Exponential inequality for autoregressive processes in adaptive tracking

B. Bercu

University of Bordeaux, France

Key Laboratory of Systems and Control,
Academy of Mathematics and Systems Science, CAS
Beijing, China, April 8, 2007

Dedicated to Professor H. F. Chen on the occasion of his seventieth birthday

Outline

Outline

1. Classical exponential inequalities
   - Hoeffding's inequality
   - Bennett's inequality
   - Bernstein's inequality
2. Exponential inequalities for martingales
   - Azuma-Hoeffding's inequality
   - Freedman's inequality
   - De la Peña's inequality
3. Application to adaptive tracking
4. Two open problems

Hoeffding's inequality

Let (X_n) be a sequence of **independent** random variables and

$$S_n = \sum_{k=1}^{n} X_k.$$

Theorem (Hoeffding's inequality)

Assume that for each $1 \leqslant k \leqslant n$, $a_k \leqslant X_k \leqslant b_k$ a.s. for some constants $a_k < b_k$. Then, for all $x \geqslant 0$,

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geqslant x) \leqslant 2 \exp\Big(-\frac{2x^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\Big).$$

Bennett's inequality

Theorem (Bennett's inequality)

Assume that (X_n) is square integrable and for each $1 \leqslant k \leqslant n$, $X_k \leqslant c$ a.s. for some constant $c > 0$. Then, for all $x \geqslant 0$,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) \leqslant \exp\Big(-\frac{V_n}{c^2} h\Big(\frac{xc}{V_n}\Big)\Big)$$

where V_n is the variance of S_n and

$$h(x) = (1+x)\log(1+x) - x.$$

Bernstein's inequality

For all $x \geqslant 0$,

$$h(x) \geqslant \frac{3x^2}{2(3+x)}.$$

Theorem (Bernstein's inequality)

Assume that (X_n) is square integrable and for each $1 \leqslant k \leqslant n$, $X_k \leqslant c$ a.s. for some constant $c > 0$. Then, for all $x \geqslant 0$,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) \leqslant \exp\Big(-\frac{x^2}{2(V_n + xc/3)}\Big).$$

Outline

Azuma-Hoeffding's inequality

Let (M_n) be a square integrable martingale adapted to $\mathbb{F} = (\mathcal{F}_n)$ with $M_0 = 0$. The **predictable** and the **total** quadratic variations of (M_n) are given by

$$< M >_n = \sum_{k=1}^{n} \mathbb{E}[\Delta M_k^2 | \mathcal{F}_{k-1}], \qquad [M]_n = \sum_{k=1}^{n} \Delta M_k^2$$

where $\Delta M_n = M_n - M_{n-1}$.

Theorem (Azuma-Hoeffding's inequality)

Assume that for each $1 \leqslant k \leqslant n$, $a_k \leqslant \Delta M_k \leqslant b_k$ a.s. for some constants $a_k < b_k$. Then, for all $x \geqslant 0$,

$$\mathbb{P}(|M_n| \geqslant x) \leqslant 2 \exp\left(- \frac{2x^2}{\sum_{k=1}^{n} (b_k - a_k)^2} \right).$$

Freedman's inequality

Theorem (Freedman's inequality)

Assume that for each $1 \leqslant k \leqslant n$, $\Delta M_k \leqslant c$ a.s. for some constant $c > 0$. Then, for all $x, y > 0$,

$$\mathbb{P}(M_n \geqslant x, <M>_n \leqslant y) \leqslant \exp\left(-\frac{x^2}{2(y+cx)}\right).$$

Theorem

Freedman's inequality also holds under the Bernstein moment condition: for all $n \geqslant 1$, $p \geqslant 2$ and for some constant $c > 0$,

$$\sum_{k=1}^{n} \mathbb{E}[|\Delta M_k|^p | \mathcal{F}_{k-1}] \leqslant \frac{p!}{2} c^{p-2} <M>_n \qquad a.s.$$

De la Peña's inequality

Definition. We shall say that (M_n) is **conditionally symmetric** if, for all $n \geqslant 1$, $\mathcal{L}(\Delta M_n | \mathcal{F}_{n-1})$ is symmetric.

