The Annals of Probability
1999, Vol. 27, No. 1, 537–564

A GENERAL CLASS OF EXPONENTIAL INEQUALITIES FOR MARTINGALES AND RATIOS[1]

By Victor H. de la Peña

Columbia University

In this paper we introduce a technique for obtaining exponential inequalities, with particular emphasis placed on results involving ratios. Our main applications consist of approximations to the tail probability of the ratio of a martingale over its conditional variance (or its quadratic variation for continuous martingales). We provide examples that strictly extend several of the classical exponential inequalities for sums of independent random variables and martingales. The spirit of this application is that, when going from results for sums of independent random variables to martingales, one should replace the variance by the conditional variance and the exponential of a function of the variance by the expectation of the exponential of the same function of the conditional variance. The decoupling inequalities used to attain our goal are of independent interest. They include a new exponential decoupling inequality with constraints and a sharp inequality for the probability of the intersection of a fixed number of dependent sets. Finally, we also present an exponential inequality that does not require any integrability conditions involving the ratio of the sum of conditionally symmetric variables to its sum of squares.

0. Introduction. In this paper we introduce a technique for obtaining (a new class of) exponential inequalities which as special cases contain several of the known results for sums of independent variables and martingales. Our approach seems to be useful in obtaining extensions of exponential inequalities that are derived based on the use of Markov's inequality and the moment generating function, including results of Hoeffding (1963), Freedman (1975), Pinelis and Utev (1989), Hitczenko (1990b) and finally, Pinelis (1992, 1994) for discrete time martingales and McKean (1962) and Khoshnevisan (1996) for continuous time martingales. In some instances our results improve on the known inequalities (under expanded conditions). In brief, what we introduce is a new technique for obtaining exponential inequalities, with special emphasis placed on results involving ratios. We provide several new results but did not attempt to include all possible applications of our method, as that would have been too time-consuming. Instead, we present a technique and several examples showing how to apply it.

The paper is divided as follows. In Section 1 we present the new exponential inequalities for discrete time martingales and show how they compare with known results. In Section 2 we present a brief introduction to the theory

Received June 1997; revised June 1998.
[1]Supported in part by NSF Grant DMS-96-26175.
AMS 1991 *subject classifications*. Primary 60E15, 60G42, 60G44; secondary 60G40.
Key words and phrases. Self-normalized, exponential martingale inequalities, decoupling.

of decoupling. In Section 3 we derive several new decoupling inequalities including a surprising sharp decoupling inequality for the probability of the intersection of dependent sets (Corollary 3.4). In this section we also obtain an exponential inequality with constraints (Corollary 3.1) that is the cornerstone of the approach used in Section 4 to obtain the proofs of several of the inequalities of Section 1. Section 5 presents exponential inequalities for continuous time martingales and Section 6 inequalities for the ratio of a sum of conditionally symmetric variables to its sum of squares, showing that the applicability of our technique is not restricted to problems involving martingales (no integrability assumptions are made). The Appendix provides further extensions. The paper is written in such a way that one can read Sections 1–4 independently of Sections 5 and 6.

1. New exponential inequalities. We will begin analyzing our approach with the special case of Bernstein's and Bennett's inequalities which we state next [cf. Chow and Teicher (1988), Exercise 4.3.14 and Bennett (1962).]

THEOREM 1.1. *Let* $\{x_i\}$ *be a sequence of independent random variables with* $S_n = \sum_{i=1}^n x_i$, $Ex_j = 0$, $Ex_j^2 < \infty$, $v_n^2 = \sum_{j=1}^n Ex_j^2$. *Furthermore, assume that* $E|x_j|^k \le (k!/2)Ex_j^2 c^{k-2}$ *or* $P(|x_j| \le c) = 1$, *for* $k > 2$, $0 < c < \infty$. *Then, for all* $x > 0$,

$$P\left(\sum_{i=1}^n x_i \ge x\right) \le \exp\left\{-\frac{x^2}{v_n^2\left(1+\sqrt{1+2cx/v_n^2}\right)+cx}\right\} \le \exp\left\{-\frac{x^2}{2\left(v_n^2+cx\right)}\right\}. \tag{1.1}$$

In turn, for martingales we can derive the following inequalities.

THEOREM 1.2A. *Let* $\{d_i, \mathscr{F}_i\}$ *be a martingale difference sequence with* $E(d_j \mid \mathscr{F}_{j-1}) = 0$, $E(d_j^2 \mid \mathscr{F}_{j-1}) = \sigma_j^2$, $V_n^2 = \sum_{j=1}^n \sigma_j^2$. *Furthermore, assume that* $E(|d_j|^k \mid \mathscr{F}_{j-1}) \le (k!/2)\sigma_j^2 c^{k-2}$ *a.e. or* $P(|d_j| \le c \mid \mathscr{F}_{j-1}) = 1$ *for* $k > 2$, $0 < c < \infty$. *Then, for all* $x, y > 0$,

$$P\left(\sum_{i=1}^n d_i \ge x, V_n^2 \le y \text{ for some } n\right) \le \exp\left\{-\frac{x^2}{y\left(1+\sqrt{1+2cx/y}\right)+cx}\right\} \le \exp\left\{-\frac{x^2}{2(y+cx)}\right\}. \tag{1.2}$$

In the special case of independent random variables, if we set $y = V_n^2$, we get exactly Theorem 1.1. This result (which might be available in the literature) should be compared to Freedman's [(1975), Theorem 1.6].

REMARK 1.1. In order to avoid potential problems with the definition of conditional expectations, in this paper we will use the following notation. Let X be a positive random variable on the probability space $(\Omega, \mathscr{F}, P)$. Let A be an $\mathscr{F}$-measurable set, then $E(X \mid A) = \int_A X\,dP/P(A)$. Our concern is that, as stated by Example 8.8 of Wise and Hall (1993), "Conditional expectations of an integrable random variable need not be obtainable from a corresponding conditional probability distribution."

Looking at the problem of extending Theorem 1.1, we considered normalizing the sums by their conditional variances and obtained the next result.

THEOREM 1.2B. *Let $\{d_i, \mathscr{F}_i\}$ be a martingale difference sequence with $E(d_j \mid \mathscr{F}_{j-1}) = 0$, $E(d_j^2 \mid \mathscr{F}_{j-1}) = \sigma_j^2$, $V_n^2 = \sum_{j=1}^n \sigma_j^2$. Furthermore, assume that $E(|d_j|^k \mid \mathscr{F}_{j-1}) \le (k!/2)\sigma_j^2 c^{k-2}$ a.e. or $P(|d_j| \le c \mid \mathscr{F}_{j-1}) = 1$ for $k > 2$, $0 < c < \infty$. Then, for all $\mathscr{F}_\infty$ measurable sets A, $x > 0$,*

$$P\left(\frac{\sum_{i=1}^n d_i}{V_n^2} \ge x,\ A\right) \le E\left[\exp\left\{-\left(\frac{x^2}{1 + cx + \sqrt{1 + 2cx}}\right)V_n^2\right\}\Bigg|\left(\frac{M_n}{V_n^2} \ge x,\ A\right)\right], \tag{1.3}$$

$$P\left(\frac{\sum_{i=1}^n d_i}{V_n^2} \ge x,\ A\right) \le \sqrt{E\left[\exp\left\{-\left(\frac{x^2}{1 + cx + \sqrt{1 + 2cx}}\right)V_n^2\right\}1_A\right]} \tag{1.4}$$

and

$$P\left(\frac{\sum_{i=1}^n d_i}{V_n^2} \ge x,\ \frac{1}{V_n^2} \le y \text{ for some } n\right) \le \exp\left\{-\frac{1}{y}\left(\frac{x^2}{1 + cx + \sqrt{1 + 2cx}}\right)\right\} \le \exp\left\{-\frac{x^2}{2y(1 + cx)}\right\}. \tag{1.5}$$

We observe that the right-hand side of (1.3) is bounded by

$$E\left[\exp\left\{-\frac{x^2 V_n^2}{2(1 + cx)}\right\}\Bigg|\left(\frac{M_n}{V_n^2} \ge x,\ A\right)\right]. \tag{1.6}$$

REMARK 1.2. We remark that (1.5) points to a different minimizing value than the one given in Theorem 3.3 of Pinelis (1994). We checked our result several times to verify its accuracy. The original version of the paper had quantity $(x/c + 1/c^2 - \sqrt{2cx + 1}/c^2)$ (which was obtained by using Mathematica) in the exponent instead of the more streamlined expression $(x^2/(1 + cx + \sqrt{1 + 2cx}))$. We are grateful to Giné and Pinelis for pointing us to the more streamlined form of the exponent.

In the case the $d_i = x_i$'s are independent (1.3), (1.5) and (1.6) can be used to obtain (1.1) by replacing V_n by the constant v_n and next letting $x = x'/v_n^2$ and $y = 1/v_n^2$. The fact that v_n is a constant produces a situation where the conditioning in (1.6) is no longer in effect.

If indeed one has a martingale, letting $V_n^2 = \|\sum_{i=1}^{\infty} \sigma_i^2\|_\infty$ (a finite constant) in the proof of (1.3) and (1.6), a change of variables gives

$$
\begin{aligned}
P\left(\sum_{i=1}^{n} d_i \geq x\right) &\leq \exp\left\{\frac{-x^2}{V_n^2 + cx + V_n\sqrt{V_n^2 + 2cx}}\right\} \\
&\leq \exp\left\{\frac{-x^2}{2\left(V_n^2 + cx\right)}\right\},
\end{aligned} \tag{1.7}
$$

from (1.3). By following the proofs given in Section 4, it is easy to see that all the inequalities presented in this section are valid when taking $V_n^2 = \|\sum_{i=1}^{n} \sigma_i^2\|_\infty$. A change of variables as done in (1.7) shows that this replacement gives us the typical martingale results. This observation validates our claim that our approach provides extensions of the typical exponential inequalities for martingales.

Related results are given next, where we divide by $\alpha + \beta V_n^2$ instead of by V_n^2 only. It will be apparent later that it is easy to extend (1.3)–(1.5) to include linear combinations.

