corollary 2

$\forall \epsilon > 0$, $n \in \mathbb{N}$, let $N_n \leq \frac{n\epsilon^2}{10}$, $a_n = \frac{\epsilon^2}{32}$, $\delta = \frac{\epsilon^2}{16}$. If $\lim_{n\to\infty} N_n = \infty$, then $(\mathcal{F}_n)_{n\geq 1}$ as defined in lemma 8 satisfies $\mathcal{K}(\mathcal{F}_n, \delta) \leq \frac{n\epsilon^2}{5}$ ultimately, thus satisfying assumption 2 part ii).

Histograms Example

An Important Special Case

- We focus on **real-valued random variables**. Formally $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and, $\forall n \in \mathbb{N},\ \omega \in \mathbb{R}$, we have $X_n : \omega \mapsto \omega$

An Important Special Case

- We focus on **real-valued random variables**. Formally $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and, $\forall n \in \mathbb{N},\ \omega \in \mathbb{R}$, we have $X_n : \omega \mapsto \omega$
- We restrict our attention to priors on the **set of distributions on [0,1]** that are absolutely continuous w.r.t. the Lebesgue measure. This is not a limitation, since, for a given cdf F_* and a random variable X with support $\mathbb{R}$, we have that $Y = F_*(X)$ has support on [0,1]. Thus, we can define priors on sets of distributions on [0,1] and then map back probabilities on the corresponding sets of distributions on $\mathbb{R}$

Histograms I

- $\mathcal{P} = \{\mu \text{ fin. meas. on } ([0,1], \mathcal{B}([0,1])) : \mu([0,1]) = 1,\ \mu \ll \lambda = Leb\}$

Histograms I

- $\mathcal{P} = \{\mu \text{ fin. meas. on } ([0,1], \mathcal{B}([0,1])) : \mu([0,1]) = 1,\ \mu \ll \lambda = \mathit{Leb}\}$
- Assume $\mathcal{I}(P_0, \lambda) < \infty$

Histograms I

- $\mathcal{P} = \{\mu \text{ fin. meas. on } ([0,1], \mathcal{B}([0,1])) : \mu([0,1]) = 1,\ \mu \ll \lambda = Leb\}$
- Assume $\mathcal{I}(P_0, \lambda) < \infty$
- $\forall n \in \mathbb{N}$, let $p_n > 0 : \sum_{n=1}^{\infty} p_n = 1$

Histograms I

- $\mathcal{P} = \{\mu \text{ fin. meas. on } ([0,1], \mathcal{B}([0,1])) : \mu([0,1]) = 1,\ \mu \ll \lambda = Leb\}$
- Assume $\mathcal{I}(P_0, \lambda) < \infty$
- $\forall n \in \mathbb{N}$, let $p_n > 0 : \sum_{n=1}^{\infty} p_n = 1$
- $\forall n \in \mathbb{N}$, let $N_n \in \mathbb{N}$ and let $\mathcal{T}_n$ be the partition:

$$\mathcal{T}_n = \left\{ \left[0, \frac{1}{N_n}\right) \left[\frac{1}{N_n}, \frac{2}{N_n}\right), \cdots, \left[\frac{N_n - 1}{N_n}, 1\right] \right\}$$

Histograms I

- $\mathcal{P} = \{\mu \text{ fin. meas. on } ([0,1], \mathcal{B}([0,1])) : \mu([0,1]) = 1, \ \mu \ll \lambda = Leb\}$
- Assume $\mathcal{I}(P_0, \lambda) < \infty$
- $\forall n \in \mathbb{N}$, let $p_n > 0 : \sum_{n=1}^{\infty} p_n = 1$
- $\forall n \in \mathbb{N}$, let $N_n \in \mathbb{N}$ and let $\mathcal{T}_n$ be the partition:

$$\mathcal{T}_n = \left\{ \left[0, \frac{1}{N_n}\right) \left[\frac{1}{N_n}, \frac{2}{N_n}\right), \cdots, \left[\frac{N_n - 1}{N_n}, 1\right] \right\}$$

- Let $\mathcal{U}_n \in \mathcal{C}$ be the collection of all distributions that have constant density on every interval in $\mathcal{T}_n$

Histograms II

- We choose a prior that places probability p_n on the set $\mathcal{U}_n$

Histograms II

- We choose a prior that places probability p_n on the set $\mathcal{U}_n$
- Let $P \in \mathcal{U}_n$. We can write $P = (f_1, \dots, f_{N_n})$, with $\sum_{i=1}^{N_n} f_i \lambda(A_i) = 1$ and $A_i = \left[\frac{i-1}{N_n}, \frac{i}{N_n}\right)$. In this way $f_i = f_P(x)\ \forall x \in A_i$

