Exponential inequalities for self-normalized martingales with applications

Bernard Bercu, Abderrahmen Touati

EXPONENTIAL INEQUALITIES FOR SELF-NORMALIZED MARTINGALES WITH APPLICATIONS

BERNARD BERCU AND ABDERRAHMEN TOUATI

Abstract. We propose several exponential inequalities for self-normalized martingales similar to those established by De la Peña. The keystone is the introduction of a new notion of random variable heavy on left or right. Applications associated with linear regressions, autoregressive and branching processes are also provided.

1. Introduction

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

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

where $\Delta M_n = M_n - M_{n-1}$. The celebrated Azuma-Hoeffding's inequality [4] is as follows.

Theorem 1.1 (Azuma-Hoeffding's inequality)**.** *Let (M_n) be a locally square integrable real martingale such that, for each $1 \leq k \leq n$, $a_k \leq \Delta M_k \leq b_k$ a.s. for some constants $a_k < b_k$. Then, for all $x \geq 0$,*

$$\mathbb{P}(|M_n| \geq x) \leq 2 \exp\Big(-\frac{2x^2}{\sum_{k=1}^{n} (b_k - a_k)^2}\Big). \tag{1.1}$$

Another result which involves the predictable quadratic variation ($<M>_n$) is the so-called Freedman's inequality [12].

Theorem 1.2 (Freedman's inequality)**.** *Let (M_n) be a locally square integrable real martingale such that, for each $1 \leq k \leq n$, $|\Delta M_k| \leq c$ a.s. for some constant $c > 0$. Then, for all $x, y > 0$,*

$$\mathbb{P}(M_n \geq x, <M>_n \leq y) \leq \exp\Big(-\frac{x^2}{2(y + cx)}\Big). \tag{1.2}$$

Over the last decade, extensive study has been made to establish exponential inequalities for (M_n) relaxing the boundedness assumption on its increments. On

1991 *Mathematics Subject Classification.* Primary:60E15; 60G42; Secondary: 60G15; 60J80.
Key words and phrases. Exponential inequalities; Martingales; Autoregressive processes; Branching processes.

the one hand, under the standard Bernstein's condition that for $n \geq 1$, $p \geq 2$ and for some constant $c > 0$,

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

De la Peña [8] recovers (1.2). Van de Geer [11] also proves (1.2) replacing $<M>_n$ by a suitable increasing process. On the other hand, if (M_n) is conditionally symmetric which means that for $n \geq 1$, the conditional distribution of ΔM_n given $\mathcal{F}_{n-1}$ is symmetric, then De la Peña [8] establishes the nice following result.

Theorem 1.3 (De la Peña's inequality)**.** *Let (M_n) be a locally square integrable and conditionally symmetric real martingale. Then, for all $x, y > 0$,*

$$\mathbb{P}(M_n \geq x, [M]_n \leq y) \leq \exp\Big(-\frac{x^2}{2y}\Big). \tag{1.3}$$

Some recent extensions of the above inequalities in a more general framework including discrete-time martingales can also be found in [9], [10] where the conditionally symmetric assumption is still required for (1.3). By a careful reading of [8], one can see that (1.3) is a two-sided exponential inequality. More precisely, if (M_n) is conditionally symmetric then, for all $x, y > 0$,

$$\mathbb{P}(|M_n| \geq x, [M]_n \leq y) \leq 2 \exp\Big(-\frac{x^2}{2y}\Big). \tag{1.4}$$

By comparing (1.4) and (1.1), we are only halfway to Azuma-Hoeffding's inequality which holds without the total quadratic variation $[M]_n$.

The purpose of this paper is to establish several exponential inequalities in the spirit of the original work of De la Peña [8]. In Section 2, we shall propose two-sided exponential inequalities involving $<M>_n$ as well as $[M]_n$ without any assumption on the martingale (M_n). Section 3 is devoted to the introduction of a new concept of random variables heavy on left or right. This notion is really useful if one is only interested in obtaining a one-sided exponential inequality for (M_n). It also provides a clearer understanding of De la Peña's conditional symmetric assumption. We shall show in Section 4 that this new concept allows us to prove (1.3). As in [8], we shall also propose exponential inequalities for (M_n) self-normalized by $[M]_n$ or $<M>_n$. Section 5 is devoted to applications on linear regressions, autoregressive and branching processes. All technical proofs are postponed in the appendices.

2. Two-sided exponential inequalities

This section is devoted to two-sided exponential inequalities involving both $<M>_n$ and $[M]_n$. We start with the following basic lemma.

Lemma 2.1. Let X be a square integrable random variable with mean zero and variance $\sigma^2 > 0$. For all $t \in \mathbb{R}$, denote

$$L(t) = \mathbb{E}\Big[\exp\Big(tX - \frac{t^2}{2} X^2\Big)\Big]. \tag{2.1}$$

Then, we have for all $t \in \mathbb{R}$,

$$L(t) \leq 1 + \frac{t^2}{2} \sigma^2. \tag{2.2}$$

Proof. The proof is given in Appendix A. □

Our first result, without any assumption on (M_n), is as follows.

Theorem 2.1. *Let (M_n) be a locally square integrable martingale. Then, for all $x, y > 0$,*

$$\mathbb{P}\Big(|M_n| \geq x, [M]_n + <M>_n \leq y\Big) \leq 2\exp\Big(-\frac{x^2}{2y}\Big). \tag{2.3}$$

For self-normalized martingales, we obtain the following result.

Theorem 2.2. *Let (M_n) be a locally square integrable martingale. Then, for all $x, y > 0$, $a \geq 0$ and $b > 0$,*

$$\mathbb{P}\left(\frac{|M_n|}{a + b <M>_n} \geq x, <M>_n \geq [M]_n + y\right) \leq 2\exp\left(-x^2\Big(ab + \frac{b^2 y}{2}\Big)\right). \tag{2.4}$$

Moreover, we also have

$$\mathbb{P}\left(\frac{|M_n|}{a + b <M>_n} \geq x, [M]_n \leq y <M>_n\right) \tag{2.5}$$

$$\leq 2 \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{(1+y)}\Big(ab + \frac{b^2}{2} <M>_n\Big)\Big)\Big]\right)^{1/p}.$$

Proof. The proof are given in Appendix B. □

Remark 2.1. It is not hard to see that (2.4) and (2.5) also hold exchanging the roles of $<M>_n$ and $[M]_n$.

3. Random variables heavy on left or right

This section deals with our new notion of random variables heavy on left or right. It allows us to improve Lemma 2.1

Definition 3.1. We shall say that an integrable random variable X is heavy on left if $\mathbb{E}[X] = 0$ and, for all $a > 0$, $\mathbb{E}[T_a(X)] \leq 0$ where

$$T_a(X) = \min(|X|, a)\text{sign}(X)$$

is the truncated version of X. Moreover, X is heavy on right if $-X$ is heavy on left.

