Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

A Note on Bernstein-type Inequalities for Martingales

K. Dzhaparidze and J.H. van Zanten

Probability, Networks and Algorithms (PNA)

PNA-R9912 November 30, 1999

Report PNA-R9912
ISSN 1386-3711

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

A Note on Bernstein-type Inequalities for Martingales

K. Dzhaparidze and J.H. van Zanten
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
kacha@cwi.nl *hvz@cwi.nl*

ABSTRACT

Several inequalities of Bernstein's type are derived in a unified manner. Some extra light is shed on the classical inequalities and implications are sought for instance for conditionally symmetric martingales and sequences of asymptotically continuous martingales.

1991 Mathematics Subject Classification: 60E15, 60G44, 60G57.
Keywords and Phrases: Locally square integrable martingale, Bernstein inequality, Multiplicative decomposition, Exponential supermartingale.
Note: Work carried out under project PNA3.3 'Stochastic Processes and Applications'.

1 Introduction

In this report we discuss Bernstein-type inequalities for locally square integrable martingales. We will use standard notations, M for a martingale, $\langle M\rangle$ and $[M]$ for its predictable and optional quadratic variation, respectively, and ΔM for its jumps. Finally, M^* will stand for the process given at any time $t \geq 0$ by

$$M^*_t = \sup_{s\leq t} |M_s|. \tag{1.1}$$

Besides, with a martingale M in question we will associate the family of processes H^a indexed by the nonnegative numbers $a \geq 0$ as follows: at each time $t \geq 0$

$$H^a_t = \sum_{s\leq t} (\Delta M_s)^2 \, 1_{\{|\Delta M_s|\geq a\}} + \langle M\rangle_t \, . \tag{1.2}$$

Our main result concerns these two processes evaluated at a finite stopping time τ. It states that for any fixed value of the index $a \geq 0$ and any $z, L \geq 0$

$$\mathrm{P}\left(M^*_\tau \geq z \, , \, H^a_\tau \leq L\right) \leq 2e^{-\frac{1}{2}\frac{z^2}{az+L}}, \tag{1.3}$$

see theorem 3.2 as well as its corollary 3.3 to trace the links to known results on the present subject. Assertion (i) of this corollary tells us that if the truncation level a is taken to be zero, we get the result (3.9) due to Barlow et al. (1986). Assertion (ii) deals with the special case of martingales with bounded jumps and shows how to reduce (1.3) to the classical Bernstein inequality by fixing the level $a \geq 0$ above all the jumps $|\Delta M| \leq a$. Since in this case the expression (1.2) for H^a retains only the second term, the quadratic variation $\langle M \rangle$, we get (3.10), cf e.g. Shorack and Wellner (1986). Observe that if M is simply a continuous local martingale, then on the right hand side of (3.10) we may substitute $a = 0$ to get the same upper bound as in (3.9):

$$\mathrm{P}\left(M_\tau^* \geq z, \langle M \rangle_\tau \leq L\right) \leq 2e^{-\frac{1}{2}\frac{z^2}{L}}. \tag{1.4}$$

Inequalities with such bounds are often called *sub-Gaussian*. In presence of jumps the sub-Gaussian inequalities (3.9) and (4.2) (the latter concerns the conditionally symmetric case) are less useful as compared to (1.4), for we are led to substitute predictable quadratic variation $\langle M \rangle$ by the optional one $[M]$ in (4.2), with the additional term $\left\langle M^d \right\rangle$ in (3.9). Moreover, we are able to restore the relationship (1.4) only asymptotically, by passing to the limit under the usual Lindeberg condition (6.1), cf corollary 6.1. This substitution causes substantial difficulties in applications. Often the alternatives (3.10) or (5.2) are sought that allow predictable characteristics under the probability sign, but at the expense of more stringent conditions and a cruder bound conform to (1.3), cf corollary 3.3, assertion (ii) and theorem 5.1.

It is our intention to pursue a unified approach and to provide transparent proofs of all the mentioned inequalities, otherwise rather scattered in the probabilistic, as well as in statistical literature (see e.g. Bennett (1962), Freedman (1975), Barlow et al. (1986), Shorack and Wellner (1986) and Van de Geer (1995)). The departure point will be theorem 2.3 on a supermartingale property of the exponential of a local martingale. The first application, corollary 3.1 yields the principal argument leading to (1.3) in theorem 3.2. For two other consequences, see sections 4 and 5. As was mentioned earlier, the report is closed by the application to the asymptotically continuous case.