THEOREM 1.3. *Let $\{d_i, \mathscr{F}_i\}$ be a supermartingale difference sequence (i.e., $E(d_j \mid \mathscr{F}_{j-1}) \leq 0$) with $E(d_j^2 \mid \mathscr{F}_{j-1}) = \sigma_j^2$, $V_n^2 = \sum_{j=1}^{n} \sigma_j^2$ or $V_n^2 = \|\sum_{i=1}^{n} \sigma_i^2\|_\infty$. Furthermore, assume that $d_j \leq 1$. Then, for all $\mathscr{F}_\infty$ measurable sets A and all $\beta > 0$, $\alpha, x \geq 0$,*

$$
\begin{aligned}
&P\left(\frac{\sum_{i=1}^{n} d_i}{\alpha + \beta V_n^2} \geq x, A\right) \\
&\quad \leq \left(\frac{1}{\beta x + 1}\right)^{\alpha x} E\left[\left(\exp\{\beta x\}\left(\frac{1}{\beta x + 1}\right)^{\beta x + 1}\right)^{V_n^2} \middle| \left(\frac{M_n}{\alpha + \beta V_n^2} \geq x, A\right)\right]
\end{aligned} \tag{1.8}
$$

$$
\begin{aligned}
&P\left(\frac{\sum_{i=1}^{n} d_i}{\alpha + \beta V_n^2} \geq x, A\right) \\
&\quad \leq \left(\frac{1}{\beta x + 1}\right)^{\alpha x} E\left[\exp\left\{-\frac{\beta^2 x^2 V_n^2}{2(1 + \beta x)}\right\} \middle| \left(\frac{M_n}{\alpha + \beta V_n^2} \geq x, A\right)\right]
\end{aligned} \tag{1.9}
$$

and

$$
P\left(\frac{\sum_{i=1}^{n} d_i}{\alpha + \beta V_n^2} \geq x, A\right) \leq \sqrt{\left(\frac{1}{\beta x + 1}\right)^{\alpha x} E \exp\left\{-\frac{\beta x^2 V_n^2}{2(1 + \beta x)}\right\} 1_A} \tag{1.10}
$$

and finally,

$$(1.11)\qquad \begin{aligned} &P\left(\frac{\sum_{i=1}^n d_i}{\alpha+\beta V_n^2}\ge x,\ \frac{1}{V_n^2}\le y \text{ for some } n\right)\\ &\quad\le\left(\frac{1}{\beta x+1}\right)^{\alpha x}\left[\exp\{\beta x\}\left(\frac{1}{\beta x+1}\right)^{\beta x+1}\right]^{1/y}\\ &\quad\le\left(\frac{1}{\beta x+1}\right)^{\alpha x}\exp\left\{-\frac{\beta^2x^2}{y(1+\beta x)}\right\}. \end{aligned}$$

THEOREM 1.4A. *Let* $\{d_i,\mathscr{F}_i\}$ *be a supermartingale difference sequence with* $E(d_j\mid\mathscr{F}_{j-1})\le 0$, $|d_j|\le c$, *for* $0<c<\infty$, $E(d_j^2\mid\mathscr{F}_{j-1})=\sigma_j^2$ *and* $V_n^2=\sum_{i=1}^n\sigma_i^2$ *or* $V_n^2=\|\sum_{i=1}^n\sigma_i^2\|_\infty$. *Then, for all* $x, y>0$,

$$(1.12)\qquad \begin{aligned} &P\left(\sum_{i=1}^n d_i\ge x,\ V_n^2\le y \text{ for some } n\right)\\ &\quad\le\exp\left\{-\left[\frac{x}{c}-\left(\frac{x}{c}+\frac{y}{c^2}\right)\ln\left(1+\frac{cx}{y}\right)\right]\right\}. \end{aligned}$$

In the case of normalized sums we get the following theorem.

THEOREM 1.4B. *Let* $\{d_i,\mathscr{F}_i\}$ *be a supermartingale sequence with* $E(d_j\mid\mathscr{F}_{j-1})\le 0$, $|d_j|\le c$, *for* $0<c<\infty$, $E(d_j^2\mid\mathscr{F}_{j-1})=\sigma_j^2$ *and* $V_n^2=\sum_{i=1}^n\sigma_i^2$ *or* $V_n^2=\|\sum_{i=1}^n\sigma_i^2\|_\infty$. *Then, for all* $\mathscr{F}_\infty$*-measurable sets* A, $\beta>0$ *and* $\alpha, x\ge 0$,

$$(1.13)\qquad \begin{aligned} &P\left(\frac{\sum_{i=1}^n d_i}{\alpha+\beta V_n^2}\ge x,\ A\right)\\ &\quad\le\exp\left\{-\frac{\alpha x}{c}\ln(1+c\beta n)\right\}\\ &\qquad\times E\left[\exp\left\{-\left(\frac{\beta x}{c}-\left(\frac{\beta x}{c}+\frac{1}{c^2}\right)\ln(1+\beta cx)\right)V_n^2\right\}\,\middle|\,\left(\sum_{i=1}^n d_i\ge\left(\alpha+\beta V_n^2\right)x,\ A\right)\right]. \end{aligned}$$

Finally,

$$(1.14)\qquad \begin{aligned} &P\left(\frac{\sum_{i=1}^n d_i}{\alpha+\beta V_n^2}\ge x,\ \frac{1}{V_n^2}\le y \text{ for some } n\right)\\ &\quad\le\exp\left\{-\frac{\alpha x}{c}\ln(1+c\beta x)\right\}\\ &\qquad\times\exp\left\{-\left(\frac{\beta x}{c}-\left(\frac{\beta x}{c}+\frac{1}{c^2}\right)\ln(1+\beta cx)\right)\frac{1}{y}\right\}. \end{aligned}$$

This result extends the one given in Freedman (1975).

Next, we extend a result of Hitczenko (1990b) and Levental (1989) which is a martingale version of Prokhorov's inequality for sum of independent random variables.

THEOREM 1.5A. *Let* $\{d_i, \mathscr{F}_i\}$ *be a martingale difference sequence with* $E(d_j \mid \mathscr{F}_{j-1}) = 0$, $|d_j| \le c$, *for* $0 < c < \infty$, $E(d_j^2 \mid \mathscr{F}_{j-1}) = \sigma_j^2$ *and* $V_n^2 = \sum_{i=1}^n \sigma_i^2$ *or* $V_n^2 = \|\sum_{i=1}^n \sigma_i^2\|_\infty$. *Then, for all* $x, y > 0$,

$$P\left(\sum_{i=1}^{n} d_i \ge x,\ V_n^2 \le y \text{ for some } n\right) \le \exp\left\{-\frac{x}{2c}\operatorname{arc\,sinh}\left(\frac{cx}{2y}\right)\right\}. \tag{1.15}$$

For normalized sums, we have the following theorem.

THEOREM 1.5B. *Let* $\{d_i, \mathscr{F}_i\}$ *be a martingale difference sequence with* $E(d_j \mid \mathscr{F}_{j-1}) = 0$, $|d_j| \le c$, *for* $0 < c < \infty$, $E(d_j^2 \mid \mathscr{F}_{j-1}) = \sigma_j^2$ *and* $V_n^2 = \sum_{i=1}^n \sigma_i^2$ *or* $V_n^2 = \|\sum_{i=1}^n \sigma_i^2\|_\infty$. *Then, for all* $\mathscr{F}_\infty$*-measurable sets* A, *and all* $\beta > 0$ *and* $\alpha, x \ge 0$,

$$\begin{aligned} &P\left(\frac{\sum_{i=1}^n d_i}{(\alpha + \beta V_n^2)} \ge x,\ A\right) \\ &\quad \le \exp\left\{-\frac{\alpha x}{c}\operatorname{arc\,sinh}\left(\frac{\beta c x}{2}\right)\right\} \\ &\qquad \times E\left[\exp\left\{\left(-\frac{\beta x}{2c}\operatorname{arc\,sinh}\left(\frac{\beta c x}{2}\right)\right)V_n^2\right\}\middle|\left(M_n \ge (\alpha + \beta V_n^2)x,\ A\right)\right], \end{aligned} \tag{1.16}$$

$$\begin{aligned} &P\left(\frac{\sum_{i=1}^n d_i}{(\alpha + \beta V_n^2)} \ge x,\ A\right) \\ &\quad \le \sqrt{\exp\left\{-\frac{\alpha x}{c}\operatorname{arc\,sinh}\left(\frac{\beta c x}{2}\right)\right\} E\exp\left\{\left(-\frac{\beta x}{2c}\operatorname{arc\,sinh}\left(\frac{\beta c x}{2}\right)\right)V_n^2\right\} 1_A} \end{aligned} \tag{1.17}$$

and finally,

$$\begin{aligned} &P\left(\frac{\sum_{i=1}^n d_i}{\alpha + \beta V_n^2} \ge x,\ \frac{1}{V_n^2} \le y \text{ for some } n\right) \\ &\quad \le \exp\left\{-\frac{\alpha x}{c}\operatorname{arc\,sinh}\left(\frac{\beta c x}{2}\right)\right\}\exp\left\{-\frac{\beta x}{2cy}\operatorname{arc\,sinh}\left(\frac{\beta c x}{2}\right)\right\}. \end{aligned} \tag{1.18}$$

Again, specializing the above, we obtain the known results for martingales and sums of independent random variables. Before we can embark on proving the above exponential inequalities, we need to review several concepts and extend results of the theory of decoupling inequalities.

2. General theory of decoupling.

DEFINITION 2.1. Let $\{d_i\}$ be a sequence of random variables adapted to $\{\mathscr{F}_i\}$ (increasing) with $\mathscr{F}_0$ the trivial σ-field. An $\{\mathscr{F}_i\}$-adapted sequence $\{e_i\}$ is $\{\mathscr{F}_i\}$-tangent to $\{d_i\}$ if for all i,

$$\mathscr{L}(d_i \mid \mathscr{F}_{i-1}) = \mathscr{L}(e_i \mid \mathscr{F}_{i-1}). \tag{2.1}$$

DEFINITION 2.2. A sequence $\{e_i\}$ of random variables adapted to an increasing sequence of σ-fields $\{\mathscr{F}_i\}$ contained in $\mathscr{F}$ is said to satisfy condition CI if there exists a σ-algebra $\mathscr{G}$ contained in $\mathscr{F}$ such that $\{e_i\}$ is a sequence of conditionally independent random variables given $\mathscr{G}$ and

$$\mathscr{L}(d_i \mid \mathscr{F}_{i-1}) = \mathscr{L}(e_i \mid \mathscr{F}_{i-1}) = \mathscr{L}(e_i \mid \mathscr{G})$$

for all i. The sequence $\{e_i\}$ is said to be DECOUPLED.

REMARK 2.1. Concerning sequences of random sets $\{D_i\}$ and $\{E_i\}$, the above terms apply whenever their indicator variables satisfy the stated conditions.

A key result in the theory of decoupling states that, for any sequence $\{d_i\}$ of variables adapted to an increasing sequence of σ-fields $\{\mathscr{F}_i\}$, one can always find a sequence $\{e_i\}$ which is tangent to $\{d_i\}$ and satisfies the CI condition. Frequently, one can take $\mathscr{G} = \sigma(\{d_i\})$ [cf. Kwapień and Woyczyński (1989, 1992)]. The general framework that we will follow consists of applying this result conditionally on $\mathscr{G}$ and use known inequalities for sums of independent random variables.

Let $\mathscr{L}(Y \mid \mathscr{H})$ denote the regular version of the conditional distribution of X given a σ-field $\mathscr{H}$. One approach for constructing a decoupled (CI) sequence to any adapted process is to proceed sequentially [see de la Peña (1994) and Pinelis (1995)]. Let $\mathscr{G} = \sigma(\bigcup \mathscr{F}_j)$. At the jth stage in the process producing the sequence $\{d_i\}$, e_j is drawn as a conditionally independent copy of d_j, given $\mathscr{F}_{j-1}$. Therefore, we obtain e_j from $\mathscr{L}(d_j \mid \mathscr{F}_{j-1})$ and e_j is $\mathscr{F}_{j-1}$-conditionally independent of $e_1, \dots, e_{j-1}, \mathscr{G}$. The following diagram taken from de la Peña (1994) illustrates the idea:

$$\begin{array}{ccccccccccccc} d_1 & \to & d_2 & \to & d_3 & \to & d_4 & \to & \cdots & \to & d_{j-1} & \to & d_j \\ & \searrow & & \searrow & & \searrow & & \searrow & & & & \searrow & \\ e_1 & & e_2 & & e_3 & & e_4 & & \cdots & & e_{j-1} & & e_j. \end{array}$$

Therefore, as long as the regular versions of the conditional distributions exist [see Shiryayev (1984)], then $\mathscr{L}(d_i \mid \mathscr{F}_{i-1}) = \mathscr{L}(e_i \mid \mathscr{F}_{i-1}) = \mathscr{L}(e_i \mid \mathscr{G})$, and it also follows that the constructed variables $\{e_i\}$ are conditionally independent given $\mathscr{G}$. In fact, the sequence $\{e_i\}$ is $\{\mathscr{F}_i'\}$-tangent to $\{d_i\}$ with $\mathscr{F}_i' = \sigma(\mathscr{F}_i; e_1, \dots, e_i)$ and moreover $\{e_i\}$ satisfies the CI condition with respect to $\{\mathscr{F}_i'\}$ and $\mathscr{G}$, which is contained in $\mathscr{F}_\infty'$. Therefore, renaming $\mathscr{F}_i'$ as $\mathscr{F}_i$, we have that $\{e_i\}$ is a decoupled $\{\mathscr{F}_i\}$-tangent sequence to $\{d_i\}$.