Remark 3.1. Let F be the cumulative distribution function associated with X. Standard calculation leads to $\mathbb{E}[T_a(X)] = -H(a)$ where H is the function defined, for all $a > 0$, by

$$H(a) = \int_0^a F(-x) - (1 - F(x-))\, dx$$

where $F(x-)$ stands for the left limit of F at point x. Consequently, X is heavy on left if $\mathbb{E}[X] = 0$ and, for all $a > 0$, $H(a) \geq 0$. Moreover, H is equal to zero at infinity as

$$\lim_{a\to\infty} H(a) = \mathbb{E}[X] = 0.$$

Furthermore, on can observe that a random variable X is symmetric if and only if X is heavy on left and on right.

The following lemma is the keystone of our one-sided exponential inequalities.

Lemma 3.2. For a random variable X and for all $t \in \mathbb{R}$, let

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

1) If X is heavy on left, then for all $t \geq 0$, $L(t) \leq 1$.
2) If X is heavy on right, then for all $t \leq 0$, $L(t) \leq 1$.
3) If X is symetric, then for all $t \in \mathbb{R}$, $L(t) \leq 1$.

Proof. The proof is given in Appendix A. □ □

We shall now provide several examples of random variables heavy on left. More details concerning those examples may be found in Appendix E. We wish to point out that most of all positive random variables centered around their mean are heavy on left. As a matter of fact, let Y be a positive integrable random variable with mean m and denote

$$X = Y - m.$$

Discrete random variables.

1) If Y has a Bernoulli distribution $\mathcal{B}(p)$ with parameter $0 < p < 1$, then X is heavy on left, heavy on right, or symetric if $p < 1/2$, $p > 1/2$, or $p = 1/2$, respectively.
2) If Y has a Geometric distribution $\mathcal{G}(p)$ with parameter $0 < p < 1$, then X is always heavy on left.
3) If Y has a Poisson distribution $\mathcal{P}(\lambda)$ with parameter $\lambda > 0$, then X is heavy on left as soon as
$$2\exp(-\lambda)\sum_{k=0}^{[\lambda]}\frac{\lambda^k}{k!} \geq 1.$$
One can observe that this condition is always fulfilled if λ is a positive integer, see lemma 1 of [1].

Continuous random variables.

1) If Y has an exponential distribution $\mathcal{E}(\lambda)$ with parameter $\lambda > 0$, then X is always heavy on left.
2) If Y has a Gamma distribution $\mathcal{G}(a, \lambda)$ with parameters $a, \lambda > 0$, then X is always heavy on left.
3) If Y has a Pareto distribution with parameters $a, \lambda > 0$ i.e. $Y = a\exp(Z)$ where Z has an exponential distribution $\mathcal{E}(\lambda)$, then X is always heavy on left.
4) If Y has a log-normal distribution with parameters $m \in \mathbb{R}$ and $\sigma^2 > 0$ i.e. $Y = \exp(Z)$ where Z has a Normal distribution $\mathcal{N}(m, \sigma^2)$, then X is always heavy on left.

4. One-sided exponential inequalities

Our next results are related to martingales heavy on left in the sense of the following definition.

Definition 4.1. Let (M_n) be a locally square integrable martingale adapted to a filtration $\mathbb{F} = (\mathcal{F}_n)$. We shall say that (M_n) is heavy on left if all its increments are conditionally heavy one left. In other words, for all $n \geq 1$ and for any $a > 0$, $\mathbb{E}[T_a(\Delta M_n)|\mathcal{F}_{n-1}] \leq 0$. Moreover, (M_n) is heavy on right if $(-M_n)$ is heavy on left.

We shall recover Theorem 1.3 under the assumption that (M_n) is heavy on left.

Theorem 4.1. *Let (M_n) be a locally square integrable martingale heavy on left. Then, for all $x, y > 0$,*

$$\mathbb{P}\Big(M_n \geq x, [M]_n \leq y\Big) \leq \exp\Big(-\frac{x^2}{2y}\Big). \tag{4.1}$$

For self-normalized martingales, our results are as follows.

Theorem 4.2. *Let (M_n) be a locally square integrable martingale heavy on left. Then, for all $x > 0$, $a \geq 0$ and $b > 0$,*

$$\mathbb{P}\left(\frac{M_n}{a + b[M]_n} \geq x\right) \leq \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big(-(p-1)x^2\Big(ab + \frac{b^2}{2}[M]_n\Big)\Big)\Big]\right)^{1/p}, \tag{4.2}$$

and, for all $y > 0$,

$$\mathbb{P}\left(\frac{M_n}{a + b[M]_n} \geq x, [M]_n \geq y\right) \leq \exp\left(-x^2\Big(ab + \frac{b^2 y}{2}\Big)\right). \tag{4.3}$$

Moreover, we also have

$$\mathbb{P}\left(\frac{M_n}{a + b <M>_n} \geq x, [M]_n \leq y <M>_n\right) \tag{4.4}$$

$$\leq \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{y}\Big(ab + \frac{b^2}{2} <M>_n\Big)\Big)\Big]\right)^{1/p}.$$

Proof. The proof are given in Appendix C. □

Remark 4.1. In the particular case $p = 2$, Theorem 4.2 is due to De la Peña [8] under the conditional symmetric assumption on (M_n). The only difference between (2.5) and (4.4) is that $(1 + y)$ is replaced by y in the upper-bound of (4.4).

Remark 4.2. A locally square integrable martingale (M_n) is Gaussian if, for all $n \geq 1$, the distribution of its increments ΔM_n given $\mathcal{F}_{n-1}$ is $\mathcal{N}(0, \Delta <M>_n)$. Moreover, (M_n) is called sub-Gaussian if there exists some constant $\alpha > 0$ such that, for all $n \geq 1$ and $t \in \mathbb{R}$,

$$\mathbb{E}\Big[\exp(t\Delta M_n)|\mathcal{F}_{n-1}\Big] \leq \exp\Big(\frac{\alpha^2 t^2}{2}\Delta <M>_n\Big). \tag{4.5}$$

It is well-known that if the increments of (M_n) are bounded or if (M_n) is Gaussian, then (M_n) is sub-Gaussian. In addition, if (M_n) satisfies (4.5), then inequalities (4.1), (4.2) and (4.3) hold with appropriate upper-bounds, replacing $[M]_n$ by $<M>_n$ everywhere. For example, (4.2) can be rewritten as

$$\mathbb{P}\left(\frac{M_n}{a + b<M>_n} \geq x\right) \leq \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{\alpha^2}\Big(ab + \frac{b^2}{2} <M>_n\Big)\Big)\Big]\right)^{1/p}. \tag{4.6}$$