2 Exponential of a martingale

We will use throughout standard notions of the general theory of stochastic processes. For more details we refer to Lipster and Shiryayev (1989) or Jacod and Shiryayev (1987). Processes under consideration will always be defined on a certain fixed stochastic basis (except in section 6 where the sequence of processes is treated each defined on its own stochastic basis). We assume, for simplicity, that the martingale M in question starts from zero, i.e. $M_0 = 0$, so that its canonical representation reads

$$M = M^c + x * (\mu - \nu)$$

where M^c is the continuous part of M and μ its jump measure with the compensator ν. When the latter characteristic ν is so that at each time $t \geq 0$

$$\varphi * \nu_t < \infty \quad \text{a.s.}, \quad \varphi(x) = e^x - 1 - x, \tag{2.1}$$

we may associate with M the so-called *cumulant* process

$$G = \tfrac{1}{2} \langle M^c \rangle + \varphi * \nu. \tag{2.2}$$

Here and elsewhere below we agree upon the standard notation $\varphi(x) * \nu_t$ for the pathwise stochastic integral $\int_0^t \int \varphi_s(x) \nu(ds \times dx)$ (x stands for a dummy variable $\in \mathbb{R}$). Since

$$\Delta G_t = \int (e^x - 1)\, \nu(\{t\} \times dx) > -1 \quad \text{a.s.} \tag{2.3}$$

at each $t \geq 0$ (cf Lipster and Shiryayev (1989), p. 346), the Doléans-Dade exponential $\mathcal{E}(G)$ is well-defined and

$$\mathcal{E}_t(G) = e^{G_t} \prod_{s \leq t} (1 + \Delta G_s) e^{-\Delta G_s}. \tag{2.4}$$

The latter process occurs as the compensator in the following multiplicative decomposition of a positive semimartingale $\exp(M)$ (cf Lipster and Shiryayev (1989), section 4.13):

Theorem 2.1. *Let M be a local martingale whose characteristic ν satisfies the condition (2.1). Then there exists a nonnegative local martingale N such that*

$$e^M = N \mathcal{E}(G). \tag{2.5}$$

Remark 2.2. The process $\exp(M - G)$ is a supermartingale, since by (2.5) the martingale part in its multiplicative decomposition is again N while the compensator equals to $\mathcal{E}(G) \exp(-G)$ that is nonincreasing. Indeed, $\log(1 + x) - x \leq 0$ for $x > -1$ and it follows from (2.4) that the Doléans-Dade exponential $\mathcal{E}(G)$ is dominated by the usual exponential $\exp(G)$.

Like in Lipster and Shiryayev (1989) it is said that a local martingale M satisfies *Cramér's condition* for $\lambda \in \mathbb{R}$ if at each time $t \geq 0$

$$\varphi(\lambda x) * \nu_t < \infty \quad \text{a.s.}, \tag{2.6}$$

where φ is again given by (2.1). Under this condition the cumulant process $G(\lambda)$ for λM is well-defined,

$$G(\lambda) = \tfrac{1}{2}\lambda^2 \langle M^c \rangle + \varphi(\lambda x) * \nu, \tag{2.7}$$

and in view of the latter remark $\exp(\lambda M - G(\lambda))$ is a supermartingale. The following theorem may be regarded as an extension of this fact (to see this, put $f \equiv 0$).

Theorem 2.3. *Let M be a local martingale whose characteristic ν is such that for a certain function f with $f(0) = 0$ and a fixed $\lambda \in \mathbb{R}$*

$$|f(\lambda x)| * \nu < \infty, \quad \varphi(\lambda x - f(\lambda x)) * \nu < \infty \quad \textit{a.s.} \tag{2.8}$$

with φ as in (2.1). Define the submartingale $S(\lambda) = M(\lambda) + A(\lambda)$ with the martingale part $M(\lambda) = f(\lambda x) * (\mu - \nu)$ and the nondecreasing compensator $A(\lambda) = \frac{1}{2}\lambda^2 \langle M^c \rangle + \varphi(\lambda x - f(\lambda x)) * \nu$. Then

$$e^{\lambda M - S(\lambda)}$$

is a supermartingale.

