On Discrete Time Ergodic Filters with Wrong Initial Data

Alexander Yu. Veretennikov\textsuperscript{1} Marina Kleptsyna\textsuperscript{2}

\textsuperscript{1}University of Leeds, UK
\textsuperscript{2}University of Le Mans, France

March, 9th 2007 / Marseille
Outline

1 Introduction
   - Problem statement
   - Historical survey
   - Reformulation of the problem, the Bayes approach

2 Assumptions and results
   - Stability with absolutely continuous initial data
   - Stability without initial data absolute continuity

3 Auxiliaries
   - Ergodic processes in $R^d$
   - Birkhoff metric

4 Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
   - Theorem 2, idea of the proof
Statement of the problem without formulas

The model

Nonobservable ergodic Markov chain \((X_n)\) with
- values in \(\mathbb{R}^d\);
- observations \((Y_n)\) from \(\mathbb{R}^\ell\);
- initial distribution \(\mu_0\) (of \(X_0\)) known with some error.

The question

Whether or not this error is forgotten by the optimal filtering algorithm in the long run.

The question for discussion

What does it mean "the optimal filtering algorithm"?
### Statement of the problem without formulas

**The model**

Nonobservable ergodic Markov chain \((X_n)\) with
- values in \(\mathbb{R}^d\);
- observations \((Y_n)\) from \(\mathbb{R}^\ell\);
- initial distribution \(\mu_0\) (of \(X_0\)) known with some error.

**The question**

Whether or not this error is forgotten by the optimal filtering algorithm in the long run.

**The question for discussion**

What does it mean "the optimal filtering algorithm"?
Statement of the problem without formulas

The model

Nonobservable ergodic Markov chain \((X_n)\) with
- values in \(\mathbb{R}^d\);
- observations \((Y_n)\) from \(\mathbb{R}^\ell\);
- initial distribution \(\mu_0\) (of \(X_0\)) known with some error.

The question

Whether or not this error is forgotten by the optimal filtering algorithm in the long run.

The question for discussion

What does it mean "the optimal filtering algorithm"?
The observation model
the precise definition

Markov chain:

\[ X_{n+1} = X_n + b(X_n) + \xi_{n+1}, \quad (n \geq 0), \]

observation:

\[ Y_n = h(X_n) + V_n \quad (n \geq 1), \]

where

\[ (\xi_n, V_n) \in \mathbb{R}^{d+\ell} \text{ – IID centered sequence;} \]
\[ b : \mathbb{R}^d \to \mathbb{R}^d; \]
\[ h : \mathbb{R}^d \to \mathbb{R}^{\ell}; \]
**The observation model**

the precise definition

- Markov chain:

  \[ X_{n+1} = X_n + b(X_n) + \xi_{n+1}, \quad (n \geq 0), \]

- observation:

  \[ Y_n = h(X_n) + V_n, \quad (n \geq 1), \]

- where

  - \((\xi_n, V_n) \in \mathbb{R}^{d+\ell}\) – IID centered sequence;
  - \(b : \mathbb{R}^d \rightarrow \mathbb{R}^d;\)
  - \(h : \mathbb{R}^d \rightarrow \mathbb{R}^\ell;\)
Stating the main question

- The **true** conditional probability:

  \[ P_{n}^{{\mu_0}},Y(\cdot) = P_{\mu_0}(X_n \in \cdot \mid \mathcal{F}_n^Y), \]

  - with \( \mathcal{F}_n^Y = \sigma(Y_k : 1 \leq k \leq n) \),
  - with the initial measure \( \mu_0 \).

- The **strange** conditional probability:

  \[ P_{n}^{{\nu_0}},Y(\cdot) = P_{n}^{{\mu_0},Y}(\cdot) \mid \mu_0 = \nu_0. \]

  - with \( \mu_0 \) replaced by \( \nu_0 \).

The main question:

True or false:

\[ \lim_{n \to \infty} E_{\mu_0} \| P_{n}^{{\mu_0}},Y(\cdot) - P_{n}^{{\nu_0}},Y(\cdot) \|_{TV} = 0? \]
Stating the main question

- **The true conditional probability:**
  \[ P_{n}^{\mu_{0}, Y}(\cdot) = P_{\mu_{0}}(X_{n} \in \cdot \mid F_{n}^{Y}), \]
  - with \( F_{n}^{Y} = \sigma(Y_{k} : 1 \leq k \leq n) \),
  - with the initial measure \( \mu_{0} \).

- **The strange conditional probability:**
  \[ P_{n}^{\nu_{0}, Y}(\cdot) = P_{n}^{\mu_{0}, Y}(\cdot) \mid \mu_{0} = \nu_{0}, \]
  - with \( \mu_{0} \) replaced by \( \nu_{0} \).

**The main question:**

True or false:

\[ \lim_{n \to \infty} E_{\mu_{0}} \| P_{n}^{\mu_{0}, Y}(\cdot) - P_{n}^{\nu_{0}, Y}(\cdot) \|_{TV} = 0? \]
Stating the main question

- The true conditional probability:
  \[ P_{n, Y}^{\mu_0} (\cdot) = P_{\mu_0} (X_n \in \cdot \mid F_n^Y) , \]
  
  - with \( F_n^Y = \sigma (Y_k : 1 \leq k \leq n) \),
  - with the initial measure \( \mu_0 \).

- The strange conditional probability:
  \[ P_{n, Y}^{\nu_0} (\cdot) = P_{n, Y}^{\mu_0} (\cdot) \mid \mu_0 = \nu_0 . \]
  
  - with \( \mu_0 \) replaced by \( \nu_0 \).

The main question:

True or false:

\[ \lim_{n \to \infty} E_{\mu_0} \| P_{n, Y}^{\mu_0} (\cdot) - P_{n, Y}^{\nu_0} (\cdot) \|_{TV} = 0 ? \]
Stability of filters

True or false:

$$\lim_{n \to \infty} E_{\mu_0} \left( \pi_{n, Y}^{\mu_0} (f) - \pi_{n, Y}^{\nu_0} (f) \right)^2 = 0? \quad \forall f \in C_b$$

where

- the true conditional expectation:
  $$\pi_{n, Y}^{\mu_0} (f) = E_{\mu_0} (f(X_n) | F_{n, Y})$$

- the strange conditional expectation:
  $$\pi_{n, Y}^{\nu_0} (f) = E_{\mu_0} (f(X_n) | F_{n, Y}) \mid \mu_0 = \nu_0.$$
Stability of filters

True or false:

\[
\lim_{{n \to \infty}} E_{\mu_0}(\pi_{n, Y}^{\mu_0}(f) - \pi_{n, Y}^{\nu_0}(f))^2 = 0? \quad \forall f \in C_b
\]

where

- the true conditional expectation:
  \[
  \pi_{n, Y}^{\mu_0}(f) = E_{\mu_0}(f(X_n) \mid \mathcal{F}_n^Y)
  \]

- the strange conditional expectation:
  \[
  \pi_{n, Y}^{\nu_0}(f) = E_{\nu_0}(f(X_n) \mid \mathcal{F}_n^Y) \mid \mu_0 = \nu_0.
  \]
Outline

1. Introduction
   - Problem statement
   - Historical survey

2. Assumptions and results
   - Stability with absolutely continuous initial data
   - Stability without initial data absolute continuity

3. Auxiliaries
   - Ergodic processes in $\mathbb{R}^d$
   - Birkhoff metric

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
   - Theorem 2, idea of the proof
The first time

D.Blackwell, 1957

The model

Nonobservable stationary ergodic finite state Markov chain \((X_n)\)

- observations \(Y_n = \Phi(X_n)\)
  - \(\Phi\) is not one-to-one.

The question

Is the stationary measure of the conditional distribution unique?

The question for discussion

What is the connection with the subject of the talk?
The first time

D. Blackwell, 1957

The model

Nonobservable **stationary** ergodic **finite** state Markov chain \((X_n)\)

- observations \(Y_n = \Phi(X_n)\)
- \(\Phi\) is not one-to-one.

The question

Is the stationary measure of the conditional distribution unique?

The question for discussion

What is the connection with the subject of the talk?
The first time

D. Blackwell, 1957

The model

Nonobservable stationary ergodic finite state Markov chain \((X_n)\)

- observations \(Y_n = \Phi(X_n)\)
- \(\Phi\) is not one-to-one.

The question

Is the stationary measure of the conditional distribution unique?

The question for discussion

What is the connection with the subject of the talk?
D. Blackwell, 1957

The model

Nonobservable **stationary** ergodic **finite** state Markov chain \((X_n)\)

- observations \(Y_n = \Phi(X_n)\)
- \(\Phi\) is not one-to-one.

The question

Is the stationary measure of the conditional distribution unique?