The following lemma, which we take from Hitczenko (1990a), provides a condensed version of the above cited result of Kwapień and Woyczyński.

LEMMA 2.1. *Let* $\{d_i\}$ *be any adapted sequence of random variables on* $(\Omega, \mathscr{F}, P)$ *and let* $\mathscr{B}_n$ *be the σ-algebra generated by the first n coordinates in* R^N, $\mathscr{B} = \sigma(\bigcup \mathscr{B}_n)$. *A new probability space* $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$ *and the sequence* $\{\tilde{\mathscr{F}}_n\}$ *combining the ones already introduced is defined as follows*:

$$\tilde{\Omega} = \Omega \times R^N, \qquad \tilde{\mathscr{F}} = \mathscr{F} \otimes \mathscr{B}, \qquad \tilde{\mathscr{F}}_n = \mathscr{F}_n \otimes \mathscr{B}_n, \tag{2.3}$$

$$\tilde{P}(A \times B) = \int_A \left(\bigotimes_{n=1}^{\infty} \mathscr{L}(d_n \mid \mathscr{F}_{n-1})(\omega) \right)(B)\, dP(\omega), \qquad A \in \mathscr{F}, B \in \mathscr{B}. \tag{2.4}$$

Letting $d_n(\omega, \{t_k\}) = d_n(\omega)$ *and* $e_n(\omega, \{t_k\}) = t_n$, *then the sequences* $\{d_n\}$ *and* $\{e_n\}$ *are* $\{\tilde{\mathscr{F}}_n\}$*-tangent with* $\{e_n\}$ *satisfying the CI condition for* $\mathscr{G} = \mathscr{F} \otimes \{\varnothing, R^N\}$.

Next, we present an example that helps illuminate the concepts presented above.

EXAMPLE 2.1. Let $\{X_i\}$ be a sequence of independent mean zero random variables. Let T be a stopping time adapted to $\sigma(\{X_i\})$. Let $\{\tilde{X}_i\}$ be an independent copy of $\{X_i\}$ and hence independent of T as well. Let $\mathscr{F}_n = \sigma(\{X_i\})$. Set $M_n = M_{T \wedge n} = \sum_{i=1}^n X_i 1(T \geq i) = \sum_{j=1}^n d_j$ and $U_n = \sum_{j=2}^n (\sum_{i=1}^{j-1} X_i) X_j = \sum_{j=2}^n d'_j$. Both $\{M_n, \mathscr{F}_n, n \geq 1\}$ and $\{U_n, \mathscr{F}_n, n \geq 2\}$ are mean zero martingales with highly dependent martingale difference sequences $d_j = X_j 1(T \geq j)$, $j \geq 1$ and $d'_j = \sum_{i=1}^{j-1} X_i X_j$, $j \geq 2$. Let $\mathscr{F}_i' = \sigma(X_1, \dots, X_i;\ \tilde{X}_1, \dots, \tilde{X}_i)$. Then, the sequences $\{e_j = \tilde{X}_j 1(T \geq j),\ j \geq 1\}$ and $\{e'_j = \sum_{i=1}^{j-1} X_i \tilde{X}_j,\ j \geq 2\}$ are $\{\mathscr{F}_i'\}$-tangent to $\{d_i\}$ and $\{d'_i\}$, respectively. Moreover, both $\{e_i\}$ and $\{e'_i\}$ are decoupled versions of the original sequences which satisfy the CI condition with respect to $\mathscr{G} = \sigma(\{X_i\})$.

The advantage of dealing with the decoupled sequences is made quite evident if we assume that the X's are standard normal random variables. In this case, $\tilde{M}_n = \sum_{j=1}^n e_j = \sum_{j=1}^{T \wedge n} \tilde{X}_j$ has the same distribution as $(T \wedge n)\tilde{X}_1$ and $\tilde{U}_n = \sum_{j=2}^n e'_j = \sum_{j=2}^n (\sum_{i=1}^{j-1} X_i) \tilde{X}_j$ has the same distribution as $\sqrt{\sum_{j=2}^n \left(\sum_{i=1}^{j-1} X_i\right)^2}\, \tilde{X}_1$, where we recall that the variables were chosen so that the pair $(T, \{X_i\})$ is independent of $\{\tilde{X}_i\}$. In the next section we provide inequalities comparing the moment generating functions of M_n and U_n to those of $\tilde{M}_n$ and $\tilde{U}_n$ and, more generally, inequalities relating the moment generating functions of the sums of two tangent sequences when one of them is decoupled.

3. New decoupling inequalities. We will begin by proving the following (constrained) decoupling inequality which extends a result in de la Peña (1994) [see also de la Peña (1995)].

THEOREM 3.1. *Let* $\{d_i\}$ *be a sequence of nonnegative, nondegenerate random variables. Then there exists a σ-field* $\mathscr{G}$ *and a* $\mathscr{G}$*-conditionally indepen-*

dent sequence $\{e_i\}$, *tangent to* $\{d_i\}$, *such that for all random variables* $g \geq 0$ *measurable with respect to* $\mathscr{G}$,

$$E\left(g\prod_{i=1}^{n} d_i\right)^{1/2} \leq \left(Eg\prod_{i=1}^{n} e_i\right)^{1/2}. \tag{3.1}$$

[Recall that $\mathscr{G}$ may be taken to equal $\sigma(\{d_i\})$].

The proof of the above result is based on the following lemma, which (in the case $K = 1$) basically consists of a variation of Jensen's inequality allowing for a change of measure.

LEMMA 3.1. *Let* X, Y *be two random variables with* $X \geq 0$, $Y \geq 0$, $X/Y \geq 0$ *a.e. and* $E(X/Y) \leq K$ *for some constant* K. *Then*,

$$E\sqrt{X} \leq \sqrt{KEY}. \tag{3.2}$$

PROOF. $E\sqrt{X} = E\sqrt{X/Y} \times \sqrt{Y} \leq \sqrt{E(X/Y)} \times \sqrt{EY} \leq \sqrt{KEY}$, by Hölder's inequality.

PROOF OF THEOREM 3.1. The proof is very similar to the main result of de la Peña (1994). From Section 2, it follows that for any sequence $\{d_i\}$ one can find a tangent sequence $\{e_i\}$ where $\{e_i\}$ is *conditionally independent* given a master σ-field $\mathscr{G}$.

Let $\mathscr{F}_0$ be the trivial σ-field and for $i \geq 1$, let $\mathscr{F}_i$ be the σ-field generated by $\{d_1, \dots, d_i; e_1, \dots, e_i\}$. In order to see that

$$E\frac{\prod_{i=1}^{n} d_i}{\prod_{i=1}^{n} E(d_i \mid \mathscr{F}_{i-1})} = 1, \tag{3.3}$$

one uses an induction argument along with the properties of conditional expectation. More formally, if we assume (3.3) is valid for $n - 1$, then

$$\begin{aligned} E\frac{\prod_{i=1}^{n} d_i}{\prod_{i=1}^{n} E(d_i \mid \mathscr{F}_{i-1})} &= E\left[\frac{\prod_{i=1}^{n-1} d_i}{\prod_{i=1}^{n-1} E(d_i \mid \mathscr{F}_{i-1})} \times \frac{E(d_n \mid \mathscr{F}_{n-1})}{E(d_n \mid \mathscr{F}_{n-1})}\right] \\ &= E\frac{\prod_{i=1}^{n-1} d_i}{\prod_{i=1}^{n-1} E(d_i \mid \mathscr{F}_{i-1})} = 1. \end{aligned}$$

Also, since $\{e_i\}$ is tangent to $\{d_i\}$ and *conditionally independent* given $\mathscr{G}$,

$$\prod_{i=1}^{n} E(d_i \mid \mathscr{F}_{i-1}) = \prod_{i=1}^{n} E(e_i \mid \mathscr{F}_{i-1}) = \prod_{i=1}^{n} E(e_i \mid \mathscr{G}) = E\left(\prod_{i=1}^{n} e_i \mid \mathscr{G}\right). \tag{3.4}$$

Note that without loss of generality we may assume that $g > 0$ a.e. Next, we replace X by $g\prod_{i=1}^{n} d_i$ and Y by $g\prod_{i=1}^{n} E(d_i \mid \mathscr{F}_{i-1})$ in Lemma 3.1 to get

$$E\sqrt{g\prod_{i=1}^{n} d_i} \leq \sqrt{1 \times E\left[gE\left(\prod_{i=1}^{n} e_i \mid \mathscr{G}\right)\right]} = \sqrt{Eg\prod_{i=1}^{n} e_i},$$

since g is $\mathscr{G}$-measurasble. This completes the proof. □

The following example due to Hitczenko and found in de la Peña (1994) shows that the above result is sharp, at least in the case $n = 2$.

EXAMPLE 3.1. Let $d_1(\omega) = 1_A(w)$, the indicator variable of the set A. Set $d_2 = d_1$, $g = 1$. Then we can take as e_1 an independent copy of d_1. Due to tangency, e_2 must equal d_1. Then, $Ed_1 = E\sqrt{d_1 d_2} \le \sqrt{Ee_1 e_2} = \sqrt{Ee_1 Ee_2} = Ed_1$.

Taking $d_i' = \exp\{td_i\}$ and $e_i' = \exp\{te_i\}$ for some finite t, we get the following corollary to Theorem 3.1. This corollary has the form we will need to provide proofs for the exponential inequalities for martingales stated in Section 1.

COROLLARY 3.1. *Let* $\{d_i\}$, $\{e_i\}$ *be* $\{\mathscr{F}_i\}$*-tangent. Assume that* $\{e_i\}$ *is decoupled* (*CI*). *Let* $g \ge 0$ *be any random variable measurable with respect to* $\sigma(\{d_i\}_{i=1}^{\infty})$. *Then for all finite* t,

$$Eg \exp\left\{t \sum_{i=1}^{n} d_i\right\} \le \sqrt{Eg^2 \exp\left\{2t \sum_{i=1}^{n} e_i\right\}}\,. \tag{3.5}$$

Applying this result twice, for t and $-t$ and using the inequality $\exp\{|x|\} \le \exp\{x\} + \exp\{-x\}$, one gets the following corollary.

COROLLARY 3.2. *Under the assumptions of Corollary* 3.1, *for all* $0 < t < \infty$,

$$Eg \exp\left\{t\left|\sum_{i=1}^{n} d_i\right|\right\} \le 2\sqrt{Eg^2 \exp\left\{2t\left|\sum_{i=1}^{n} e_i\right|\right\}}\,.$$

Corollary 3.1 can also be extended to deal with hyperbolic functions.