Histograms II

- We choose a prior that places probability p_n on the set $\mathcal{U}_n$
- Let $P \in \mathcal{U}_n$. We can write $P = (f_1, \dots, f_{N_n})$, with $\sum_{i=1}^{N_n} f_i \lambda(A_i) = 1$ and $A_i = \left[\frac{i-1}{N_n}, \frac{i}{N_n}\right)$. In this way $f_i = f_P(x)\ \forall x \in A_i$
- Let $a_n > 0$. Conditional on $P \in \mathcal{U}_n$ we assign P/N_n the Dirichlet distribution $Dir(a_n, \dots, a_n)$

Histograms II

- We choose a prior that places probability p_n on the set $\mathcal{U}_n$
- Let $P \in \mathcal{U}_n$. We can write $P = (f_1, \ldots, f_{N_n})$, with $\sum_{i=1}^{N_n} f_i \lambda(A_i) = 1$ and $A_i = \left[\frac{i-1}{N_n}, \frac{i}{N_n}\right)$. In this way $f_i = f_P(x)\ \forall x \in A_i$
- Let $a_n > 0$. Conditional on $P \in \mathcal{U}_n$ we assign P/N_n the Dirichlet distribution $Dir(a_n, \ldots, a_n)$
- By careful choice of N_n and p_n this prior distribution satisfies the conditions of Theorem 1. In particular, we set $N_n = 2^{m_n}$, with $m_n = \lfloor log_2(n) - log_2(log(n)) \rfloor$, and $p_n = (1-a)a^n$ for some $0 < a < 1$. Notice that the choice for N_n implies that, when going from n to $n+1$, each time we either split each interval into two or keep the same partition.

Histograms III

- It can be shown that the prior we constructed satisfies Assumption 1 proof in Appendix I.

Histograms III

- It can be shown that the prior we constructed satisfies Assumption 1 proof in Appendix I.
- Consider, now, the sets $(\mathcal{F}_n)_{n=1}^{\infty}$ as in Corollary 2

Histograms III

- It can be shown that the prior we constructed satisfies Assumption 1 proof in Appendix I.
- Consider, now, the sets $(\mathcal{F}_n)_{n=1}^{\infty}$ as in Corollary 2
- Since each $a_n > 0$ and the probability distributions in $\mathcal{U}_n$ have constant density on every $A \in \mathcal{T}_n$, then $\mathcal{U}_n \subseteq \mathcal{F}_n$. Also, $\mathcal{U}_n \subseteq \mathcal{U}_{n+1}$ for all n, so $\pi(\mathcal{F}_n^C) \leq \sum_{l=n+1}^{\infty} p_l$

Histograms III

- It can be shown that the prior we constructed satisfies Assumption 1 proof in Appendix I.

- Consider, now, the sets $(\mathcal{F}_n)_{n=1}^{\infty}$ as in Corollary 2

- Since each $a_n > 0$ and the probability distributions in $\mathcal{U}_n$ have constant density on every $A \in \mathcal{T}_n$, then $\mathcal{U}_n \subseteq \mathcal{F}_n$. Also, $\mathcal{U}_n \subseteq \mathcal{U}_{n+1}$ for all n, so $\pi(\mathcal{F}_n^C) \leq \sum_{l=n+1}^{\infty} p_l$

- Since $p_n = (1-a)a^n$ for some $0 < a < 1$, then, Assumption 2 part i) holds

Histograms III

- It can be shown that the prior we constructed satisfies Assumption 1 proof in Appendix I.
- Consider, now, the sets $(\mathcal{F}_n)_{n=1}^{\infty}$ as in Corollary 2
- Since each $a_n > 0$ and the probability distributions in $\mathcal{U}_n$ have constant density on every $A \in \mathcal{T}_n$, then $\mathcal{U}_n \subseteq \mathcal{F}_n$. Also, $\mathcal{U}_n \subseteq \mathcal{U}_{n+1}$ for all n, so $\pi(\mathcal{F}_n^C) \leq \sum_{l=n+1}^{\infty} p_l$
- Since $p_n = (1-a)a^n$ for some $0 < a < 1$, then, Assumption 2 part i) holds
- Finally, given that N_n behaves asymptotically as $\frac{n}{log(n)}$, we can apply Corollary 2, so that also Assumption 2 part ii) holds

Main Proof

Proof I

- We will use many preliminary lemmas that are stated in the Appendix Appendix II

Proof I

- We will use many preliminary lemmas that are stated in the Appendix Appendix II
- Let $\epsilon > 0$ and let $(\mathcal{F}_n)_{n \geq 1}$, c, c_1, c_2, and δ be as guaranteed by Assumption 2. We'll show that $\pi(A_\epsilon^C | x^n)$ goes to 0 a.s. $[P_0^\infty]$

