(\(3-\alpha\))-MCMC - Approximation under Drift Condition

Krzysztof Latuszynski  
(presenting author)

Institute of Econometrics,  
Warsaw School of Economics  
latush@gmail.com

Wojciech Niemiro

Faculty of Mathematics and Computer Science,  
Nicolaus Copernicus University  
wniem@mat.uni.torun.pl
An often problem encountered in Bayesian inference is the computation of

\[ I = \int_{\mathbb{C}} f(x) \Pi(x) \, dx \]
An often problem encountered in Bayesian inference is the computation of

\[ I = \int_{x} f(x) \, \Pi(x) \, dx \]

region in high-dimensional space (e.g. \( x \in \mathbb{R}^d \))
An often problem encountered in Bayesian inference is the computation of the target function of interest.

\[ I = \int_{\text{region in high-dimensional space (e.g. } \mathcal{X} \subseteq \mathbb{R}^d)} f(x) T(x) \, dx \]
An often problem encountered in Bayesian inference is the computation of the target function of interest.

\[ I = \int_{\mathcal{X}} f(x) \Pi(x) dx \]

Region in high-dimensional space (e.g. \( \mathcal{X} \subseteq \mathbb{R}^d \))

Prob. dist. on \( \mathcal{X} \), usually known up to a normalizing constant.
An often problem encountered in Bayesian inference is the computation of target function of interest:

\[ I = \int_{\mathcal{X}} f(x) \pi(x) \, dx \]

Region in high-dimensional space (e.g. \( \mathcal{X} \subseteq \mathbb{R}^d \))

Common approach: simulate an ergodic Markov chain \((X_n)_{n \geq 0}\) using a transition kernel \(P\) with stationary distribution \(\pi\).
• An often problem encountered in Bayesian inference is the computation of the target function of interest:

\[ I = \int_{\mathcal{X}} f(x) \Pi(x) \, dx \]

region in high-dimensional space (e.g. \( \mathcal{X} \subseteq \mathbb{R}^d \))

• Common approach: simulate an ergodic Markov chain \((X_n)_{n \geq 0}\) using a transition kernel \(P\) with stationary distribution \(\Pi\).

• Once you have the chain, there are two basic strategies for estimation:
An often problem encountered in Bayesian inference is the computation of

\[ I = \int_{x} f(x) \pi(x) \, dx \]

\( x \) is the region in high-dimensional space (e.g. \( x \in \mathbb{R}^d \)).

Common approach: simulate an ergodic Markov chain \( (X_n)_{n \geq 0} \) using a transition kernel \( P \) with stationary distribution \( \pi \).

Once you have the chain, there are two basic strategies for estimation:

1. Use average over final states of multiple independent runs of the chain. e.g.

\[ X_0^{(1)}, X_1^{(1)} \ldots X_t^{(1)} \]

\[ X_0^{(2)}, X_1^{(2)} \ldots X_t^{(2)} \]

\[ \vdots \]

\[ X_0^{(n)}, X_1^{(n)} \ldots X_t^{(n)} \]
• An often problem encountered in Bayesian inference is the computation of the target function of interest.

\[ I = \int f(x) \pi(x) \, dx \]

Region in high-dimensional space (e.g., \( x \in \mathbb{R}^d \)).

• Common approach: simulate an ergodic Markov chain \((X_n)_{n \geq 0}\) using a transition kernel \(P\) with stationary distribution \(\pi\).

• Once you have the chain, there are two basic strategies for estimation:

  1. Use average over final states of multiple independent runs of the chain, i.e.

\[ \hat{I} = \frac{1}{n} \sum_{k=1}^{n} f(X_t^{(k)}) \]

   \[ X_0^{(1)}, X_1^{(1)}, \ldots, X_t^{(1)} \]

   \[ X_0^{(2)}, X_1^{(2)}, \ldots, X_t^{(2)} \]

   \[ \vdots \]

   \[ X_0^{(n)}, X_1^{(n)}, \ldots, X_t^{(n)} \]
An often problem encountered in Bayesian inference is the computation of a target function of interest.

\[ I = \int_{\mathbb{R}^d} f(x) \pi(x) \, dx \]

prob. distr. on \( X \), usually known up to a normalizing constant.

Common approach: simulate an ergodic Markov chain \( (X_n)_{n \geq 0} \) using a transition kernel \( \pi \) with stationary distribution \( \pi \).