The question for discussion

What is the connection with the subject of the talk?
### Stability and uniqueness
Don’t trouble trouble until trouble troubles you

<table>
<thead>
<tr>
<th>Two measure-valued processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Our distribution (finite observation) – Markov process $P_n^i = P^{\mu}(x_n = i \mid y_n \ldots y_0)$ with $\mu = \mu^{st}$ or not.</td>
</tr>
<tr>
<td>- Blackwell’s distribution (infinite observation) – stationary Markov process $\bar{P}<em>n^i = P^{\mu}(x_n = i \mid y_n \ldots y_0, y</em>{-1} \ldots)$</td>
</tr>
</tbody>
</table>

#### Their common properties

- They have the same generator.
- Let $Q$ be the distribution of $\bar{P}_0^i$. It is a common stationary distribution of $\bar{P}$ and $P$. $Q$ – a stationary measure on the space of measures for the process of conditional measures.
- $Q$ can be continuous, atomic etc.
### Stability and uniqueness

Don’t trouble trouble until trouble troubles you

#### Two measure-valued processes

- Our distribution (finite observation) – Markov process
  
  \[ P_n^i = P^\mu(x_n = i \mid y_n \ldots y_0) \]  
  with \( \mu = \mu^{st} \) or not.

- Blackwell’s distribution (infinite observation) – stationary Markov process
  
  \[ \bar{P}_n^i = P^\mu(x_n = i \mid y_n \ldots y_0, y_{-1} \ldots) \]

### Their common properties

- They have the same generator.

- Let \( Q \) be the distribution of \( \bar{P}_0^i \). It is a common stationary distribution of \( \bar{P} \) and \( P \). \( Q \) – a stationary measure on the space of measures for the process of conditional measures.

- \( Q \) can be continuous, atomic etc.
Stability and uniqueness
Don’t trouble trouble until trouble troubles you

Two measure-valued processes

- Our distribution (finite observation) – Markov process
  \( P_n^i = P^\mu(x_n = i \mid y_n \ldots y_0) \) with \( \mu = \mu^{st} \) or not.
- Blackwell’s distribution (infinite observation) – stationary Markov process
  \( \bar{P}_n^i = P^\mu(x_n = i \mid y_n \ldots y_0, y_{-1} \ldots) \).

Their common properties

- They have the same generator.
  - Let \( Q \) be the distribution of \( \bar{P}_0^i \). It is a common stationary distribution of \( \bar{P} \) and \( P \). \( Q \) – a stationary measure on the space of measures for the process of conditional measures.
  - \( Q \) can be continuous, atomic etc.
Stability and uniqueness
Don’t trouble trouble until trouble troubles you

Two measure-valued processes

- Our distribution (finite observation) – Markov process
  \[ P_n^i = P^\mu(x_n = i \mid y_n \ldots y_0) \] with \( \mu = \mu^{st} \) or not.

- Blackwell’s distribution (infinite observation) – stationary Markov process
  \[ \bar{P}_n^i = P^\mu(x_n = i \mid y_n \ldots y_0, y_{-1} \ldots) \]

Their common properties

- They have the same generator.

- Let \( Q \) be the distribution of \( \bar{P}_0^i \). It is a common stationary distribution of \( \bar{P} \) and \( P \). \( Q \) – a stationary measure on the space of measures for the process of conditional measures.

- \( Q \) can be continuous, atomic etc.
Stability and uniqueness
Don’t trouble trouble until trouble troubles you

Two measure-valued processes

- Our distribution (finite observation) – Markov process
  \[ P_n^i = P^\mu(x_n = i \mid y_n \ldots y_0) \] with \( \mu = \mu^{st} \) or not.

- Blackwell’s distribution (infinite observation) – stationary Markov process
  \[ \bar{P}_n^i = P^\mu(x_n = i \mid y_n \ldots y_0, y_{-1} \ldots) \].

Their common properties

- They have the same generator.

- Let \( Q \) be the distribution of \( \bar{P}_0^i \). It is a common stationary distribution of \( \bar{P} \) and \( P \). \( Q \) – a stationary measure on the space of measures for the process of conditional measures.

- \( Q \) can be continuous, atomic etc.
## Two measure-valued processes

- Our distribution (finite observation) – Markov process
  \[ P_n^i = P^{\mu}(x_n = i \mid y_n \ldots y_0) \] with \( \mu = \mu^{st} \) or not.

- Blackwell’s distribution (infinite observation) – stationary Markov process
  \[ \bar{P}_n^i = P^{\mu}(x_n = i \mid y_n \ldots y_0, y_{-1} \ldots) \].

## Their common properties

- They have the same generator.

- Let \( Q \) be the **distribution** of \( \bar{P}_0^i \). It is a common stationary distribution of \( \bar{P} \) and \( P \). \( Q \) – a stationary measure on the space of measures for the process of conditional measures.

- \( Q \) can be continuous, atomic etc.
Stability and uniqueness 2

Two related questions

- Blackwell: Is a stationary measure unique (only Q)?
- We: Is the filter stable?

Fact

**Stability** of filter $\Rightarrow$ **uniqueness** of stationary measure. (A. Budhiraja, H.J.Kushner).
Two related questions

- Blackwell: Is a stationary measure unique (only $Q$)?
- We: Is the filter stable?

Fact

**Stability of filter $\Rightarrow$ uniqueness of stationary measure.** (A. Budhiraja, H.J.Kushner).
The first time: **often** the answer is "yes"

- **1971, 1991, H. Kunita: "yes" in diffusion model.**
  - **Model:**
    Signal $X_t$ — ergodic Markov process valued in a locally compact space.
  - **Observations:**
    \[ dY_t = h(X_t)dt + dW_t \]
  - **Claim:**
    \[ \lim_{t \to \infty} E_\mu_0 (f(X_t) - \pi_t^{\mu_0, Y}(f))^2 \]
    does non depend on $\mu_0$,
    the invariant measure of the filtering process is unique.

- **L. Stettner, 1989, 1991:** generalization; discrete time included.
The first time: **often** the answer is "yes"

  - **Model:**
    Signal $X_t$ — ergodic Markov process valued in a locally compact space.
  - **Observations:**
    \[ dY_t = h(X_t)dt + dW_t \]
  - **Claim:**
    \[ \lim_{t \to \infty} E_{\mu_0}(f(X_t) - \pi_t^{\mu_0, Y}(f))^2 \]
    does non depend on $\mu_0$,
    the invariant measure of the filtering process is unique.

The first time: often the answer is "yes"

  - **Model:**
    Signal $X_t$ — ergodic Markov process valued in a locally compact space.
  - **Observations:**
    
    $$dY_t = h(X_t)dt + dW_t$$

- **Claim:**
  
  $$\lim_{t \to \infty} E_{\mu_0}(f(X_t) - \pi_t^{\mu_0, Y}(f))^2$$

  does non depend on $\mu_0$, the invariant measure of the filtering process is unique.

- **L.Stettner, 1989, 1991:** generalization; discrete time included.
The first time: **often** the answer is "yes"

  - **Model:**
    Signal $X_t$ — ergodic Markov process valued in a locally compact space.
  - **Observations:**
    \[
    dY_t = h(X_t)dt + dW_t
    \]
  - **Claim:**
    \[
    \lim_{t \to \infty} E_{\mu_0}(f(X_t) - \pi_{t,\mu_0}^Y(f))^2
    \]
    does non depend on $\mu_0$,
    the invariant measure of the filtering process is unique.

The first time: often the answer is "yes"

  - Model:
    Signal $X_t$ — ergodic Markov process valued in a locally compact space.
  - Observations:
    \[ dY_t = h(X_t)dt + dW_t \]
  - Claim:
    \[ \lim_{t \to \infty} E_{\mu_0} (f(X_t) - \pi_{\mu_0}^t, Y(f))^2 \]
    does non depend on $\mu_0$,
    the invariant measure of the filtering process is unique.

The first time, sometimes the answer is "no"

1974, Kaijser: a counter-example

- $X_n$ - an ergodic Markov chain with $\mathbb{S} = \{1, 2, 3, 4\}$
- transition matrix
  \[
  \Lambda = \frac{1}{2} \begin{pmatrix}
  1 & 1 & 0 & 0 \\
  0 & 1 & 1 & 0 \\
  0 & 0 & 1 & 1 \\
  1 & 0 & 0 & 1
  \end{pmatrix}
  \]
- observation (noiseless): $Y_n = 1_{X_n=1} + 1_{X_n=3}$

Result: there is no uniqueness, no stability

$$\lim_{n \to \infty} E_{\mu_0}(\pi_{\mu_0}^{\mu_0}(x) - \pi_{\nu_0}^{\nu_0}(x))^2 \geq C(\mu_0, \nu_0) > 0.$$
The first time, sometimes the answer is "no"

1974, Kaijser: a counter-example

- $X_n$ - an **ergodic** Markov chain with $\mathcal{S} = \{1, 2, 3, 4\}$
- transition matrix

$$
\Lambda = \frac{1}{2} \begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 \\
1 & 0 & 0 & 1
\end{pmatrix}
$$

- observation (noiseless): $Y_n = 1_{X_n=1} + 1_{X_n=3}$

Result: there is no uniqueness, no stability

$$
\lim_{n \to \infty} E_{\mu_0}(\pi_n^{\mu_0}, Y(x) - \pi_n^{\nu_0}, Y(x))^2 \geq C(\mu_0, \nu_0) > 0.
$$
At the same time, independently, I

1991, Delyon & Zeitouni:
- consider finite state space ergodic signal or linear case;
- introduce the term "memory length" of filters;
- propose a programme of analysis of exponential stability of filters using Lyapounov exponents.