COROLLARY 3.3. *Let* $\{d_i\}$, $\{e_i\}$ *be* $\{\mathscr{F}_i\}$*-tangent. Assume that* $\{e_i\}$ *is decoupled* (*CI*). *Let* $g \ge 0$ *be any random variable measurable with respect to* $\sigma(\{d_i\}_{i=1}^{\infty})$. *Then for all finite* t,

$$Eg \cosh\left(t \sum_{i=1}^{n} d_i\right) \le \sqrt{Eg^2 \cosh\left(2t \sum_{i=1}^{n} e_i\right)}\,. \tag{3.6}$$

To see this, recall that $\cosh(x) = (\exp\{x\} + \exp\{-x\})/2$ and apply Corollary 3.1 twice and use the fact that for $a, b \ge 0$, $(\sqrt{a} + \sqrt{b})/2 \le \sqrt{(a+b)/2}$.

COROLLARY 3.4. *Let* $\{D_i\}$, $\{E_i\}$ *be* $\{\mathscr{F}_i\}$*-tangent sets* (*see Remark* 2.1). *Assume that* $\{E_i\}$ *is decoupled* (*CI*). *Then*, *for any set* G, *measurable with respect to* $\sigma(\{D_i\}_{i=1}^{\infty})$,

$$P\left(\bigcap_{i=1}^{n} D_i\right) \le P\left(\bigcap_{i=1}^{n} E_i \,\middle|\, \bigcap_{i=1}^{n} D_i\right), \tag{3.7}$$

and more generally,

$$P\left(\bigcap_{i=1}^{n} D_i \cap G\right) \le P\left(\bigcap_{i=1}^{n} E_i \cap G \,\middle|\, \bigcap_{i=1}^{n} D_i \cap G\right). \tag{3.8}$$

The following example shows that Corollary 3.4 is sharp for every n.

EXAMPLE 3.2. Let $\{D_i\}$ be a sequence of independent sets. Let $\{E_i\}$ be an independent copy of $\{D_i\}$. Then, the above sequences are $\{\mathscr{F}_i\}$-tangent with respect to $\mathscr{F}_n = \sigma(\{D_i\}, \{E_i\}, i = 1, \dots, n)$. Observe that in this case,

$$P\left(\bigcap_{i=1}^{n} D_i\right) = P\left(\bigcap_{i=1}^{n} E_i\right) = P\left(\bigcap_{i=1}^{n} E_i \,\middle|\, \bigcap_{i=1}^{n} D_i\right), \tag{3.9}$$

since the two sequences are independent. Compare this with Example 3.1.

4. Proofs of results of Section 1. Armed with the results from the last section, we proceed to extend Bernstein's and Bennett's inequalities to martingales. The proof of (1.1) is based on the following inequality for sums of independent random variables.

LEMMA 4.1. *Under the assumptions of Theorem* 1.1, *for all* $x > 0$ *and* $0 < r < 1/c$,

$$E\exp\left\{r\sum_{i=1}^{n} x_i\right\} \le \exp\left\{\frac{v_n^2 r^2}{2(1-cr)}\right\}. \tag{4.1}$$

PROOF. The conditions imposed along with an expansion of the exponential gives

$$E\exp\left\{\frac{x\sum_{i=1}^{n} x_i}{v_n^2 + cx}\right\} \le \exp\left\{\frac{x^2}{2\left(v_n^2 + cx\right)}\right\},$$

which is valid for all $x > 0$ [cf. Chow and Teicher (1988)]. More precisely, observe that $\exp\{rx_j\} \le 1 + rx_j + \sum_{k=2}^{\infty} r^k(|x_j|^k/k!)$ for $r > 0$ implies that $E\exp\{rx_j\} \le \exp\{\sigma_j^2 r^2/2(1-cr)\}$ for $0 < cr < 1$. Letting $r = x/(v_n^2 + cx)$, we have $E\exp\{r\sum_{i=1}^{n} x_i\} \le \exp\{(xr)/2\} = \exp\{r^2 v_n^2/(2(1-cr))\}$, which is (4.1) upon replacing r by its assigned value.

In what follows, we will use the notation $M_n = \sum_{i=1}^{n} d_i$ and $\mathscr{G} = \sigma\{d_i\}$.

PROOF OF THEOREM 1.2A. Let $\tau = \inf\{n\colon M_n \ge x\}$ where $\inf\varnothing = \infty$ if this does not happen. Taking A to be the set ($M_n \ge x$ and $V_n^2 \le y$ for some n) on A, we have that $\tau < \infty$, $M_\tau \ge x$ and $V_\tau^2 \le y$. Moreover, observe that $1(A) = 1(A)1(\tau < \infty)1(M_\tau \ge x)$ implies $P(A) = P(M_\tau \ge x, A)$. Applying Markov's inequality first, followed by Fatou's lemma (valid since $\tau < \infty$ on A) and a use of

Corollary 3.1 with $g = \exp\{-(\lambda/2)x\}1(M_{\tau\wedge n} \geq x,\ A)$ we get

$$\begin{aligned}
P(A) &= P\left(\sum_{i=1}^{\tau} d_i \geq x,\ A\right)\\
&\leq \inf_{\lambda>0} E\exp\left\{\frac{\lambda}{2}\left(\sum_{i=1}^{\tau} d_i - x\right)\right\}1(M_\tau \geq x,\ A)\\
&= \inf_{\lambda>0} E\liminf_{n} \exp\left\{\frac{\lambda}{2}\left(\sum_{i=1}^{\tau\wedge n} d_i - x\right)\right\}1(M_\tau \geq x,\ A)\\
&\leq \inf_{\lambda>0} \liminf_{n\to\infty} E\exp\left\{\frac{\lambda}{2}\left(\sum_{i=1}^{\tau\wedge n} d_i - x\right)\right\}1(M_\tau \geq x,\ A)\\
&\leq \inf_{\lambda>0} \liminf_{n\to\infty} \sqrt{E\exp\left\{\lambda\left(\sum_{i=1}^{\tau\wedge n} e_i - x\right)\right\}1(M_\tau \geq x,\ A)}\\
&= \inf_{\lambda>0} \liminf_{n\to\infty} \sqrt{E\left[1(M_\tau \geq x,\ A)\exp\{-\lambda x\}E\left(\exp\left\{\lambda\sum_{i=1}^{\tau\wedge n} e_i\right\}\middle|\mathscr{G}\right)\right]},
\end{aligned}$$

where the last equality follows since the variables outside the conditional expectation are $\mathscr{G}$ measurable. Observe that since $\{d_i\}$ and $\{e_i\}$ are tangent and $\{e_i\}$ is conditionally independent given $\mathscr{G}$, the moment assumptions on the distribution of d_i transfer to conditions on the e_i's and therefore we can apply Lemma 4.1 to obtain

$$E\left[\exp\left\{\lambda\sum_{i=1}^{\tau\wedge n} e_i\right\}\,\middle|\,\mathscr{G}\right] \leq \exp\{h(\lambda)V_{\tau\wedge n}^2\}, \tag{4.2}$$

where $h(\lambda) = \lambda^2/(2(1-\lambda c))$ Replacing this in the above bound, one obtains

$$P\left(\sum_{i=1}^{\tau} d_i \geq x,\ A\right) \leq \inf_{\lambda>0} \liminf_{n\to\infty} \sqrt{E\exp\{-(\lambda x - h(\lambda)V_{\tau\wedge n}^2)\}1(M_\tau \geq x,\ A)}\,.$$

Since the variable inside the expectation is dominated by

$$\exp\{-(\lambda x - h(\lambda)V_\tau^2)\}1(M_\tau \geq x,\ A),$$

and $V_\tau \leq y$ on A, using the dominated convergence theorem, we get

$$P\left(\sum_{i=1}^{\tau} d_i \geq x,\ A\right) \leq \inf_{\lambda>0} \sqrt{E\exp\{-(\lambda x - h(\lambda)V_\tau^2)\}1(M_\tau \geq x,\ A)}\,.$$

Dividing both sides by $\sqrt{P(M_\tau \geq x,\ A)}$ gives

$$P\left(\sum_{i=1}^{\tau} d_i \geq x,\ A\right) \leq \inf_{\lambda>0} E\left[\exp\{-(\lambda x - h(\lambda)V_\tau^2)\} \mid (M_\tau \geq x,\ A)\right].$$

Then, since $\tau < \infty$, $M_\tau \ge x$ and $V_\tau^2 \le y$ on A, we have

$$P\left(\sum_{i=1}^n d_i \ge x, V_n^2 \le y \text{ for some } n\right) \le \inf_{\lambda > 0} \exp\{-(\lambda x - h(\lambda) y)\}.$$

The proof is finished by using calculus to obtain the value minimizing the above expression.

Proof of Theorem 1.2B. Applying Markov's inequality first, followed by a use of Corollary 3.1 with $g = \exp\{-(\lambda/2)V_n^2 x\} 1(M_n/V_n^2 \ge x,\ A)$, we obtain

$$\begin{aligned}
P&\left(\sum_{i=1}^n d_i \ge V_n^2 x,\ A\right) \\
&\le \inf_{\lambda>0} E\exp\left\{\frac{\lambda}{2}\left(\sum_{i=1}^n d_i - V_n^2 x\right)\right\} 1(M_n \ge V_n^2 x,\ A) \\
&\le \inf_{\lambda>0} \sqrt{E\exp\left\{\lambda\left(\sum_{i=1}^n e_i - V_n^2 x\right)\right\} 1(M_n \ge V_n^2 x,\ A)} \\
&= \inf_{\lambda>0} \sqrt{E\left[1(M_n \ge V_n^2 x,\ A)\exp\{-\lambda V_n^2 x\} E\left(\exp\left\{\lambda \sum_{i=1}^n e_i\right\}\middle|\mathscr{G}\right)\right]},
\end{aligned}$$

where the last equality follows since the variables outside the conditional expectation are $\mathscr{G}$ measurable. Observe that since $\{d_i\}$ and $\{e_i\}$ are tangent and $\{e_i\}$ is conditionally independent given $\mathscr{G}$, the moment assumptions on the distribution of d_i transfer to conditions on the e_i's and therefore we can apply Lemma 4.1 to show that

$$E\left(\exp\left\{\lambda \sum_{i=1}^n e_i\right\}\middle|\mathscr{G}\right) \le \exp\left\{\frac{\lambda^2}{2(1-\lambda c)} V_n^2\right\}. \tag{4.3}$$

Replacing this in the above bound one obtains,

$$\begin{aligned}
P&\left(\sum_{i=1}^n d_i \ge V_n^2 x,\ A\right) \\
&\le \inf_{\lambda>0} \sqrt{E\exp\left\{-\left(\lambda x - \frac{\lambda^2}{2(1-\lambda c)}\right)V_n^2\right\} 1(M_n \ge V_n^2 x,\ A)} \\
&\le \sqrt{E\exp\left\{-\left(\frac{x}{c} + \frac{1}{c^2} - \frac{\sqrt{1+2cx}}{c^2}\right)V_n^2\right\} 1(M_n \ge V_n^2 x,\ A)} \\
&= \sqrt{E\exp\left\{-\left(\frac{x^2}{1+cx+\sqrt{1+2cx}}\right)V_n^2\right\} 1(M_n \ge V_n^2 x,\ A)}.
\end{aligned}$$

Dividing both sides by $\sqrt{P(M_n \geq V_n^2 x,\ A)}$ gives (1.3), while (1.5) is obtained by adapting the stopping time argument of Freedman (1975), already used in the previous theorem, by letting $A = \{M_n/V_n \geq x,\ 1/V_n^2 \leq y \text{ for some } n\}$. This is possible since Corollary 3.1 also works once we replace n by $\tau \wedge n$, where

$$\tau = \left\{\inf n \geq 1\colon \frac{\sum_{i=1}^n d_i}{V_n^2} \geq x\right\}, \tag{4.4}$$

with $\inf\varnothing = \infty$. To complete the proof, we use the bounded convergence theorem considering the fact that we can choose λ so that $-(\lambda x - \lambda^2/(2(1 - \lambda c))) \leq 0$.