5. Applications

5.1. **Linear regressions.** Consider the stochastic linear regression given, for all $n \geq 0$, by

$$X_{n+1} = \theta\phi_n + \varepsilon_{n+1} \tag{5.1}$$

where X_n, ϕ_n and ε_n are the observation, the regression variable and the driven noise, respectively. We assume that (ϕ_n) is a sequence of independent and identically distributed random variables. We also assume that (ε_n) is a sequence of identically distributed random variables, with mean zero and variance $\sigma^2 > 0$. Moreover, we suppose that, for all $n \geq 0$, the random variable ε_{n+1} is independent of $\mathcal{F}_n$ where $\mathcal{F}_n = \sigma(\phi_0, \varepsilon_1, \ldots, \phi_{n-1}, \varepsilon_n)$. In order to estimate the unknown parameter θ, we make use of the least-squares estimator $\widehat{\theta}_n$ given, for all $n \geq 1$, by

$$\widehat{\theta}_n = \frac{\sum_{k=1}^n \phi_{k-1} X_k}{\sum_{k=1}^n \phi_{k-1}^2} \tag{5.2}$$

It immediately follows from (5.1) and (5.2) that

$$\widehat{\theta}_n - \theta = \sigma^2 \frac{M_n}{<M>_n} \tag{5.3}$$

where

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

Let H and L the cumulant generating functions of the sequences (ϕ_n^2) and (ε_n^2), respectively given, for all $t \in \mathbb{R}$, by

$$H(t) = \log \mathbb{E}[\exp(t\phi_n^2)] \qquad \text{and} \qquad L(t) = \log \mathbb{E}[\exp(t\varepsilon_n^2)]$$

Corollary 5.1. Assume that L is finite on some interval $[0, c]$ with $c > 0$ and denote by I its Fenchel-Legendre transform on $[0, c]$,

$$I(x) = \sup_{0 \leq t \leq c} \{xt - L(t)\}.$$

Then, for all $n \geq 1$, $x > 0$ and $y > 0$, we have

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2 \inf_{p>1} \exp\Big(\frac{n}{p} H\Big(-\frac{(p-1)x^2}{2\sigma^2(1+y)}\Big)\Big) + \exp\Big(-nI\Big(\frac{\sigma^2 y}{n}\Big)\Big). \tag{5.4}$$

Remark 5.1. Corollary 5.1 is also true if (ϕ_n, ε_n) is a sequence of independent and identically distributed random vectors of $\mathbb{R}^2$ such that the marginal distribution of ε_n is symmetric. By use of (4.4), inequality (5.4) holds replacing $(1 + y)$ by y in the argument of H.

Remark 5.2. As soon as the sequence (ε_n) is bounded, the right-hand side of (5.4) vanishes since we may directly compare $[M_n]$ with $<M>_n$. For example, assume that (ε_n) is distributed as a centered Bernoulli $\mathcal{B}(p)$ distribution with parameter $0 < p < 1$. If $r = \max(p, q)$, we clearly have for all $n \geq 0$,

$$[M]_n \leq \frac{r^2}{pq} <M>_n .$$

Consequently, we immediately infer from (2.5) that for all $n \geq 1$ and $x > 0$,

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2\exp\Big(\frac{n}{2} H\Big(-\frac{x^2}{4r^2}\Big)\Big).$$

Furthermore, assume that (ϕ_n) is distributed as a normal $\mathcal{N}(0, \tau^2)$ distribution with variance $\tau^2 > 0$. Then, we deduce that for all $n \geq 1$ and $x > 0$,

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2\exp\Big(-\frac{n}{4}\log\Big(1 + \frac{\tau^2 x^2}{2r^2}\Big)\Big).$$

Proof. It follows from (2.5) that, for all $n \geq 1$, $x > 0$ and $y > 0$,

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) = \mathbb{P}\Big(|M_n| \geq \frac{x}{\sigma^2} <M>_n\Big) \leq P_n(x, y) + Q_n(y)$$

where $Q_n(y) = \mathbb{P}([M]_n > y <M>_n)$ and

$$\begin{aligned} P_n(x,y) &= 2 \inf_{p>1} \left(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{2\sigma^4(1+y)} <M>_n\Big)\Big]\right)^{1/p}, \\ &= 2 \inf_{p>1} \exp\Big(\frac{n}{p} H\Big(-\frac{(p-1)x^2}{2\sigma^2(1+y)}\Big)\Big). \end{aligned}$$

In addition, for all $y > 0$ and $0 \leq t \leq c$,

$$\begin{aligned} Q_n(y) &\leq \mathbb{P}\Big(\sum_{k=1}^n \varepsilon_k^2 > \sigma^2 y\Big) \leq \exp(-\sigma^2 t y)\mathbb{E}\Big[\exp\Big(t\sum_{k=1}^n \varepsilon_k^2\Big)\Big], \\ &\leq \exp(-\sigma^2 t y + nL(t)) \leq \exp\Big(-nI\Big(\frac{\sigma^2 y}{n}\Big)\Big), \end{aligned}$$

which achieves the proof of Corollary 5.1. □

5.2. **Autoregressive processes.** Consider the autoregressive process given, for all $n \geq 0$, by

$$X_{n+1} = \theta X_n + \varepsilon_{n+1} \tag{5.5}$$

where X_n and ε_n are the observation and the driven noise, respectively. We assume that (ε_n) is a sequence of independent and identically distributed random variables with standard $\mathcal{N}(0, \sigma^2)$ distribution where $\sigma^2 > 0$. The process is said to be stable if $|\theta| < 1$, unstable if $|\theta| = 1$ and explosive if $|\theta| > 1$. We can estimate the unknown parameter θ by the least-squares or the Yule-Walker estimators given, for all $n \geq 1$, by

$$\widehat{\theta}_n = \frac{\sum_{k=1}^n X_{k-1}X_k}{\sum_{k=1}^n X_{k-1}^2} \qquad \text{and} \qquad \widetilde{\theta}_n = \frac{\sum_{k=1}^n X_{k-1}X_k}{\sum_{k=0}^n X_k^2}. \tag{5.6}$$

It is well-known that $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ both converge almost surely to θ and their fluctuations can be found in [20]. In the stable case $|\theta| < 1$, the large deviation principles were established in [6]. More precisely, set

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

Assume that X_0 is independent of (ε_n) with $\mathcal{N}(0, \sigma^2/(1-\theta^2))$ distribution. Then, $(\widehat{\theta}_n)$ and $(\widetilde{\theta}_n)$ satisfy large deviation principles with good rate functions respectively

given by

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

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

It is only recently that sharp large deviation principles were established for the Yule-Walker estimator $\widetilde{\theta}_n$ in the stable, unstable and explosive cases [7]. Much of work remains to be done for the least-squares estimator $\widehat{\theta}_n$. Our goal is to propose, whatever the value of θ is, a very simple exponential inequality for both $\widehat{\theta}_n$ and $\widetilde{\theta}_n$. For the sake of simplicity, we assume that X_0 is independent of (ε_n) with $\mathcal{N}(0,\tau^2)$ distribution where $\tau^2 \geq \sigma^2$.