1996, D. Ocone, E. Pardoux:
- consider Kunita’s model.
- **Claim**: The optimal filter is stable:

\[
\lim_{n \to \infty} E_{\mu_0}(\pi_n^{\mu_0,Y}(f) - \pi_n^{\nu_0,Y}(f))^2 = 0 \quad \forall f \in C_b, \; \nu_0 \sim \mu_0.
\]

(the proof is *crucially* based on the H. Kunita result)
At the same time, independently, I

1991, Delyon & Zeitouni:
- consider finite state space ergodic signal or linear case;
- introduce the term "memory length" of filters;
- propose a programme of analysis of exponential stability of filters using Lyapunov exponents.

1996, D.Ocone, E.Pardoux:
- consider Kunita’s model.
- **Claim**: The optimal filter is stable:

\[
\lim_{n \to \infty} E_{\mu_0}(\pi_n^{\mu_0}, Y(f) - \pi_n^{\nu_0}, Y(f))^2 = 0 \quad \forall f \in C_b, \nu_0 \sim \mu_0.
\]

(the proof is **crucially** based on the H. Kunita result)
1991, Delyon & Zeitouni:
- consider finite state space ergodic signal or linear case;
- introduce the term "memory length" of filters;
- propose a programme of analysis of exponential stability of filters using Lyapounov exponents.

1996, D. Ocone, E. Pardoux:
- consider Kunita’s model.
- Claim: The optimal filter is stable:
  \[ \lim_{n \to \infty} E_{\mu_0} \left( \pi_{\mu_0}^n, Y(f) - \pi_{\nu_0}^n, Y(f) \right)^2 = 0 \quad \forall f \in C_b, \nu_0 \sim \mu_0. \]

(The proof is crucially based on the H. Kunita result)
At the same time, independently, Il

- **1997, Atar & Zeitouni**
  - consider discrete and **continuous** time, **compact** valued Markov signal;
  - propose **Birkhoff contraction** principle.

- **1998, Atar**
  - considers **continuous time, one dimensional non-compact** case, with **linear** observations and sufficiently small noise in observations.
At the same time, independently, II

- **1997, Atar & Zeitouni**
  - consider discrete and continuous time, compact valued Markov signal;
  - propose Birkhoff contraction principle.

- **1998, Atar**
  - considers continuous time, one dimensional non-compact case, with linear observations and sufficiently small noise in observations.
2004 P.Baxendale, P.Chiganskii, R.Liptser
Serious gap in Kunita’s proof.

The Kunita’s proof was based on the following:

True or false

$$\bigcap_{n \geq 1} \mathcal{F}^Y_{[0,\infty)} \cup \mathcal{F}^X_{[n,\infty)} = \mathcal{F}^Y_{[0,\infty)}$$

for an ergodic Markov process $X_t$?
2004 P.Baxendale, P.Chiganskii, R.Liptser
Serious gap in Kunita’s proof.
The Kunita’s proof was based on the following:

True or false

\[ \bigcap_{n \geq 1} \mathcal{F}^Y_{[0,\infty)} \bigvee \mathcal{F}^X_{[n,\infty)} = \mathcal{F}^Y_{[0,\infty)} \]

for an **ergodic** Markov process \( X_t \)?
Counterexample for the proof

an **ergodic** Markov process $X_t$ with
- state space $\mathcal{S} = \{1, 2, 3, 4\}$;
- transition intensity matrix:

$$\Lambda = \frac{1}{2} \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix};$$

- **noiseless** observation: $Y_n = 1_{X_n=1} + 1_{X_n=3}$.

Result: the answer is negative. Filter is unstable, the invariant measure of the filtering process is not unique.
Counterexample for the proof

an **ergodic** Markov process $X_t$ with
- state space $S = \{1, 2, 3, 4\}$;
- transition intensity matrix:

$$\Lambda = \frac{1}{2} \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix};$$

- **noiseless** observation: $Y_n = 1_{X_n=1} + 1_{X_n=3}$.

**Result:** the answer is **negative**. Filter is unstable, the invariant measure of the filtering process is not unique.


W. Stannat (2005) Continuous time case, a gradient type drift and linear observation part under additional assumptions.

Today

- W. Stannat (2005) **Continuous time case**, a gradient type drift and linear observation part under additional assumptions.
Today

- W. Stannat (2005) **Continuous time case**, a gradient type drift and linear observation part under additional assumptions.
**Outline**

1. **Introduction**
   - Problem statement
   - Historical survey
   - Reformulation of the problem, the Bayes approach

2. **Assumptions and results**
   - Stability with absolutely continuous initial data
   - Stability without initial data absolute continuity

3. **Auxiliaries**
   - Ergodic processes in \( R^d \)
   - Birkhoff metric

4. **Sketch of the proof**
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
   - Theorem 2, idea of the proof
The Bayes formula, part 1

- **Changing of measure:**
  Let $L_n(\overline{X^n}, \overline{Y^n})$ be the conditional density of $\overline{Y^n} = (Y_1, Y_2, \ldots, Y_n)$ given by $\overline{X^n} = (X_1, X_2, \ldots, X_n)$.

- Then the Bayes formula holds:

$$P(X_n \in \cdot \mid F_n^Y) = \frac{\hat{E}(1(X_n \in \cdot)L_n(\overline{X^n}, \overline{Y^n}) \mid F_n^Y)}{\hat{E}L_n(\overline{X^n}, \overline{Y^n}) \mid F_n^Y},$$

with

$$\frac{dP}{d\hat{P}} = \frac{L_n(\overline{X^n}, \overline{Y^n})}{f(Y^n)}$$
The Bayes formula, part 1

- **Changing of measure:**
  Let $L_n(X^n, Y^n)$ be the conditional density of $Y^n = (Y_1, Y_2, \ldots, Y_n)$ given by $X^n = (X_1, X_2, \ldots, X_n)$.

- Then **the Bayes formula** holds:

$$P(X_n \in \cdot \mid \mathcal{F}_n^Y) = \frac{\hat{E}(1(X_n \in \cdot) L_n(X^n, Y^n) \mid \mathcal{F}_n^Y)}{\hat{E} L_n(X^n, Y^n) \mid \mathcal{F}_n^Y},$$

with

$$\frac{dP}{d\hat{P}} = \frac{L_n(X^n, Y^n)}{f(Y^n)}$$
Remark 1. Processes \((\overline{X}^n, \overline{Y}^n)\) are independent w.r.t. \(\hat{P}\), and the law of \(\overline{X}^n\) has not been changed.

Remark 2.

\[
L_n(\overline{X}^n, \overline{Y}^n) = \prod_{i=1}^{n} \psi(x_i, Y_i)
\]

\[
\psi(x_i, y_i) = q_V(y_i - h(x_i))
\]

(where \(q_V\) denotes the density of \(V_1\))
Remark 1. Processes \((\overline{X}^n, \overline{Y}^n)\) are independent w.r.t. \(\hat{P}\), and the law of \(\overline{X}^n\) has not been changed.

Remark 2.

\[
L_n(\overline{X}^n, \overline{Y}^n) = \prod_{i=1}^{n} \psi(x_i, Y_i)
\]

\[
\psi(x_i, y_i) = q_v(y_i - h(x_i))
\]

(where \(q_v\) denotes the density of \(V_1\))
Conditional probability via a nonlinear operator

- Introduce the nonlinear operator

\[ P_{n, Y}^{\mu_0}(dx_n) =: \mu_0 \bar{S}_n^{Y, \mu_0}(dx_n) \]

- Its explicit form:

\[
\mu_0 \bar{S}_n^{Y, \mu_0}(dx_n) = d_{n, \mu_0}^{\mu_0} \int \prod_{i=1}^{n} Q(x_{i-1}, dx_i) \psi(x_i, Y_i) \mu_0(dx_0).
\]

- \( Q(x, dx') \) – the transition kernel of \( X_n \), \( q_\xi \) – the density of \( \xi_1 \).

\[
Q(x, dx') = q_\xi(x' - x - b(x)) dx'.
\]

- \( d_{n, \mu_0}^{\mu_0} \) - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).
Conditional probability via a nonlinear operator

- Introduce the nonlinear operator
  \[
  P_n^{\mu_0, Y}(dx_n) =: \mu_0 \bar{S}_n^{Y, \mu_0}(dx_n)
  \]

- Its explicit form:
  \[
  \mu_0 \bar{S}_n^{Y, \mu_0}(dx_n) = d_n^{\mu_0} \int \prod_{i=1}^n Q(x_{i-1}, dx_i) \psi(x_i, Y_i) \mu_0(dx_0).
  \]

- \(Q(x, dx')\) – the transition kernel of \(X_n\), \(q_\xi\) – the density of \(\xi_1\).
  \[
  Q(x, dx') = q_\xi(x' - x - b(x))\, dx'.
  \]

- \(d_n^{\mu_0}\) - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).
Introduction

Problem statement

Historical survey

Reformulation of the problem, the Bayes approach

Assumptions and results

Stability with absolutely continuous initial data

Stability without initial data absolute continuity

Auxiliaries

Ergodic processes in $\mathbb{R}^d$

Birkhoff metric

Sketch of the proof

Coupling and separation

The main inequality

Sketch of the proof, part 2

Conditional probability via a nonlinear operator

- Introduce the nonlinear operator

\[ P_{n}^{\mu_0, Y}(dx_n) =: \mu_0 \bar{S}_n^{Y, \mu_0}(dx_n) \]