Proof I

- We will use many preliminary lemmas that are stated in the Appendix
  Appendix II
- Let $\epsilon > 0$ and let $(\mathcal{F}_n)_{n\geq 1}$, c, c_1, c_2, and δ be as guaranteed by Assumption 2. We'll show that $\pi(A_\epsilon^C|x^n)$ goes to 0 a.s. $[P_0^\infty]$
- Write:

$$\pi(A_\epsilon^C|x^n) = \pi(A_\epsilon^C \cap \mathcal{F}_n|x^n) + \pi(A_\epsilon^C \cap \mathcal{F}_n^C|x^n)$$

  Next, observe that, by Assumption 2, $\pi(\mathcal{F}_n^C) \leq c_1 exp(-nc_2)$ ultimately. Lemma 5, then, assures us that $\pi(A_\epsilon^C \cap \mathcal{F}_n^C|x^n)$ goes to 0 a.s. $[P_0^\infty]$.

Proof I

- We will use many preliminary lemmas that are stated in the Appendix Appendix II
- Let $\epsilon > 0$ and let $(\mathcal{F}_n)_{n\geq 1}$, c, c_1, c_2, and δ be as guaranteed by Assumption 2. We'll show that $\pi(A_\epsilon^C|x^n)$ goes to 0 a.s. $[P_0^\infty]$
- Write:

$$\pi(A_\epsilon^C|x^n) = \pi(A_\epsilon^C \cap \mathcal{F}_n|x^n) + \pi(A_\epsilon^C \cap \mathcal{F}_n^C|x^n)$$

  Next, observe that, by Assumption 2, $\pi(\mathcal{F}_n^C) \leq c_1 exp(-nc_2)$ ultimately. Lemma 5, then, assures us that $\pi(A_\epsilon^C \cap \mathcal{F}_n^C|x^n)$ goes to 0 a.s. $[P_0^\infty]$.
- Let then $C_n = A_\epsilon^C \cap \mathcal{F}_n$. We can focus on $\pi(C_n|x^n)$

Proof II

- By applying the Bayesian updating formula and by multiplying and dividing for the true density, we can write:

Proof II

- By applying the Bayesian updating formula and by multiplying and dividing for the true density, we can write:

$$\pi(C_n|x^n) = \frac{\int_{C_n} \prod_{i=1}^{n} f_P(x_i) d\pi(P)}{\int \prod_{i=1}^{n} f_P(x_i) d\pi(P)}$$

Proof II

- By applying the Bayesian updating formula and by multiplying and dividing for the true density, we can write:

$$
\begin{aligned}
\pi(C_n|x^n) &= \frac{\int_{C_n} \prod_{i=1}^n f_P(x_i) d\pi(P)}{\int \prod_{i=1}^n f_P(x_i) d\pi(P)} \\
&= \frac{p_n(x^n)}{m_n(x^n)} \int_{C_n} \prod_{i=1}^n \frac{f_P(x_i)}{f_0(x_i)} d\pi(P)
\end{aligned}
$$

Proof II

- By applying the Bayesian updating formula and by multiplying and dividing for the true density, we can write:

$$\begin{aligned}\pi(C_n|x^n) &= \frac{\int_{C_n}\prod_{i=1}^n f_P(x_i)d\pi(P)}{\int\prod_{i=1}^n f_P(x_i)d\pi(P)}\\ &= \frac{p_n(x^n)}{m_n(x^n)}\int_{C_n}\prod_{i=1}^n \frac{f_P(x_i)}{f_0(x_i)}d\pi(P)\end{aligned}$$

- Lemma 4 will take care of the ratio multiplying the integral, so we can focus on the latter, for now.

Proof III

- Assumption 2, again, allows us to define $r \equiv r(n, \delta) = exp(\mathcal{K}(\mathcal{F}_n, \delta))$. Let $\{f_1^U, \ldots, f_r^U\}$. For $j = 1, \ldots, r$ define:
$$\tilde{E}_j = \{P \in \mathcal{F}_n \ : \ f_P \leq f_j^U \ a.e. \ [\lambda]\}$$

Proof III

- Assumption 2, again, allows us to define $r \equiv r(n, \delta) = exp(\mathcal{K}(\mathcal{F}_n, \delta))$. Let $\{f_1^U, \ldots, f_r^U\}$. For $j = 1, \ldots, r$ define:

$$\tilde{E}_j = \{P \in \mathcal{F}_n \ : \ f_P \leq f_j^U \ a.e. \ [\lambda]\}$$

