Topic 6
Self-normalization II: Pseudo-maximization

Victor H. de la Peña

Professor of Statistics, Columbia University

Artificial Intelligence Institute for Advances in Optimization
Georgia Institute of Technology 2024

Topics Preview

1. Background on Pseudo-maximization (Method of Mixtures)
2. Gaussian Bounds for $\frac{A}{\sqrt{B^2+\mathbb{E}^2 B}}$
3. Matrix-Normalized Processes
4. Gaussian Bounds for $\frac{A}{\sqrt{(B^2+(\mathbb{E}|A|^p)^{2/p})}}$
5. Boundary-crossing Problem

Topics Preview

1. Background on Pseudo-maximization (Method of Mixtures)
2. Gaussian Bounds for $\frac{A}{\sqrt{B^2+\mathbb{E}^2 B}}$
3. Matrix-Normalized Processes
4. Gaussian Bounds for $\frac{A}{\sqrt{(B^2+(\mathbb{E}|A|^p)^{2/p})}}$
5. Boundary-crossing Problem

Canonical Assumption

Let us carefully recall the canonical assumption in the last class. We assume that, for a pair of random variables A, B with $B > 0$,

$$\mathbb{E}\exp(\lambda A - \lambda^2 B^2/2) \leq 1, \tag{1}$$

holds

- for all real λ;
- for all $\lambda \geq 0$;
- for all $0 \leq \lambda < \lambda_0$, where $0 < \lambda < \infty$.

If the global maximizer $\hat{\lambda} := \frac{A}{B^2}$ lies in the regime we are interested, and is of course, deterministic, then by CHEBYSHEV inequality, we have

$$\mathbb{P}\left(\frac{|A|}{|B|} > x\right) = \mathbb{P}\left(\frac{A^2}{2B^2} > \frac{x^2}{2}\right) \leq e^{-\frac{x^2}{2}} \mathbb{E} e^{\frac{A^2}{2B^2}} \leq e^{-\frac{x^2}{2}}.$$

a beautiful Gaussian bound.

Framework of Pseudo-maximization

Unfortunately, since $\frac{A}{B^2}$ is random, we need an alternative method for dealing with this maximization. And today, I introduce one celebrated approach: pseudo-maximization. An informal framework of this method can be stated as follows:

(i)For $\lambda \in \Lambda$ a measurable set, we construct a finite (probability) measure of λ, with distribution function F independent of A and B.

(ii) Now we have that

$$\mathbb{E}\left[\exp\left(\lambda A - \frac{\lambda^2 B^2}{2}\right)\right] = \frac{1}{F(+\infty)}\int_{\mathbb{R}} \mathbb{E}\left[\exp\left(\lambda A - \frac{\lambda^2 B^2}{2}\right)\right] dF$$
$$= \frac{1}{F(+\infty)}\mathbb{E}\left[\int_{\mathbb{R}} \exp\left(\lambda A - \frac{\lambda^2 B^2}{2}\right) dF\right]$$

by FUBINI. And we note that $F(+\infty) = 1$ when we assume a probability measure F for λ.

To be effective for all possible pairs, the F chosen would need to be as uniform as possible so as to include the maximum value of $\exp(\lambda A - \lambda^2 B^2/2)$. And for (1) that holds for all real λ or $\lambda \geq 0$, since all finite measures vanish at infinity, we need to construct a measure decaying as slowly as we can manage.

Topics Preview

1. Background on Pseudo-maximization (Method of Mixtures)
2. Gaussian Bounds for $\frac{A}{\sqrt{B^2+\mathbb{E}^2 B}}$
3. Matrix-Normalized Processes
4. Gaussian Bounds for $\frac{A}{\sqrt{(B^2+(\mathbb{E}|A|^p)^{2/p})}}$
5. Boundary-crossing Problem

One application of pseudo-maximization is to construct a Gaussian bound for $\frac{A}{\sqrt{B^2+\mathbb{E}^2 B}}$, with the theorem stated below.

Theorem (de la Peña, Klass & Lai [4])

Let A, B with $B > 0$ be random variables satisfying the canonical assumption (7) for all $\lambda \in \mathbb{R}$. Then

$$\mathbb{P}\left(\frac{|A|}{\sqrt{B^2 + \mathbb{E}^2 B}} \geq x\right) \leq \sqrt{2}\exp\left(-\frac{x^2}{4}\right). \tag{2}$$

The following lemma, established through pseudo-maximization, plays a pivotal role in proving Theorem 1.