- Its explicit form:

\[ \mu_0 \bar{S}_n^{Y, \mu_0}(dx_n) = d_n^{\mu_0} \int \prod_{i=1}^{n} Q(x_{i-1}, dx_i) \psi(x_i, Y_i) \mu_0(dx_0). \]

- $Q(x, dx')$ – the transition kernel of $X_n$, $q_\xi$ – the density of $\xi_1$.

\[ Q(x, dx') = q_\xi(x' - x - b(x)) \, dx'. \]

- $d_n^{\mu_0}$ - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).
Conditional probability via a nonlinear operator

- Introduce the nonlinear operator
  \[ P_{n0}^{\mu_0, Y}(dx_n) =: \mu_0 \bar{S}_n^{Y, \mu_0}(dx_n) \]

- Its explicit form:
  \[
  \mu_0 \bar{S}_n^{Y, \mu_0}(dx_n) = d_n^{\mu_0} \int \prod_{i=1}^{n} Q(x_{i-1}, dx_i) \psi(x_i, Y_i) \mu_0(dx_0).
  \]

- \( Q(x, dx') \) – the transition kernel of \( X_n \), \( q_\xi \) – the density of \( \xi_1 \).
  \[
  Q(x, dx') = q_\xi(x' - x - b(x)) \, dx'.
  \]

- \( d_n^{\mu_0} \) - normalizing coefficient, gives the nonlinearity, (the denominator in the Bayes formula).
Strange conditional probability via the same operator:

\[ P_{n}^{\nu_0, Y}(dx_n) =: \nu_0 \bar{S}_n^{Y, \nu_0}(dx_n) \]

\[ P_n^{\nu_0, Y}(dx_n) = d_n^{\nu_0} \int \prod_{i=1}^{n} Q(x_{i-1}, dx_i) \psi(x_i, Y_i) \nu_0(dx_0). \]
Main question - reformulation

True or false:

$$\lim_{n \to \infty} E_{\mu_0} \| \mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0} \|_{TV} = 0?$$
Assumptions, I

(A0) $b$ and $h$ are locally bounded;

(A1p) : recurrence & moments. 
(Khasminskii-Veretennikov conditions):

$p = 0 : \limsup_{|x| \to \infty} \left\langle b(x), \frac{x}{|x|} \right\rangle \leq -r, \ r > 0$

$E \exp(c|\xi|) < \infty$

or

$p = 1 : \lim_{|x| \to \infty} \left\langle b(x), x \right\rangle = -\infty.$

$E|\xi|^m < \infty \ \forall \ m > 0;$
Assumptions, I

(A0) \( b \) and \( h \) are locally bounded;

(A1\(p \)) : recurrence & moments. (Khasminskii-Veretennikov conditions):

\[
\begin{align*}
\rho = 0 : & \quad \limsup_{|x| \to \infty} \left< b(x), \frac{x}{|x|} \right> \leq -r, \; r > 0 \\
& \quad E \exp(c|\xi|) < \infty \\
\text{or} & \quad \rho = 1 : \quad \lim_{|x| \to \infty} \left< b(x), x \right> = -\infty. \\
& \quad E|\xi|^m < \infty \; \forall \; m > 0;
\end{align*}
\]
Assumptions, I, continued

Examples

\((p = 0)\) : \(b(x) = -\text{sign}(x), \ b(x) = -x; \ldots\)

\((p = 1)\) : \(b(x) = -\frac{\arctan(x)}{\sqrt{1 + |x|}}; \ldots\)
Assumptions, I, continued

Examples

\((p = 0) : b(x) = -\text{sign}(x), b(x) = -x; \ldots\)

\((p = 1) : b(x) = -\frac{\arctan(x)}{\sqrt{1 + |x|}}; \ldots\)
Assumptions, I, continued

\[(A2)\] local mixing. Density \(q_v > 0\) and

\[C_R := \sup_{|x|,|v| \leq R} \frac{q_\xi(x)}{q_\xi(v)} < \infty,\]

\[(A3)\] absolute continuity of initial data.

\[\left\| \frac{d\mu_0}{d\nu_0} \right\|_{L_\infty(\nu_0)} < \infty.\]

\[(A4)\] (initial moments)

\[\int e^{c|x|} \mu_0(dx) < \infty.\]
Assumptions, I, continued

(A2) : **local mixing.** Density \( q_v > 0 \) and

\[
C_R := \sup_{|x|, |v| \leq R} \frac{q_{\xi}(x)}{q_{\xi}(v)} < \infty,
\]

(A3) : **absolute continuity** of initial data.

\[
\left\| \frac{d\mu_0}{d\nu_0} \right\|_{L^\infty(\nu_0)} < \infty.
\]

(A4) *(initial moments)*

\[
\int e^{c|x|} \mu_0(dx) < \infty.
\]
(A2) : local mixing. Density \( q_v > 0 \) and

\[
C_R := \sup_{|x|,|v| \leq R} \frac{q_\xi(x)}{q_\xi(v)} < \infty,
\]

(A3) : absolute continuity of initial data.

\[
\left\| \frac{d\mu_0}{d\nu_0} \right\|_{L_\infty(\nu_0)} < \infty.
\]

(A4) (initial moments)

\[
\int e^{c|x|} \mu_0(dx) < \infty.
\]
Stability, Theorem I

Under Assumptions (A0) – (A4) the following bounds hold:

$$E_{\mu_0} \| \mu_0 \bar{S}_n^\mathcal{Y},\mu_0 - \nu_0 \bar{S}_n^\mathcal{Y},\nu_0 \|_{TV} \leq \begin{cases} C_m n^{-m}, & p = 1, \forall m > 0, \\ C \exp(-cn), & p = 0. \end{cases}$$
Outline

1. Introduction
   - Problem statement
   - Historical survey
   - Reformulation of the problem, the Bayes approach

2. Assumptions and results
   - Stability with absolutely continuous initial data
   - Stability without initial data absolute continuity

3. Auxiliaries
   - Ergodic processes in $\mathbb{R}^d$
   - Birkhoff metric

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
   - Theorem 2, idea of the proof
(A'0) The function $b$ is locally bounded; the function $h$ is **bounded**.

(A'1) (Recurrence)

$$\lim_{|x| \to \infty} (|x + b(x)| - |x|) = -\infty, \ (p = -\infty)$$

(A'2) (Gaussian noises) The noise $(\xi_n, V_n)$ is an IID standard Gaussian random sequence.

(A'3) (Moments)

$$\int e^{c|x|} (\mu_0(dx) + \nu_0(dx)) < \infty.$$
Stability, Theorem II

Theorem

Under the assumptions \( (A'0) - (A'3) \) \( \exists c_0 > 0 \) such that

\[
E_{\mu_0} \| \mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0} \|_{TV} \leq C \exp(-c_0 n).
\]

Example. \( b(x) = -\frac{x}{5} \).

Remark. Non-Gaussian noises in both components of the system could be considered too.
Ergodic Filters

Veretennikov, Kleptsyna

Introduction

Problem statement

Historical survey

Reformulation of the problem, the Bayes approach

Assumptions and results

Stability with absolutely continuous initial data

Stability without initial data absolute continuity

Auxiliaries

Ergodic processes in $\mathbb{R}^d$

Birkhoff metric

Sketch of the proof

Coupling and separation

The main inequality

Sketch of the proof, part 2

Stability, Theorem II

**Theorem**

*Under the assumptions $(A’0) – (A’3)$ $\exists c_0 > 0$ such that*

$$E_{\mu_0} \| \mu_0 \bar{S}_n^{\bar{Y},\mu_0} - \nu_0 \bar{S}_n^{\bar{Y},\nu_0} \|_{TV} \leq C \exp(-c_0 n).$$

**Example.** $b(x) = -\frac{x}{5}$.

**Remark.** Non-Gaussian noises in both components of the system could be considered too.
Stability, Theorem II

**Theorem**

*Under the assumptions (A’0) – (A’3) ∃c₀ > 0 such that*

\[ E_{µ_0} \| µ_0 \bar{S}_n^{Y,µ_0} - ν_0 \bar{S}_n^{Y,ν_0} \|_{TV} \leq C \exp(-c_0 n). \]

**Example.** \( b(x) = -\frac{x}{5} \).

**Remark.** Non-Gaussian noises in both components of the system could be considered too.
Outline

1 Introduction
   - Problem statement
   - Historical survey
   - Reformulation of the problem, the Bayes approach

2 Assumptions and results
   - Stability with absolutely continuous initial data
   - Stability without initial data absolute continuity

3 Auxiliaries
   - Ergodic processes in $R^d$
   - Birkhoff metric

4 Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
   - Theorem 2, idea of the proof
Hitting time estimates (Veretennikov, 1987, 2005):
For $\hat{\tau} = \inf(n \geq 0 : |X_n| \leq R)$

\[
\begin{align*}
E_x\hat{\tau}^k & \leq C_m(1 + |x|^m) \quad (\forall m > 2k; p = 1), \\
E_x \exp(\alpha \hat{\tau}) & \leq C \exp(c|x|) \quad (p = 0).
\end{align*}
\]