Once you have the chain, there are two basic strategies for estimation:

I. Use average over final states of multiple independent runs of the chain, i.e.,

\[ \bar{I} = \frac{1}{n} \sum_{k=1}^{n} f(X^{(k)}_t) \]

II. Use average along a single trajectory, e.g.,

\[ X_0, X_t, X_{t+1}, \ldots, X_t, X_{t+n-1} \]
An often problem encountered in Bayesian inference is the computation of

\[ I = \int_{\mathbb{R}^d} f(x) \, \Pi(x) \, dx \]

the target function of interest

\[ \mathbb{P}(\text{target function}) \]

region in high-dimensional space (e.g. \( \mathbb{X} \subseteq \mathbb{R}^d \)).

Common approach: simulate an ergodic Markov chain \((X_n)_{n \geq 0}\) using a transition kernel \(P\) with stationary distribution \(\Pi\).

Once you have the chain, there are two basic strategies for estimation:

I Use average over final states of multiple independent runs of the chain (i.e. \(X_0^{(1)}, X_1^{(1)}, \ldots, X_T^{(1)}\))

\[ I = \frac{1}{n} \sum_{k=1}^{n} f(X^{(k)}_t) \]

II Use average along a single trajectory \(X_0, X_1, \ldots, X_T, X_{T+1}, \ldots, X_{T+n-1}\)

\[ \hat{I}_{t,n} = \frac{1}{n} \sum_{i=t}^{t+n-1} f(X_i) \]
I - easier to analyze, $X_t^{(k)}$ - iid

II - practitioners choice, but $X_t, X_{t+1}, \ldots, X_{t+n-1}$ are dependent
- easier to analyze, $X_t^{(k)}$ - iid

- practitioners choice, but
  $X_t, X_{t+1}, \ldots, X_{t+n-1}$ are dependent

Our Goal: $\hat{I}_{t,n} = \frac{1}{n} \sum_{k=t}^{t+n-1} f(X_k)$

give lower bounds for $t$ and $n$

that ensure $(\varepsilon - \alpha)$-approximation

$P \left( \left| \hat{I}_{t,n} - I \right| \leq \varepsilon \right) \geq 1 - \alpha$

precision of estimation

confidence level

Under the assumption of a Drift Condition
- easier to analyze, $X_t^{(k)} - iid$

- practitioners choice, but
  $X_t, X_{t+1}, \ldots, X_{t+n-1}$ are dependent

Our Goal: $\hat{I}_{t,n} = \frac{1}{n} \sum_{k=t}^{t+n-1} f(X_k)$

give lower bounds for $t$ and $n$

that ensure $(\varepsilon - \alpha)$-approximation

$P(\left| \hat{I}_{t,n} - I \right| \leq \varepsilon) \geq 1 - \alpha$

precision of estimation confidence level

Under the Assumption of a Drift Condition
- easier to analyze, $X_t^{(k)} \sim u.d$

- practitioners choice, but $X_t, X_{t+1}, \ldots, X_{t+n-1}$ are dependent

**Our Goal**: $\hat{I}_{t,n} = \frac{1}{n} \sum_{k=t}^{t+n-1} f(X_k)$

give lower bounds for $t$ and $n$ that ensure $(\varepsilon - \alpha)$-approximation

$P\left( |\hat{I}_{t,n} - I| \leq \varepsilon \right) \geq 1 - \alpha$

Precision of estimation, confidence level

**Similar Results**

1. Finite state space $\mathcal{X}$ + bounded $f$
   - Aldous 87
   - Gillman 98
   - Niemiro, Pokarowski 03 - relative precision

2. Léon, Perron 04 - Hoeffding type bounds

Under the assumption of a drift condition
I - easier to analyze, \( X_t^{(k)} - i.i.d \)

II - practitioner's choice, but \( X_t, X_{t+1}, \ldots, X_{t+n-1} \) are dependent

**Our Goal:** \( \hat{I}_{t,n} = \frac{1}{n} \sum_{k=t}^{t+n-1} f(X_k) \)

give lower bounds for \( t \) and \( n \) that ensure \((\varepsilon - \alpha)\)-approximation

\[ P \left( \left| \hat{I}_{t,n} - I \right| \leq \varepsilon \right) \geq 1 - \alpha \]

precision of estimation confidence level

**Under the Assumption of a Drift Condition**

**Similar Results**

1. **Finite state space** \( \mathcal{X} \) + bounded \( f \)
   - Aldous 87; Gillman 98
   - Niemiro, Pokarowski 03 - relative precision
   - Léon, Perron 04 - Hoeffding type bounds

2. **Doeblin chains** + bounded \( f \)
   - Glynn, Ormonaid 02
   - Heyn, Kontoyiannis 05 - Hoeffding type bounds

Lasstas-Mantano
- easier to analyze, $X_t^{(k)}$ - i.i.d.