PROOF OF THEOREM 1.3. Observe first that with (3.5) from Freedman (1975), it follows that for variables satisfying $x_i \leq 1$, $Ex_i \leq 0$ and all $\lambda \geq 0$,

$$E\exp\{\lambda x_i\} \leq 1 + h(\lambda)\mathrm{Var}(x_i) \leq \exp\{h(\lambda)\mathrm{Var}(x_i)\}, \tag{4.5}$$

where $h(\lambda) = \{\exp\{\lambda\} - 1 - \lambda\}$. From this, it follows that, if $\{x_i\}$ is a sequence of independent random variables under the above conditions, then

$$E\exp\left\{\lambda\sum_{i=1}^n x_i\right\} \leq \exp\left\{h(\lambda)\sum_{i=1}^n \mathrm{Var}(x_i)\right\}. \tag{4.6}$$

Using this fact after applying Markov's inequality and Corollary 3.1 with

$$g = \exp\left\{-\frac{\lambda}{2}V_n^2 x\right\}1\left(\frac{M_n}{(\alpha + \beta V_n^2)} \geq x,\ A\right),$$

we get

$$\begin{aligned}
&P\left(\sum_{i=1}^n d_i \geq (\alpha + \beta V_n^2)x,\ A\right)\\
&\quad\leq \inf_{\lambda>0}\exp\left\{-\frac{\lambda}{2}\alpha x\right\}E\exp\left\{\frac{\lambda}{2}\left(\sum_{i=1}^n d_i - V_n^2\beta x\right)\right\}1\big(M_n \geq (\alpha+\beta V_n^2)x,\ A\big)\\
&\quad\leq \inf_{\lambda>0}\left[\exp\{-\lambda\alpha x\}E\exp\left\{\lambda\left(\sum_{i=1}^n e_i - V_n^2\beta x\right)\right\}\right.\\
&\qquad\qquad\left.\times 1\big(M_n \geq (\alpha + \beta V_n^2)x,\ A\big)\right]^{1/2}\\
&\quad= \inf_{\lambda>0}\left\{\exp\{-\lambda\alpha x\}E\Big[1\big(M_n \geq (\alpha + \beta V_n^2)x,\ A\big)\right.\\
&\qquad\qquad\left.\times\exp\{-\lambda V_n^2\beta x\}E\left(\exp\left\{\lambda\sum_{i=1}^n e_i\right\}\middle|\mathscr{G}\right)\Big]\right\}^{1/2}
\end{aligned}$$

$$\le \inf_{\lambda>0} \big[\exp\{-\lambda\alpha x\} E 1\big(M_n \ge p\big(\alpha + \beta V_n^2\big)x,\, A\big) \times \exp\big\{-\lambda V_n^2 \beta x\big\} \exp\big\{h(\lambda) V_n^2\big\}\big]^{1/2}$$
$$\le \inf_{\lambda>0} \big[\exp\{-\lambda\alpha x\} E 1\big(M_n \ge \big(\alpha + \beta V_n^2\big)x,\, A\big) \times \exp\big\{-(\lambda\beta x - h(\lambda)) V_n^2\big\}\big]^{1/2}.$$

The minimum inside the expectation is attained at $\lambda = \log(\beta x + 1)$. The remainder of the proof follows as in the previous proofs.

The proof of Theorem 1.4A is very similar to that of Theorem 1.2A and will not be given here. However, we include next the proof of its parallel result.

PROOF OF THEOREM 1.4B. Let $\{x_i\}$ be a sequence of independent random variables with $S_n = \sum_{i=1}^n x_i$ and $\mathrm{Var}(S_n) = v_n^2$, $Ex_i \le 0$ and $|x_i| \le c$ for all i; then it is easy to see that [as shown in Pinelis (1994), Theorem 3.4]

$$Z_n = \exp\left\{\lambda \sum_{i=1}^n x_i - \frac{\exp\{\lambda c\} - 1 - \lambda c}{c^2} v_n^2\right\} \tag{4.7}$$

is a supermartingale and hence

$$E \exp\left\{\lambda \sum_{i=1}^n x_i\right\} \le \exp\left\{\frac{\exp\{\lambda c\} - 1 - \lambda c}{c^2} v_n^2\right\}. \tag{4.8}$$

To see this observe that

$$Z_n = \frac{\exp\{\lambda \sum_{i=1}^n x_i\}}{\prod_{i=1}^n \big(1 + E\exp\{\lambda x_j\} - 1 - \lambda x_j\big)} \tag{4.9}$$

is a supermartingale and consider the function $f(r) = (\exp\{r\} - 1 - r)/r^2$, where $r \ne 0$ and $g(0)$ is taken to be $1/2$. Then, since $f(r)$ is increasing in r, and we are assuming that $|x_i| \le c$, one has that

$$E\big(\exp\{\lambda x_i\} - 1 - \lambda x_i\big) \le \frac{\exp\{\lambda c\} - 1 - \lambda c}{c^2} Ex_i^2. \tag{4.10}$$

To complete the proof of (4.7) and hence (4.8), use the fact that $1 + r \le \exp\{r\}$.

Using Markov's inequality and Corollary 3.1, we obtain

$$P\left(\sum_{i=1}^n d_i \ge \big(\alpha + \beta V_n^2\big)x,\, A\right)$$
$$= P\left(\exp\left\{\frac{\lambda}{2}\sum_{i=1}^n d_i\right\} \ge \exp\left\{\frac{\lambda}{2}\big(\alpha + \beta V_n^2\big)x\right\},\, A\right)$$
$$\le \inf_{\lambda>0} E \frac{\exp\{(\lambda/2)\sum_{i=1}^n d_i\}}{\exp\{(\lambda/2)\big(\alpha + \beta V_n^2\big)x\}} 1\left(\sum_{i=1}^n d_i \ge \big(\alpha + \beta V_n^2\big)x,\, A\right)$$

$$\leq \inf_{\lambda>0}\left[E\exp\left\{\lambda\sum_{i=1}^{n}e_i\right\}\exp\left\{-\left(\lambda\left(\alpha+\beta V_n^2\right)x\right)\right\}\right.$$
$$\left.\times 1\left(\sum_{i=1}^{n}d_i \geq \left(\alpha+\beta V_n^2\right)x,\ A\right)\right]^{1/2}$$
$$\leq \inf_{\lambda>0}\left[E\exp\left\{\frac{\exp\{\lambda c\}-1-\lambda c}{c^2}V_n^2\right\}\exp\left\{-\left(\lambda\left(\alpha+\beta V_n^2\right)x\right)\right\}\right.$$
$$\left.\times 1\left(\sum_{i=1}^{n}d_i \geq \left(\alpha+\beta V_n^2\right)x,\ A\right)\right]^{1/2}$$
$$= \inf_{\lambda>0}\left[\exp\{-\lambda\alpha x\}E\exp\left\{\left(\frac{\exp\{\lambda c\}-1-\lambda c}{c^2}-\lambda\beta x\right)V_n^2\right\}\right.$$
$$\left.\times 1\left(\sum_{i=1}^{n}d_i \geq \left(\alpha+\beta V_n^2\right)x,\ A\right)\right]^{1/2}.$$

The line previous to last follows using (4.8) conditionally on $\mathscr{G}$. Observe also that the λ that minimizes the function inside the expectation is $\lambda_o = (1/c)\ln(1+\beta cx)$. Substituting we obtain the bound,

$$\left[\exp\left\{-\frac{\alpha x}{c}\ln(1+c\beta x)\right\}\right.$$
$$\times E\exp\left\{-\left(\frac{\beta x}{c}-\left(\frac{\beta x}{c}+\frac{1}{c^2}\right)\ln(1+\beta cx)\right)V_n^2\right\}$$
$$\left.\times 1\left(\sum_{i=1}^{n}d_i \geq \left(\alpha+\beta V_n^2\right)x,\ A\right)\right]^{1/2}.$$

The proof is complete by dividing by

$$\sqrt{P\left(\sum_{i=1}^{n}d_i \geq \left(\alpha+\beta V_n^2\right)x,\ A\right)}. \qquad \square$$

In order to avoid unnecessary repetitions, the proof of Theorem 1.5A is omitted. We instead include the proof of its parallel result.

PROOF OF THEOREM 1.5B. We start with some preliminary facts. It is easy to see that for independent mean zero random variables $\{x_i\}$ with $S_n = \sum_{i=1}^{n} x_i$, and $\mathrm{Var}(S_n) = v_n^2$ and $|x_i| \leq c$ for all i, then

$$Z_n = \exp\left\{\lambda\sum_{i=1}^{n}x_i - \frac{\lambda}{c}\sinh \lambda c v_n^2\right\} \tag{4.11}$$

is a supermartingale which gives that

$$(4.12) \qquad E\exp\left\{\lambda\sum_{i=1}^{n} x_i\right\} \le \exp\left\{\frac{\lambda}{c}\sinh \lambda c v_n^2\right\}.$$

[See the proof of Proposition 3.1 in Hitczenko (1990b)]. This fact will be used later. Uisng Markov's inequality and Corollary 3.1, we obtain

$$P\left(\sum_{i=1}^{n} d_i \ge (\alpha+\beta V_n^2)x,\ A\right)$$
$$\le \inf_{\lambda>0}\exp\left\{\frac{\lambda}{2}\alpha x\right\}E\exp\left\{\frac{\lambda}{2}\left(\sum_{i=1}^{n} d_i - V_n^2\beta x\right)\right\}1\left(M_n \ge (\alpha+\beta V_n^2)x\right)$$
$$\le \inf_{\lambda>0}\sqrt{\exp\{-\lambda\alpha x\}E\exp\left\{\lambda\left(\sum_{i=1}^{n} e_i - V_n^2\beta x\right)\right\}1\left(M_n \ge (\alpha+\beta V_n^2)x,\ A\right)}\,.$$

Working conditionally on $\mathscr{G}$ and using (4.12), we get that the above is bounded by

$$\inf_{\lambda>0}\left\{\exp\{-\lambda\alpha x\}\right.$$
$$\left.\times E\left[E\left(\exp\left\{\left(\frac{\lambda}{c}\sinh\lambda c - \lambda\beta x\right)V_n^2\right\}1\left(M_n \ge (\alpha+\beta V_n^2)x,\ A\right)\Bigg|\ \mathscr{G}\right)\right]\right\}^{1/2}.$$

Taking $\lambda_o = (1/c)\operatorname{arc\,sinh}(c\beta x/2)$, we get $\beta x/2 = \sinh(\lambda_o c)/c$ gives the bound

$$\left\{\exp\left\{-\frac{\alpha x}{c}\operatorname{arc\,sinh}\frac{c\beta x}{2}\right\}\right.$$
$$\left.\times E\left[E\left(\exp\left\{-\left(\frac{\beta x}{2c}\operatorname{arc\,sinh}\left(\frac{\beta c x}{2}\right)\right)V_n^2\right\}1\left(M_n \ge (\alpha+\beta V_n^2)x,\ A\right)\Bigg|\ \mathscr{G}\right)\right]\right\}^{1/2}.$$

The proof is complete by removing the conditional expectation and dividing by

$$\sqrt{P\left(M_n \ge (\alpha+\beta V_n^2)x,\ A\right)}\,. \qquad \square$$

REMARK 4.1. As might have been surmised by the reader, we could have avoided the use of decoupling when obtaining the proofs of the results by invoking instead appropriate exponential inequalities. However, one of the strengths of our approach is that by using decoupling, we are ready to improve on our inequalities in cases where new bounds on the moment generating function for sums of independent random variables are available, hence saving us the task of proving those results for martingales. Moreover, decoupling is the guiding force behind the intuitive drive that lead us to the results included in this work.