Corollary 5.2. For all $n \geq 1$ and $x > 0$, we have

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2\exp\Big(-\frac{nx^2}{2(1+y_x)}\Big) \tag{5.7}$$

where y_x is the unique positive solution of the equation $h(y_x) = x^2$ and h is the function $h(x) = (1+x)\log(1+x) - x$. Moreover, for all $n \geq 1$ and $x > 0$, we also have

$$\mathbb{P}(|\widetilde{\theta}_n - \theta| \geq x + |\theta|) \leq 2\exp\Big(-\frac{nx^2}{2(1+y_x)}\Big). \tag{5.8}$$

Proof. The proof is given in Appendix D. □

Remark 5.3. Inequality (5.7) can be very simple if x is small enough. As a matter of fact, one can easily see that for all $0 < x < 1$, $h(x) < x^2/4$. Consequently, it immediately follows from (5.7) that, for all $0 < x < 1/2$,

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

Moreover, if $\theta > 0$, we can deduce from (5.6) that, for all $x > 0$,

$$\mathbb{P}(\widetilde{\theta}_n - \theta \geq x) \leq \exp\Big(-\frac{nx^2}{2(1+y_x)}\Big).$$

5.3. **Branching processes.** Consider the Galton-Watson process starting from $X_0 = 1$ and given, for all $n \geq 1$, by

$$X_n = \sum_{k=1}^{X_{n-1}} Y_{n,k} \tag{5.9}$$

where $(Y_{n,k})$ is a sequence of independent and identically distributed, nonnegative integer-valued random variables. The distribution of $(Y_{n,k})$, with finite mean m and variance σ^2, is commonly called the offspring or reproduction distribution. Hereafter, we shall assume that $m > 1$. In order to estimate the offspring mean m, we can make use of the Lotka-Nagaev or the Harris estimators given, for all $n \geq 1$, by

$$\widetilde{m}_n = \frac{X_n}{X_{n-1}} \qquad \text{and} \qquad \widehat{m}_n = \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n X_{k-1}}. \tag{5.10}$$

Without loss of generality, we can suppose that the set of extinction of the process (X_n) is negligeable. Consequently, the Lokta-Nagaev estimator $\widetilde{m}_n$ is always well-defined. It is well-known that $\widetilde{m}_n$ and $\widehat{m}_n$ both converge almost surely to m and their fluctuations are given in [2], [13], [14]. Moreover, the large deviation properties associated with $(\widetilde{m}_n)$ may be found in [3], [18], [19]. Our goal is now to establish, as in the previous sections, exponential inequalities for both $\widetilde{m}_n$ and $\widehat{m}_n$. Denote by L the cumulant generating function associated with the centered offspring distribution given, for all $t \in \mathbb{R}$, by $L(t) = \log \mathbb{E}[\exp(t(Y_{n,k} - m))]$.

Corollary 5.3. Assume that L is finite on some interval $[-c, c]$ with $c > 0$ and let I be its Fenchel-Legendre transform,

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

Then, for all $n \geq 1$ and $x > 0$,

$$\mathbb{P}(|\widetilde{m}_n - m| \geq x) \leq 2\mathbb{E}[\exp(-J(x)X_{n-1})] \tag{5.11}$$

where $J(x) = \min(I(x), I(-x))$. Moreover, we also have

$$\mathbb{P}(|\widetilde{m}_n - m| \geq x) \leq 2 \inf_{p>1} \Big(\mathbb{E}\Big[\exp(-(p-1)J(x)X_{n-1})\Big]\Big)^{1/p}. \tag{5.12}$$

In addition, if $S_n = \sum_{k=0}^{n} X_k$, we have for all $n \geq 1$ and $x > 0$,

$$\mathbb{P}(|\widehat{m}_n - m| \geq x) \leq 2 \inf_{p>1} \Big(\mathbb{E}\Big[\exp(-(p-1)J(x)S_{n-1})\Big]\Big)^{1/p}. \tag{5.13}$$

Proof. The proof is given in Appendix E. □

Remark 5.4. On the one hand, inequality (5.12) obviously holds for the Harris estimator $\widehat{m}_n$ since we always have $S_n \geq X_n$. On the other hand, in order to precise the right-hand side of (5.11), (5.12) or (5.13), it is necessary to find an upper-bound or to provide an explicit expression of the moment generating function of X_n. One can easily carry out this calculation when the offspring distribution is the geometric $\mathcal{G}(p)$ distribution with parameter $0 < p < 1$. As a matter of fact, in that particular case, the offspring mean $m = 1/p$ and it follows from formula (7.3) of [14] that for all $0 < s < 1$,

$$\mathbb{E}[s^{X_n}] \leq \frac{p^n s}{1 - s}.$$

Consequently, for all $n \geq 1$ and $x > 0$, we obtain the simple inequality

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

If the offspring distribution is not geometric, on can precisely estimate the moment generating function of X_n using Theorem 1, page 80 of [2] which gives a good approximation of the distribution of X_n based on the limiting distribution

$$W = \lim_{n \to \infty} \frac{X_n}{m^n} \qquad \text{a.s.}$$

Appendix A.

This appendix is devoted to the proofs of Lemma 2.1 and Lemma 3.2. Lemma 2.1 immediately follows from Jensen's inequality. As a matter of fact, (2.1) implies that for all $t \in \mathbb{R}$,

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

Consequently, we obtain that for all $t \in \mathbb{R}$,

$$L(t) \geq 1 - \frac{t^2}{2}\sigma^2. \tag{A.1}$$

Furthermore, for all $t \in \mathbb{R}$,

$$L(t) + L(-t) = 2\,\mathbb{E}\Big[\exp\Big(-\frac{t^2}{2}X^2\Big)\cosh(tX)\Big] \leq 2 \tag{A.2}$$

by the well-known inequality $\cosh(x) \leq \exp(x^2/2)$. Hence, we obtain from (A.1) together with (A.2) that for all $t \in \mathbb{R}$,

$$L(t) \leq 2 - L(-t) \leq 1 + \frac{t^2}{2}\sigma^2.$$

Lemma 3.2 is much more difficult to prove. Let f be the function defined, for all $x \in \mathbb{R}$, by

$$f(x) = \exp\Big(x - \frac{x^2}{2}\Big).$$

We clearly have $f'(x) = (1-x)f(x)$ and $f'(-x) = (1+x)f(-x)$. We shall also make use of the functions a and b defined, for all $x \in \mathbb{R}$, by $a(x) = f'(-x)$ and $b(x) = f'(-x) - f'(x)$. One can realize that, for all $x > 0$, $0 < a(x) < 1$, $0 < b(x) < 2$ and $a'(x) < 0$ as $a'(x) = -(2x + x^2)f(-x)$. After those simple preliminaries, we are in position to prove Lemma 3.2. For all $t \in \mathbb{R}$,

$$L(t) = \mathbb{E}\Big[\exp\Big(tX - \frac{t^2X^2}{2}\Big)\Big] = \int_{\mathbb{R}} f(tx)\,dF(x)$$

where F is the distribution function associated with X. Integrating by parts, we have for all $t \in \mathbb{R}$,

$$\begin{aligned} L(t) &= -t\int_{\mathbb{R}} f'(tx)F(x)\,dx = -t\int_{-\infty}^{0} f'(tx)F(x)\,dx - t\int_{0}^{+\infty} f'(tx)F(x)\,dx, \\ &= -t\int_{0}^{+\infty} f'(-tx)F(-x)\,dx - t\int_{0}^{+\infty} f'(tx)F(x)\,dx. \end{aligned} \tag{A.3}$$