Theorem (De la Peña's inequality)

If (M_n) is conditionally symmetric, then for all $x, y > 0$,

$$\mathbb{P}(M_n \geqslant x, [M]_n \leqslant y) \leqslant \exp\left(-\frac{x^2}{2y}\right).$$

Self-normalized martingales

Theorem (De la Peña's inequality)

If (M_n) is conditionally symmetric, then for all $x, y > 0$,

$$\mathbb{P}\Big(\frac{M_n}{[M]_n} \geqslant x\Big) \leqslant \sqrt{\mathbb{E}\Big[\exp\Big(-\frac{x^2}{2}[M]_n\Big)\Big]},$$

$$\mathbb{P}\Big(\frac{M_n}{[M]_n} \geqslant x, [M]_n \geqslant \frac{1}{y}\Big) \leqslant \exp\Big(-\frac{x^2}{2y}\Big).$$

Goal. Normalize by $<M>_n$ instead of $[M]_n$.

Self-normalized martingales

Theorem (De la Peña's inequality)

If (M_n) is conditionally symmetric, then for all $x, y > 0$,

$$\mathbb{P}\Big(\frac{M_n}{[M]_n} \geqslant x\Big) \leqslant \sqrt{\mathbb{E}\Big[\exp\Big(-\frac{x^2}{2}[M]_n\Big)\Big]},$$

$$\mathbb{P}\Big(\frac{M_n}{[M]_n} \geqslant x, [M]_n \geqslant \frac{1}{y}\Big) \leqslant \exp\Big(-\frac{x^2}{2y}\Big).$$

Goal. Normalize by $<M>_n$ instead of $[M]_n$.

Self-normalized martingales

Definition. We shall say that (M_n) is **conditionally Gaussian** if for all $n \geqslant 1$,

$$\mathcal{L}(\Delta M_n | \mathcal{F}_{n-1}) = \mathcal{N}(0, \Delta{<}M{>}_n)$$

where $\Delta{<}M{>}_n = {<}M{>}_n - {<}M{>}_{n-1}$.

Theorem (Bercu)

If (M_n) is conditionally Gaussian, then for all $x > 0$,

$$\mathbb{P}\Big(\frac{M_n}{{<}M{>}_n} \geqslant x\Big) \leqslant \inf_{p>1} \left(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{2}{<}M{>}_n\Big)\Big]\right)^{1/p}.$$

Outline

Autoregressive process

Consider the autoregressive process given, for all $n \geqslant 0$, by

$$X_{n+1} = \theta X_n + U_n + \varepsilon_{n+1}$$

- $X_n \longrightarrow$ the system output,
- $U_n \longrightarrow$ the adaptive control that can be chosen,
- $\varepsilon_n \longrightarrow$ the driven noise.

We assume that the noise (ε_n) is iid with $\mathcal{N}(0, \sigma^2)$ distribution.
Our goal is to

- Estimate the unknown parameter θ,
- Control the dynamic of the process (X_n).

Autoregressive process

Consider the autoregressive process given, for all $n \geqslant 0$, by

$$X_{n+1} = \theta X_n + U_n + \varepsilon_{n+1}$$

- $X_n \longrightarrow$ the system output,
- $U_n \longrightarrow$ the adaptive control that can be chosen,
- $\varepsilon_n \longrightarrow$ the driven noise.

We assume that the noise (ε_n) is iid with $\mathcal{N}(0, \sigma^2)$ distribution. Our goal is to

- Estimate the unknown parameter θ,
- Control the dynamic of the process (X_n).

Autoregressive process

Consider the autoregressive process given, for all $n \geqslant 0$, by

$$X_{n+1} = \theta X_n + U_n + \varepsilon_{n+1}$$

- $X_n \longrightarrow$ the system output,
- $U_n \longrightarrow$ the adaptive control that can be chosen,
- $\varepsilon_n \longrightarrow$ the driven noise.

We assume that the noise (ε_n) is iid with $\mathcal{N}(0, \sigma^2)$ distribution. Our goal is to

- Estimate the unknown parameter θ,
- Control the dynamic of the process (X_n).

Autoregressive process

Consider the autoregressive process given, for all $n \geqslant 0$, by

$$X_{n+1} = \theta X_n + U_n + \varepsilon_{n+1}$$

- $X_n \longrightarrow$ the system output,
- $U_n \longrightarrow$ the adaptive control that can be chosen,
- $\varepsilon_n \longrightarrow$ the driven noise.

We assume that the noise (ε_n) is iid with $\mathcal{N}(0, \sigma^2)$ distribution. Our goal is to

- Estimate the unknown parameter θ,
- Control the dynamic of the process (X_n).

Autoregressive process

Consider the autoregressive process given, for all $n \geqslant 0$, by

$$X_{n+1} = \theta X_n + U_n + \varepsilon_{n+1}$$

- $X_n \longrightarrow$ the system output,
- $U_n \longrightarrow$ the adaptive control that can be chosen,
- $\varepsilon_n \longrightarrow$ the driven noise.