- practitioners choice, but $X_t, X_{t+1}, \ldots, X_{t+n-1}$ are dependent

**Our Goal:** $\hat{I}_{t,n} = \frac{1}{n} \sum_{k=t}^{t+n-1} f(X_k)$

give lower bounds for $t$ and $n$

that ensure $(\varepsilon - \alpha)$-approximation

$P(\left| \hat{I}_{t,n} - I \right| \leq \varepsilon) \geq 1 - \alpha$

precision of estimation confidence level

**Similar Results**

1. Finite state space $\mathcal{X}$ + bounded $f$
   - Aldous 87'; Gillman 98';
   - Niemiro, Pokarowski 03' - relative precision
   - Léon, Perron 04' - Hoeffding type bounds

2. Doeblin chains + bounded $f$
   - Glynn, Ormonait 02';
   - Heyn, Kontoyiannis 05' \{ Hoeffding type bounds

3. General setting
   - $X$ - not compact, $f$ - not bounded

Under the Assumption of a Drift Condition
- easier to analyze, $X_t^{(k)} - i.i.d.$

- practitioners choice, but

$X_t, X_{t+1}, \ldots, X_{t+n-1}$ are dependent

**Our Goal:**

$$\hat{I}_{t,n} = \frac{1}{n} \sum_{k=t}^{t+n-1} f(X_k)$$

give lower bounds for $t$ and $n$

that ensure $(\varepsilon - \alpha)$-approximation

$$P\left( \left| \hat{I}_{t,n} - I \right| \leq \varepsilon \right) \geq 1 - \alpha$$

**Similar Results**

1. **Finite state space** $\mathcal{X}$ + bounded $f$
   - Aldous 87', Gillman 98'
   - Niemiro, Pokarowski 03' - relative precision
   - Léon, Perron 04' - Hoeffding type bounds

2. Doeblin chains + bounded $f$
   - Glynn, Ormonait 02'
   - Heyn, Kontoyiannis 05' - type bounds
   - Lastras-Mantano

3. **General setting**
   - $X$ - not compact, $f$ - not bounded

Under the **Assumption of a Drift Condition**
Our approach to \((\varepsilon-\delta)\)-approx.

MSE bounds \(\int_{\varepsilon} \Rightarrow (\varepsilon-\delta)\)-approximation

Chebyshev's Ineq
Our approach to \((\epsilon, \alpha)\)-approx.

MSE bounds \(\int_0^1 \epsilon \Rightarrow (\epsilon, \alpha)\)-approximation

+ Chebyshev's Ineq

\[ \text{MSE}(\hat{I}_{t,n}) = E(\hat{I}_{t,n} - I)^2 \leq b(t,n) \]

\[ P(|I - \hat{I}_{t,n}| \leq \epsilon) = 1 - P(|I - \hat{I}_{t,n}| \geq \epsilon) \geq \text{Chebyshev} \]

\[ 1 - \frac{\text{MSE}(\hat{I}_{t,n})}{\epsilon^2} \geq 1 - \alpha \]

if \( \text{MSE}(\hat{I}_{t,n}) \leq \epsilon^2 \alpha \)
Our approach to \((\varepsilon-\alpha)\)-approx.

MSE bounds

\[ \text{Chebyshev's Inequality} \]

\[ \text{MSE}(\hat{I}_{t,n}) = E(\hat{I}_{t,n} - \hat{I})^2 \leq b(t, n) \]

\[ P(|\hat{I} - \hat{I}_{t,n}| \leq \varepsilon) = 1 - P(1 - \hat{I}_{t,n} \geq \varepsilon) \geq \text{Chebyshev} \]

\[ 1 - \left( \frac{\text{MSE}(\hat{I}_{t,n})}{\varepsilon^2} \right) \geq 1 - \alpha \]

if \[ \text{MSE}(\hat{I}_{t,n}) \leq \varepsilon^2 \alpha \]

From now on we concentrate on MSE bounds!!
Our approach to $(\epsilon, \alpha)$-approx.

