Concentration inequalities for sums and martingales

Bernard Bercu

Bordeaux University, France

IFCAM Summer School

Bangalore, India, July 2015

Outline

Outline

Hoeffding's inequality

Let $X_1, \ldots, X_n$ be a finite sequence of **independent** random variables. Denote

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

Theorem (Hoeffding's inequality, 1963)

Assume that for all $1 \leqslant k \leqslant n$, $\boldsymbol{a_k \leqslant X_k \leqslant b_k}$ *a.s. for some constants* $a_k < b_k$. *Then, for any positive* x,

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geqslant x) \leqslant 2 \exp\Big(-\frac{2x^2}{D_n}\Big)$$

where

$$D_n = \sum_{k=1}^{n} (b_k - a_k)^2.$$

A keystone lemma

The proof of Hoeffding's inequalitiy relies on the following keystone lemma.

Lemma

Let X be a square integrable random variable with mean zero and variance σ^2. Assume that $\boldsymbol{a \leqslant X \leqslant b}$ a.s. *for some real constants a and b. Then,*

$$\sigma^2 \leqslant -ab \leqslant \frac{(b-a)^2}{4}.$$

In addition, for any real t,

$$\mathbb{E}\big[\exp(tX)\big] \leqslant \exp\Big(\frac{t^2}{8}(b-a)^2\Big).$$

Proof.

The convexity of the square function implies that $X^2 \leqslant (a+b)X - ab$ a.s. By taking the expectation on both sides,

$$\sigma^2 = \mathbb{E}[X^2] \leqslant -ab \leqslant \frac{(b-a)^2}{4}.$$

The convexity of the exponential function also implies that for all $t \in \mathbb{R}$,

$$\boldsymbol{\exp(tX) \leqslant \frac{\big(\exp(tb) - \exp(ta)\big)X}{b-a} + \frac{b\exp(ta) - a\exp(tb)}{b-a}} \quad \text{a.s.}$$

By taking the expectation on both sides,

$$\begin{aligned}\mathbb{E}[\exp(tX)] &\leqslant \frac{b}{b-a}\exp(ta) - \frac{a}{b-a}\exp(tb),\\ &\leqslant (1-p)\exp(-py) + p\exp((1-p)y)\end{aligned}$$

where $p = -a/(b-a)$ and $y = (b-a)t$. □

Proof.

One can observe that $0 < p < 1$ as $a < 0 < b$. Therefore, for all $t \in \mathbb{R}$,

$$\mathbb{E}[\exp(tX)] \leqslant \exp(h(y))$$

where $h(y) = -py + \log(1 - p + p\exp(y))$. Furthermore, it follows from straighforward calculation that

$$\begin{aligned} h'(y) &= -p + \frac{p}{p + (1-p)\exp(-y)}, \\ h''(y) &= \frac{p(1-p)\exp(-y)}{(p + (1-p)\exp(-y))^2} \leqslant \frac{1}{4}. \end{aligned}$$

As $h(0) = 0$ and $h'(0) = 0$, Taylor's formula implies that for all $y \in \mathbb{R}$

$$h(y) \leqslant \frac{y^2}{8} = \frac{t^2}{8}(b-a)^2,$$

which completes the proof of the lemma. □

Proof of Hoeffding's inequality

Proof.

It follows from Markov's inequality that for any positive x and t,

$$\begin{aligned}
\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) &= \mathbb{P}(\exp(t(S_n - \mathbb{E}[S_n]) \geqslant \exp(tx)), \\
&\leqslant \exp(-tx)\mathbb{E}\Big[\exp(t(S_n - \mathbb{E}[S_n]))\Big], \\
&\leqslant \exp(-tx)\mathbb{E}\Big[\exp(t\sum_{k=1}^{n} Y_k)\Big]
\end{aligned}$$

where $Y_k = X_k - \mathbb{E}[X_k]$. One can observe that (Y_n) is a sequence of independent random variables such that, for all $1 \leqslant k \leqslant n$,

$$c_k \leqslant Y_k \leqslant d_k \qquad \text{a.s}$$

where $c_k = a_k - \mathbb{E}[X_k]$ and $d_k = b_k - \mathbb{E}[X_k]$, $d_k - c_k = b_k - a_k$. □

Proof.

Hence, we deduce from the above lemma that

$$\mathbb{E}\Big[\exp(t\sum_{k=1}^{n} Y_k)\Big] = \prod_{k=1}^{n} \mathbb{E}\Big[\exp(tY_k)\Big] \leqslant \exp\Big(\frac{t^2}{8}\sum_{k=1}^{n}(b_k - a_k)^2\Big).$$

Consequently, for any positive x and t,

$$\begin{aligned}\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) &\leqslant \exp\Big(-tx + \frac{t^2}{8}D_n\Big),\\ &\leqslant \exp\left(-\frac{2x^2}{D_n}\right)\end{aligned}$$

by taking the optimal value $t = 4x/D_n$. Replacing X_k by $-X_k$, we obtain by the same token that, for any positive x,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \leqslant -x) \leqslant \exp\left(-\frac{2x^2}{D_n}\right)$$

which completes the proof of Hoeffding's inequality. □

Improvement of Hoeffding's inequality

Theorem (B-Delyon-Rio, 2015)

Assume that for all $1 \leqslant k \leqslant n$, $a_k \leqslant X_k \leqslant b_k$ *a.s. for some constants* $a_k < b_k$. *Then, for any positive* x,

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geqslant x) \leqslant 2 \exp\left(-\frac{3x^2}{D_n + 2V_n}\right)$$

where

$$D_n = \sum_{k=1}^{n} (b_k - a_k)^2 \qquad \textit{and} \qquad V_n = Var(S_n).$$

$\longrightarrow$ One can observe that $D_n \geqslant 4V_n$ which means that this result improves Hoeffding's inequality.