Consequently, as

$$-t\int_{0}^{\infty} f'(tx)\,dx = \Big[-\exp\Big(tx - \frac{t^2x^2}{2}\Big)\Big]_{0}^{+\infty} = 1,$$

we obtain from (A.3) that, for all $t \in \mathbb{R}$, $L(t) = 1 - tI(t)$ where

$$I(t) = \int_{0}^{+\infty} f'(-tx)F(-x)dx - \int_{0}^{+\infty} f'(tx)(1-F(x))dx = A(t) + B(t) \tag{A.4}$$

with

$$\begin{aligned} A(t) &= \int_0^{+\infty} a(tx)(F(-x) - (1 - F(x)))\, dx, \\ B(t) &= \int_0^{+\infty} b(tx)(1 - F(x))\, dx. \end{aligned}$$

First of all, assume that X is heavy on left. Our goal is to show that, for all $t > 0$, the integral $I(t)$ is nonnegative. We obviously have, for all $t > 0$, $B(t) \geq 0$ as $b(tx) > 0$. In addition, for any $a > 0$, let

$$H(a) = \int_0^a F(-x) - (1 - F(x))\, dx.$$

Since $H'(a) = F(-a) - (1 - F(a))$ almost everywhere, integrating once again by parts, we find that

$$\begin{aligned} A(t) &= \int_0^{+\infty} a(tx)H'(x)\, dx = \Big[a(tx)H(x)\Big]_0^{+\infty} - \int_0^{+\infty} ta'(tx)H(x)\, dx, \\ &= -t\int_0^{+\infty} a'(tx)H(x)\, dx \qquad \text{(A.5)} \end{aligned}$$

as $H(0) = 0$ and H vanishes at infinity. Hereafter, as X is heavy on left, $H(a) \geq 0$ for all $a \geq 0$. Moreover, we recall that, for all $x > 0$, $a'(x) < 0$. Hence, we immediately deduce from (A.5) that, for all $t > 0$, $A(t) \geq 0$. Consequently, relation (A.4) leads to $I(t) \geq 0$ and $L(t) \leq 1$ for all $t > 0$, which completes the proof of part 1) of Lemma 3.2. Next, if X is heavy on right, $-X$ is heavy on left. Hence, we immediately infer from (2.1) and part 1) of Lemma 3.2 that $L(t) \leq 1$ for all $t < 0$. Finally, part 3) of Lemma 3.2 follows from the conjunction of parts 1) and 2). Another straighforward way to prove part 3) is as follows. If X is symetric, we have for all $t \in \mathbb{R}$,

$$\begin{aligned} L(t) &= \int_{\mathbb{R}} f(tx)\, dF(x) = \int_0^{+\infty} (f(tx) + f(-tx))\, dF(x), \\ &= 2\int_0^{+\infty} \exp(-t^2x^2/2)\cosh(tx)\, dF(x) \leq 1 \end{aligned}$$

by the well-known inequality $\cosh(x) \leq \exp(x^2/2)$.

Appendix B.

In order to prove Theorem 2.1 and Theorem 2.2, we shall often make use of the following lemma.

Lemma B.1. Let (M_n) be a locally square integrable martingale. For all $t \in \mathbb{R}$ and $n \geq 0$, denote

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

Then for all $t \in \mathbb{R}$, $(V_n(t))$ is a positive supermartingale with $\mathbb{E}[V_n(t)] \leq 1$.

Proof. For all $t \in \mathbb{R}$ and $n \geq 1$, we have

$$V_n(t) = V_{n-1}(t)\exp\Big(t\Delta M_n - \frac{t^2}{2}\Big(\Delta[M]_n + \Delta <M>_n\Big)\Big)$$

where $\Delta M_n = M_n - M_{n-1}$, $\Delta[M]_n = \Delta M_n^2$ and $\Delta <M>_n = \mathbb{E}[\Delta M_n^2 | \mathcal{F}_{n-1}]$. Hence, we deduce from Lemma 2.1 that for all $t \in \mathbb{R}$,

$$\begin{aligned}\mathbb{E}[V_n(t)|\mathcal{F}_{n-1}] &\leq V_{n-1}(t) \exp\Big(-\frac{t^2}{2}\Delta <M>_n\Big)\Big(1 + \frac{t^2}{2}\Delta <M>_n\Big),\\ &\leq V_{n-1}(t).\end{aligned}$$

Consequently, for all $t \in \mathbb{R}$, $(V_n(t))$ is a positive supermartingale such that $\mathbb{E}[V_n(t)] \leq \mathbb{E}[V_{n-1}(t)]$ which implies that $\mathbb{E}[V_n(t)] \leq \mathbb{E}[V_0(t)] = 1$. □

We are now in position to prove Theorem 2.1 and Theorem 2.2 inspired by the original work of De la Peña [8]. First of all, denote

$$Z_n = [M]_n + <M>_n .$$

For all $x, y > 0$, let

$$A_n = \{|M_n| \geq x, Z_n \leq y\}.$$

We have the decomposition $A_n = A_n^+ \cup A_n^-$ where $A_n^+ = \{M_n \geq x, Z_n \leq y\}$ and $A_n^- = \{M_n \leq -x, Z_n \leq y\}$. By Markov's inequality, we have for all $t > 0$,

$$\begin{aligned}\mathbb{P}(A_n^+) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{tx}{2}\Big)\mathbb{1}_{A_n^+}\Big],\\ &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{t^2}{4}Z_n\Big)\exp\Big(\frac{t^2}{4}Z_n - \frac{tx}{2}\Big)\mathbb{1}_{A_n^+}\Big],\\ &\leq \exp\Big(\frac{t^2 y}{4} - \frac{tx}{2}\Big)\sqrt{\mathbb{E}[V_n(t)]\mathbb{P}(A_n^+)}.\end{aligned}$$

Hence, we deduce from Lemma B.1 that for all $t > 0$,

$$\mathbb{P}(A_n^+) \leq \exp\Big(\frac{t^2 y}{4} - \frac{tx}{2}\Big)\sqrt{\mathbb{P}(A_n^+)}. \tag{B.1}$$

Dividing both sides of (B.1) by $\sqrt{\mathbb{P}(A_n^+)}$ and choosing the value $t = x/y$, we find that

$$\mathbb{P}(A_n^+) \leq \exp\Big(-\frac{x^2}{2y}\Big).$$

We also find the same upper-bound for $\mathbb{P}(A_n^-)$ which immediately leads to (2.3). We next proceed to the proof of Theorem 2.2 in the special case $a = 0$ and $b = 1$ inasmuch as the proof for the general case follows exactly the same lines. For all $x, y > 0$, let

$$B_n = \{|M_n| \geq x <M>_n, <M>_n - [M]_n \geq y\} = B_n^+ \cup B_n^-$$

where

$$\begin{aligned}B_n^+ &= \{M_n \geq x <M>_n, <M>_n - [M]_n \geq y\},\\ B_n^- &= \{M_n \leq -x <M>_n, <M>_n - [M]_n \geq y\}.\end{aligned}$$