MSE bounds $\int (\epsilon, \alpha)$-approximation

+ Chebyshev's lueg

\[
\text{MSE} (\hat{I}_{t,n}) = E(\hat{I}_{t,n} - I)^2 \leq b(t, n)
\]

\[
P(|I - \hat{I}_{t,n}| \leq \epsilon) = 1 - P(|I - \hat{I}_{t,n}| \geq \epsilon) \geq 1 - \frac{\text{MSE}(\hat{I}_{t,n})}{\epsilon^2} \geq 1 - \alpha
\]

if $\text{MSE}(\hat{I}_{t,n}) \leq \epsilon^2 \alpha$

From now on we concentrate on MSE bounds!!
Our approach to \((\varepsilon, \alpha)\)-approx.

MSE bounds

\[ \int_0^t (\varepsilon - x) \approx (\varepsilon, \alpha) \text{-approximation} \]

Chebyshev’s Inequality

\[ \text{MSE}(\hat{I}_{t,n}) = E(\hat{I}_{t,n} - I)^2 \leq b(t, n) \]

\[ P(|I - \hat{I}_{t,n}| \leq \varepsilon) = 1 - P(|I - \hat{I}_{t,n}| \geq \varepsilon) \geq 1 - \frac{\text{MSE}(\hat{I}_{t,n})}{\varepsilon^2} \geq 1 - \alpha \]

Chebyshev

\[ \text{if } \text{MSE}(\hat{I}_{t,n}) \leq \varepsilon^2 \alpha \]

From now on we concentrate on MSE bounds!!
Our approach to $(\varepsilon, \alpha)$-approx.

- MSE bounds $\int E(\varepsilon, \alpha)$-approximation

- Chebyshev's inequality + Chebyshev's inequality

- $\text{MSE}(\hat{I}_{t,n}) = E(\hat{I}_{t,n} - I)^2 \leq b(t, n)$

- $P(|I - \hat{I}_{t,n}| \leq \varepsilon) = 1 - P(|I - \hat{I}_{t,n}| \geq \varepsilon) \geq 1 - \frac{\text{MSE}(\hat{I}_{t,n})}{\varepsilon^2}$

- Chebyshev

- if $\text{MSE}(\hat{I}_{t,n}) \leq \varepsilon^2 \alpha$

From now on we concentrate on MSE bounds!!

A Drift Condition

(See Heym-Tweedle for various drift conditions and their relation)

A.1 (Small Set)

There exists $C \subseteq \mathcal{X}$, $\beta > 0$, $\nu^*$-prob. measure on $\mathcal{X}$

$\forall x \in C, A \subseteq \mathcal{X} \quad \nu^*(x, A) \geq \beta \nu(A)$

A2. (Drift)

There exist a function $V: \mathcal{X} \to [1, \infty)$, constants $\lambda < 1$, $K < \infty$, satisfying

$PV(x) \leq \begin{cases} 2V(x), & \text{if } x \notin C \\ K, & \text{if } x \in C \end{cases}$
Our approach to \((\varepsilon, \alpha)\)-approx.

- MSE bounds
  
  \[ \text{Chebychev's Ineq.} \]

- \( \text{MSE}(\hat{T}_{t,n}) = \text{E}(\hat{T}_{t,n} - I)^2 \leq b(t, n) \)

- \( P(\mid I - \hat{T}_{t,n} \mid < \varepsilon) = 1 - P(\mid I - \hat{T}_{t,n} \mid \geq \varepsilon) \geq 1 - \alpha \)

- \( \text{Chebychev} \geq 1 - \frac{\text{MSE}(\hat{T}_{t,n})}{\varepsilon^2} \geq 1 - \alpha \)

- \( \text{if } \text{MSE}(\hat{T}_{t,n}) \leq \varepsilon^2 \alpha \)

From now on we concentrate on MSE bounds!!