A second keystone lemma

Lemma

*Let X be a square integrable random variable with mean zero and variance σ^2 such that $\sigma^2 \leqslant v$. Assume that **X** $\leqslant$ **b** a.s. for some positive constant b. Then, for any positive t,*

$$\mathbb{E}[\exp(tX)] \leqslant p\exp\big(s(1-p)\big) + (1-p)\exp(-sp),$$

$$\leqslant \exp\left(\frac{(1-2p)s^2}{4\log((1-p)/p)}\right)$$

where

$$p = \frac{v}{b^2+v} \qquad \text{and} \qquad s = \frac{tv}{bp}.$$

$\longrightarrow$ In the special case $|X| \leqslant b$ a.s., we clearly have $\sigma^2 \leqslant b^2$, $v = b^2$,

$$p = \frac{1}{2} \qquad \text{and} \qquad s = 2tb.$$

A second keystone lemma

Moreover, it follows from **L'Hospital's rule** that

$$\lim_{p \to 1/2} \frac{(1-2p)}{\log((1-p)/p)} = \frac{1}{2}.$$

The convexity of the exponential function implies that for all $t \in \mathbb{R}$,

$$\exp(tX) \leqslant \frac{1}{2}\sinh(tb)X + \cosh(tb).$$

By taking the expectation on both sides, we obtain that for all $t \in \mathbb{R}$,

$$\mathbb{E}[\exp(tX)] \leqslant \cosh(tb) \leqslant \exp\Big(\frac{t^2b^2}{2}\Big)$$

which is exactly the second inequality of the lemma.

Proof of the second keystone lemma

Proof.

Using **integration by parts**, we can prove that for any positive t,

$$\mathbb{E}[\exp(tX)] \leqslant \mathbb{E}[\exp(tZ)]$$

where Z is a two-value random variable with mean zero and variance v

$$Z = \begin{cases} b & p \\ a & 1-p \end{cases}$$

where

$$p = \frac{v}{b^2 + v} \qquad \text{and} \qquad a = -\frac{v}{b}.$$

□

Proof of the second keystone lemma, continued

Proof.

It is not hard to see that

$$\begin{aligned} \mathbb{E}[\exp(tZ)] &\leqslant p\exp(tb) + (1-p)\exp(ta) \\ &= p\exp\big(s(1-p)\big) + (1-p)\exp(-sp) \end{aligned}$$

where

$$s = \frac{tv}{bp}.$$

We can show via the **minimax theorem** that for any positive s,

$$p\exp\big(s(1-p)\big) + (1-p)\exp(-sp) \leqslant \exp\left(\frac{(1-2p)s^2}{4\log((1-p)/p)}\right)$$

which completes the proof of the lemma. □

Proof of the improvement of Hoeffding's inequality

Proof.

We already saw from Markov's inequality that for any positive x and t,

$$\begin{aligned}
\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) &= \mathbb{P}(\exp(t(S_n - \mathbb{E}[S_n]) \geqslant \exp(tx)), \\
&\leqslant \exp(-tx)\mathbb{E}\Big[\exp(t(S_n - \mathbb{E}[S_n]))\Big], \\
&\leqslant \exp(-tx)\mathbb{E}\Big[\exp(t\sum_{k=1}^{n} Y_k)\Big]
\end{aligned}$$

where $Y_k = X_k - \mathbb{E}[X_k]$,

$$c_k \leqslant Y_k \leqslant d_k \qquad \text{a.s}$$

with $c_k = a_k - \mathbb{E}[X_k]$ and $d_k = b_k - \mathbb{E}[X_k]$, $d_k - c_k = b_k - a_k$. □

Proof.

For all $1 \leqslant k \leqslant n$, let $v_k = \text{Var}(Y_k) \leqslant -c_k d_k$. It follows from the above lemma that for any positive t,

$$\mathbb{E}[\exp(tY_k)] \leqslant \exp\Big(\frac{t^2 d_k^2}{4}\varphi\Big(\frac{v_k}{d_k^2}\Big)\Big)$$

where

$$\varphi(v) = \frac{v^2 - 1}{\log v}.$$

It is not hard to see that, for any positive v,

$$\varphi(v) = \frac{1}{3}(1 + 4v + v^2).$$

Consequently,

$$\mathbb{E}\Big[\exp(t\sum_{k=1}^{n} Y_k)\Big] = \prod_{k=1}^{n} \mathbb{E}\Big[\exp(tY_k)\Big] \leqslant \exp\Big(\frac{t^2}{12}\sum_{k=1}^{n} d_k^2\Big(1 + 4\frac{v_k}{d_k^2} + \frac{v_k^2}{d_k^4}\Big)\Big).$$

Proof of the improvement of Hoeffding's inequality

Proof.