By Cauchy-Schwarz's inequality, we have for all $t > 0$,

$$\begin{aligned}\mathbb{P}(B_n^+) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{tx}{2} <M>_n\Big)\mathbb{1}_{B_n^+}\Big],\\ &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{t^2}{4}Z_n\Big)\exp\Big(\frac{t}{4}(t-2x) <M>_n + \frac{t^2}{4}[M]_n\Big)\mathbb{1}_{B_n^+}\Big]. \qquad \text{(B.2)}\end{aligned}$$

Consequently, we obtain from (B.2) with the particular choice $t = x$ that

$$\mathbb{P}(B_n^+) \leq \exp\Big(-\frac{x^2 y}{4}\Big)\sqrt{\mathbb{P}(B_n^+)}. \tag{B.3}$$

Therefore, if we divide both sides of (B.3) by $\sqrt{\mathbb{P}(B_n^+)}$, we find that

$$\mathbb{P}(B_n^+) \leq \exp\Big(-\frac{x^2 y}{2}\Big).$$

The same upper-bound holds for $\mathbb{P}(B_n^-)$ which clearly implies (2.4). Furthermore, for all $x, y > 0$, let

$$C_n = \{|M_n| \geq x <M>_n, [M]_n \leq y <M>_n\} = C_n^+ \cup C_n^-$$

where

$$\begin{aligned} C_n^+ &= \{M_n \geq x <M>_n, [M]_n \leq y <M>_n\}, \\ C_n^- &= \{M_n \leq -x <M>_n, [M]_n \leq y <M>_n\}. \end{aligned}$$

By Holder's inequality, we have for all $t > 0$ and $q > 1$,

$$\begin{aligned} \mathbb{P}(C_n^+) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{q}M_n - \frac{tx}{q} <M>_n\Big)\mathbb{1}_{C_n^+}\Big], \\ &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{q}M_n - \frac{t^2}{2q}Z_n\Big)\exp\Big(\frac{t}{2q}(t - 2x + ty) <M>_n\Big)\mathbb{1}_{C_n^+}\Big], \\ &\leq \Big(\mathbb{E}\Big[\exp\Big(\frac{tp}{2q}(t - 2x + ty) <M>_n\Big)\Big]\Big)^{1/p}. \end{aligned} \tag{B.4}$$

Consequently, as $p/q = p - 1$, we can deduce from (B.4) and the particular choice $t = x/(1 + y)$ that

$$\mathbb{P}(C_n^+) \leq \inf_{p>1}\Big(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{2(1+y)} <M>_n\Big)\Big]\Big)^{1/p}.$$

We also find the same upper-bound for $\mathbb{P}(C_n^-)$ which completes the proof of Theorem 2.2.

Appendix C.

The proofs of Theorem 4.1 and Theorem 4.2 are based on the following lemma.

Lemma C.2. Let (M_n) be a locally square integrable martingale. For all $t \in \mathbb{R}$ and $n \geq 0$, denote

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

1) If (M_n) is heavy on left, then for all $t \geq 0$, $(W_n(t))$ is a supermartingale with $\mathbb{E}[W_n(t)] \leq 1$.
2) If (M_n) is heavy on right, then for all $t \leq 0$, $(W_n(t))$ is a supermartingale with $\mathbb{E}[W_n(t)] \leq 1$.
3) If (M_n) is conditionally symmetric, then for all $t \in \mathbb{R}$, $(W_n(t))$ is a supermartingale with $\mathbb{E}[W_n(t)] \leq 1$.

Proof. Lemma C.2 part 3) is due to De la Peña [8], Lemma 6.1. Our approach is totally different as it mainly relies on Lemma 3.2. Assume that (M_n) is heavy on left. For all $t \in \mathbb{R}$ and $n \geq 1$, we have

$$W_n(t) = W_{n-1}(t) \exp\Big(t\Delta M_n - \frac{t^2}{2}\Delta[M]_n\Big)$$

where $\Delta[M]_n = \Delta M_n^2$. We infer from Lemma 3.2 part 1) that for all $n \geq 1$ and for all $t \geq 0$,

$$\mathbb{E}\left[\exp\Big(t\Delta M_n - \frac{t^2}{2}\Delta M_n^2\Big)|\mathcal{F}_{n-1}\right] \leq 1.$$

Consequently, for all $t \geq 0$, $(W_n(t))$ is a positive supermartingale such that $\mathbb{E}[W_n(t)] \leq \mathbb{E}[W_{n-1}(t)]$ which leads to $\mathbb{E}[W_n(t)] \leq \mathbb{E}[W_0(t)] = 1$. The rest of the proof is also a straightforward application of Lemma 3.2. □

By use of Lemma C.2, the proof of Theorem 4.1 is quite analogous to that of Theorem 2.1 and therefore is left to the reader. We shall proceed to the proof of Theorem 4.2 in the special case $a = 0$ and $b = 1$. For all $x > 0$, let $A_n = \{M_n \geq x[M]_n\}$. By Holder's inequality, we have for all $t > 0$ and $q > 1$,

$$\begin{aligned}
\mathbb{P}(A_n) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{q}M_n - \frac{tx}{q}[M]_n\Big)\mathbb{1}_{A_n}\Big], \\
&\leq \mathbb{E}\Big[\exp\Big(\frac{t}{q}M_n - \frac{t^2}{2q}[M]_n\Big)\exp\Big(\frac{t}{2q}(t-2x)[M]_n\Big)\mathbb{1}_{A_n}\Big], \\
&\leq \left(\mathbb{E}\Big[\exp\Big(\frac{tp}{2q}(t-2x)[M]_n\Big)\Big]\right)^{1/p}\Big(\mathbb{E}\Big[W_n(t)\Big]\Big)^{1/q}. \qquad \text{(C.1)}
\end{aligned}$$

Since (M_n) is heavy on left, it follows from lemma C.2 that for all $t \geq 0$, $\mathbb{E}[W_n(t)] \leq 1$. Consequently, as $p/q = p-1$, we can deduce from (C.1) and the particular choice $t = x$ that

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

Furthermore, for all $x, y > 0$, let $B_n = \{M_n \geq x[M]_n, [M]_n \geq y\}$. As before, we find that for all $0 < t < 2x$,

$$\begin{aligned}
\mathbb{P}(B_n) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{t^2}{4}[M]_n\Big)\exp\Big(\frac{t}{4}(t-2x)[M]_n\Big)\mathbb{1}_{B_n}\Big], \\
&\leq \exp\Big(\frac{ty}{4}(t-2x)\Big)\mathbb{E}\Big[\exp\Big(\frac{t}{2}M_n - \frac{t^2}{4}[M]_n\Big)\mathbb{1}_{B_n}\Big], \\
&\leq \exp\Big(\frac{ty}{4}(t-2x)\Big)\sqrt{\mathbb{P}(B_n)}, \\
&\leq \exp\Big(-\frac{x^2y}{2}\Big)
\end{aligned}$$

choosing the value $t = x$. Finally, the last inequality of Theorem 4.2 is left to the reader as its proof follows exactly the same arguments as (4.3).