Corollary. Let $\#1(X)_R := \sum_{k=0}^n 1(|X_k| \leq R)$. Then (for $R$ large enough)

\[
E_{\mu_0} 1(\#1(X)_R < \varepsilon n) \leq \begin{cases} 
C_m n^{-m}, & (p = 1), \\
C \exp(-cn), & (p = 0, \exists c), \\
C \exp(-cn), & (\forall c, p = -\infty)
\end{cases}
\]
Hitting time estimates (Veretennikov, 1987, 2005):
For $\hat{\tau} = \inf(n \geq 0 : |X_n| \leq R)$

\[
\begin{cases}
E_x \hat{\tau}^k \leq C_m(1 + |x|^m) & (\forall m > 2k; p = 1), \\
E_x \exp(\alpha \hat{\tau}) \leq C \exp(c|x|) & (p = 0).
\end{cases}
\]

**Corollary.** Let $\#1(X)_R := \sum_{k=0}^{n} 1(|X_k| \leq R)$. Then (for $R$ large enough)

\[
E_{\mu_0} 1(\#1(X)_R < \varepsilon n) \leq \begin{cases}
C_m n^{-m}, & (p = 1), \\
C \exp(-cn), & (p = 0, \exists c), \\
C \exp(-cn), & (\forall c, p = -\infty)
\end{cases}
\]
Outline

1. Introduction
   - Problem statement
   - Historical survey
   - Reformulation of the problem, the Bayes approach

2. Assumptions and results
   - Stability with absolutely continuous initial data
   - Stability without initial data absolute continuity

3. Auxiliaries
   - Ergodic processes in $R^d$
   - Birkhoff metric

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
   - Theorem 2, idea of the proof
Birkhoff metric

Definition

Let $\mu$ and $\nu$ be positive measures. The Birkhoff distance $\rho(\mu, \nu)$ is defined by:

$$\rho(\mu, \nu) = \begin{cases} 
\ln \sup \left( \frac{d\mu}{d\nu} \right) + \ln \sup \left( \frac{d\nu}{d\mu} \right), & \text{if finite}, \\
+\infty, & \text{otherwise}.
\end{cases}$$

Remark. It is a pseudo-distance, measuring the difference between directions.
Birkhoff metric

Definition

Let $\mu$ and $\nu$ be positive measures. The Birkhoff distance $\rho(\mu, \nu)$ is defined by:

$$\rho(\mu, \nu) = \begin{cases} 
\ln \sup \left( \frac{d\mu}{d\nu} \right) + \ln \sup \left( \frac{d\nu}{d\mu} \right), & \text{if finite}, \\
+\infty, & \text{otherwise}.
\end{cases}$$

Remark. It is a pseudo-distance, measuring the difference between directions.
Comparison of total variation distance and Birkhoff distance

(Christophe Leuriden, private communication)

- For normalized measures $\mu$ and $\nu$:

$$\|\mu - \nu\|_{TV} \leq \rho(\mu, \nu)$$

- The converse statement does not hold.

Example

$$q_\mu(x) = \begin{cases} 1 & (x \in [-1/2, 1/2]) \\ 0 & \text{otherwise} \end{cases}$$

$$q_\nu(x) = \frac{1}{2} \cdot 1(|x| \in [\varepsilon, 1/2]) + C \cdot 1(x \in [-\varepsilon, \varepsilon])$$

Then $$\|\mu - \nu\|_{TV} = 1 - 2\varepsilon, \quad \rho(\mu, \nu) = \ln(1 + \frac{2}{\varepsilon})$$
Comparison of total variation distance and Birkhoff distance

(Christophe Leuriden, private communication)

- For normalized measures $\mu$ and $\nu$:
  \[
  \|\mu - \nu\|_{TV} \leq \rho(\mu, \nu)
  \]

- The converse statement does not hold.

Example

\[
q_\mu(x) = \begin{cases} 
1 & (x \in [-1/2, 1/2]) \\
\end{cases}
\]

\[
q_\nu(x) = \frac{1}{2} \cdot 1(|x| \in [\varepsilon, 1/2]) + C \cdot 1(x \in [-\varepsilon, \varepsilon])
\]

Then \[
\|\mu - \nu\|_{TV} = 1 - 2\varepsilon, \quad \rho(\mu, \nu) = \ln(1 + \frac{2}{\varepsilon})
\]
Two important properties, I

Comparison of total variation distance and Birkhoff distance

(Christophe Leuriden, private communication)

- For normalized measures $\mu$ and $\nu$:

$$\|\mu - \nu\|_{TV} \leq \rho(\mu, \nu)$$

- The converse statement does not hold.

Example

$$q_\mu(x) = \begin{cases} 1 & (x \in [-1/2, 1/2]) \\ 0 & \text{otherwise} \end{cases}$$

$$q_\nu(x) = \frac{1}{2} \cdot \begin{cases} 1 & (|x| \in [\varepsilon, 1/2]) \\ 0 & \text{otherwise} \end{cases} + C \cdot \begin{cases} 1 & (x \in [-\varepsilon, \varepsilon]) \\ 0 & \text{otherwise} \end{cases}$$

Then $\|\mu - \nu\|_{TV} = 1 - 2\varepsilon$, $\rho(\mu, \nu) = \ln(1 + \frac{2}{\varepsilon})$
Two important properties,

### Comparison of total variation distance and Birkhoff distance

(Christophe Leuriden, private communication)
- For normalized measures $\mu$ and $\nu$:
  \[
  \|\mu - \nu\|_{TV} \leq \rho(\mu, \nu)
  \]
- The converse statement does not hold.

### Example

\[
q_\mu(x) = 1 \left( x \in \left[ -\frac{1}{2}, \frac{1}{2} \right] \right)
\]
\[
q_\nu(x) = \frac{1}{2} \cdot 1 \left( |x| \in [\varepsilon, \frac{1}{2}] \right) + C \cdot 1 \left( x \in [-\varepsilon, \varepsilon] \right)
\]

Then $\|\mu - \nu\|_{TV} = 1 - 2\varepsilon$, $\rho(\mu, \nu) = \ln(1 + \frac{2}{\varepsilon})$
Birkhoff contraction for nonnegative kernels:

Let $Q : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$ s.t.: $\mu Q(dy) = \int_{\mathbb{R}^d} Q(x, dy) \mu(dx)$.

Contraction

$$\rho(\mu Q, \nu Q) \leq \frac{C^2 - 1}{C^2 + 1} \rho(\mu, \nu), \text{ with}$$

- $(\text{Krasnosel'skii, M. A., Lifshits, E. A., Sobolev, A. V.})$

$$C = \sup_{x, z, y} \frac{q(x, y)}{q(z, y)}, \quad Q(x, dy) = q(x, y)dy.$$

- $(\text{Le Gland, F., Oudjane, N.})$

$$C = \sup_{x, z, A} \frac{Q(x, A)}{Q(z, A)}.$$
Birkhoff contraction for nonnegative kernels:
Let $Q : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$ s.t.: $\mu Q(dy) = \int_{\mathbb{R}^d} Q(x, dy) \mu(dx)$.

Contraction

$$\rho(\mu Q, \nu Q) \leq \frac{C^2 - 1}{C^2 + 1} \rho(\mu, \nu), \text{ with}$$

- (Krasnosel’skii, M. A., Lifshits, E. A., Sobolev, A. V.)

$$C = \sup_{x, z, y} \frac{q(x, y)}{q(z, y)}, \ Q(x, dy) = q(x, y)dy.$$  

- (Le Gland, F., Oudjane, N.)

$$C = \sup_{x, z, A} \frac{Q(x, A)}{Q(z, A)}.$$
Two important properties, II

**Birkhoff contraction for nonnegative kernels:**
Let $Q : \mathcal{M}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$ s.t.: $\mu Q(dy) = \int_{\mathbb{R}^d} Q(x, dy) \mu(dx)$.

**Contraction**

\[
\rho(\mu Q, \nu Q) \leq \frac{C^2 - 1}{C^2 + 1} \rho(\mu, \nu), \quad \text{with}
\]

- (Krasnosel’skii, M. A., Lifshits, E. A., Sobolev, A. V.)

\[
C = \sup_{x, z, y} \frac{q(x, y)}{q(z, y)}, \quad Q(x, dy) = q(x, y)dy.
\]

- (Le Gland, F., Oudjane, N.)

\[
C = \sup_{x, z, A} \frac{Q(x, A)}{Q(z, A)}.
\]
Outline

1. Introduction
   - Problem statement
   - Historical survey
   - Reformulation of the problem, the Bayes approach

2. Assumptions and results
   - Stability with absolutely continuous initial data
   - Stability without initial data absolute continuity

3. Auxiliaries
   - Ergodic processes in $\mathbb{R}^d$
   - Birkhoff metric

4. Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
   - Theorem 2, idea of the proof
Doubling the space

Consider **independent** couples \((X, Y)\) and \((\tilde{X}, \tilde{Y})\) with initial laws \(\mathcal{L}(X_0) = \mu_0, \mathcal{L}(\tilde{X}_0) = \nu_0\).

Strange conditional probability, continued

We can interpret **strange** conditional probability as:

\[
P_{\nu_0, Y}^n(\cdot) = P_{\nu_0}(\tilde{X}_n \in \cdot \mid \mathcal{F}_n^{\tilde{Y}}) \mid \tilde{Y} = Y.
\]

with \(\tilde{Y}\) replaced by \(Y\).
Doubling the space

Consider **independent** couples \((X, Y)\) and \((\tilde{X}, \tilde{Y})\) with initial laws \(\mathcal{L}(X_0) = \mu_0, \mathcal{L}(\tilde{X}_0) = \nu_0\).