Hence, as $v_k \leqslant -c_k d_k$ and $d_k - c_k = b_k - a_k$, we obtain that

$$
\begin{aligned}
\mathbb{E}\Big[\exp(t\sum_{k=1}^{n} Y_k)\Big] &\leqslant \exp\Big(\frac{t^2}{12}\sum_{k=1}^{n}\Big(d_k^2 + 4v_k + \frac{v_k^2}{d_k^2}\Big)\Big), \\
&\leqslant \exp\Big(\frac{t^2}{12}\sum_{k=1}^{n}(d_k^2 + 4v_k + c_k^2)\Big), \\
&\leqslant \exp\Big(\frac{t^2}{12}\sum_{k=1}^{n}(d_k^2 + 2v_k - 2c_k d_k + c_k^2)\Big), \\
&\leqslant \exp\Big(\frac{t^2}{12}\sum_{k=1}^{n}((d_k - c_k)^2 + 2v_k)\Big), \\
&\leqslant \exp\Big(\frac{t^2}{12}\sum_{k=1}^{n}((b_k - a_k)^2 + 2v_k)\Big).
\end{aligned}
$$

Proof of the improvement of Hoeffding's inequality

Proof.

Consequently, for any positive x and t,

$$\begin{aligned}\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) &\leqslant \exp\Big(-tx + \frac{t^2}{12}(D_n + 2V_n)\Big),\\ &\leqslant \exp\left(-\frac{3x^2}{D_n + 2V_n}\right)\end{aligned}$$

by taking the optimal value $t = 6x/(D_n + 2V_n)$. Replacing X_k by $-X_k$, we obtain by the same token that, for any positive x,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \leqslant -x) \leqslant \exp\left(-\frac{3x^2}{D_n + 2V_n}\right)$$

which completes the proof. □

Bernstein's inequality

Let $X_1, \ldots, X_n$ be a finite sequence of centered and **independent** random variables. Denote

$$S_n = \sum_{k=1}^{n} X_k, \qquad V_n = \mathrm{Var}(S_n), \qquad v_n = \frac{V_n}{n}.$$

We shall say that $X_1, \ldots, X_n$ satisfy **Bernstein's condition** if it exists some positive constant c such that, for any integer $p \geqslant 3$,

$$\sum_{k=1}^{n} \mathbb{E}\big[(\max(0, X_k))^p\big] \leqslant \frac{p! c^{p-2}}{2} V_n.$$

Theorem (Bernstein's inequality)

Under **Bernstein's condition**, *we have for any positive x,*

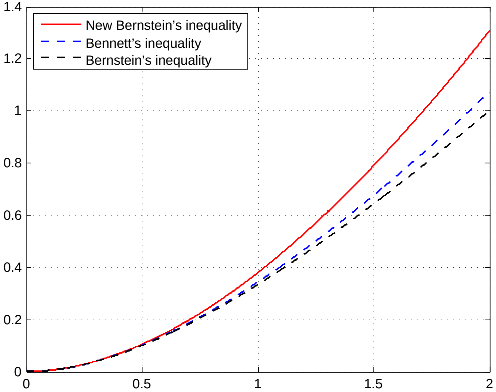
$$\mathbb{P}(S_n \geqslant nx) \leqslant \Big(1 + \frac{x^2}{2(v_n + cx)}\Big)^n \exp\Big(-\frac{nx^2}{v_n + cx}\Big)$$
$$\leqslant \exp\Big(-\frac{nx^2}{2(v_n + cx)}\Big).$$

In addition, we also have for any positive x,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp\Big(-\frac{nx^2}{v_n + cx + \sqrt{v_n(v_n + 2cx)}}\Big).$$

$\longrightarrow$ The last inequality is due to Bennett while the second inequality in blue is known as Bernstein's inequality.

Comparisons in Bernstein's inequalities



Proof of Bernstein's inequalities

Proof.

It follows from Markov's inequality that for any positive x and t,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp(-ntx)\mathbb{E}\big[\exp(tS_n)\big].$$

The **concavity of the logarithm function** implies that

$$\mathbb{E}\big[\exp(tS_n)\big] \leqslant \exp\big(n\ell(t)\big) \quad \text{where} \quad \ell(t) = \log\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\exp(tX_k)]\Big).$$

However, it is not hard to see that for any real x,

$$\exp(x) \leqslant 1 + x + \frac{x^2}{2} + \sum_{p=3}^{\infty}\frac{(\max(0,x))^p}{p!}.$$

□

Proof.