Appendix D.

We shall now focus our attention on the proof of Corollary 5.2. It immediately follows from (5.5) together with (5.6) that for all $n \geq 1$,

$$\widehat{\theta}_n - \theta = \sigma^2 \frac{M_n}{<M>_n} \tag{D.1}$$

where

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

The driven noise (ε_n) is a sequence of independent and identically distributed random variables with $\mathcal{N}(0, \sigma^2)$ distribution. Consequently, for all $n \geq 1$, the distribution of the increments $\Delta M_n = X_{n-1}\varepsilon_n$ given $\mathcal{F}_{n-1}$ is $\mathcal{N}(0, \sigma^2 X_{n-1}^2)$ which implies that (M_n) is a Gaussian martingale. Therefore, we infer from inequality (4.6) that for all $n \geq 1$ and $x > 0$,

$$\begin{aligned} \mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) &= \mathbb{P}\Big(|M_n| \geq \frac{x}{\sigma^2} <M>_n\Big) = 2\mathbb{P}\Big(M_n \geq \frac{x}{\sigma^2} <M>_n\Big), \\ &\leq 2 \inf_{p>1} \left(\mathbb{E}\Big[\exp\Big(-(p-1)\frac{x^2}{2\sigma^4}<M>_n\Big)\Big]\right)^{1/p}. \end{aligned} \tag{D.2}$$

Similar result may be found in [16], [21], [22]. We are now halfway to our goal and it remains to find a suitable upper-bound for the right-hand side of (D.2). For all $t \in \mathbb{R}$ such that $1 - 2\sigma^2 t > 0$, if $\alpha = 1/\sqrt{1 - 2\sigma^2 t}$, we deduce from (5.5) that, for all $n \geq 1$,

$$\begin{aligned} \mathbb{E}[\exp(tX_n^2)|\mathcal{F}_{n-1}] &= \exp(t\theta^2 X_{n-1}^2)\mathbb{E}[\exp(2\theta t X_{n-1}\varepsilon_n + t\varepsilon_n^2)|\mathcal{F}_{n-1}], \\ &= \frac{\exp(t\theta^2 X_{n-1}^2)}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} \exp\Big(-\frac{x^2}{2\alpha^2\sigma^2}\Big) \exp(2\theta t X_{n-1} x) dx. \end{aligned}$$

Hence, if $\beta = 2t\alpha\sigma\theta X_{n-1}$, we find via the change of variables $y = x/\alpha\sigma$ that

$$\begin{aligned} \mathbb{E}[\exp(tX_n^2)|\mathcal{F}_{n-1}] &= \frac{\alpha \exp(t\theta^2 X_{n-1}^2)}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-\frac{y^2}{2} + \beta y)\, dy, \\ &= \alpha \exp\Big(t\theta^2 X_{n-1}^2 + \frac{\beta^2}{2}\Big) = \alpha \exp(t\alpha^2\theta^2 X_{n-1}^2) \end{aligned}$$

which implies that, for all $t < 0$ and $n \geq 1$,

$$\mathbb{E}[\exp(tX_n^2)|\mathcal{F}_{n-1}] \leq \alpha. \tag{D.3}$$

Furthermore, as X_0 is $\mathcal{N}(0, \tau^2)$ distributed with $\tau^2 \geq \sigma^2$, $\mathbb{E}[\exp(tX_0^2)] \leq \alpha$. It immediately follows from (D.3) together with the tower property of the conditional expectation that for all $t < 0$ and $n \geq 0$,

$$\mathbb{E}[\exp(t <M>_n)] \leq (1 - 2\sigma^4 t)^{-n/2}. \tag{D.4}$$

Consequently, we deduce from the conjunction of (D.2) and (D.4) with the value $t = -(p-1)x^2/2\sigma^4$ and the change of variables $y = (p-1)x^2$ that for all $x > 0$ and $n \geq 1$,

$$\mathbb{P}(|\widehat{\theta}_n - \theta| \geq x) \leq 2 \inf_{y>0} \exp\Big(-\frac{nx^2}{2}\ell(y)\Big)$$

where the function ℓ is given by

$$\ell(y) = \frac{\log(1+y)}{x^2+y}.$$

We clearly have

$$\ell'(y) = \frac{x^2 - h(y)}{(1+y)(x^2+y)^2}$$

where $h(y) = (1+y)\log(1+y) - y$. One can observe that the function h is the Cramer transform of the centered Poisson distribution with parameter one. Let y_x be the unique positive solution of the equation $h(y_x) = x^2$. The value y_x maximises the function ℓ and this natural choice clearly leads to (5.7). Finally, it follows from (5.6) and (5.7) that for all $x > 0$ and $n \geq 1$,

$$\mathbb{P}(|\widetilde{\theta}_n - \theta + \theta f_n| \geq x) \leq 2\exp\Big(-\frac{nx^2}{2(1+y_x)}\Big) \tag{D.5}$$

where the random variable $0 \leq f_n \leq 1$. Hence, (D.5) implies (5.8) which completes the proof of Corollary 5.2.

Appendix E.

We shall now proceed to the proof of Corollary 5.3. We only focus our attention on the Harris estimator inasmuch as the proof for the Lotka-Nagaev estimator follows essentially the same lines. First of all, relation (5.9) can be rewritten as

$$X_n = mX_{n-1} + \xi_n \tag{E.1}$$

where $\xi_n = X_n - \mathbb{E}[X_n|\mathcal{F}_{n-1}]$. Consequently, we obtain from (5.10) together with (E.1) that for all $n \geq 1$,

$$\widehat{m}_n - m = \frac{M_n}{S_{n-1}} \qquad \text{where} \qquad M_n = \sum_{k=1}^{n} \xi_k.$$

Moreover, for all $n \geq 1$ and $0 \leq t \leq c$, $\mathbb{E}[\exp(t\xi_n)|\mathcal{F}_{n-1}] = \exp(X_{n-1}L(t))$ which implies that

$$\mathbb{E}[\exp(tM_n - L(t)S_{n-1})] = 1. \tag{E.2}$$

We are in position to prove (5.13). For all $x > 0$, let $D_n = \{|\widehat{m}_n - m| \geq x\}$. We have the decomposition $D_n = D_n^+ \cup D_n^-$ where $D_n^+ = \{\widehat{m}_n - m \geq x\}$ and $D_n^- = \{\widehat{m}_n - m \leq -x\}$. By Holder's inequality together with (E.2), we have for all $0 \leq t \leq c$ and $q > 1$,

$$\begin{aligned}
\mathbb{P}(D_n^+) &\leq \mathbb{E}\Big[\exp\Big(\frac{t}{q}M_n - \frac{tx}{q}S_{n-1}\Big)\mathbb{1}_{D_n^+}\Big], \\
&\leq \mathbb{E}\Big[\exp\Big(\frac{t}{q}M_n - \frac{L(t)}{q}S_{n-1}\Big)\exp\Big(\frac{1}{q}(L(t)-tx)S_{n-1}\Big)\mathbb{1}_{D_n^+}\Big], \\
&\leq \Big(\mathbb{E}\Big[\exp\Big(\frac{p}{q}(L(t)-tx)S_{n-1}\Big)\Big]\Big)^{1/p}. \qquad \text{(E.3)}
\end{aligned}$$