A Drift Condition

(See Heym-Tweedie for various drift conditions and their relation)

A1 (Small Set)

There exists \( C \subseteq X, \beta > 0, \nu \)-prob. measure on \( X \)

\[ \forall x \in C, A \subseteq X \quad P(x, A) \geq \beta \nu(A) \]

A2 (Drift)

There exist a function \( V: X \rightarrow [1, \infty) \), constants \( \beta < 1, K < \infty \), satisfying

\[ P^V(x) \leq \begin{cases} 2V(x), & \text{if } x \notin C \\ K, & \text{if } x \in C \end{cases} \]

A3 (Aperiodicity)

There exists \( \beta > 0 \), such that

\[ \beta \nu(C) \geq \beta \]
Thm (Baxendale 05')
Under this Drift Condition, \( (X_n)_{n \geq 0} \) has a unique stationary distribution \( \pi \), and \( \pi V < \infty \). Moreover, there exists \( g < 1 \), depending only and explicitly on \( \beta, \hat{\beta}, \alpha, K \), such that whenever \( g < g < 1 \), there exists \( M < \infty \), depending only and explicitly on \( g, \beta, \hat{\beta}, \alpha, K \), such that for all \( n \geq 0 \)

\[ \| p^n - \pi \|_V \leq M g^n \]
Thm (Baxendale 05')
Under this Drift Condition, \((X_n)_{n \geq 0}\) has a unique stationary distribution \(\pi\), and \(\pi V < \infty\). Moreover, there exists \(\gamma < 1\), depending only and explicitly on \(\beta, \bar{\beta}, \lambda, K\), such that whenever

\[ \gamma < \gamma' < 1, \text{ there exists } M < \infty, \text{ depending only and explicitly on } \gamma, \beta, \bar{\beta}, \lambda, K, \]

such that for all \(n > 0\)

\[ ||| p^n - \pi |||_V \leq M \gamma^n \]

\[ ||| \mu_1 - \mu_2 |||_V := \sup_{|g| \leq V} |\mu_1 g - \mu_2 g| \quad \text{for prob. measures} \]

\[ |g|_V := \sup_{x \in X} \frac{|g(x)|}{V(x)} \quad \text{for functions} \]

\[ ||| Q - R |||_V := \sup_{x \in X} \frac{||| Q(x, \cdot) - R(x, \cdot) |||_V}{V(x)} \quad \text{for transition kernels} \]
Thm (Baxendale 05)
Under this Drift Condition, \((X_n)_{n \geq 0}\) has a unique stationary distribution \(\pi\), and \(\pi V < \infty\). Moreover, there exists \(\delta < 1\), depending only and explicitly on \(\delta, \bar{\delta}, \bar{\theta}, k\), such that whenever \(\delta < \gamma < 1\), there exists \(M < \infty\), depending only and explicitly on \(\gamma, \bar{\delta}, \bar{\theta}, k\), such that for all \(n \geq 0\)

\[ \|P^n - \pi\|_V \leq M \gamma^n \]

Lemma 1
If for a Markov chain \((X_n)_{n \geq 0}\) with transition kernel \(P\) the Drift Condition holds with parameters \(\beta, V(x), \bar{\alpha}, \bar{\kappa}, \bar{\beta}\),

\[ \|\mu_1 - \mu_2\|_V := \sup_{|g| \leq V} |\mu_1 g - \mu_2 g| \] - for prob. measures

\[ |g|_V := \sup_{x \in \mathcal{X}} \left| \frac{g(x)}{V(x)} \right| \] - for functions

\[ \|Q - R\|_V := \sup_{x \in \mathcal{X}} \frac{\|Q(x, \cdot) - R(x, \cdot)\|_V}{V(x)} \] - for transition kernels
**Thm (Baxendale 05)**

Under this Drift Condition, \((X_n)_{n \geq 0}\) has a unique stationary distribution \(\pi\), and \(\pi V < \infty\). Moreover, there exists \(s < 1\), depending only and explicitly on \(\gamma, \beta, \lambda, \kappa\), such that whenever \(s < \gamma < 1\), there exists \(M < \infty\), depending only and explicitly on \(s, \beta, \lambda, \kappa\), such that for all \(n \geq 0\)

\[
\| P^n - \pi \|_V \leq M \gamma^n
\]

**Lemma 1**

If for a Markov chain \((X_n)_{n \geq 0}\) with transition kernel \(P\) the Drift Condition holds with parameters \(\beta, V(x), \lambda, \kappa, \beta\), it holds also with \(\beta_r = \beta; V_r(x) = V(\gamma_r x); A_r = \gamma_r A; K_r = K\gamma_r;\) for every \(r > 1\).
**Thm (Baxendale 05)**