Hence, it follows from the **monotone convergence theorem** that for all $1 \leqslant k \leqslant n$ and for any positive t,

$$\mathbb{E}[\exp(tX_k)] \leqslant 1 + t\mathbb{E}[X_k] + \frac{t^2\mathbb{E}[X_k^2]}{2} + \sum_{p=3}^{\infty} \frac{t^p\mathbb{E}[(\max(0, X_k))^p]}{p!}.$$

Consequently, we deduce from Bernstein's condition that

$$\sum_{k=1}^{n} \mathbb{E}[\exp(tX_k)] \leqslant n + \frac{t^2}{2} V_n + \frac{V_n}{2} \sum_{p=3}^{\infty} c^{p-2} t^p = n + \frac{V_n}{2} \sum_{p=2}^{\infty} c^{p-2} t^p.$$

Therefore, as soon as $0 < tc < 1$,

$$\begin{aligned} \exp(\ell(t)) &= \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[\exp(tX_k)] \leqslant 1 + \frac{v_n t^2}{2} \sum_{p=0}^{\infty} (tc)^p, \\ &\leqslant 1 + \frac{v_n t^2}{2(1 - tc)}. \end{aligned}$$

Proof of Bernstein's inequalities

Proof.

It leads to

$$\begin{aligned}\mathbb{P}(S_n \geqslant nx) &\leqslant \exp\Big(-ntx + n\log\Big(1 + \frac{v_n t^2}{2(1-tc)}\Big)\Big),\\ &\leqslant \exp\Big(-\frac{nx^2}{v_n + cx}\Big)\Big(1 + \frac{x^2}{2(v_n + cx)}\Big)^n\end{aligned}$$

by taking the optimal value

$$t = \frac{x}{v_n + cx}.$$

Finally, the elementary inequality $\mathbf{1 + x \leqslant \exp(x)}$ where x is positive, ensures that

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp\Big(-\frac{nx^2}{2(v_n + cx)}\Big),$$

which completes the proofs of Bernstein's inequalities. □

Outline

1. Concentration inequalities for sums
   - Hoeffding's inequality
   - Bernstein's inequality
2. Concentration inequalities for martingales
   - Azuma-Hoeffding's inequality
   - Bernstein's inequality
   - De la Peña's inequalities
   - Two-sided exponential inequalities
   - One-sided exponential inequalities
3. Statistical applications
   - Autoregressive process
   - Branching process
   - Random permutations

Azuma-Hoeffding's inequality

Let (M_n) be a square integrable martingale adapted to $\mathbb{F} = (\mathcal{F}_n)$ with $M_0 = 0$. Its **increasing process** is defined by

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

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

Theorem (Azuma-Hoeffding's inequality, 1967)

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

$$\mathbb{P}(|M_n| \geqslant x) \leqslant 2 \exp\left(-\frac{2x^2}{D_n}\right)$$

where

$$D_n = \sum_{k=1}^{n} (b_k - a_k)^2.$$

Azuma-Hoeffding's inequality

Theorem (B-Delyon-Rio, 2015)

Assume that for all $1 \leqslant k \leqslant n$,

$$A_k \leqslant \Delta M_k \leqslant B_k \qquad a.s.$$

where (A_k, B_k) *is a couple of bounded and* $\mathcal{F}_{k-1}$*-measurable random variables. Then, for any positive* x *and* y,

$$\mathbb{P}(M_n \geqslant x, 2 < M >_n + \mathcal{D}_n \leqslant y) \leqslant \exp\Big(-\frac{3x^2}{y}\Big)$$

where

$$\mathcal{D}_n = \sum_{k=1}^{n} (B_k - A_k)^2.$$

Van de Geer's inequality

The convexity of the square function implies that almost surely

$$\Delta M_k^2 \leqslant (A_k + B_k)\Delta M_k - A_k B_k \leqslant (A_k + B_k)\Delta M_k + \frac{1}{4}(B_k - A_k)^2.$$

By taking the conditional expectation on both sides,

$$< M >_n = \sum_{k=1}^{n} \mathbb{E}[\Delta M_k^2 | \mathcal{F}_{k-1}] \leqslant \frac{1}{4} \sum_{k=1}^{n} (B_k - A_k)^2 = \frac{1}{4}\mathcal{D}_n.$$

Consequently, we can deduce Van de Geer's inequality which says that, for any positive x and y,

$$\mathbb{P}(M_n \geqslant x, \mathcal{D}_n \leqslant y) \leqslant \exp\Big(-\frac{2x^2}{y}\Big).$$

Theorem (Bernstein's inequality)

Assume that it exists some positive constant c such that, for any integer $p \geqslant 3$ and for all $1 \leqslant k \leqslant n$,

$$\mathbb{E}\Big[(\max(0, \Delta M_k))^p | \mathcal{F}_{k-1}\Big] \leqslant \frac{p! c^{p-2}}{2} \Delta < M >_k \qquad a.s.$$

Then, for any positive x and y,

$$\mathbb{P}(M_n \geqslant nx, < M >_n \leqslant ny) \leqslant \Big(1 + \frac{x^2}{2(y+cx)}\Big)^n \exp\Big(-\frac{nx^2}{y+cx}\Big)$$

$$\leqslant \exp\Big(-\frac{nx^2}{2(y+cx)}\Big).$$

In addition, we also have for any positive x and y,

$$\mathbb{P}(M_n \geqslant nx, < M >_n \leqslant ny) \leqslant \exp\Big(-\frac{nx^2}{y + cx + \sqrt{y(y+2cx)}}\Big).$$

De la Peña's inequalities

Definition. We 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, 1999)

If (M_n) is conditionally symmetric, then for any positive x and y,

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

where

$$[M]_n = \sum_{k=1}^{n} \Delta M_k^2.$$

Self-normalized martingales

Theorem (De la Peña, 1999)

If (M_n) is conditionally symmetric, then for any positive x and y, and for all $a \geqslant 0$ and $b > 0$,

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

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

Goal. Self-normalized by $< M >_n$ instead of $[M]_n$. In addition, avoid the symmetric condition on the distribution of M_n.