- Let $E_1 = \tilde{E}_1$ and for $j > 1$ let $E_j = \{P \in \tilde{E}_j \ : \ P \notin \tilde{E}_s \ \forall \ s < j\}$, so that the sets $\{E_1, \ldots, E_r\}$ are disjoint and cover C_n. Hence, we have constructed a partition of C_n, Moreover, if $P \in E_j$, then $f_p \leq f_j^U$ a.e. $[\lambda]$. Using these two facts, we can write:

Proof III

- Assumption 2, again, allows us to define $r \equiv r(n,\delta) = exp(\mathcal{K}(\mathcal{F}_n,\delta))$. Let $\{f_1^U,\dots,f_r^U\}$. For $j = 1,\dots,r$ define:
$$\tilde{E}_j = \{P \in \mathcal{F}_n \ : \ f_P \le f_j^U \ a.e. \ [\lambda]\}$$
- Let $E_1 = \tilde{E}_1$ and for $j > 1$ let $E_j = \{P \in \tilde{E}_j \ : \ P \notin \tilde{E}_s \ \forall \ s < j\}$, so that the sets $\{E_1,\dots,E_r\}$ are disjoint and cover C_n. Hence, we have constructed a partition of C_n, Moreover, if $P \in E_j$, then $f_p \le f_j^U$ a.e. $[\lambda]$. Using these two facts, we can write:
$$\int_{C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) =$$

Proof III

- Assumption 2, again, allows us to define $r \equiv r(n,\delta) = exp(\mathcal{K}(\mathcal{F}_n,\delta))$. Let $\{f_1^U,\dots,f_r^U\}$. For $j = 1,\dots,r$ define:
$$\tilde{E}_j = \{P \in \mathcal{F}_n \ : \ f_P \leq f_j^U \ a.e. \ [\lambda]\}$$
- Let $E_1 = \tilde{E}_1$ and for $j > 1$ let $E_j = \{P \in \tilde{E}_j \ : \ P \notin \tilde{E}_s \ \forall \ s < j\}$, so that the sets $\{E_1,\dots,E_r\}$ are disjoint and cover C_n. Hence, we have constructed a partition of C_n, Moreover, if $P \in E_j$, then $f_p \leq f_j^U$ a.e. $[\lambda]$. Using these two facts, we can write:
$$\int_{C_n} \prod_{i=1}^n \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) = \sum_{j=1}^r \int_{E_j \cap C_n} \prod_{i=1}^n \frac{f_P(x_i)}{f_0(x_i)} d\pi(P)$$

Proof III

- Assumption 2, again, allows us to define $r \equiv r(n, \delta) = exp(\mathcal{K}(\mathcal{F}_n, \delta))$. Let $\{f_1^U, \ldots, f_r^U\}$. For $j = 1, \ldots, r$ define:
$$\tilde{E}_j = \{P \in \mathcal{F}_n \ : \ f_P \leq f_j^U \ a.e. \ [\lambda]\}$$
- Let $E_1 = \tilde{E}_1$ and for $j > 1$ let $E_j = \{P \in \tilde{E}_j \ : \ P \notin \tilde{E}_s \ \forall \ s < j\}$, so that the sets $\{E_1, \ldots, E_r\}$ are disjoint and cover C_n. Hence, we have constructed a partition of C_n, Moreover, if $P \in E_j$, then $f_p \leq f_j^U$ a.e. $[\lambda]$. Using these two facts, we can write:

$$\begin{aligned}\int_{C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) &= \sum_{j=1}^{r} \int_{E_j \cap C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \\ &\leq \sum_{j=1}^{r} \int_{E_j \cap C_n} \prod_{i=1}^{n} \frac{f_j^U(x_i)}{f_0(x_i)} d\pi(P)\end{aligned}$$

Proof III

- Assumption 2, again, allows us to define $r \equiv r(n, \delta) = exp(\mathcal{K}(\mathcal{F}_n, \delta))$. Let $\{f_1^U, \ldots, f_r^U\}$. For $j = 1, \ldots, r$ define:
$$\tilde{E}_j = \{P \in \mathcal{F}_n \ : \ f_P \leq f_j^U \ a.e. \ [\lambda]\}$$
- Let $E_1 = \tilde{E}_1$ and for $j > 1$ let $E_j = \{P \in \tilde{E}_j \ : \ P \notin \tilde{E}_s \ \forall \ s < j\}$, so that the sets $\{E_1, \ldots, E_r\}$ are disjoint and cover C_n. Hence, we have constructed a partition of C_n, Moreover, if $P \in E_j$, then $f_p \leq f_j^U$ a.e. $[\lambda]$. Using these two facts, we can write:

$$\begin{aligned}
\int_{C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) &= \sum_{j=1}^{r} \int_{E_j \cap C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \\
&\leq \sum_{j=1}^{r} \int_{E_j \cap C_n} \prod_{i=1}^{n} \frac{f_j^U(x_i)}{f_0(x_i)} d\pi(P) \\
&\leq \sum_{j=1}^{r} \prod_{i=1}^{n} \frac{f_j^U(x_i)}{f_0(x_i)} \pi(E_j \cap C_n) \ a.e. \ [\lambda^n]
\end{aligned}$$