We assume that the noise (ε_n) is iid with $\mathcal{N}(0, \sigma^2)$ distribution. Our goal is to

- Estimate the unknown parameter θ,
- Control the dynamic of the process (X_n).

Estimation

We estimate the parameter θ by the **least squares** estimator

$$\widehat{\theta}_n = \frac{\sum_{k=1}^n X_{k-1}(X_k - U_{k-1})}{\sum_{k=1}^n X_{k-1}^2}.$$

$$\boldsymbol{\widehat{\theta}_n - \theta = \sigma^2 \frac{M_n}{<M>_n}}$$

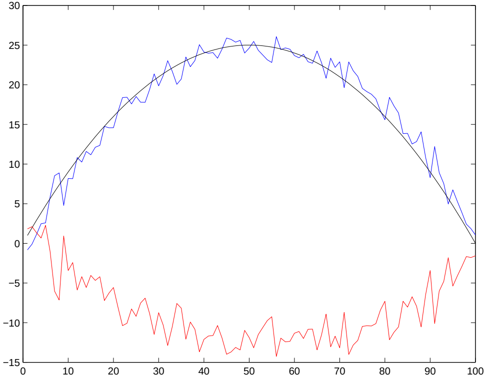
where

$$M_n = \sum_{k=1}^n X_{k-1}\varepsilon_k \qquad \text{and} \qquad <M>_n = \sigma^2 \sum_{k=1}^n X_{k-1}^2.$$

Adaptive Control

The role of U_n is to force X_n to track, step by step, a given reference trajectory (x_n). We use the **adaptive control**

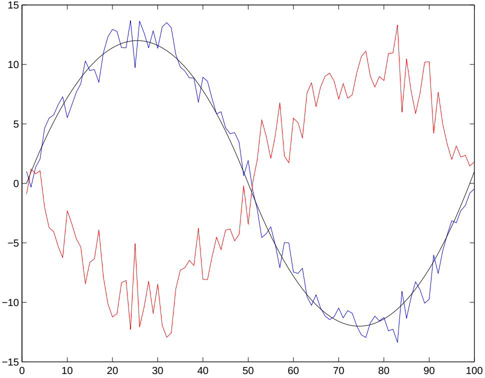
$$U_n = x_{n+1} - \widehat{\theta}_n X_n.$$



Adaptive Control

The role of U_n is to force X_n to track, step by step, a given reference trajectory (x_n). We use the **adaptive control**

$$U_n = x_{n+1} - \widehat{\theta}_n X_n.$$



Exponential Inequality

Theorem (Bercu)

For all $n \geqslant 1$ and $x > 0$, we have

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geqslant x) \leqslant 2\exp\Big(-\frac{nx^2}{2(1+y_x)}\Big)$$

where y_x is the unique positive solution of the equation

$$(1+y)\log(1+y) - y = x^2.$$

Remark. For all $0 < x < 1/2$, $y_x < 2x$ so that

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geqslant x) \leqslant 2\exp\Big(-\frac{nx^2}{2(1+2x)}\Big).$$

Outline

Large Deviations

Consider the stable autoregressive process without control. Let

$$a = \frac{\theta - \sqrt{\theta^2 + 8}}{4} \qquad \text{and} \qquad b = \frac{\theta + \sqrt{\theta^2 + 8}}{4}.$$

Theorem (Bercu-Gamboa-Rouault)

The sequence $(\widehat{\theta}_n)$ satisfies an **LDP** *with rate function*

$$I(x) = \begin{cases} \frac{1}{2} \log \left(\frac{1 + \theta^2 - 2\theta x}{1 - x^2} \right) & \textit{if } \ x \in [a, b], \\ \log |\theta - 2x| & \textit{otherwise.} \end{cases}$$

In particular, for all $x > \theta$, $\displaystyle \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\widehat{\theta}_n \geqslant x) = -I(x).$

Goal. Investigate the **LDP** for $(\widehat{\theta}_n)$ in adpative tracking.

Moderate Deviations

Goal. Let (a_n) be a sequence of positive numbers increasing to infinity such that $a_n = o(n)$. Denote

$$\boldsymbol{V_n = \sqrt{\frac{n}{a_n}}(\widehat{\theta}_n - \theta).}$$

One can conjecture that (V_n) satisfies an **LDP** with speed a_n and rate function $I(x) = x^2/2$.