Self-normalized martingales

Theorem (De la Peña, 1999)

If (M_n) is conditionally symmetric, then for any positive x and y, and for all $a \geqslant 0$ and $b > 0$,

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

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

Goal. Self-normalized by $< M >_n$ instead of $[M]_n$. In addition, avoid the symmetric condition on the distribution of M_n.

Theorem (B-Touati, 2008)

For any positive x and y, we always have

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

Theorem (Delyon, 2009)

For any positive x and y, we always have

$$\mathbb{P}(|M_n| \geqslant x, [M]_n + 2<M>_n \leqslant y) \leqslant 2\exp\Big(-\frac{3x^2}{2y}\Big).$$

Remark. For any positive x and y,

$$\begin{aligned}\mathbb{P}\Big(M_n \geqslant x, [M]_n + <M>_n \leqslant y\Big) &\leqslant \mathbb{P}\Big(M_n \geqslant x, [M]_n + 2<M>_n \leqslant 2y\Big),\\ &\leqslant \exp\Big(-\frac{3x^2}{4y}\Big) \leqslant \exp\Big(-\frac{x^2}{2y}\Big),\end{aligned}$$

which means that Delyon's inequality improves the previous one.

Two elementary inequalities

The proof of the first result relies on the fact that for any real x,

$$f(x) = \exp\Big(x - \frac{x^2}{2}\Big) \leqslant g(x) = 1 + x + \frac{x^2}{2},$$

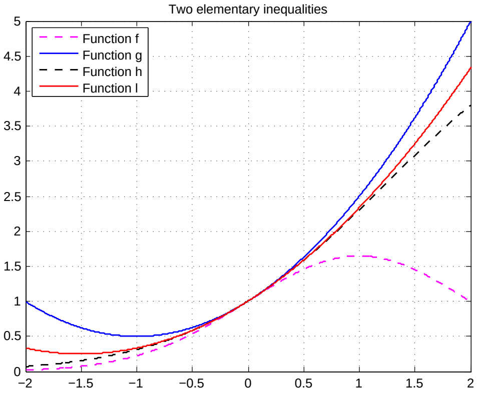
whereas that of the second one is based on the fact that for any real x,

$$h(x) = \exp\Big(x - \frac{x^2}{6}\Big) \leqslant \ell(x) = 1 + x + \frac{x^2}{3}.$$

One can observe that for any real x,

$$f(x) \leqslant h(x) \leqslant \ell(x) \leqslant g(x).$$

Two elementary inequalities



Two keystone lemma

Lemma

Let X be a square integrable random variable with mean zero and variance σ^2. Then, for any real t,

$$\boldsymbol{L(t) = \mathbb{E}\left[\exp\left(tX - \frac{t^2}{6}X^2\right)\right] \leqslant 1 + \frac{t^2}{3}\sigma^2.}$$

Lemma

For any real t and for all $n \geqslant 0$, denote

$$V_n(t) = \exp\left(tM_n - \frac{t^2}{6}[M]_n - \frac{t^2}{3} < M >_n\right).$$

Then, $(V_n(t))$ is a positive supermartingale such that $\boldsymbol{\mathbb{E}[V_n(t)] \leqslant 1}$.

Heavy on left or right

Definition. Let X be a centered random variable on $(\Omega, \mathcal{A}, \mathbb{P})$.

- X is heavy on left if, for any positive a, $\mathbb{E}[T_a(X)] \leqslant 0$,
- X is heavy on right if, for any positive a, $\mathbb{E}[T_a(X)] \geqslant 0$.

where

$$T_a(x) = \begin{cases} a & \text{if} \quad x \geqslant a, \\ x & \text{if} \quad -a \leqslant x \leqslant a, \\ -a & \text{if} \quad x \leqslant -a. \end{cases}$$

X is symmetric $\Longleftrightarrow$ X is heavy on left and on right.

Heavy on left or right

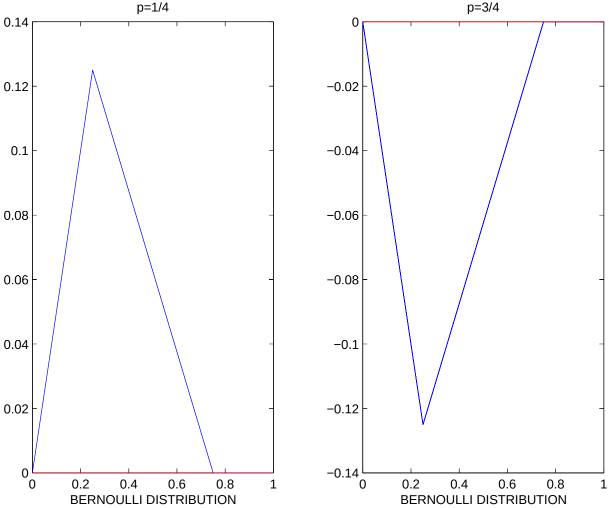
Denote by F the distribution function of X and