Lemma

Under the canonical assumption (1) for all $\lambda \in \mathbb{R}$, for every $y > 0$, we have

$$\mathbb{E}\left[\frac{y}{\sqrt{B^2 + y^2}} \exp\left(\frac{A^2}{2(B^2 + y^2)}\right)\right] \leq 1 \qquad (3)$$

Considering that $\lambda \in R$, we here let $\lambda \sim \mathcal{N}(0, \frac{1}{y^2})$. Multiplying both sides of (1) by $(2\pi^{-1/2})y\exp(-\lambda^2 y^2/2)$ and integrating over λ, we obtain that

$$1 \geq \int_{\mathbb{R}} \mathbb{E}\frac{y}{\sqrt{2\pi}}\exp\left(\lambda A - \frac{\lambda^2}{2}B^2\right)\exp\left(-\frac{\lambda^2 y^2}{2}\right)d\lambda$$

$$= \mathbb{E}[\frac{y}{\sqrt{B^2+y^2}}\exp\left(\frac{A^2}{2(B^2+y^2)}\right)$$

$$\int_{\mathbb{R}}\frac{\sqrt{B^2+y^2}}{\sqrt{2\pi}}\exp\{-\frac{B^2+y^2}{2}\left(\lambda^2 - 2\frac{A}{B^2+y^2} + \frac{A^2}{(B^2+y^2)^2}\right)\}d\lambda] \tag{4}$$

$$= \mathbb{E}\left[\frac{y}{\sqrt{B^2+y^2}}\exp\left(\frac{A^2}{2(B^2+y^2)}\right)\right],$$

where in step (4) we fix A, B inside $\mathbb{E}(\cdot)$ and change the measure to $\lambda \sim \mathcal{N}(\frac{A}{B^2+y^2}, \frac{1}{B^2+y^2})$.

Now we have prepared to prove the theorem, with the inequality in the lemma above:

$$\mathbb{E}\left[\frac{y}{\sqrt{B^2+y^2}}\exp\left(\frac{A^2}{2(B^2+y^2)}\right)\right]\leq 1 \tag{5}$$

By CAUCHY-SCHWARZ inequality and the inequality above, we have that

$$\begin{aligned}\mathbb{E}\exp\left(\frac{A^2}{4(B^2+y^2)}\right) &\leq \sqrt{\mathbb{E}\frac{y\exp\left(\frac{A^2}{2(B^2+y^2)}\right)}{\sqrt{B^2+y^2}}\ \ \mathbb{E}\sqrt{\frac{B^2+y^2}{y^2}}} \\ &\leq \sqrt{\mathbb{E}\sqrt{\frac{B^2+y^2}{y^2}}}.\end{aligned}$$

Recall that B is nonnegative a.s., and we have that $\mathbb{E}\sqrt{\frac{B^2+y^2}{y^2}} \leq \mathbb{E}(\frac{B}{y}+1)$, which means that we can set $y = \mathbb{E}B$ so that

$$\mathbb{E}\exp(\frac{A^2}{4(B^2+y^2)}) \leq \sqrt{\mathbb{E}\sqrt{\frac{B^2+y^2}{y^2}}} \leq \sqrt{2}.$$

Finally, combining Markov's inequality, we have that

$$\mathbb{P}\left(\frac{|A|}{\sqrt{B^2+\mathbb{E}^2B}} \geq x\right) = \mathbb{P}\left(\frac{A^2}{4(B^2+\mathbb{E}^2B)} \geq \frac{x^2}{4}\right) \leq \sqrt{2}\exp\left(-\frac{x^2}{4}\right).$$

Topics Preview

We then extend the canonical assumption to the setting of a random vector A and the canonical assumption on a random vector A and a symmetric, positive definite random matrix C:

$$\mathbb{E}\exp\left(\theta^T A - \frac{1}{2}\theta^T C\theta\right), \quad \forall\theta \in \mathbb{R}^d. \tag{6}$$

The following two lemmas below are the extensions of the corresponding results in Lecture 6. More examples can be seen in Section 14.1.2 of the book by LAI, SHAO, and me [5].

Examples

Lemma

Let M_t be a continuous, square-integrable martingale taking values in $\mathbb{R}^d$, with $M_0 = \mathbf{0}$. Then $\exp\{\theta^T M_t - \theta^T \langle M \rangle_t \theta/2\}$ is a supermartingale for all $\theta \in \mathbb{R}^d$, and therefore

$$\mathbb{E} \exp\{\theta^T M_t - \theta^T \langle M \rangle_t \theta/2\} \leq 1.$$

Lemma

Let $\{d_i\} \subseteq \mathbb{R}^d$ be a sequence of variables adapted to an increasing sequence of $\sigma-$fields $\{\mathcal{F}_i\}$. Assume that the d_i's are conditionally symmetric (i.e., $\mathcal{L}(d_i|\mathcal{F}_{i-1}) = \mathcal{L}(-d_i|\mathcal{F}_{i-1})$). Then $\exp(\theta^T(\sum_{i=1}^n d_i) - \theta^T(\sum_{i=1}^n d_i^2)\theta/2)$, $n \geq 1$, is a supermartingale with mean ≤ 1, for all $\theta \in \mathbb{R}^d$.