Under this Drift Condition, \((X_n)_{n \geq 0}\) has a unique stationary distribution \(\pi\), and \(\pi V < \infty\). Moreover, there exists \(\delta < 1\), depending only and explicitly on \(\beta, \beta, \lambda, K\), such that whenever \(\delta < \delta < 1\), there exists \(M < \infty\), depending only and explicitly on \(\delta, \beta, \beta, \lambda, K\), such that for all \(n \geq 0\)

\[
\|P^n - \pi\|_V \leq M \delta^n
\]

**Lemma 1**

If for a Markov chain \((X_n)_{n \geq 0}\) with transition kernel \(P\) the Drift Condition holds with parameters \(\beta, V(x), \lambda, K\), it holds also with

\[
\beta_r = \beta; \quad V_r(x) := V(x)^r; \quad \lambda_r := \lambda_r; \quad K_r := K_r,
\]

for every \(r > 1\).

**Corollary**

Under the initial Drift Condition Assumption

\[
\|P^n - \pi\|_{V_r} \leq M_r \delta^n_r
\]
**Theorem (Baxendale 05)**

Under this **Drift Condition**, \((X_n)_{n \geq 0}\) has a unique stationary distribution \(\pi\), and \(\pi V < \infty\). Moreover, there exists \(\delta < 1\), depending only and explicitly on \(\beta, \wp, \alpha, K\), such that whenever \(\delta' < \delta < 1\), there exists \(M < \infty\), depending only and explicitly on \(\delta, \beta, \wp, \alpha, K\), such that for all \(n > 0\)

\[
\| P^n - \pi \|_V \leq M \delta^n
\]

**Lemma 1**

If for a Markov chain \((X_n)_{n \geq 0}\) with transition kernel \(P\) the Drift Condition holds with parameters \(\beta, V(x), \alpha, K\), it holds also with \(\beta_r = \beta; V_r(x) = V(x)^r; \alpha_r = \alpha^r; K_r = K^r\), for every \(r > 1\).

**Corollary**

Under the initial Drift Condition Assumption

\[
\| P^n - \pi \|_V \|_V^{1/r} \leq M_r \delta_r^n
\]

\[
\| P^n - \pi \|_V \leq M \delta^n
\]

For prob. measures

\[
\| \mu_1 - \mu_2 \|_V := \sup_{|g| \leq V} \| \mu_1 g - \mu_2 g \|
\]

For functions

\[
\| g \|_V := \sup_{x \in X} \frac{|g(x)|}{V(x)}
\]

For transition kernels

\[
\| Q - R \|_V := \sup_{x \in X} \frac{\| Q(x, \cdot) - R(x, \cdot) \|_V}{V(x)}
\]

We need this corollary to deal with autocorrelation of the chain.
Corollary (from Baxendale's Thm)
Under the Drift Condition

$$\| \pi_0 P^n - \pi \|_V \leq \min \left\{ \| \pi_0 V \|, \| \pi_0 - \pi \|_V \right\} \cdot M \gamma^n$$

This enables us to deal with the burn-in.
Corollary (from Baxendale's Thin)

Under the Drift Condition

\[ \| \pi_0 P^n - \pi \|_V \leq \min \left\{ \pi_0 V, \| \pi_0 - \pi \|_V \right\} \cdot MP^n \]

This enables us to deal with the burn-in.

**MSE bound**

Assume the Drift Condition holds, and \( X_0 \sim \pi_0 \). Then for every \( f: X \to \mathbb{R} \), every \( P \geq 2 \), every \( r \in \left[ \frac{P}{P-1}, P \right] \),

\[ \text{MSE}(\hat{I}_{t,m}) \leq \]
Corollary (from Baxendale's Thm)

Under the Drift Condition

\[ \| \pi_0 P^n - \pi \|_\nu \leq \min \left\{ \pi_0 \nu, \| \pi_0 - \pi \|_\nu \right\} \cdot M \rho^n \]

This enables us to deal with the burn-in.

MSE bound

Assume the Drift Condition holds, and \( X_0 \sim \pi_0 \). Then for every \( f: X \to \mathbb{R} \), every \( p \geq 2 \), every \( r \in \left[ \frac{p}{p-1}, p \right] \),

\[ \text{MSE} \left( \hat{I}_{t,n} \right) \leq \ldots \]

Elementary computation, use Baxendale's Thm, Baxendale's \( r \)-Thm, the last corollary...
Corollary (from Baxendale's Thm)

Under the Drift Condition

\[ \| \pi_0 P^n - \pi \| V \leq \min \left\{ \| \pi_0 V, \| \pi_0 - \pi \| V \right\} \cdot M \gamma^n \]

This enables us to deal with the burn-in.