$$H(a) = \int_0^a F(-x) - (1 - F(x))\, dx = -\mathbb{E}[T_a(X)].$$

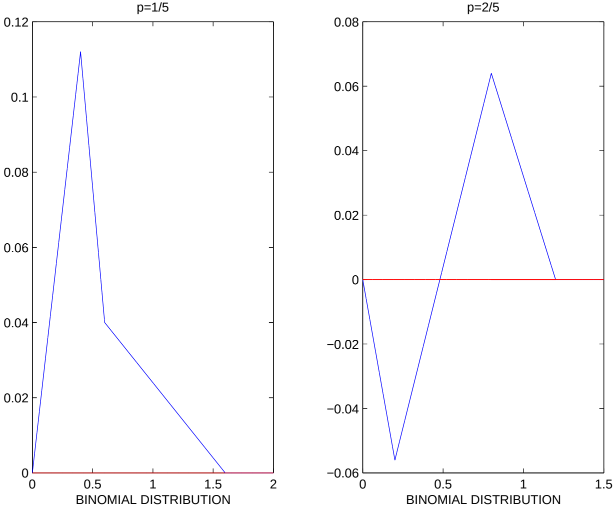
- X is heavy on left if, for any positive a, $H(a) \geqslant 0$,
- X is heavy on right if, for any positive a, $H(a) \leqslant 0$.

X is symmetric $\Longleftrightarrow$ For any positive a, $H(a) = 0$.