Proof IV

- Since $\delta < \frac{\epsilon^2}{4}$, by Assumption 2, $\exists\ \beta$ and c such that:

$$0 < \beta < (\epsilon - \sqrt{\delta})^2 - \delta - 2c$$

Proof IV

- Since $\delta < \frac{\epsilon^2}{4}$, by Assumption 2, $\exists\ \beta$ and c such that:

$$0 < \beta < (\epsilon - \sqrt{\delta})^2 - \delta - 2c$$

- Define:

$$F_{n,j} = \left\{ x^n : \prod_{i=1}^{n} \frac{f_j^U(x_i)}{f_0(x_i)} \geq exp\left(- \frac{n\beta}{2} \right) \right\}$$

Proof IV

- Since $\delta < \frac{\epsilon^2}{4}$, by Assumption 2, $\exists\ \beta$ and c such that:

$$0 < \beta < (\epsilon - \sqrt{\delta})^2 - \delta - 2c$$

- Define:

$$F_{n,j} = \left\{ x^n : \prod_{i=1}^{n} \frac{f_j^U(x_i)}{f_0(x_i)} \geq exp\left(-\frac{n\beta}{2} \right) \right\}$$

- By Lemma 7, if $P \in E_j$, then $d(f_P, f_j^U) < \sqrt{\delta}$. On the other hand, by definition, if $P \in C_n$, then $d(f_0, f_P) > \epsilon$.

Proof IV

- Since $\delta < \frac{\epsilon^2}{4}$, by Assumption 2, $\exists\ \beta$ and c such that:

$$0 < \beta < (\epsilon - \sqrt{\delta})^2 - \delta - 2c$$

- Define:

$$F_{n,j} = \left\{ x^n : \prod_{i=1}^{n} \frac{f_j^U(x_i)}{f_0(x_i)} \geq exp\left(-\frac{n\beta}{2} \right) \right\}$$

- By Lemma 7, if $P \in E_j$, then $d(f_P, f_j^U) < \sqrt{\delta}$. On the other hand, by definition, if $P \in C_n$, then $d(f_0, f_P) > \epsilon$.

- Suppose $\exists\ P \in C_n \cap E_j$. By the triangle inequality, $d(f_0, f_j^U) \geq d(f_0, f_P) - d(f_j^U, f_P) \geq \epsilon - \sqrt{\delta}$.

Proof IV

- Since $\delta < \frac{\epsilon^2}{4}$, by Assumption 2, $\exists\ \beta$ and c such that:

$$0 < \beta < (\epsilon - \sqrt{\delta})^2 - \delta - 2c$$

- Define:

$$F_{n,j} = \left\{ x^n : \prod_{i=1}^{n} \frac{f_j^U(x_i)}{f_0(x_i)} \geq exp\left(- \frac{n\beta}{2} \right) \right\}$$

- By Lemma 7, if $P \in E_j$, then $d(f_P, f_j^U) < \sqrt{\delta}$. On the other hand, by definition, if $P \in C_n$, then $d(f_0, f_P) > \epsilon$.
- Suppose $\exists\ P \in C_n \cap E_j$. By the triangle inequality, $d(f_0, f_j^U) \geq d(f_0, f_P) - d(f_j^U, f_P) \geq \epsilon - \sqrt{\delta}$.
- By contraposition, then , we can conclude that, for all j s.t. $d(f_0, f_j^U) < \epsilon - \sqrt{\delta}$, then $E_j \cap C_n = \varnothing$

Proof V

- $\forall$ n, j such that $d(f_0, f_j^U) \geq \epsilon - \sqrt{\delta}$, we can apply Lemma 6 with $g = f_j^U$, $\gamma = d(f_0, f_j^U)^2 \geq [\epsilon - \sqrt{\delta}]^2$, and β as defined above.