Proof. We apply theorem 2.1 to the local martingale $M'(\lambda) = \lambda M - M(\lambda)$. Since $\langle M'(\lambda)^c \rangle = \lambda^2 \langle M^c \rangle$ and the compensator $\nu'(\lambda)$ to the jump measure of $M'(\lambda)$ is so that $\varphi * \nu'(\lambda) = \varphi(\lambda x - f(\lambda x)) * \nu$, the cumulant process $G'(\lambda)$ for $M'(\lambda)$ is given by

$$\begin{aligned} G'(\lambda) &= \tfrac{1}{2} \left\langle M'(\lambda)^c \right\rangle + \varphi * \nu'(\lambda) \\ &= \tfrac{1}{2}\lambda^2 \langle M^c \rangle + \varphi(\lambda x - f(\lambda x)) * \nu = A(\lambda). \end{aligned}$$

In view of remark 2.2 we thus have that $\exp(M'(\lambda) - G'(\lambda)) = \exp(\lambda M - S(\lambda))$ is a supermartingale. □

3 Inequality for a square integrable martingale

The present section is devoted to the proof of our inequality (1.3), see theorem 3.2 below. It is preceded by an application of theorem 2.3 under the special choice (3.2) of the function f.

Corollary 3.1. *Let M be a locally square integrable martingale and $a \geq 0$. Then for all $\lambda \in (0, 1/a)$ the process*

$$e^{\lambda M - \psi_a(\lambda)\, H^a} \tag{3.1}$$

is a supermartingale, where $\psi_a(\lambda) = \lambda^2/(2 - 2\lambda a)$ and H^a is given by (1.2).

Proof. We need to verify the conditions of theorem 2.3 with the special choice

$$f(x) = \tfrac{1}{2} x^2 1_{\{|x| \geq \lambda a\}}. \tag{3.2}$$

Obviously $|f(\lambda x)| * \nu = \frac{1}{2}\lambda^2 x^2 1_{\{|x|\geq a\}} * \nu < \infty$ a.s. since M is locally square integrable. To show that $\varphi(\lambda x - f(\lambda x)) * \nu < \infty$ a.s. as well, we have to take into consideration the following arguments: for $|x| < a$

$$\varphi(\lambda x) \leq \sum_{n\geq 2} \frac{\lambda^n |x|^n}{n!} \leq \tfrac{1}{2} \sum_{n\geq 2} \lambda^n |x|^n \leq \psi_a(\lambda) x^2, \tag{3.3}$$

and for $|x| \geq a$

$$|e^{\lambda x - \frac{1}{2}\lambda^2 x^2} - 1 - \lambda x| \leq \tfrac{1}{2}\lambda^2 x^2 \leq \psi_a(\lambda) x^2 \tag{3.4}$$

with the same $\psi_a(\lambda)$ as in (3.1). Indeed, $|\exp(x - \frac{1}{2}x^2) - 1 - x| \leq \frac{1}{2}x^2$ for all $x \in \mathbb{R}$. The second inequality in (2.8) is then a straightforward consequence of the local square integrability of M. By theorem 2.3 $\exp(\lambda M - S(\lambda))$ is a supermartingale. To complete thus the proof, i.e. to show that for $\lambda \in (0, 1/a)$ the process (3.1) is a supermartingale as well, it remains only to verify that with the special substitution (3.2) of the function f the difference $\psi_a(\lambda)\, H^a - S(\lambda)$ is decreasing. To this end, apply again the inequalities (3.3) and (3.4). □

Theorem 3.2. *Let M be a locally square integrable martingale. The associated processes M^* and H^a with $a \geq 0$ (cf (1.1) and (1.2)), evaluated at a finite stopping time τ, satisfy the Bernstein inequality (1.3).*