In the next two sections we present results for which we do not have a decoupling argument. They include inequalities for continuous time martingales and for sums of conditionally symmetric variables.

5. Continuous time martingales. We were partly inspired to work on this problem by the following result that we learned from Khoshnevisan (1996), which he attributes (essentially) to McKean (1962).

THEOREM 5.1. *Let $\{M_t, \mathscr{F}_t, t > 0\}$ be a continuous martingale with $M_0 = 0$, including the terminal point at infinity. Assume further that $E \exp^{\eta M_\infty} < \infty$ for all $\eta \in R^1$. Then, for all $\alpha, \beta, x > 0$,*

$$P(M_\infty \geq (\alpha + \beta \langle M \rangle_\infty) x) \leq \exp\{-2\alpha\beta x^2\}. \tag{5.1}$$

Khoshnevisan (1996) shows that this result is asymptotically optimal. In this paper we provide an extension to the case where α might be zero. As a result of our extension, we arrive at conditions on $\langle M \rangle_t$ under which one can improve upon Theorem 5.1.

More precisely, we can prove the following theorem.

THEOREM 5.2. *Let $\{M_t, \mathscr{F}_t, t \geq 0\}$, $M_0 = 0$ be a continuous martingale for which $\exp\{\lambda M_t - (\lambda^2 \langle M \rangle_t / 2\}$ is a supermartingale for all $\lambda > 0$. Let A be a set measurable with respect to $\mathscr{F}_\infty$. Then, for all $0 < t < \infty$, $\beta > 0$, α, $x \geq 0$,*

$$\begin{aligned} &P(M_t \geq (\alpha + \beta\langle M \rangle_t) x, A) \\ &\quad \leq E\left[\exp\left\{-x^2\left[\frac{\beta^2}{2}\langle M \rangle_t + \alpha\beta\right]\right\}\middle|(M_t \geq (\alpha + \beta\langle M \rangle_t) x, A)\right], \end{aligned} \tag{5.2}$$

$$\begin{aligned} &P\left(M_t \geq (\alpha + \beta\langle M \rangle_t) x, \frac{1}{\langle M \rangle_t} < y \text{ for some } t < \infty\right) \\ &\quad \leq \exp\left\{-x^2\left(\frac{\beta^2}{2y} + \alpha\beta\right)\right\}. \end{aligned} \tag{5.3}$$

It is easy to see that for finite t, this result improves on Theorem 5.1 in several situations, the more trivial one being when $\langle M \rangle_t > 2\alpha/\beta$, since then we get an improved bound by letting $y = \beta/2\alpha$ in (5.3).

PROOF OF THEOREM 5.2. Using Markov's inequality we obtain that for all $\lambda > 0$,

$$\begin{aligned} &P(M_t \geq (\alpha + \beta\langle M \rangle_t) x, A) \\ &\quad \leq E \exp\left\{\frac{\lambda}{2} M_t - \left(\frac{\lambda\alpha x}{2} - \frac{\lambda\beta x}{2}\langle M \rangle_t\right)\right\} 1\left(\frac{M_t}{(\alpha + \beta\langle M \rangle_t)} \geq x, A\right). \end{aligned} \tag{5.4}$$

This in turn equals

$$\exp\left\{-\frac{\lambda\alpha x}{2}\right\}E\left[\exp\left\{\frac{\lambda}{2}M_t-\frac{\lambda^2}{4}\langle M\rangle_t+\frac{\lambda^2}{4}\langle M\rangle_t-\frac{\lambda\beta x}{2}\langle M\rangle_t\right\}\right.$$
$$\left.\times 1\left(\frac{M_t}{(\alpha+\beta\langle M\rangle_t)}\geq x,\ A\right)\right]$$

$$(5.5)\quad \leq \exp\left\{-\frac{\lambda\alpha x}{2}\right\}\left[\sqrt{E\exp\left\{\lambda M_t-\frac{\lambda^2}{2}\langle M\rangle_t\right\}}\right.$$
$$\left.\times\sqrt{E\exp\left\{\left(\frac{\lambda^2}{2}-\lambda\beta x\right)\langle M\rangle_t\right\}1\left(\frac{M_t}{(\alpha+\beta\langle M\rangle_t)}\geq x,\ A\right)}\right]$$
$$=\exp\left\{-\frac{\lambda\alpha x}{2}\right\}\sqrt{E\exp\left\{\left(\frac{\lambda^2}{2}-\lambda\beta x\right)\langle M\rangle_t\right\}1\left(\frac{M_t}{(\alpha+\beta\langle M\rangle_t)}\geq x,\ A\right)}\,.$$

Taking $\lambda=\beta x$ minimizes the expression inside the square root sign and we get

$$P(M_t\geq(\alpha+\beta\langle M\rangle_t)x,\ A)$$
$$\leq\exp\left\{-\frac{\alpha\beta x^2}{2}\right\}\sqrt{E\exp\left\{-\frac{\beta^2x^2}{2}\langle M\rangle_t\right\}1\left(\frac{M_t}{(\alpha+\beta\langle M\rangle_t)}\geq x,\ A\right)}\,.$$

Dividing over $\sqrt{P(M_t\geq(\alpha+\beta\langle M\rangle_t)x,\ A)}$, we obtain

$$P(M_t\geq(\alpha+\beta\langle M\rangle_t)x,\ A)$$
$$(5.6)\quad \leq\exp\{-\alpha\beta x^2\}E\left[\exp\left\{-\frac{\beta^2x^2}{2}\langle M\rangle_t\right\}\left|\left(\frac{M_t}{(\alpha+\beta\langle M\rangle_t)}\geq x,\ A\right)\right]\right..$$

Concerning (5.3) we use a similar argument to the ones used in the previous section. Namely, let $\tau=\{\text{int}\, t>0\colon\ M_t/(\alpha+\beta\langle M\rangle_t)\geq x\}$, with $\inf\varnothing=\infty$, and $A=\{M_t/(\alpha+\beta\langle M\rangle_t)\geq x,\ 1/\langle M_t\rangle\leq y \text{ for some } t\}$ and observe that Fatou's lemma gives that since $\tau<\infty$ on A, and we have a supermartingale,

$$(5.7)\qquad E\exp\left\{\lambda M_\tau-\frac{\lambda^2}{2}\langle M\rangle_\tau\right\}1(A)\leq 1.$$

Replacing t for τ in (5.4), using the expectation of (5.7) in equation (5.5) instead of the expectation given there and following the steps that continue, we arrive at the bound

$$P(M_\tau\geq(\alpha+\beta\langle M\rangle_\tau)x,\ A)$$
$$(5.8)\quad \leq\exp\left\{-\frac{\alpha\beta x^2}{2}\right\}E\left[\exp\left\{-\frac{\beta^2x^2}{2}\langle M\rangle_\tau\right\}\left|\left(\frac{M_\tau}{(\alpha+\beta\langle M\rangle_\tau)}\geq x,\ A\right)\right]\right..$$

Next, observe that

$$1(A) = 1(\{\tau < \infty\})1(A)1\left(\left\{\frac{M_\tau}{(\alpha + \beta\langle M\rangle_\tau)} \geq x\right\}\right).$$

Therefore,

$$\begin{aligned} P(A) &= P\left(M_t \geq (\alpha + \beta\langle M\rangle_t)x, \frac{1}{\langle M\rangle_t} < y \text{ for some } t < \infty\right) \\ &\leq P\left(\frac{M_\tau}{(\alpha + \beta\langle M\rangle_\tau)} \geq x, A\right) \\ &\leq \exp\left\{-\frac{\alpha\beta x^2}{2}\right\} E\left[\exp\left\{-\frac{\beta^2 x^2}{2}\langle M\rangle_\tau\right\}\middle|\left(\frac{M_\tau}{(\alpha + \beta\langle M\rangle_\tau)} \geq x, A\right)\right] \\ &\leq \exp\left\{-x^2\left(\frac{\beta^2}{2y} + \alpha\beta\right)\right\}, \end{aligned}$$

where the last inequality follows since on A, $\langle M\rangle_\tau > (1/y)$, therefore completing the proof of the theorem. □

As follows easily from Proposition 5.1, which we take from Barlow, Jacka and Yor (1986), Proposition 4.2.1, $\exp\{\lambda M_t - (\lambda^2\langle M\rangle_t)/2\}$ is a supermartingale for all $0 < \lambda < \infty$; hence the related condition of Theorem 5.2 always holds. We will use this proposition to provide an extension of Theorem 5.2 to the case of martingales with jumps.

Proposition 5.1. *Let $\{M_t, t \geq 0\}$ be a locally square-integrable martingale, with $M_0 = 0$. Let $\langle M^c\rangle$ denote the quadratic variation of its continuous part. Let $\{V_t\}$ be an increasing process, which is adapted, purely discontinuous and locally integrable; let $V^{(p)}$ be its dual predictable projection. Set $X_t = M_t + V_t$, and*

$$C_t = \sum_{s \leq t}\left((\Delta X_s)^+\right)^2, \qquad D_t = \left\{\sum_{s \leq t}\left((\Delta X_s)^-\right)^2\right\}_t^{(p)}, \tag{5.9}$$
$$H_t = \langle M^c\rangle_t + C_t + D_t.$$

Then $\exp\{X_t - V_t^{(p)} - \frac{1}{2}H_t\}$ is a supermartingale.

Applying a special case of Proposition 5.1, we obtain the following inequality.

Theorem 5.3. *Let $\{M_t, \mathscr{F}_t, t \geq 0\}$ be a locally square-integrable martingale, with $M_0 = 0$. In the notation of Proposition 5.1, let $\{V_t = 0\}$. Then for any* *set A measurable with respect to* $\mathscr{F}_\infty$, *and for all* $\beta > 0$, $\alpha, x \geq 0$,

$$P(M_t \geq (\alpha + \beta H_t)x, A) \leq E\left(\exp\left\{-x^2\left[\frac{\beta^2}{2}H_t + \alpha\beta\right]\right\}\Big|(M_t \geq (\alpha + \beta H_t)x, A)\right). \tag{5.10}$$

$$P\left(M_t \geq (\alpha + \beta H_t)x, \frac{1}{H_t} < y \text{ for some } t < \infty\right) \leq \exp\left\{-x^2\left(\frac{\beta^2}{2y} + \alpha\beta\right)\right\}. \tag{5.11}$$

PROOF. The proof is almost identical to the one of Theorem 5.2. To see this, take $M'_t = \lambda M_t$ in Proposition 5.1 to obtain the appropriate supermartingale. □

It is easy to see that in Theorem 5.3 we could remove the condition $\{V_t = 0\}$ by replacing M_t by $M_t + V_t^{(p)}$.

The natural question that arises is what happens when we have variables which are not truncated if we do not want to impose on them an integrability condition. The next section provides an example where this problem is treated.