Taking the infimum over the interval $[0, c]$, we infer from (E.3) that

$$\mathbb{P}(D_n^+) \leq \Big(\mathbb{E}\Big[\exp\Big(-(p-1)I(x)S_{n-1}\Big)\Big]\Big)^{1/p}. \tag{E.4}$$

Via the same lines, we also find that

$$\mathbb{P}(D_n^-) \leq \Big(\mathbb{E}\Big[\exp\Big(-(p-1)I(-x)S_{n-1}\Big)\Big]\Big)^{1/p}. \tag{E.5}$$

Finally, (5.13) immediately follows from (E.4) and (E.5).

Appendix F.

This appendix is devoted to some justifications about the examples of random variables heavy on left or right. Consider an integrable random variable X with zero mean and denote by F its cumulative distribution function. Let H be the function defined, for all $a > 0$, by

$$H(a) = \int_0^a F(-x) - (1 - F(x-))\, dx.$$

We already saw that X is heavy on left if, for all $a > 0$, $H(a) \geq 0$ while X is heavy on right if, for all $a > 0$, $H(a) \leq 0$. Let Y be a positive integrable random variable with mean m and denote

$$X = Y - m.$$

Discrete random variables. Assume that Y is a discrete random variable taking its values in $\mathbb{N}$. For all $n \geq 0$, let

$$s_n = \sum_{k=0}^{n} \mathbb{P}(Y = k).$$

After some straightforward calculations, we obtain that, for all $a > 0$,

$$H(a) = -a + \sum_{k=[m-a]}^{[m+a]} s_k - s_{[m+a]} + \{m+a\}s_{[m+a]} - \{m-a\}s_{[m-a]}$$

where, for all $x \in \mathbb{R}$, $[x]$ stands for the integer part of x and its fractional part $\{x\}$ is given by $\{x\} = x - [x]$ and, of course, $s_n = 0$ for all $n < 0$.

Continuous random variables. Assume that Y is a real random variable absolutely continuous with respect to the Lebesgue measure. Denote by g its probability density function. It is not hard to see that, for all $a > 0$,

$$H(a) = -a + 2a \int_0^{a_m} g(x)\, dx + \int_{a_m}^{m+a} (m + a - x)g(x)\, dx$$

where $a_m = \inf\{m - a, 0\}$. Consequently, in order to check that X is heavy on left, it is only necessary to show that, for all $a > 0$, $H(a) \geq 0$.

Acknowledgements

The authors would like to thank V. De la Peña, T. L. Lai and M. Ledoux for fruitful discussions.

References

[1] Adell, J. A. and Jodrá, P. The median of the Poisson distribution, Metrika, vol. 61, p. 337-346, 2005.

[2] Athreya, K. B. and Ney P. E. Branching processes, Springer-Verlag, Berlin, 1972.

[3] Athreya, K. B. and Vidyashankar A. N. Large deviation rates for branching processes I. Single type case, Annals of Applied Probability, vol. 4, p. 779-790, 1994.

[4] Azuma, K. Weighted sums of certain dependent random variables, Tôkuku Mathematical Journal, vol. 19, p. 357-367, 1967.

[5] Bennett, G. Probability inequalities for the sum of independent random variables, Journal of the Amer. Statist. Assoc., vol. 57, p. 33-45, 1962.

[6] Bercu, B., Gamboa, F. and Rouault, A. Large deviations for quadratic forms of stationary Gaussian processes. Stochastic Processes and their Applications, vol. 71, p. 75-90, 1997.

[7] Bercu, B. On large deviations in the Gaussian autoregressive process: stable, unstable and explosive cases, Bernoulli, vol. 7, p. 299-316, 2001.

[8] De la Peña, V. H. A general class of exponential inequalities for martingales and ratios. Annals of Probability, vol. 27, p. 537-564, 1999.

[9] De la Peña, V. H., Klass, M. J. and Lai, T. L. Self-normalized processes: exponential inequalities, moments bounds and iterated logarithm law, Annals of Probability, vol. 32, p. 1902-1933, 2004.

[10] Dzhaparidze, K. and Van Zanten, J. H. On Bernstein-type inequalities for martingales. Stochastic Processes and their Applications, vol. 93, p. 109-117, 2001.

[11] Van de Geer, S. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. Annals of Statistics, vol. 23, p. 1779-1801, 1995.

[12] Freedman, D. A. On tail probabilities for martingales. Annals of Probability, vol. 3, p. 100-118, 1975.

[13] Guttorp, P. Statistical inference for branching processes, John Wiley, New York, 1991.

[14] Harris, T.E. The theory of branching processes, Springer-Verlag, Berlin, 1963.

[15] Hoeffding, W. J. Probability inequalities sums of bounded random variables, Journal of the Amer. Statist. Assoc., vol. 58, p. 713-721, 1963.

[16] Liptser, R. and Spokoiny, V. Deviation probability bound for martingales with applications to statistical estimation. Statistic and Probability Letters, vol. 46, p. 347-357, 2000.

[17] McDiarmid, C. Concentration, in Probabilistic methods for algorithmic discrete mathematics, Springer-Verlag, Berlin, p. 195-248, 1998.

[18] Ney, P. E. and Vidyashankar A. N. Harmonic moments and large deviation rates for supercritical branching processes, Annals of Applied Probability, vol. 13, p. 475-489, 2003.

[19] Ney, P. E. and Vidyashankar A. N. Local limit theory and large deviations for supercritical branching processes, Annals of Applied Probability, vol. 14, p. 1135-1166, 2004.

[20] White, J. S. The limit distribution of the serial correlation in the explosive case. Annals of Mathematical Statistics, vol. 29, p. 1188-1197, 1958.
[21] Worms, J. Moderate deviations for stable Markov chains and regression models. Electronic Journal of Probability, vol. 4, p. 1-28, 1999.
[22] Worms, J. Large and moderate deviations upper bounds for the Gaussian autoregressive process. Statistic and Probability Letters, vol. 51, p. 235-243, 2001.

Université Bordeaux 1, Institut de Mathématiques de Bordeaux, UMR 5251, 351 cours de la libération, 33405 Talence cedex, France.
E-mail address: Bernard.Bercu@math.u-bordeaux1.fr

Faculté des Sciences de Bizerte, Département de Mathématiques, 7021 Zarzouna, Tunisie.
E-mail address: Abder.Touati@fsb.rnu.tn