Proof. Since H^a is nondecreasing we have for each $\lambda \in (0, 1/a)$ and $z, L \geq 0$ that

$$\mathrm{P}\left(\sup_{t\leq\tau} M_t \geq x, H^a_\tau \leq L\right) \leq \mathrm{P}\left(\sup_{t\leq\tau} Z_t(\lambda) \geq e^{\lambda z - \psi_a(\lambda) L}\right) \tag{3.5}$$

where $Z(\lambda)$ stands for the positive supermartingale (3.1). Since $Z_0(\lambda) = 1$, the optional sampling theorem yields for any stopping time σ the inequality

$$\mathrm{E}\, Z_\sigma(\lambda) 1_{\{\sigma<\infty\}} \le 1. \tag{3.6}$$

By applying this to a particular stopping time, namely to $\sigma = \inf\{t : Z_t(\lambda) \ge \exp(\lambda z - \psi_a(\lambda)L)\}$, we may extend (3.5) as follows:

$$\mathrm{P}\left(\sup_{t\le\tau} Z_t(\lambda) \ge e^{\lambda z - \psi_a(\lambda)L}\right) \le \mathrm{P}(\sigma \le \tau) \le \mathrm{P}(\sigma < \infty) \le e^{\psi_a(\lambda)L - \lambda z}.$$

We have first taken into consideration that τ is finite and then applied inequality (3.6). So (3.5) turns into

$$\mathrm{P}\left(\sup_{t\le\tau} M_t \ge z, H^a_\tau \le L\right) \le e^{\psi_a(\lambda)L - \lambda z}$$

for each $\lambda \in (0, 1/a)$. Clearly, the same inequality holds with M substituted by $-M$, for the process H^a remains unaltered. Thus

$$\mathrm{P}\left(M^*_\tau \ge z\,,\, H^a_\tau \le L\right) \le 2e^{\psi_a(\lambda)L - \lambda z} \tag{3.7}$$

for each $\lambda \in (0, 1/a)$, in particular for the choice

$$\lambda = \frac{x/L}{1 + ax/L} \tag{3.8}$$

that yields the desired inequality (1.3). $\square$

The following simple corollary of theorem 3.2 clarifies the relationship between the classical Bernstein inequality for martingales with bounded jumps and a result of Barlow et al. (1986). These two inequalities may be viewed as the extreme cases of (1.3).

Corollary 3.3.
(i) Let M be a locally square integrable martingale and let τ be a finite stopping time. Then

$$\mathrm{P}\left(M^*_\tau \ge z,\, H^0_\tau \le L\right) \le 2e^{-\frac{1}{2}\frac{z^2}{L}} \tag{3.9}$$

for all $z, L \ge 0$, where $H^0 = [M] + \langle M - M^c\rangle$.
(ii) Besides, if $|\Delta M| \le a$, then

$$\mathrm{P}\left(M^*_\tau \ge z, \langle M\rangle_\tau \le L\right) \le 2e^{-\frac{1}{2}\frac{z^2}{az+L}} \tag{3.10}$$

for all $z, L \ge 0$.

Proof. (i) Clearly $x^2 * \mu + \langle M \rangle = H^0$, and the assertion follows directly from theorem 3.2 as $a = 0$. (ii) Since $|\Delta M| \le a$, the jump measure μ of M is almost surely concentrated on $\mathbb{R}^+ \times [-a, a]$. So

$$H^{a+\varepsilon} = x^2 1_{\{|x| \ge a+\varepsilon\}} * \mu + \langle M \rangle = \langle M \rangle$$

as $\varepsilon > 0$. Theorem 3.2 thus implies that for each $\varepsilon > 0$

$$\mathrm{P}\left(M_\tau^* \ge z, \langle M \rangle_\tau \le L\right) = \mathrm{P}\left(M_\tau^* \ge z, H_\tau^{a+\varepsilon} \le L\right) \le 2e^{-\frac{1}{2}\frac{z^2}{(a+\varepsilon)z+L}}.$$