Strange conditional probability, continued

We can interpret **strange** conditional probability as:

\[
P_{\nu_0}^{\nu_0, Y}(\cdot) = P_{\nu_0}(\tilde{X}_n \in \cdot \mid \mathcal{F}_n \tilde{Y}) \mid \tilde{Y} = Y.
\]

with \(\tilde{Y}\) replaced by \(Y\).
For fixed $R$, $n$, and any non-random vector $\delta \in \Delta = \{0; 1\}^{n+1}$ we define (with a convention $0^0 = 1$)

$$1_\delta(X, \tilde{X}) := \prod_{i=0}^{n-1} (1 - 1(D_i))^{\delta_i} \times (1 - 1(D_i))^{1-\delta_i},$$

where

$$D_i := \left\{ \max \left( |X_i|, |\tilde{X}_i| \right) \leq R; \right\}$$
Separation

Partition of unity

For fixed $R, n$, and any non-random vector $\delta \in \Delta = \{0; 1\}^{n+1}$ we define (with a convention $0^0 = 1$)

$$1_\delta(X, \tilde{X}) := \prod_{i=0}^{n-1} (1 (D_i))^{\delta_i} \times (1 - 1 (D_i))^{1-\delta_i},$$

where

$$D_i := \left\{ \max \left( |X_i|, |\tilde{X}_i| \right) \leq R; \right\}$$
Partition of unity, II

Multiplicative decomposition

\[ 1_\delta(X, \tilde{X}) := \prod_{i=0}^{n-1} 1_{\delta}(D_i) \]

with

\[ 1_{\delta_i}(D_i) = 1(\delta_i = 1)1(D_i) + 1(\delta_i = 0)(1 - 1(D_i)). \]
Partition of unity, II

**Multiplicative decomposition**

\[ 1_\delta(X, \tilde{X}) := \prod_{i=0}^{n-1} 1_{\delta_i}(D_i) \]

with

\[ 1_{\delta_i}(D_i) = 1(\delta_i = 1)1(D_i) + 1(\delta_i = 0)(1 - 1(D_i)). \]

**Partition of unity**

\[ 1 = \sum_{\delta \in \Delta} 1_\delta(X, \tilde{X}) \]
Denote by \( \#1(\delta) \) the total number of pairs of ones in \( \delta \) and by

\[
\#1(X)_R := \sum_{k=0}^{n-1} 1(|X_k| \leq R)
\]

The following inequality holds:

\[
\sum_{\delta: \#1(\delta) < \varepsilon n} 1_{\delta}(X, \tilde{X}) \leq 1(\#1(X)_R < \frac{3 + \varepsilon}{4} n) + 1(\#1(\tilde{X})_R < \frac{3 + \varepsilon}{4} n)
\]
Separation of pairs

Denote by \#1(\delta) the total number of pairs of ones in \delta and by

\[
\#1(X)_R := \sum_{k=0}^{n-1} 1(|X_k| \leq R)
\]

The following inequality holds:

\[
\sum_{\delta: \#1(\delta) < \varepsilon n} 1_\delta(X, \tilde{X}) \leq 1(\#1(X)_R < \frac{3}{4} n) + 1(\#1(\tilde{X})_R < \frac{3}{4} n)
\]
Let introduce the central object of the following study:

Strange probability separator

\[ e_n^{Y;\delta;\mu_0,\nu_0} := E_{\mu_0,\nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) \bigg| \tilde{Y} = Y. \]
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Subsections</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>Problem statement, Historical survey</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reformulation of the problem, the Bayes approach</td>
</tr>
<tr>
<td>2</td>
<td>Assumptions and results</td>
<td>Stability with absolutely continuous initial data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stability without initial data absolute continuity</td>
</tr>
<tr>
<td>3</td>
<td>Auxiliaries</td>
<td>Ergodic processes in $R^d$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Birkhoff metric</td>
</tr>
<tr>
<td>4</td>
<td>Sketch of the proof</td>
<td>Coupling and separation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The main inequality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sketch of the proof, part 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theorem 2, idea of the proof</td>
</tr>
</tbody>
</table>
An estimate to prove

Our goal is to prove the following inequality:

The main inequality

\[ E_{\mu_0} \| \mu_0 \bar{S}_n^{Y, \mu_0} - \nu_0 \bar{S}_n^{Y, \nu_0} \|_{TV} \leq C \sum_{\delta \in \Delta} k_R \#^1(\delta) E_{\mu_0, \nu_0} e_n^{Y; \delta; \mu_0, \nu_0}, \]

\[ k_R := \frac{C^2_R - 1}{C^2_R + 1} < 1, \]

\[ C_R := \sup_{|x|, |v| \leq R} \frac{q_\xi(x)}{q_\xi(v)} < \infty. \]
Our goal is to prove the following inequality:

The main inequality

\[
E_{\mu_0} \| \mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0} \|_{TV} \leq C \sum_{\delta \in \Delta} \kappa_R \#(\delta) E_{\mu_0,\nu_0} e_n^{Y,\delta;\mu_0,\nu_0},
\]

where

\[
\kappa_R := \frac{C_R^2 - 1}{C_R^2 + 1} < 1,
\]

\[
C_R := \sup_{|x|,|v| \leq R} \frac{q_\xi(x)}{q_\xi(v)} < \infty.
\]
The sum in the main inequality we can split into two terms $(\forall \varepsilon > 0)$:

$$\sum_{\delta: \#1(\delta) \geq \varepsilon n} + \sum_{\delta: \#1(\delta) < \varepsilon n}$$

and we have:

$$\sum_{\delta: \#1(\delta) \geq \varepsilon n} \kappa_R \#1(\delta) E_{\mu_0} e_{\varepsilon n}^{Y; \delta; \mu_0, \nu_0} \leq \kappa_R \varepsilon n$$

$$\sum_{\delta: \#1(\delta) < \varepsilon n} \kappa_R \#1(\delta) E_{\mu_0} \left( E_{\mu_0, \nu_0}(1_\delta(X, \tilde{X}) | Y, \tilde{Y}) \right)_{\tilde{Y} = Y}$$

$$\leq \sum_{\delta: \#1(\delta) < \varepsilon n} E_{\mu_0} \left( E_{\mu_0, \nu_0}(1_\delta(X, \tilde{X}) | Y, \tilde{Y}) \right)_{\tilde{Y} = Y}.$$
The sum in the main inequality we can split into two terms \((\forall \varepsilon > 0):
\sum_{\delta: \#1(\delta) \geq \varepsilon n} + \sum_{\delta: \#1(\delta) < \varepsilon n}\)

and we have:

\[
\sum_{\delta: \#1(\delta) \geq \varepsilon n} \kappa_R \#1(\delta) E_{\mu_0} e_n^{Y;\delta;\mu_0,\nu_0} \leq \kappa_R^\varepsilon n
\]

\[
\sum_{\delta: \#1(\delta) < \varepsilon n} \kappa_R \#1(\delta) E_{\mu_0} \left( E_{\mu_0,\nu_0}(1_{\delta}(X, \tilde{X}) \mid Y, \tilde{Y}) \bigg| \tilde{Y} = Y \right)
\]

\[
\leq \sum_{\delta: \#1(\delta) < \varepsilon n} E_{\mu_0} \left( E_{\mu_0,\nu_0}(1_{\delta}(X, \tilde{X}) \mid Y, \tilde{Y}) \bigg| \tilde{Y} = Y \right)
\]
The sum in the main inequality we can split into two terms
\( \forall \varepsilon > 0 \):
\[
\sum_{\delta : \#1(\delta) \geq \varepsilon n} + \sum_{\delta : \#1(\delta) \leq \varepsilon n}
\]
and we have:

\[
\sum_{\delta : \#1(\delta) \geq \varepsilon n} \kappa_R \#1(\delta) E_{\mu_0} e_n^{Y;\delta;\mu_0,\nu_0} \leq \kappa_R^{\varepsilon n}
\]

\[
\sum_{\delta : \#1(\delta) \leq \varepsilon n} \kappa_R \#1(\delta) E_{\mu_0} \left( E_{\mu_0,\nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) \bigg| \tilde{Y} = Y \right)
\]
\[
\leq \sum_{\delta : \#1(\delta) \leq \varepsilon n} E_{\mu_0} \left( E_{\mu_0,\nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) \bigg| \tilde{Y} = Y \right).
\]
Theorem 1, sketch of the proof, 2

We can finish the proof:

\[
E_{\mu_0} \left( E_{\mu_0,\nu_0} \left( \sum_{\delta : \#1(\delta) < \varepsilon n} 1_{\delta(X, \tilde{X}) | Y, \tilde{Y}} \right) \bigg| \tilde{Y} = Y \right)
\]

\[
\leq E_{\mu_0} \left( E_{\mu_0} \left( 1(\#1(X)_R < \frac{3 + \varepsilon}{4} n) | Y \right) \right)
\]

\[
+ E_{\mu_0} \left( E_{\nu_0} \left( 1(\#1(\tilde{X})_R < \frac{3 + \varepsilon}{4} n) | \tilde{Y} \right) \bigg| \tilde{Y} = Y \right)
\]

(because \( X \) does not depend on \( \tilde{Y} \), nor \( \tilde{X} \) depends on \( Y \)).