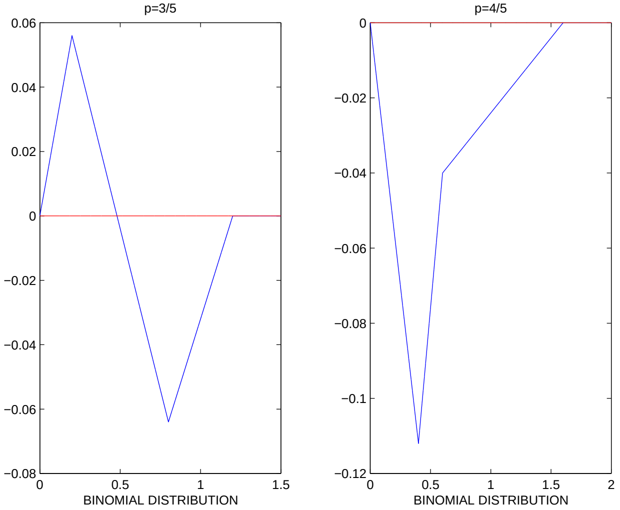
Centered Bernoulli $\mathcal{B}(p)$



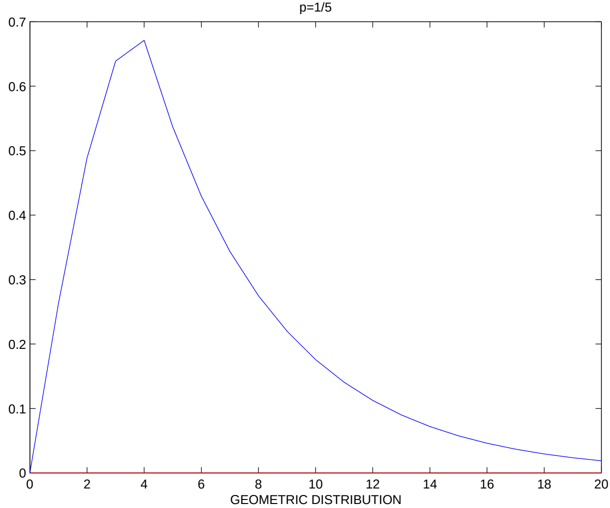
Centered Binomial $\mathcal{B}(2, p)$



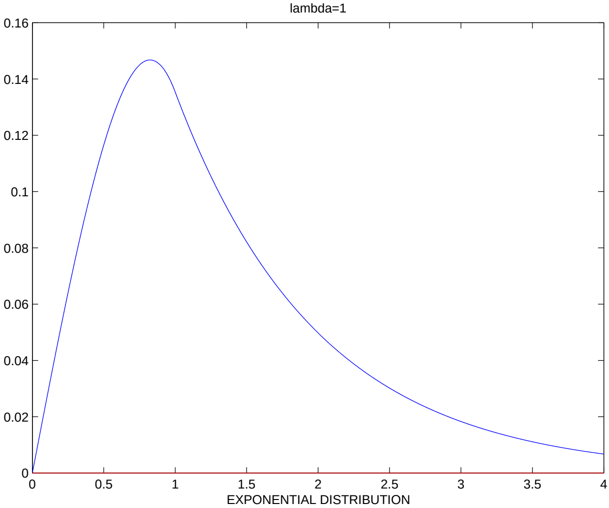
Centered Binomial $\mathcal{B}(2, p)$



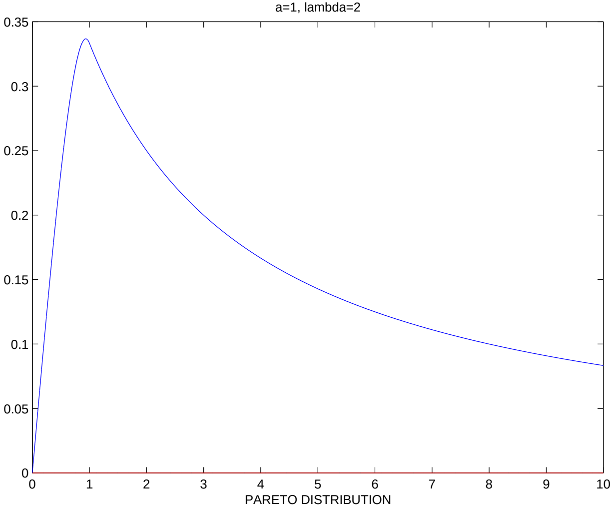
Centered Geometric $\mathcal{G}(p)$



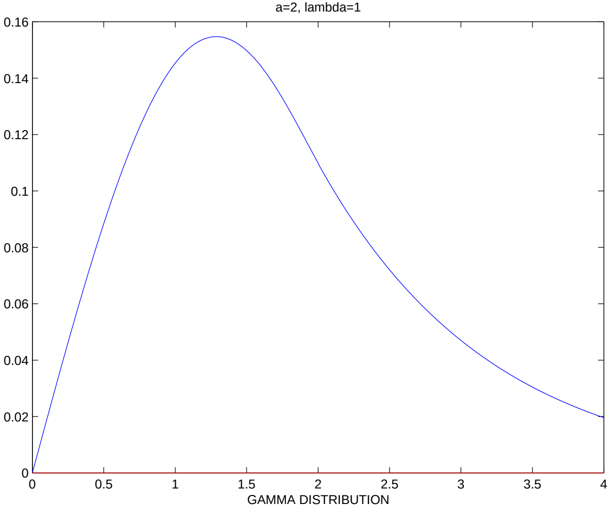
Centered Exponential $\mathcal{E}(\lambda)$



Centered Pareto $\mathcal{P}(a, \lambda)$



Centered Gamma $\mathcal{G}(a, \lambda)$



Martingales heavy on left or right

Definition. We say that (M_n) is **conditionally heavy on left** if, for all $n \geqslant 1$ and for any positive a,

$$\mathbb{E}[T_a(\Delta M_n)|\mathcal{F}_{n-1}] \leqslant 0 \qquad \text{a.s.}$$

Theorem (B-Touati, 2008)

If (M_n) is conditionally heavy on left, then for any positive x and y,

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

$\longrightarrow$ De la Peña's inequality holds true without the assumption that (M_n) is conditionally symmetric.

Self-normalized martingales

Theorem (B-Touati, 2008)

If (M_n) is conditionally heavy on left, then for any positive x and y, and for all $a \geqslant 0$ and $b > 0$,

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

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

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

Self-normalized martingales

Theorem (B-Touati, 2008)

If (M_n) is conditionally heavy on left, then for any positive x and y, and for all $a \geqslant 0$ and $b > 0$,

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

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

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

Self-normalized martingales

Theorem (B-Touati, 2008)

If (M_n) is conditionally heavy on left, then for any positive x and y, and for all $a \geqslant 0$ and $b > 0$,

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

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

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

Two keystone lemmas

Lemma

For a random variable X and for any real t, let

$$L(t) = \mathbb{E}\Big[\exp\Big(tX - \frac{t^2}{2}X^2\Big)\Big].$$

- *X is heavy on left $\Longrightarrow$ For any positive t, $L(t) \leqslant 1$,*
- *X is heavy on right $\Longrightarrow$ For any negative t, $L(t) \leqslant 1$,*
- *X is symmetric $\Longrightarrow$ For any real t, $L(t) \leqslant 1$.*

Lemma

For any real t and for all $n \geqslant 0$, denote

$$W_n(t) = \exp\Big(tM_n - \frac{t^2}{2}[M]_n\Big).$$

Then, $(W_n(t))$ is a positive supermartingale such that $\mathbb{E}[W_n(t)] \leqslant 1.$

Outline

Autoregressive process

Consider the stable autoregressive process

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}, \qquad |\theta| < 1$$

where (ε_n) is **iid** $\mathcal{N}(0, \sigma^2)$ with positive variance σ^2 and the initial state X_0 is independent of (ε_n) with $\mathcal{N}(0, \sigma^2/(1-\theta^2))$ distribution. Denote by $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ the **least squares** and the **Yule-Walker** estimators of θ

$$\widehat{\theta}_n = \frac{\sum_{k=1}^{n} X_k X_{k-1}}{\sum_{k=1}^{n} X_{k-1}^2} \qquad \text{and} \qquad \widetilde{\theta}_n = \frac{\sum_{k=1}^{n} X_k X_{k-1}}{\sum_{k=0}^{n} X_k^2}.$$

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

Theorem (B-Gamboa-Rouault, 1997)