Proof V

- $\forall\ n,\ j$ such that $d(f_0, f_j^U) \geq \epsilon - \sqrt{\delta}$, we can apply Lemma 6 with $g = f_j^U$, $\gamma = d(f_0, f_j^U)^2 \geq [\epsilon - \sqrt{\delta}]^2$, and β as defined above.
- Thus, $P_0^\infty(F_{n,j}) \leq exp(-nv)$, with $v > c$. Combined with the previous inequality, Assumption 2 part ii) and the fact that $\sum_j \pi(E_j \cap \mathcal{C}_n) \leq 1$, this implies that, for all but finitely many n:

Proof V

- $\forall\ n,\ j$ such that $d(f_0, f_j^U) \geq \epsilon - \sqrt{\delta}$, we can apply Lemma 6 with $g = f_j^U$, $\gamma = d(f_0, f_j^U)^2 \geq [\epsilon - \sqrt{\delta}]^2$, and β as defined above.
- Thus, $P_0^\infty(F_{n,j}) \leq exp(-nv)$, with $v > c$. Combined with the previous inequality, Assumption 2 part ii) and the fact that $\sum_j \pi(E_j \cap \mathcal{C}_n) \leq 1$, this implies that, for all but finitely many n:

$$P_0^n\left(x^n : \int_{C_n} \prod_{i=1}^n \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \geq exp\left\{-\frac{n\beta}{2}\right\}\right)$$

Proof V

- $\forall\ n,\ j$ such that $d(f_0, f_j^U) \geq \epsilon - \sqrt{\delta}$, we can apply Lemma 6 with $g = f_j^U$, $\gamma = d(f_0, f_j^U)^2 \geq [\epsilon - \sqrt{\delta}]^2$, and β as defined above.
- Thus, $P_0^\infty(F_{n,j}) \leq exp(-nv)$, with $v > c$. Combined with the previous inequality, Assumption 2 part ii) and the fact that $\sum_j \pi(E_j \cap \mathcal{C}_n) \leq 1$, this implies that, for all but finitely many n:

$$P_0^n\left(x^n : \int_{C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \geq exp\left\{ - \frac{n\beta}{2} \right\}\right)$$

$$\leq P_0^n\left(x^n : \sum_{j=1}^{r} \left[\prod_{i=1}^{n} \frac{f_j^U(x_i)}{f_0(x_i)} \right] \pi(E_j \cap C_n) \geq exp\left\{ - \frac{n\beta}{2} \right\}\right)$$

Proof V

- $\forall\ n,\ j$ such that $d(f_0, f_j^U) \geq \epsilon - \sqrt{\delta}$, we can apply Lemma 6 with $g = f_j^U$, $\gamma = d(f_0, f_j^U)^2 \geq [\epsilon - \sqrt{\delta}]^2$, and β as defined above.
- Thus, $P_0^\infty(F_{n,j}) \leq exp(-nv)$, with $v > c$. Combined with the previous inequality, Assumption 2 part ii) and the fact that $\sum_j \pi(E_j \cap \mathcal{C}_n) \leq 1$, this implies that, for all but finitely many n:

$$P_0^n\left(x^n : \int_{C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \geq exp\left\{-\frac{n\beta}{2}\right\}\right)$$

$$\leq P_0^n\left(x^n : \sum_{j=1}^{r} \left[\prod_{i=1}^{n} \frac{f_j^U(x_i)}{f_0(x_i)}\right] \pi(E_j \cap C_n) \geq exp\left\{-\frac{n\beta}{2}\right\}\right)$$

$$\leq \sum_{j=1}^{r} P_0^n(F_{n,j})$$

Proof V

- $\forall\ n,\ j$ such that $d(f_0, f_j^U) \geq \epsilon - \sqrt{\delta}$, we can apply Lemma 6 with $g = f_j^U$, $\gamma = d(f_0, f_j^U)^2 \geq [\epsilon - \sqrt{\delta}]^2$, and β as defined above.
- Thus, $P_0^\infty(F_{n,j}) \leq exp(-nv)$, with $v > c$. Combined with the previous inequality, Assumption 2 part ii) and the fact that $\sum_j \pi(E_j \cap \mathcal{C}_n) \leq 1$, this implies that, for all but finitely many n:

$$P_0^n\left(x^n : \int_{C_n} \prod_{i=1}^n \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \geq exp\left\{-\frac{n\beta}{2}\right\}\right)$$

$$\leq P_0^n\left(x^n : \sum_{j=1}^r \left[\prod_{i=1}^n \frac{f_j^U(x_i)}{f_0(x_i)}\right] \pi(E_j \cap C_n) \geq exp\left\{-\frac{n\beta}{2}\right\}\right)$$