Inequality "Separation of pairs" has been used.
We estimate the first term

\[ E_{\mu_0} \left( E_{\mu_0} \left( 1(\#1(X)_R < \frac{3 + \varepsilon}{4} n) \mid Y \right) \right) \]

\[ = E_{\mu_0} \left( 1(\#1(X)_R < \frac{3 + \varepsilon}{4} n) \right). \]

we can use the hitting time estimates.
We estimate the first term

$$E_{\mu_0} \left( E_{\mu_0} \left( 1(\#1(X)_R < \frac{3 + \varepsilon}{4} n) \mid Y \right) \right)$$

$$= E_{\mu_0} \left( 1(\#1(X)_R < \frac{3 + \varepsilon}{4} n) \right).$$

we can use the hitting time estimates.
Next, we estimate the other term, using the **of absolute continuity the initial measures**:

\[
E_{\mu_0} \left( E_{\nu_0} \left( 1(\#1(\tilde{X})_R < \frac{3 + \varepsilon}{4} n) \mid \tilde{Y} \right) \mid \tilde{Y} = Y \right)
\]

\[\leq C_2 \ E_{\nu_0} \left( E_{\nu_0} \left( 1(\#1(\tilde{X})_R < \frac{3 + \varepsilon}{4} n) \mid \tilde{Y} \right) \mid \tilde{Y} = Y \right)\]

\[= C_2 \ E_{\nu_0} \left( 1(\#1(\tilde{X})_R < \frac{3 + \varepsilon}{4} n) \right),\]

Again, the **hitting time estimates**.
Next, we estimate the other term, using the **of absolute continuity** the initial measures:

\[
E_{\mu_0} \left(E_{\nu_0} \left(1(\#1(\tilde{X})_R < \frac{3 + \varepsilon}{4} n) \mid \tilde{Y} \right) \mid \tilde{Y} = Y \right)
\leq C_2 E_{\nu_0} \left(E_{\nu_0} \left(1(\#1(\tilde{X})_R < \frac{3 + \varepsilon}{4} n) \mid \tilde{Y} \right) \mid \tilde{Y} = Y \right)
\]

\[
= C_2 E_{\nu_0} \left(1(\#1(\tilde{X})_R < \frac{3 + \varepsilon}{4} n) \right)
\]

Again, the **hitting time estimates**.
Coupling method, part 2
New operators, 1

How we can prove the main inequality?

the space doubled, 2

Define new operators on the spaces of normalized and non-normalized measures on $R^{2d}$

$$(\mu_0, \nu_0) \tilde{S}^{Y;\mu_0,\nu_0}_n (A \times B) = d_n^{\mu_0} d_n^{\nu_0} \int \int_{R^{2d}} 1(x_n \in A, \tilde{x}_n \in B) \times \prod_{i=1}^{n} \psi(x_i, Y_i) \psi(\tilde{x}_i, Y_i) Q(x_{i-1}, dx_i) Q(\tilde{x}_{i-1}, d\tilde{x}_i)$$

$\mu_0(dx_0) \nu_0(d\tilde{x}_0),$

(Because of $d_n^{\mu_0} d_n^{\nu_0}$ it is a nonlinear operator.)
Coupling method, part 2
New operators, 1

How we can prove the main inequality?

the space doubled, 2

- Define new operators on the spaces of normalized and non-normalized measures on $R^{2d}$

$$
(\mu_0, \nu_0) \tilde{S}_n^{Y; \mu_0, \nu_0} (A \times B) = d_n^{\mu_0} d_n^{\nu_0} \int \int_{R^{2d}} 1(x_n \in A, \tilde{x}_n \in B) \times
$$

$$
\left( \prod_{i=1}^{n} \psi(x_i, Y_i) \psi(\tilde{x}_i, Y_i) Q(x_{i-1}, dx_i) Q(\tilde{x}_{i-1}, d\tilde{x}_i) \right)
$$

$$
\mu_0(dx_0) \nu_0(d\tilde{x}_0),
$$

(Because of $d_n^{\mu_0} d_n^{\nu_0}$ it is a nonlinear operator.)
Comparison of measures and distances

The following properties hold:
- \((\mu_0)\bar{S}_n^{Y;\mu_0}(A) = (\mu_0, \nu_0)\bar{S}_n^{Y;\mu_0,\nu_0}(A \times R^d)\)
- \((\nu_0)\bar{S}_n^{Y;\nu_0}(A) = (\nu_0, \mu_0)\bar{S}_n^{Y;\mu_0,\nu_0}(A \times R^d)\)

The first property

\[\|\mu_0\bar{S}_n^{Y;\mu_0} - \nu_0\bar{S}_n^{Y;\nu_0}\|_{TV} \leq \|(\mu_0, \nu_0)\bar{S}_n^{Y;\mu_0,\nu_0} - (\nu_0, \mu_0)\bar{S}_n^{Y;\nu_0,\mu_0}\|_{TV}\]
Comparison of measures and distances

The following properties hold:

- \((\mu_0)\bar{S}_n^{Y;\mu_0}(A) = (\mu_0, \nu_0)\bar{S}_n^{Y;\mu_0,\nu_0}(A \times \mathbb{R}^d)\)
- \((\nu_0)\bar{S}_n^{Y;\nu_0}(A) = (\nu_0, \mu_0)\bar{S}_n^{Y;\mu_0,\nu_0}(A \times \mathbb{R}^d)\)

The first property

\[ \|\mu_0 \bar{S}_n^{Y;\mu_0} - \nu_0 \bar{S}_n^{Y;\nu_0}\|_{TV} \leq \|(\mu_0, \nu_0)\bar{S}_n^{Y;\mu_0,\nu_0} - (\nu_0, \mu_0)\bar{S}_n^{Y;\nu_0,\mu_0}\|_{TV} \]
Coupling method, part 2

Using partition of unity

\[(\mu_0, \nu_0)\overline{S}_n^{Y;\mu_0,\nu_0}(A \times B) = \sum_{\delta \in \Delta} \overline{S}_n^{Y;R;\delta;\mu_0,\nu_0}(A \times B)\]

with

The (non-normalized) decomposition

\[(\mu_0, \nu_0)\overline{S}_n^{Y;R;\delta;\mu_0,\nu_0}(A \times B) = d_n^{\mu_0} d_n^{\nu_0} \int 1(x_n \in A, \tilde{x}_n \in B) 1_\delta(x, \tilde{x}) \times \left( \prod_{i=1}^n \psi(x_i, Y_i) \psi(\tilde{x}_i, Y_i) Q(x_{i-1}, dx_i) Q(\tilde{x}_{i-1}, d\tilde{x}_i) \right) \mu_0(dx_0) \nu_0(d\tilde{x}_0),\]
Coupling method, part 2

Using partition of unity

\[(\mu_0, \nu_0) \tilde{S}^Y_{\mu_0, \nu_0} (A \times B) = \sum_{\delta \in \Delta} \tilde{S}^Y_{\delta; \mu_0, \nu_0} (A \times B)\]

with

The (non-normalized) decomposition

\[(\mu_0, \nu_0) \tilde{S}^Y_{\delta; \mu_0, \nu_0} (A \times B) = d^\mu_0 d^\nu_0 \int 1(x_n \in A, \tilde{x}_n \in B) 1_\delta(x, \tilde{x}) \times \left( \prod_{i=1}^n \psi(x_i, Y_i) \psi(\tilde{x}_i, Y_i) Q(x_{i-1}, dx_i) Q(\tilde{x}_{i-1}, d\tilde{x}_i) \right) \mu_0(dx_0) \nu_0(d\tilde{x}_0), \]
Probability separator, again

Normalization:

\[(\mu, \nu) \hat{S}_n^{Y; R; \delta; \mu_0; \nu_0}(A \times B) := (e_n^{Y; \delta; \mu_0; \nu_0})^{-1} (\mu, \nu) \bar{S}_n^{Y; R; \delta; \mu_0; \nu_0}(A \times B).\]

We see that the normalizing coefficient is exactly the \(e_t^{Y; \delta; \mu_0; \nu_0}\):

Probability separator, II

\[e_n^{Y; \delta; \mu_0; \nu_0} := E_{\mu_0, \nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) \bigg|_{\tilde{Y}=Y}\]

\[= d_{\mu_0}^{\nu_0} d_{\mu_0}^{\nu_0} \bar{S}_n^{Y; R; \delta; \mu_0; \nu_0}(R^{2d}; (\mu_0, \nu_0)).\]
Define new **linear** operators on the space of non-normalized measures on $R^{2d}$

**linear operator**

\[
(\mu, \nu) S_n^{Y; R; \delta}(A \times B) = \int \int 1(x_n \in A, \tilde{x}_n \in B) 1_\delta(x, \tilde{x}) \times \left( \prod_{i=1}^{n} \psi(x_i, Y_i) \psi(\tilde{x}_i, Y_i) Q(x_{i-1}, dx_i) Q(\tilde{x}_{i-1}, \tilde{d}x_i) \right) \times \mu(dx_0) \nu(d\tilde{x}_0),
\]