Now let $\varepsilon \downarrow 0$. □

4 Conditionally symmetric martingales

It is said that a local martingale is *conditionally symmetric* if the predictable characteristic of its jumps ν is such that for all integrable functions f

$$f(x) * \nu = f(-x) * \nu.$$

This notion occurs usually in the discrete time setup, in particular, within the theory of decoupling, cf e.g. de la Peña and Giné (1999) or de la Peña (1999). Our next result is in fact an extension of the discrete time result in the latter paper, section 6. Note that under the present conditions (2.7) reduces to

$$G(\lambda) = \tfrac{1}{2}\lambda^2 \langle M^c \rangle + (\cosh(\lambda x) - 1) * \nu \tag{4.1}$$

and assertion (i) of corollary 3.3 simplifies to

Theorem 4.1. *Let M be a locally square integrable and conditionally symmetric martingale. Then at each finite stopping time τ*

$$\mathrm{P}\left(M_\tau^* \ge z,\, [M]_\tau \le L\right) \le 2e^{-\frac{1}{2}\frac{z^2}{L}} \tag{4.2}$$

for all $z, L \ge 0$.

Proof. Apply theorem 2.3 with $f(x) = \frac{1}{2}x^2$. Taking also into consideration the conditional symmetry, we get

$$S(\lambda) - \tfrac{1}{2}\lambda^2[M] = (e^{-\frac{1}{2}\lambda^2 x^2}\cosh(\lambda x) - 1) * \nu$$

which is decreasing since $\exp(-x^2/2)\cosh x \le 1$ for all $x \in \mathbb{R}$. This means, in particular, that

$$e^{\lambda M - \frac{1}{2}\lambda^2 [M]}$$

is a supermartingale. Apply now the arguments like in the course of proving theorem 3.2 but with the latter supermartingale in the place of $Z(\lambda)$. Depart namely from the inequality

$$\mathrm{P}\left(\sup_{t\leq\tau} M_t \geq x, [M]_\tau \leq L\right) \leq \mathrm{P}\left(\sup_{t\leq\tau} e^{\lambda M_t - \frac{1}{2}\lambda^2 [M]_t} \geq e^{\lambda z - \frac{1}{2}\lambda^2 L}\right)$$

and obtain

$$\mathrm{P}\left(M_\tau^* \geq z\,,\, [M]_\tau \leq L\right) \leq 2e^{\frac{1}{2}\lambda^2 L - \lambda z},$$

cf (3.5) and (3.7), respectively. To complete the proof, select λ as to render the right-hand side as small as possible. □

5 Inequality under Cramér's condition

In attempt to relax the assumption of bounded jumps in assertion (ii) of corollary 3.3, Bennet (1962) has treated the sum of independent random variables possessing certain exponential moments and obtained the inequality similar to (3.10) but with the predictable characteristic H as in (5.2) in place of $\langle M\rangle$ (see Van de Geer (1995) and Nishiyama (1998) for the extension to discrete-time martingales and point processes with continuous intensity). It will be shown next how to extend this to the general situation of our interest.

Theorem 5.1. *Let M be a square integrable martingale. Suppose that there exists a positive constant $a > 0$ so that at each time $t \geq 0$*

$$\varphi(|x|/a) * \nu_t < \infty \quad a.s. \tag{5.1}$$

(with φ as in (2.1)) and define the nondecreasing predictable process

$$H = \langle M^c\rangle + 2a^2\varphi(|x|/a) * \nu.$$

Then at each finite stopping time τ

$$\mathrm{P}\left(M_\tau^* \geq z, H_\tau \leq L\right) \leq 2e^{-\frac{1}{2}\frac{z^2}{az+L}} \tag{5.2}$$

for every $z, L \geq 0$.

Proof. Assumption (5.1) implies that Cramér's condition (2.6) holds for every $\lambda \in (0, 1/a)$. Indeed, for $\lambda \in (0, 1/a)$ and $x \in \mathbb{R}$ we have

$$\varphi(\lambda x) \leq \sum_{n \geq 2} a^n \lambda^n \frac{(|x|/a)^n}{n!} \leq 2a^2 \psi_a(\lambda) \varphi(|x|/a)$$

with the same $\psi_a(\lambda)$ as in (3.1). By the latter inequality and the fact that $\psi_a(\lambda) \geq \lambda^2/2$, the difference between the cumulant process $G(\lambda)$ for λM (cf (2.7)) and the process $\psi_a(\lambda)H$ is decreasing. It follows then by the same arguments as in the course of proving corollary 3.1 that $\exp(\lambda M - \psi_a(\lambda)H)$ is a supermartingale. Repeat now the arguments proving theorem 3.2 but use the latter supermartingale instead of (3.1). We arrive at the inequality

$$\mathrm{P}\left(M_\tau^* \geq z, H_\tau \leq L\right) \leq 2e^{\psi_a(\lambda)L - \lambda z}$$

for each $\lambda \in (0, 1/a)$, cf (3.7). Finally, select λ as in (3.8) to arrive at the desired inequality. □