6. Sums of conditionally symmetric variables. In what follows we consider the case of conditionally symmetric variables. Let $\{d_i\}$ be a sequence of variables adapted to $\mathscr{F}_n$. Then we say that the d_i's are conditionally symmetric if $\mathscr{L}(d_i \mid \mathscr{F}_{i-1}) = \mathscr{L}(-d_i \mid \mathscr{F}_{i-1})$.

Under these assumptions we can prove the following results.

THEOREM 6.1. *Let* $\{d_i\}$ *be a sequence of random variables adapted to* $\mathscr{F}_n$. *Assume the variables are conditionally symmetric. Then, for all* $x, y > 0$,

$$P\left(\sum_{i=1}^n d_i \geq x, \sum_{i=1}^n d_i^2 \leq y \text{ for some } n\right) \leq \exp\left\{-\frac{x^2}{2y}\right\}. \tag{6.1}$$

For self-normalized sums we get Theorem 6.2.

THEOREM 6.2. *Let* $\{d_i\}$ *be a sequence of random variables adapted to* $\mathscr{F}_n$. *Assume the variables are conditionally symmetric. Then, for all sets* $A \in F_\infty$ *and all* $\beta > 0 \alpha$, $x \geq 0$,

$$P\left(\frac{\sum_{i=1}^n d_i}{\alpha + \beta\sum_{i=1}^n d_i^2} \geq x, A\right) \leq E\left[\exp\left\{-x^2\left(\frac{\beta^2}{2}\sum_{i=1}^n d_i^2 + \alpha\beta\right)\right\}\Big|\left(\frac{\sum_{i=1}^n d_i}{\alpha + \beta\sum_{i=1}^n d_i^2} \geq x, A\right)\right], \tag{6.2}$$

$$(6.3)\quad P\left(\frac{\sum_{i=1}^n d_i}{\alpha+\beta\sum_{i=1}^n d_i^2}\geq x,\ A\right)\leq\sqrt{E\exp\left\{-x^2\left(\frac{\beta^2}{2}\sum_{i=1}^n d_i^2+\alpha\beta\right)\right\}},$$

and finally, for all $x, y > 0$,

$$(6.4)\quad \begin{aligned} &P\left(\frac{\sum_{i=1}^n d_i}{\alpha+\beta\sum_{i=1}^n d_i^2}\geq x,\ \frac{1}{\sum_{i=1}^n d_i^2}\leq y \text{ for some } n\right)\\ &\qquad\leq\exp\left\{-x^2\left(\frac{\beta^2}{2y}+\alpha\beta\right)\right\}.\end{aligned}$$

PROOF OF THEOREM 6.2. The proof of Theorem 6.2 is based on a lemma which we give first. We were inspired by the work of Burkholder (1991), Lemma 10.2, and Pinelis (1994) when attempting this problem. Hitczenko (1990a) presents the case of conditionally symmetric martingale differences through an approximation scheme. He refers to Wang (1989) for the key idea of using the σ-algebra ($\mathscr{H}_n$) introduced below.

LEMMA 6.1. *Let* $\{d_i\}$ *be a sequence of conditionally symmetric random variables. Then, for all* $\lambda > 0$,

$$(6.5)\quad \frac{\exp\{\sum_{i=1}^n \lambda d_i\}}{\exp\{(\lambda^2/2)\sum_{i=1}^n d_i^2\}},$$

is a supermartingale.

PROOF. Let $\mathscr{H}_0$ be the trivial σ-field and for $n \geq 1$, let $\mathscr{H}_n$ be the sigma field generated by $d_1, \dots, d_{n-1}, |d_n|$. Similarly, let $\mathscr{F}_0$ be the trivial σ-field and for $n \geq 2$, let $\mathscr{F}_{n-1}$ be the sigma field generated by $d_1, \dots, d_{n-1}$. For all sets $H_n \in \mathscr{H}_n$ and $F_{n-1} \in \mathscr{F}_{n-1}$, define

$$H_n^{-1} = \{\omega\colon d_1(\omega), \dots, d_{n-1}(\omega), |d_n(\omega)| \in H_n\},$$

and similarly $F_{n-1}^{-1} = \{\omega\colon d_1(\omega), \dots, d_{n-1}(\omega) \in F_{n-1}\}$. Then, the conditional symmetry of $\{d_i\}$ implies that the conditional distributions of d_n given $\mathscr{H}_n$ and that of $-d_n$ given $\mathscr{H}_n$ are the same. In what follows we show in detail that

$$E[\exp\{\lambda d_n\} \mid \mathscr{H}_n] = E[\exp\{-\lambda d_n\} \mid \mathscr{H}_n].$$

To attain this goal, observe that for all sets $H_n \in \mathscr{H}_n$, $F_{n-1} \in \mathscr{F}_{n-1}$ and all $\lambda > 0$ we have that

$$\begin{aligned}
&\int_{d_1,\dots,d_{n-1},|d_n|\in H_n}\exp\{\lambda d_n\}\,dP\\
&\quad=\int_{\omega\in H_n^{-1}}\exp\{\lambda d_n\}\,dP\\
&\quad=\int_{\omega\in H_n^{-1}\cap F_{n-1}^{-1}}\exp\{\lambda d_n\}\,dP+\int_{\omega\in H_n^{-1}\cap F_{n-1}^{-1}}\exp\{\lambda d_n\}\,dP\\
&\quad=\int_{\omega\in F_{n-1}^{-1}}\exp\{\lambda d_n\}1(d_1,\dots,d_{n-1},|d_n|\in H_n)\,dP
\end{aligned}$$

$$
\begin{aligned}
&+\int_{\omega \in F_{n-1}^{c^{-1}}} \exp\{\lambda d_n\} 1(d_1, \dots, d_{n-1}, |d_n| \in H_n)\, dP \\
&= \int_{\omega \in F_{n-1}^{-1}} \exp\{-\lambda d_n\} 1(d_1, \dots, d_{n-1}, |-d_n| \in H_n)\, dP \\
&\quad + \int_{\omega \in F_{n-1}^{c^{-1}}} \exp\{-\lambda d_n\} 1(d_{1,\dots, dn-1}, |-d_n| \in H_n)\, dP \\
&= \int_{\omega \in H_n^{-1}} \exp\{-\lambda d_n\} \\
&= \int_{d_1, \dots, d_{n-1}, |d_n| \in H_n} \exp\{-\lambda d_n\}\, dP.
\end{aligned}
$$

Moreover, $(\exp\{\lambda d_n\} + \exp\{-\lambda d_n\})/2$ is measurable with respect to $\mathscr{H}_n$. Therefore,

$$
\begin{aligned}
E[\exp\{\lambda d_n\} \mid \mathscr{H}_n] &= E\left[\frac{\exp\{\lambda d_n\} + \exp\{-\lambda d_n\}}{2} \,\middle|\, \mathscr{H}_n \right] \\
&= \frac{\exp\{\lambda d_n\} + \exp\{-\lambda d_n\}}{2} \le \exp\left\{ \frac{\lambda^2 d_n^2}{2} \right\}.
\end{aligned}
$$

Using these observations, we will show that for all $\lambda > 0$,

$$
E \frac{\exp\{\sum_{i=1}^n \lambda d_i\}}{\exp\{(\lambda^2/2)\sum_{i=1}^n d_i^2\}} \le 1.
$$

By conditioning on $\mathscr{H}_n$ we have that

$$
\begin{aligned}
&E\left[\frac{\exp\{\sum_{i=1}^{n-1} \lambda d_i\}}{\exp\{(\lambda^2/2)\sum_{i=1}^n d_i^2\}} E(\exp\{\lambda d_n\} \mid \mathscr{H}_n) \right] \\
&\quad \le E\left[\frac{\exp\{\sum_{i=1}^{n-1} \lambda d_i\}}{\exp\{(\lambda^2/2)\sum_{i=1}^n d_i^2\}} \exp\left\{ \frac{\lambda^2}{2} d_n^2 \right\} \right] \\
&\quad = E\left[\frac{\exp\{\sum_{i=1}^{n-1} \lambda d_i\}}{\exp\{(\lambda^2/2)\sum_{i=1}^{n-1} d_i^2\}} \right].
\end{aligned}
$$

To complete the proof use induction.

We are now ready to complete the proof of Theorems 6.1 and 6.2.

PROOF OF THEOREMS 6.1 AND 6.2. The proof of Theorem 6.1 is quite analogous to that of Theorem 5.1 and therefore is left to the reader. Concerning Theorem 6.2, we will deal with the case $\alpha = 0$, $\beta = 1$, because the general case follows similarly. For all $A \in \mathscr{F}_\infty$ and all $\lambda > 0$,

$$
\begin{aligned}
P&\left(\frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n d_i^2} \geq x,\, A \right) \\
&\leq E \exp\left\{ \frac{\lambda}{2} \sum_{i=1}^n d_i - \frac{\lambda x}{2} \sum_{i=1}^n d_i^2 \right\} 1\left(\frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n d_i^2} \geq x,\, A \right) \\
&= E \frac{\exp\{(\lambda/2)\sum_{i=1}^n d_i\}}{\exp\{(\lambda_4^2)\sum_{i=1}^n d_i^2\}} \exp\left\{ \frac{\lambda^2}{4} \sum_{i=1}^n d_i^2 - \frac{\lambda x}{2} \sum_{i=1}^n d_i^2 \right\} 1\left(\frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n d_i^2} \geq x,\, A \right) \\
&\leq 1 \sqrt{E \exp\left\{ \frac{\lambda^2}{2} \sum_{i=1}^n d_i^2 - \lambda x \sum_{i=1}^n d_i^2 \right\} 1\left(\frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n d_i^2} \geq x,\, A \right)} \\
&\leq \sqrt{E \exp\left\{ -\frac{x^2}{2} \sum_{i=1}^n d_i^2 \right\} 1\left(\frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n d_i^2} \geq x,\, A \right)}\,,
\end{aligned}
$$

where the last step follows by minimizing over λ, taking $\lambda = x$. The proof is complete dividing both sides by $\sqrt{P(\sum_{i=1}^n d_i / \sum_{i=1}^n d_i^2 \geq x,\, A)}$. □

The proof of (6.4) is very analogous to that of Theorem 5.2, so it is left to the reader to fill the gaps.

APPENDIX

When we were about to submit this paper (six days prior to the actual submission date) we found the preprint Caballero, Fernández and Nualart (1996) that is related to this work. We decided to write this Appendix to show what is needed to take the new information into account. Caballero, Fernández and Nualart (1996) was apparently written at roughly the same time as de la Peña (1996a, b). It includes the following inequality.

Let $\{M_t, \mathscr{F}_t\}$ be a continuous martingale null at 0, for all $p > 1$,

$$
\text{(A.1)} \qquad P(M_t \geq x\langle M \rangle_t) \leq \left(E\left[\exp\left\{ -\frac{(p-1)}{2} x^2 \langle M \rangle_t \right\} \right] \right)^{1/p}.
$$

The authors used it to obtain "estimates for the density of a random variable on the Wiener space that satisfies a nondegeneracy condition." They also refer to Exercise 4.18 of Revuz and Yor (1991), where related estimates for the ratio of the martingale over a function of its quadratic variation are given. We observe that their exponential inequality in the case $p = 2$ is a special case of Theorem 5.3. The methods they employ are similar to ours, and the combination of the two approaches can be applied to derive extensions of their result in line with the type of inequality presented in this work. In particular, we will derive the following extension of their work.