MSE bound

Assume the Drift Condition holds, and

\[ X_0 \sim \pi_0. \text{ Then for every } f: X \rightarrow \mathbb{R}, \text{ every } P \geq 2, \text{ every } r \in \left[ \frac{P}{P-1}, P \right], \]

\[
\text{MSE}(\hat{T}_{t,n}) \leq \frac{1}{n} \left( \frac{1}{1 - M \gamma^n} \right) \left( \frac{2 M r^{-1}}{1 - r^{-1}} \right) \left( \frac{M^2 \gamma^{n+1} \min \{ \| \pi_0 V, \| \pi_0 - \pi \| V \} }{n (1 - \gamma)} \right)
\]
Corollary (from Bassendales Thm)

Under the Drift Condition

\[ \| \pi_0 P^n - \pi \|_V \leq \min \{ \pi_0 V, \| \pi_0 - \pi \|_V \} \cdot M^n \]

This enables us to deal with the burn-in.

**MSE bound**

Assume the Drift Condition holds, and \( X_0 \sim \pi_0 \). Then for every \( f: X \to \mathbb{R} \), every \( p \geq 2 \), every \( r \in \left[ \frac{p}{p-1}, p \right] \),

\[
\text{MSE} \left( \tilde{I}_{t,n} \right) \leq \frac{1}{n} \left( 1 + \frac{2 M}{1 - \delta_r} \right) \left( \pi_0 V + \frac{M^2 \gamma^t \min \{ \pi_0 V, \| \pi_0 - \pi \|_V \}^2}{n (1 - \delta)} \right)
\]

interpretation:

\[ 1 f_{\pi_0}^{2/p} \pi_0 V \approx \text{Var}_{\pi_0} f \quad (\text{for } p=2 \text{ and appropriate choice of } V) \]
**Corollary (from Baxendale's Thm)**

Under the Drift Condition

\[ \| \pi_0 P^n - \pi \|_V \leq \min \{ \pi_0 V, \| \pi_0 - \pi \|_V \} \cdot M \gamma^n \]

This enables us to deal with the burn-in.

**MSE bound**

Assume the Drift Condition holds, and \( X_0 \sim \pi_0 \). Then for every \( f: X \to \mathbb{R} \), every \( p \geq 2 \), every \( r \in \left[ \frac{p-1}{p}, p \right] \),

\[
\text{MSE}(\hat{I}_{t,n}) \leq \frac{\| f \|^2_p}{n} \left( 1 + \frac{2Mr}{1 - \delta r} \right) \left( \pi V + \frac{M^2 \gamma^t \min\{\pi_0 V, \| \pi_0 - \pi \|_V \}^2}{n (1 - \delta)} \right)
\]

**Interpretation:**

\[ \frac{\| f \|^2_p}{n} \pi V \approx \text{Var}_{\pi} f \] (for \( p = 2 \) and appropriate choice of \( V \))

\[ \frac{2Mr}{1 - \delta r} \] \( \sim \) autocorrelation of the chain
Corollary (from Baxendale's Thin)

Under the Drift Condition

\[ \| \pi_0 P^n - \pi \|_V \leq \min \{ \pi_0 V, \| \pi_0 - \pi \|_V \} \cdot M r^n \]

This enables us to deal with the burn-in.

MSE bound

Assume the Drift Condition holds, and \( x_0 \sim \pi_0 \). Then for every \( f: X \to \mathbb{R} \), every \( p \geq 2 \), every \( r \in \left[ \frac{P}{P-1}, P \right] \),

\[
\text{MSE} \left( \mathbb{I}_{t,n} \right) \leq \frac{1 \cdot p^{2/p}}{\pi} \left( 1 + \frac{2M r}{1 - \delta r} \right) \left( \pi V + \frac{M^2 \gamma^t \min \{ \pi_0 V, \| \pi_0 - \pi \|_V \}^3}{n (1 - \delta)} \right)
\]

interpretation:

\[ \frac{1 \cdot p^{2/p}}{\pi} \pi V \approx \text{Var}_{\pi} f \] (for \( p=2 \) and appropriate choice of \( V \))

\[ \frac{2M r}{1 - \delta r} \sim \text{autocorrelation of the chain} \]

\[ \text{price for nonstationarity of the initial distribution.} \]