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

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

- $(\widetilde{\theta}_n)$ *satisfies an* **LDP** *with rate function*

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

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

Theorem (B-Gamboa-Rouault, 1997)

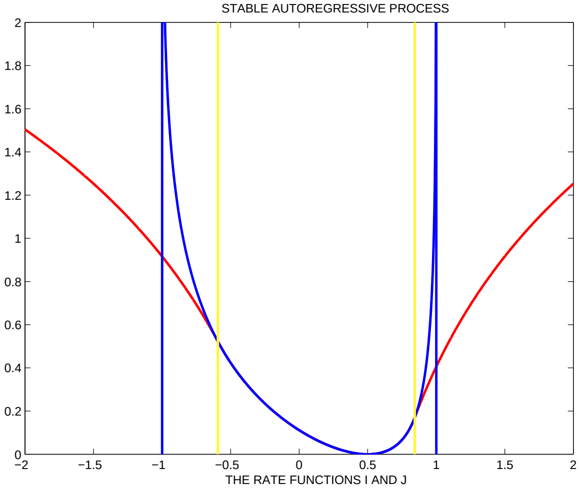
- $(\widehat{\theta}_n)$ *satisfies an* **LDP** *with rate function*

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

- $(\widetilde{\theta}_n)$ *satisfies an* **LDP** *with rate function*

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

Least squares and Yule-Walker



Corollary (B-Touati, 2008)

For all $n \geqslant 1$ and for any positive x,

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

where y_x is the unique positive solution of

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

$\longrightarrow$ For any $0 < x < 1/2$, we have $y_x < 2x$, which implies that

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

Branching process

Consider the Galton-Watson process starting from $X_0 = 1$

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k}$$

where $(Y_{n,k})$ is **iid** taking values in $\mathbb{N}$, with finite mean $m > 1$ and positive variance σ^2. We assume that the set of extinction of (X_n) is negligeable. Let $\widetilde{m}_n$ and $\widehat{m}_n$ be the **Lotka-Nagaev** and the **Harris** estimators of m

$$\widetilde{m}_n = \frac{X_n}{X_{n-1}} \qquad \text{and} \qquad \widehat{m}_n = \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n X_{k-1}}.$$

Let L be the cumulant generating function

$$L(t) = \log \mathbb{E}\Big[\exp(t(Y_{n,k} - m))\Big]$$

and denote by I its Cramér transform

$$I(x) = \sup_{-c \leqslant t \leqslant c} \Big(xt - L(t)\Big).$$

Corollary (B-Touati, 2008)

Assume that L is finite on $[-c, c]$ with $c > 0$. Then, for all $n \geqslant 1$ and for any positive x, if $J(x) = \min(I(x), I(-x))$,

$$\mathbb{P}(|\widetilde{m}_n - m| \geqslant x) \leqslant 2\mathbb{E}\Big[\exp(-J(x)X_{n-1})\Big],$$

$$\mathbb{P}(|\widetilde{m}_n - m| \geqslant x) \leqslant 2\inf_{p>1} \Big(\mathbb{E}\Big[\exp(-(p-1)J(x)X_{n-1})\Big]\Big)^{1/p}.$$

Corollary (B-Touati, 2008)

Assume that L is finite on $[-c, c]$ with $c > 0$. Then, for all $n \geqslant 1$ and for any positive x,

$$\mathbb{P}(|\widehat{m}_n - m| \geqslant x) \leqslant 2 \inf_{p>1} \left(\mathbb{E}\left[\exp(-(p-1)J(x)S_{n-1})\right]\right)^{1/p}$$

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

$\longrightarrow$ If the offspring distribution is Geometric $\mathcal{G}(p)$

$$\mathbb{P}(|\widetilde{m}_n - m| \geqslant x) \leqslant \frac{2p^n \exp(-J(x))}{p(1-\exp(-J(x)))}.$$

Random permutations

Let $(a_n(i,j))$ be an $n \times n$ array of real numbers from $[-m_a, m_a]$ where $m_a > 0$. Let π_n be chosen uniformly at random from the set of all permutations of $\{1, \ldots, n\}$. Denote

$$\boldsymbol{S_n = \sum_{i=1}^{n} a_n(i, \pi_n(i)).}$$

We clearly have

$$\mathbb{E}[S_n] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_n(i,j) \quad \text{and} \quad \mathrm{Var}(S_n) = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n} d_n^2(i,j)$$

where $d_n(i,j) = a_n(i,j) - a_n(i,*) - a_n(*,j) + a_n(*,*)$. In addition, under standard conditions,

$$\boldsymbol{\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).}$$

Random permutations

Theorem (Delyon, 2015)

For any positive x,

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geqslant x) \leqslant 4\exp\Big(-\frac{x^2}{16(\theta v_n + x m_a/3)}\Big)$$

where θ is an explicit constant and

$$v_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_n^2(i,j).$$

⟶ It was proven by Chatterjee that for any positive x,

$$\mathbb{P}(|S_n - E[S_n]| \geqslant x) \leqslant 2\exp\Big(-\frac{x^2}{4m_a\mathbb{E}[S_n] + 2x m_a}\Big).$$

This upper bound has better constants but v_n is replaced with $m_a\mathbb{E}[S_n]$.