THEOREM 5.2A. *Under the assumptions of Theorem* 5.2, *for all* $\mathscr{F}_\infty$ *sets* A,

$$P(M_t \ge x\langle M\rangle_t, A) \le \left(E\left[\exp\left\{-\frac{(p-1)}{2}x^2\langle M\rangle_t\right\}1\left(\frac{M_t}{\langle M\rangle_t} \ge x, A\right)\right]\right)^{1/p}. \tag{A.2}$$

In what follows we will prove (A.2) and will also provide the key to extending the related results in this paper in light of this new information. The basic idea is that the results change only slightly in the case where A is arbitrary. The only difference is that the variable inside the expectation is raised to a $(p-1)$ power and we take the p-th root of the resulting expectation just as in (A.2). Concerning the case of inequalities where A is fixed ahead of time like (6.4), there is no difference in the results obtained. On our way to proving (A.2) we will first extend Lemma 3.1.

LEMMA 3.1A. *Let* X, Y *be two random variables with* $X \ge 0$, $Y \ge 0$, $X/Y \ge 0$ *a.e., and* $E(X/Y) \le K$ *for some constant* K. *Then*

$$EX^{1/q} \le K^{1/q}(EY^{(p-1)})^{1/p}. \tag{A.3}$$

PROOF. By Hölder's inequality, $EX^{1/q} = E(X/Y)^{1/q}Y^{1/q} \le (E(X/Y))^{1/q}(EY^{p/q})^{1/p}$.

To get (A.2), use (A.3) with

$$X = \exp\{\lambda M_t - \lambda x\langle M\rangle_t\}1(M_t \ge x\langle M\rangle_t, A)$$

and

$$Y = \exp\left\{\frac{\lambda^2}{2}\langle M\rangle_t - \lambda x\langle M\rangle_t\right\}1(M_t \ge x\langle M\rangle_t, A).$$

To complete the proof, let $\lambda = x$ and observe that $p/q = p - 1$. □

Several of the results of Section 1 can be extended by using the following alternates to Theorem 3.1 and Corollary 3.1.

THEOREM 3.1A. *In the context of Theorem* 3.1, *for all* $p, q > 1$, *with* $1/p + 1/q = 1$,

$$E\left(g\prod_{i=1}^n d_i\right)^{1/q} \le \left\{E\left[gE\left(\prod_{i=1}^n e_i \,\middle|\, \mathscr{G}\right)\right]^{(p-1)}\right\}^{1/p}. \tag{A.4}$$

PROOF. Take $X = g\prod_{i=1}^n d_i$ and $Y = gE(\prod_{i=1}^n e_i \mid \mathscr{G})$ in (A.3). □

COROLLARY 3.1A. *Under the assumptions of Corollary* 3.1, *for all* $p, q > 1$ *with* $1/p + 1/q = 1$,

$$E\left(g\exp\left\{\lambda\sum_{i=1}^n d_i\right\}\right)^{1/q} \le \left\{E\left(gE\left[\exp\left\{\lambda\sum_{i=1}^n e_i\right\}\middle|\mathscr{G}\right]\right)^{(p-1)}\right\}^{1/p}. \tag{A.5}$$

In what follows we provide an extension to (1.8) in Theorem 1.3, which provides an example of the type of result one can derive using Corollary 3.1A.

THEOREM 1.3A. *Under the assumptions of Theorem* 1.3, *for all* $p > 1$,

$$P\left(\sum_{i=1}^{n} d_i \geq \left(\alpha + \beta V_n^2\right)x,\ A\right)$$

$$\leq \left[E\left\{\left(\frac{1}{\beta x + 1}\right)^{\alpha x}\left[\exp\{\beta x\}\left(\frac{1}{\beta x + 1}\right)^{\beta x + 1}\right]V_n^2\right\}^{(p-1)} \times 1\left(\frac{M_n}{\alpha + \beta V_n^2} \geq x,\ A\right)\right]^{1/p}. \tag{A.6}$$

PROOF. Applying Markov's inequality and Corollary 3.1A with $g = \exp\{-(\lambda/2)V_n^2 x - \lambda\alpha x\}1(M_n/(\alpha + \beta V_n^2) \geq x,\ A)$

$$\begin{aligned}
&P\left(\sum_{i=1}^{n} d_i \geq \left(\alpha + \beta V_n^2\right)x,\ A\right)\\
&\quad\leq \inf_{\lambda > 0} E\exp\left\{-\frac{1}{q}\left(V_n^2\beta x + \lambda\alpha x\right)\right\}\exp\left\{\frac{\lambda}{q}\left(\sum_{i=1}^{n} d_i\right)\right\}\\
&\qquad\times 1\left(M_n \geq \left(\alpha + \beta V_n^2\right)x,\ A\right)\\
&\quad\leq \inf_{\lambda > 0}\left(E\left[\exp\left\{-(p-1)\left(\lambda V_n^2\beta x + \lambda\alpha x\right)\right\}\right.\right.\\
&\qquad\left.\left.\times\left(E\left[\exp\left\{\lambda\sum_{i=1}^{n} e_i\right\} \mid \mathscr{G}\right]\right)^{(p-1)} 1\left(M_n \geq \left(\alpha + \beta V_n^2\right)x,\ A\right)\right]^{1/p}\right)\\
&\quad\leq \inf_{\lambda > 0}\left(E\left[\exp\left\{-(p-1)\left(\lambda V_n^2\beta x + \lambda\alpha x\right)\right\}\right.\right.\\
&\qquad\left.\left.\times\left(\exp\left\{h(\lambda)V_n^2\right\}\right)^{(p-1)}1\left(M_n \geq \left(\alpha + \beta V_n^2\right)x,\ A\right)\right]\right)^{1/p}\\
&\quad= \inf_{\lambda > 0}\left(\exp\{-(p-1)\lambda\alpha x\}\right.\\
&\qquad\left.\times E\left[\exp\left\{-\left(\lambda\beta x - h(\lambda)\right)V_n^2\right\}1\left(M_n \geq \left(\alpha + \beta V_n^2\right)x,\ A\right)\right]^{(p-1)}\right)^{1/p}.
\end{aligned}$$

The minimum inside the expectation is attained at $\lambda = \log(\beta x + 1)$, providing (A.6). □

Observe that when specializing to the case $A = \{1/V_n^2 \le y\}$, replacing V_n^2 by its lower bound and solving gives,

$$P\left(\frac{M_n}{\alpha + \beta V_n^2} \ge x, \frac{1}{V_n^2} \le y\right) \le \left(\frac{1}{\beta x + 1}\right)^{\alpha x}\left[\exp\{\beta x\}\left(\frac{1}{\beta x + 1}\right)^{\beta x + 1}\right]^{1/y}, \tag{A.7}$$

which is a special case of (1.11). Therefore, there is no difference in the resulting inequality if V_n^2 is replaced by its bound.

The same phenomenon repeats throughout the remaining results of the paper.

Acknowledgments. This work (except for the Appendix) gathers in polished form the results of two works by de la Peña (1996a, b), which were written while the author was visiting the Department of Statistics of Stanford University. We thank M. J. Klass and T. L. Lai, who during that time provided valuable comments on the contents of those two papers. We also thank them and Guillermo Gallego and B. Narasimhan for stimulating conversations. We were guided by a remark of J. L. Doob who, on a hiking trip around 1991, told us that one of the important problems in probability was the study of the properties of the ratio of a martingale over a submartingale and by a question of M. Talagrand, who around the same time asked us if we could do decoupling with constraints. Finally, we thank the Departments of Statistics of Stanford and Berkeley where we spent the fall of 1996 and the Department of Mathematical Statistics of Aahrhus University (where the final draft was completed) for their support and hospitality.

REFERENCES

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

BENNETT, G. (1962). Probability inequalities for sums of independent random variables, *J. Amer. Statist. Assoc.* **57** 33–45.

BURKHOLDER, D. L. (1991). Exploration in martingale theory and its applications. *Ecole d'Eté de Probabilités de Saint-Flour XIX. Lecture Notes in Math.* **1464** 1–66. Springer, Berlin.

CABALLERO, M. E., FERNÁNDEZ, B. and NUALART, D. (1996). Estimation of densities and applications. Univ. Barcelona Math. Preprint Ser. 222.

CHOW, Y. S. and TEICHER, H. (1988). *Probability Theory: Independence, Interchangeability, Martingales*, 2nd ed. Springer, New York.

DE LA PEÑA, H. (1994). A bound on the moment generating function of a sum of dependent variables with an application to simple random sampling without replacement. *Ann. Inst. H. Poincaré Probab. Statist.* **30** 197–211.

DE LA PEÑA, V. H. (1995). A bound on the moment generating function of a sum of dependent variables with an application to simple random sampling without replacement. (Correction). *Ann. Inst. H. Poincaré. Probab. Statist.* **31** 703–704.

DE LA PEÑA, V. H. (1996a). A new class of exponential inequalities I. Preprint.

DE LA PEÑA, V. H. (1996b). A new class of exponential inequalities II. Preprint.

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

HITCZENKO, P. (1990a). Upper bounds for the L_p-norms of martingales. *Probab. Theory Related Fields* **86** 225–238.

HITCZENKO, P. (1990b). Best constants in martingale version of Rosenthal's inequality. *Ann. Probab.* **18** 1656–1668.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

KHOSHNEVISAN, D. (1996). Deviation inequalities for continuous martingales. *Stochastic Process. Appl.* To appear.

KWAPIEŃ, S. and WOYCZYŃSKI, W. A. (1989). Tangent sequences of random variables: basic inequalities and their applications. In *Proceedings of Conference on Almost Everywhere Convergence in Probability and Ergodic Theory* (G. A. Edgar and L. Sucheston, eds.) 237–265. Academic Press, New York.

KWAPIEŃ, S. and WOYCZYŃSKI, W. A. (1992). *Random Series and Stochastic Integrals: Single and Multiple*. Birkhäuser, Boston.

LEVENTHAL, S. A. (1989). A uniform CLT for uniformly bounded families of martingale differences. *J. Theoret. Probab.* **2** 271–287.

MCKEAN, H. P. (1962). A Hölder condition for Brownian local time. *J. Math. Kyoto Univ.* **1** 195–201.

PINELIS, I. (1992). An approach to inequalities for the distributions of infinite-dimensional martingales. In *Probability in Banach Spaces 8* (R. M. Dudley, M. G. Hahn and J. Kuelbs, eds.) 128–134. Birkhäuser, Boston.

PINELIS, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679–1706.

PINELIS, I. (1995). Sharp exponential inequalities for the martingales in the 2-smooth Banach spaces and applications to "scalarizing" decoupling. In *Probability in Banach Spaces 9* 55–70 (J. Hoffmann-Jorgensen, J. Kuelbs and M. Marcus, eds.) Birkhäuser, Boston.

PINELIS, I. and UTEV, S. A. (1989). Exact exponential bounds for sums of independent random variables. *Theory Probab. Appl.* **34** 304–346.

REVUZ, D. and YOR, M. (1991). *Continuous Martingales and Brownian Motion*. Springer, Berlin.

SHIRYAYEV, A. N. (1984). *Probability*. Springer, Berlin.

WANG, G. (1989). Some sharp inequalities for conditionally symmetric martingales. Ph.D. dissertation. Univ. Illinois, Urbana-Champaign.

WISE, G. L. and HALL, E. B. (1993). *Counterexamples in Probability and Real Analysis*. Oxford Univ. Press, New York.

DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027
E-MAIL: vp@wald.stat.columbia.edu