6 Asymptotic inequality

The principal result in this section, inequality (6.2) may be regarded as an asymptotic Bernstein inequality, for it resembles (1.4) for continuous local martingales and requires the asymptotic continuity of the sequence of locally square integrable martingales. This requirement is expressed as usual in the form of Lindeberg condition (6.1), cf e.g. Jacod and Shiryaev (1987) or Liptser and Shiryayev (1989).

Thus, we will deal here with the sequence of locally square integrable martingales $\{M^n\}_{n=1,2,\ldots}$. Therefore all its characteristics and associated processes will be indexed as well. Otherwise, the previous notational conventions will be retained. Theorem 3.2 yields

Corollary 6.1. *Let $\{M^n\}_{n=1,2,\ldots}$ be the sequence of locally square integrable martingales. Suppose the corresponding sequence of predictable characteristics $\{\nu^n\}_{n=1,2,\ldots}$ is so that at a certain stopping times τ_n*

$$x^2 1_{\{|x|>\varepsilon\}} * \nu^n_{\tau_n} \xrightarrow{\mathrm{P}} 0 \tag{6.1}$$

for all $\varepsilon > 0$. Then

$$\limsup_{n\to\infty} \mathrm{P}\left(M^{n*}_{\tau_n} \geq z, \langle M^n\rangle_{\tau_n} \leq L\right) \leq 2e^{-\frac{1}{2}\frac{x^2}{L}} \tag{6.2}$$

for all $z, L \geq 0$.

Proof. Obviously, the probability in (6.2) can be bounded from above by the sum

$$\mathrm{P}\left(M^{n*}_{\tau_n} \geq z, H^{n\varepsilon}_{\tau_n} \leq K + L\right) + \mathrm{P}\left(x^2 1_{\{|x| \geq \varepsilon\}} * \mu^n_{\tau_n} > K\right).$$

for every $\varepsilon, K > 0$ (the process H^{na} is again given by (1.2)). By theorem 3.2 the first term does not exceed $2\exp(-z^2/2(\varepsilon z + L + K))$, while the second term vanishes as $n \to \infty$ in virtue of (6.1) and the Lenglart inequality. Hence

$$\limsup_{n\to\infty} \mathrm{P}\left(M^{n*}_{\tau_n} \geq z, \langle M^n \rangle_{\tau_n} \leq L\right) \leq 2e^{-\frac{1}{2}\frac{z^2}{\varepsilon z + L + K}}.$$

for every $\varepsilon, K > 0$. The proof is completed by letting $\varepsilon \downarrow 0$ and $K \downarrow 0$. $\square$

References

M.T. Barlow, S.D. Jacka and M. Yor (1986), Inequalities for a pair of processes stopped at a random time, *Proc. London Math. Soc.* **3**, 52, 142–172.

G. Bennett (1962), Probability inequalities for the sum of independent random variables, *JASA* **57**, 33–45.

V.H. de la Peña and E. Giné (1999), *Decoupling*, Springer.

V.H. de la Peña (1999), A general class of exponential inequalities for martingales and ratios, *Ann. Probab.* **27**, 537–564.

D.A. Freedman (1975), On tail probabilities for martingales, *Ann. Probab.* **3**, 100–118.

J. Jacod and A.N. Shiryaev (1987), *Limit Theorems for Stochastic Processes*, Springer.

R.Sh. Lipster and A.N. Shiryayev (1989). *Theory of Martingales*, Kluwer.

Y. Nishiyama (1998). *Entropy Methods for Martingales*, PhD-thesis, University of Utrecht.

G.R. Shorack and J.A. Wellner (1986). *Empirical Processes with Applications to Statistics*, Wiley.

S. van de Geer (1995), Exponential inequalities for martingales, with applications to maximum likelihood estimation for counting processes, *Ann. Stat.* **23**, 1779–1801.