$$\leq \sum_{j=1}^r P_0^n(F_{n,j}) \leq r \cdot exp(-nv)$$

Proof V

- $\forall\ n,\ j$ such that $d(f_0, f_j^U) \geq \epsilon - \sqrt{\delta}$, we can apply Lemma 6 with $g = f_j^U$, $\gamma = d(f_0, f_j^U)^2 \geq [\epsilon - \sqrt{\delta}]^2$, and β as defined above.
- Thus, $P_0^\infty(F_{n,j}) \leq exp(-nv)$, with $v > c$. Combined with the previous inequality, Assumption 2 part ii) and the fact that $\sum_j \pi(E_j \cap \mathcal{C}_n) \leq 1$, this implies that, for all but finitely many n:

$$P_0^n\left(x^n : \int_{C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \geq exp\left\{-\frac{n\beta}{2}\right\}\right)$$

$$\leq P_0^n\left(x^n : \sum_{j=1}^{r} \left[\prod_{i=1}^{n} \frac{f_j^U(x_i)}{f_0(x_i)}\right] \pi(E_j \cap C_n) \geq exp\left\{-\frac{n\beta}{2}\right\}\right)$$

$$\leq \sum_{j=1}^{r} P_0^n(F_{n,j}) \leq r \cdot exp(-nv) = exp(\mathcal{K}(\mathcal{F}_n, \delta) - nv)$$

Proof V

- $\forall\ n,\ j$ such that $d(f_0, f_j^U) \geq \epsilon - \sqrt{\delta}$, we can apply Lemma 6 with $g = f_j^U$, $\gamma = d(f_0, f_j^U)^2 \geq [\epsilon - \sqrt{\delta}]^2$, and β as defined above.
- Thus, $P_0^\infty(F_{n,j}) \leq exp(-nv)$, with $v > c$. Combined with the previous inequality, Assumption 2 part ii) and the fact that $\sum_j \pi(E_j \cap \mathcal{C}_n) \leq 1$, this implies that, for all but finitely many n:

$$P_0^n\left(x^n : \int_{C_n} \prod_{i=1}^n \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \geq exp\left\{-\frac{n\beta}{2}\right\}\right)$$

$$\leq P_0^n\left(x^n : \sum_{j=1}^r \left[\prod_{i=1}^n \frac{f_j^U(x_i)}{f_0(x_i)}\right]\pi(E_j \cap C_n) \geq exp\left\{-\frac{n\beta}{2}\right\}\right)$$

$$\leq \sum_{j=1}^r P_0^n(F_{n,j}) \leq r \cdot exp(-nv) = exp(\mathcal{K}(\mathcal{F}_n, \delta) - nv) \leq exp(-n[v - c])$$

Proof VI

- First Borel-Cantelli lemma, then, implies:

$$P_0^\infty\left(x^n : \int_{C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \geq exp\left\{-\frac{n\beta}{2}\right\}, i.o.\right) = 0$$

Proof VI

- First Borel-Cantelli lemma, then, implies:

$$P_0^\infty\left(x^n : \int_{C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \geq exp\left\{ -\frac{n\beta}{2} \right\}, i.o. \right) = 0$$

- Lemma 4 grants:

$$P_0^\infty\left(x^\infty : \frac{p_n(x^n)}{m_n(x^n)} \geq exp\left\{ \frac{n\beta}{4} \right\}, i.o. \right) = 0$$

Proof VI

- First Borel-Cantelli lemma, then, implies:

$$P_0^\infty\left(x^n : \int_{C_n} \prod_{i=1}^{n} \frac{f_P(x_i)}{f_0(x_i)} d\pi(P) \geq exp\left\{ -\frac{n\beta}{2} \right\}, i.o.\right) = 0$$

- Lemma 4 grants:

$$P_0^\infty\left(x^\infty : \frac{p_n(x^n)}{m_n(x^n)} \geq exp\left\{ \frac{n\beta}{4} \right\}, i.o.\right) = 0$$

- Combining these last two equations with the previous decomposition of the posterior, we finally get:

$$P_0^\infty\left(x^\infty : \pi(C_n|x^n) \geq exp\left\{ -\frac{n\beta}{4}, i.o.\right\}\right) = 0$$

Proof VI

- First Borel-Cantelli lemma, then, implies:

$$P_0^\infty\left(x^n : \int_{C_n}\prod_{i=1}^{n}\frac{f_P(x_i)}{f_0(x_i)}d\pi(P) \geq exp\left\{-\frac{n\beta}{2}\right\}, i.o.\right) = 0$$

- Lemma 4 grants:

$$P_0^\infty\left(x^\infty : \frac{p_n(x^n)}{m_n(x^n)} \geq exp\left\{\frac{n\beta}{4}\right\}, i.o.\right) = 0$$

- Combining these last two equations with the previous decomposition of the posterior, we finally get:

$$P_0^\infty\left(x^\infty : \pi(C_n|x^n) \geq exp\left\{-\frac{n\beta}{4}, i.o.\right\}\right) = 0$$