(non-normalized conditional probability).
It can be equivalently presented as

\[(\mu, \nu) S_n^{Y; R; \delta}(A \times B) = (\mu, \nu) \prod_{i=0}^{n-1} S_{i:i+1}^{Y; R; \delta}(A \times B).\]

with

\[(\mu_i, \nu_i) S_{i:i+1}^{Y; R; \delta}(A \times B)
= \int \int 1(x_{i+1} \in A, \tilde{x}_{i+1} \in B) 1_\delta(x_i, \tilde{x}_i) 1_\delta(x_{i+1}, \tilde{x}_{i+1})
\times \psi(x_{i+1}, Y_{i+1}) \psi(\tilde{x}_{i+1}, Y_{i+1}) Q(x_i, dx_{i+1}) Q(\tilde{x}_i, d\tilde{x}_{i+1})
\mu_i(dx_i) \nu_i(d\tilde{x}_i).\]
It can be equivalently presented as

$$(\mu, \nu)S_{n}^{Y;R;\delta}(A \times B) = (\mu, \nu) \prod_{i=0}^{n-1} S_{i:i+1}^{Y;R;\delta}(A \times B).$$

with

$$(\mu_i, \nu_i)S_{i:i+1}^{Y;R;\delta}(A \times B)$$

$$= \int \int 1(x_{i+1} \in A, \tilde{x}_{i+1} \in B) 1_{\delta}(x_i, \tilde{x}_i) 1_{\delta}(x_{i+1}, \tilde{x}_{i+1})$$

$$\times \psi(x_{i+1}, Y_{i+1})\psi(\tilde{x}_{i+1}, Y_{i+1})Q(x_{i}, dx_{i+1})Q(\tilde{x}_{i}, d\tilde{x}_{i+1}) \mu_i(dx_i)\nu_i(d\tilde{x}_i).$$
We can estimate the total variation norm:

\[
\| (\mu_0, \nu_0) \bar{S}_n^{Y; \mu_0, \nu_0} - (\nu_0, \mu_0) \bar{S}_n^{Y; \nu_0, \mu_0} \|_{TV}
\]

\[
\leq d^\mu_0 d^\nu_0 \sum_{\delta \in \Delta} \| (\mu_0, \nu_0) \bar{S}_n^{Y; R; \delta; \mu_0, \nu_0} - (\nu_0, \mu_0) \bar{S}_n^{Y; R; \delta; \mu_0, \nu_0} \|_{TV}
\]

\[
= \sum_{\delta \in \Delta} e_n^{Y; \delta; \mu_0, \nu_0} \| (\mu_0, \nu_0) \bar{S}_n^{Y; R; \delta; \mu_0, \nu_0} - (\nu_0, \mu_0) \bar{S}_n^{Y; R; \delta; \mu_0, \nu_0} \|_{TV}
\]
We can estimate the total variation norm:

\[ \| (\mu_0, \nu_0) \tilde{S}_n^{Y; \mu_0, \nu_0} - (\nu_0, \mu_0) \tilde{S}_n^{Y; \nu_0, \mu_0} \|_{TV} \leq d_n^{\mu_0} d_n^{\nu_0} \sum_{\delta \in \Delta} \| (\mu_0, \nu_0) \tilde{S}_n^{Y; R; \delta; \mu_0, \nu_0} - (\nu_0, \mu_0) \tilde{S}_n^{Y; R; \delta; \mu_0, \nu_0} \|_{TV} \]

\[ = \sum_{\delta \in \Delta} e_n^{Y; \delta; \mu_0, \nu_0} \| (\mu_0, \nu_0) \hat{S}_n^{Y; R; \delta; \mu_0, \nu_0} - (\nu_0, \mu_0) \hat{S}_n^{Y; R; \delta; \mu_0, \nu_0} \|_{TV} \]
We can estimate the total variation norm:

$$\| (\mu_0, \nu_0) \tilde{S}_n^{Y; \mu_0, \nu_0} - (\nu_0, \mu_0) \tilde{S}_n^{Y; \nu_0, \mu_0} \|_{TV}$$

$$\leq d_{\mu_0}^{n} d_{\nu_0}^{n} \sum_{\delta \in \Delta} \| (\mu_0, \nu_0) \tilde{S}_n^{Y; R; \delta; \mu_0, \nu_0} - (\nu_0, \mu_0) \tilde{S}_n^{Y; R; \delta; \mu_0, \nu_0} \|_{TV}$$

$$= \sum_{\delta \in \Delta} e_n^{Y; \delta; \mu_0, \nu_0} \| (\mu_0, \nu_0) \hat{S}_n^{Y; R; \delta; \mu_0, \nu_0} - (\nu_0, \mu_0) \hat{S}_n^{Y; R; \delta; \mu_0, \nu_0} \|_{TV}$$
Using the Birkhoff metric, 1

Using the properties of the Birkhoff metric we see that

Birkhoff metric, 1st property.

\[
\|((\mu_0, \nu_0) \hat{S}_n^{Y;R;\delta;\mu_0,\nu_0} - (\nu_0, \mu_0) \hat{S}_n^{Y;R;\delta;\mu_0,\nu_0})\|_{TV} \leq \\
\rho((\mu_0, \nu_0) \hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}, (\nu_0, \mu_0) \hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}).
\]
Using the Birkhoff metric, 2

and that

**Birkhoff metric, 2nd property.**

\[
\rho((\mu_0, \nu_0) \hat{S}_n^{Y; R; \delta; \mu_0, \nu_0}, (\nu_0, \mu_0) \hat{S}_n^{R; \delta; \mu_0, \nu_0}) \\
\equiv \rho \left( (\mu_0, \nu_0) S_n^{Y; R; \delta}, (\nu_0, \mu_0) S_n^{Y; R; \delta} \right) \\
\leq \kappa^{\delta n} R \rho \left( (\mu_0, \nu_0) S_{n-1}^{Y; R; \delta}, (\nu_0, \mu_0) S_{n-1}^{Y; R; \delta} \right) \\
\leq C \kappa^k R, \\
\text{with} \\
k = \#1(\delta)
\]

which gives the desired inequality.
Outline

1 Introduction
   - Problem statement
   - Historical survey
   - Reformulation of the problem, the Bayes approach

2 Assumptions and results
   - Stability with absolutely continuous initial data
   - Stability without initial data absolute continuity

3 Auxiliaries
   - Ergodic processes in $\mathbb{R}^d$
   - Birkhoff metric

4 Sketch of the proof
   - Coupling and separation
   - The main inequality
   - Sketch of the proof, part 2
   - Theorem 2, idea of the proof
The Bayes formula, again

In order to estimate the second term in the main inequality we use the Bayes formula

\[ E_{\mu_0,\nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) = \frac{\hat{E}_{\mu_0,\nu_0}(1_\delta(X, \tilde{X})L_n \mid Y, \tilde{Y})}{\hat{E}_{\mu_0,\nu_0}(L_n \mid Y, \tilde{Y})}, \]

and due to the Cauchy inequality, we see that:

\[ E_{\mu_0}(E_{\mu_0,\nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) \mid \tilde{Y} = Y) \leq \exp(Cn)E_{\mu_0,\nu_0}1_\delta(X, \tilde{X}) \]

Again, the hitting time estimates!!!
The Bayes formula, again

In order to estimate the second term in the main inequality we use the Bayes formula

\[ E_{\mu_0,\nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) = \frac{\hat{E}_{\mu_0,\nu_0}(1_\delta(X, \tilde{X})L_n \mid Y, \tilde{Y})}{\hat{E}_{\mu_0,\nu_0}(L_n \mid Y, \tilde{Y})}, \]

and due to the Cauchy inequality, we see that:

\[ E_{\mu_0}(E_{\mu_0,\nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) \mid \tilde{Y} = Y) \leq \exp(Cn)E_{\mu_0,\nu_0}1_\delta(X, \tilde{X}) \]

Again, the hitting time estimates!!!!
In order to estimate the second term in the main inequality, we use the Bayes formula

\[
E_{\mu_0, \nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) = \frac{\hat{E}_{\mu, \nu_0} \left(1_\delta(X, \tilde{X})L_n \mid Y, \tilde{Y}\right)}{\hat{E}_{\mu_0, \nu_0}^{\gamma}(L_n \mid Y, \tilde{Y})},
\]

and due to the Cauchy inequality, we see that:

\[
E_{\mu_0}(E_{\mu_0, \nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) \mid \tilde{Y} = Y) \leq \exp(Cn)E_{\mu_0, \nu_0}1_\delta(X, \tilde{X})
\]

Again, the hitting time estimates!!!
The Bayes formula, again

In order to estimate the second term in the main inequality we use the Bayes formula

\[
E_{\mu_0, \nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) = \frac{\hat{E}_{\mu_0, \nu_0}(1_\delta(X, \tilde{X})L_n \mid Y, \tilde{Y})}{\hat{E}_{\mu_0, \nu_0}(L_n \mid Y, \tilde{Y})},
\]

and due to the Cauchy inequality, we see that:

\[
E_{\mu_0}(E_{\mu_0, \nu_0}(1_\delta(X, \tilde{X}) \mid Y, \tilde{Y}) \mid \tilde{Y} = Y) \leq \exp(Cn)E_{\mu_0, \nu_0}1_\delta(X, \tilde{X})
\]

Again, the hitting time estimates!!!
Conclusion
Open questions

- Can we do something with unbounded $h$ without initial data absolute continuity?
- Can we do something with degenerated densities?
- Can we do something with nonergodic signals?