- Which implies that $\lim_{n\to\infty}\pi(C_n|x^n) = 0$, a.s. $[P_0^\infty]$

Appendix I: Histograms Assumption 1

Lemma 9

Let $(\mathcal{X}, \mathcal{B}, \lambda)$ be a probability space, and let $\mathcal{R}$ be a collection of measurable real-valued functions defined on $\mathcal{X}$. $\forall b > 0$, define

$$\mathcal{R}_b = \{f \in \mathcal{R} : esssup|f| \leq b\}$$
$$L_b = \{f : esssup|f| \leq b\}$$

where the essential supremum is relative to λ. Suppose $\exists\ r > 1$ such that $\mathcal{R}_{rb}$ is dense [in the sense of $L_1(\lambda)$] in L_b $\forall b$ large. Let $P_0 \ll \lambda$ be another probability on $(\mathcal{X}, \mathcal{B})$ such that $\mathcal{I}(P_0, \lambda) < \infty$. Then $\forall \epsilon > 0$, $\exists$ a bounded function g $\in \mathcal{R}$ such that $\mathcal{I}(P_0, P_g) < \epsilon$, where P_g is the distribution with density

$$p_g(x) = \frac{exp(g(x))}{\int exp(g(y))d\lambda(y)}$$

Histograms Assumption 1 II

- Let $\mathcal{R}$ be the set of all step functions that are constant on all of the intervals in at least one of the $\mathcal{T}_n$ partitions, and let λ be Lebesgue measure.
- Step functions are dense in the collection of bounded measurable functions and $\mathcal{R}$ is dense in the collection of step functions.
- Thus, $\forall \epsilon$, $\exists$ n and $P_\epsilon \in \mathcal{U}_n$ such that $\mathcal{I}(P_0, P_\epsilon) < \frac{\epsilon}{2}$
- Since the Dirichlet distribution over $\mathcal{U}_n$ assigns positive probability to every open neighborhood of P_ϵ and $\mathcal{I}(P_0, P)$ is continuous as a function of P for distributions with densities in $\mathcal{R}$, then Assumption 1 holds.

Back

Appendix II: Preliminary Lemmas

Preliminary Lemmas I

Lemmas 1 and 2

Under Assumption 1,

$$P_0^\infty(x^\infty : \exists\ n \textit{ such that } m_n(x^n) \in \{0, \infty\}) = 0$$
$$P_0^\infty(x^\infty : \exists\ n \textit{ such that } p_n(x^n) \in \{0, \infty\}) = 0$$

where m_n is the predictive density of X^n and p_n the density of the n-fold product measure of P_0.

Lemma 3

$\exists$ a set $B \subseteq \mathcal{X}^\infty$ such that $P_0^\infty(B) = 1$ and such that $\forall x^\infty \in B$, there is a set $G_{x^\infty} \in \mathcal{C}$ such that $\pi(G_{x^\infty}) = 1$ and $\forall P \in G_{x^\infty}$, $\lim_{n\to\infty} D_n(x^n, P) = \mathcal{I}(P_0, P)$.

Preliminary Lemmas II

Lemma 4

Under Assumption 1, $\forall \epsilon > 0$,

$$P_0^\infty\left(x^\infty : \frac{m_n(x^n)}{p_n(x^n)} \leq exp(n\epsilon),\ i.o.\right) = 0$$

where, again, m_n is the predictive density of X^n and p_n the density of the n-fold product measure of P_0.

Lemma 5

Suppose that Assumption 1 holds. Let $c_1, c_2 > 0$. Suppose that $(B_n)_{n=1}^\infty$ is a sequence of subset of $\mathcal{P}$ such that $\pi(B_n) < c_1 exp(-c_2 n)$ for all but finitely many n. Then $\lim_{n\to\infty} \pi(B_n|x^n) = 0$ a.s. $[P_0^\infty]$

Preliminary Lemmas III

Lemma 6

Let g be a non-negative, integrable function, $\beta > 0$, $d(f_0, g) = \sqrt{\gamma}$, $\int g(x) d\lambda(x) \leq 1 + \delta$ and $\delta \leq \gamma$. Then

$$P_0^n\left(x^n : \prod_{i=1}^{n} \frac{g(x_i)}{f_0(x_i)} \geq exp[-n\beta]\right) \leq exp\left(-n\frac{\gamma - \beta - \delta}{2}\right)$$

Lemma 7

Let $P \in \mathcal{P}$ and $g \in \mathcal{G}$ be such that $f_p \leq g$ a.e. $[\lambda]$ and $\int g(x) d\lambda(x) \leq 1 + \delta$. Then $d(f_p, g) \leq \sqrt{\delta}$.