Recall this aforementioned lemma, which is the one-dimensional version.

Lemma

Under the canonical assumption for a pair of random variables A, B *with* $B > 0$,

$$\mathbb{E}\exp(\lambda A - \lambda^2 B^2/2) \leq 1, \tag{7}$$

for all $\lambda \in \mathbb{R}$, *for every* $y > 0$, *we have*

$$\mathbb{E}\left[\frac{y}{\sqrt{B^2 + y^2}}\exp\left(\frac{A^2}{2(B^2 + y^2)}\right)\right] \leq 1 \tag{8}$$

We next prove the following lemma, which is the multi-variate version of Lemma before.

Lemma (de la Peña, Klass & Lai [9])

Let a random vector A and a symmetric, positive definite random matrix C satisfy the following canonical assumption

$$\mathbb{E}\exp\left(\theta^T A - \frac{1}{2}\theta^T C\theta\right) \leq 1, \quad \forall \theta \in \mathbb{R}^d. \tag{9}$$

Let V be a positive definite nonrandom matrix, then

$$\mathbb{E}\left[\sqrt{\frac{\det(V)}{\det(C+V)}}\exp\left(\frac{1}{2}A^T(C+V)^{-1}A\right)\right] \leq 1, \tag{10}$$

$$\mathbb{E}\exp\left(\frac{1}{4}A^T(C+V)^{-1}A\right) \leq \sqrt{\mathbb{E}\sqrt{\det(I+V^{-1}C)}}. \tag{11}$$

We set $\theta \sim \mathcal{N}(\mathbf{0}, V^{-1})$, i.e., the density function of θ is

$$f(\theta) = (2\pi)^{-d/2}\sqrt{\det V}\exp(-\theta^T V\theta/2), \theta \in \mathbb{R}^d.$$

After multiplying both sides of (9) by $f(\theta)$ and integrating over θ. By FUBINI's theorem,

$$1 \geq \mathbb{E}\left[\frac{\sqrt{\det V}}{(2\pi)^{-\frac{d}{2}}} e^{A^T(C+V)^{-1}A/2} \int_{\mathbb{R}^d} e^{-[\theta-(C+V)^{-1}A]^T(C+V)[\theta-(C+V)^{-1}A]} d\theta\right]$$

$$= \mathbb{E}\sqrt{\frac{\det(V)}{\det(C+V)}} e^{A^T(C+V)^{-1}A/2},$$

proving the first inequality (10).

To prove the second inequality (11), apply the first inequality (10) to the upper bound in the CAUCHY–SCHWARZ inequality

$$\mathbb{E}\exp\left(\frac{1}{4}A^T(C+V)^{-1}A\right) \leq \left[\mathbb{E}\sqrt{\frac{\det(V)}{\det(C+V)}}e^{A^T(C+V)^{-1}A/2}\right]^{\frac{1}{2}}\left[\mathbb{E}\sqrt{\frac{\det(C+V)}{\det(V)}}\right]^{\frac{1}{2}}.$$

We will discuss the application of this lemma (high-dimensional version) on the construction of confidence intervals later.

Topics Preview

1. Background on Pseudo-maximization (Method of Mixtures)
2. Gaussian Bounds for $\frac{A}{\sqrt{B^2+\mathbb{E}^2 B}}$
3. Matrix-Normalized Processes
4. Gaussian Bounds for $\frac{A}{\sqrt{(B^2+(\mathbb{E}|A|^p)^{2/p})}}$
5. Boundary-crossing Problem

De la Peña and Pang then provided the Gaussian bound for the tail probability of a similar self-normalized statistic $\frac{A}{\sqrt{(B^2+(\mathbb{E}|A|^p)^{2/p})}}$, where $p \geq 1$ can be any number such that $A \in \mathbb{L}^p$.

Theorem (de la Peña & Pang, [2])

Under the canonical assumption for all real λ, suppose $\mathbb{E}|A|^p < \infty$ for some $p \geq 1$. Then for any $x > 0$ and $q \geq 1$ such that $1/p + 1/q = 1$,

$$\mathbb{P}\left(\frac{|A|}{\sqrt{\frac{2q-1}{q}(B^2 + (\mathbb{E}|A|^p)^{2/p})}} \geq x\right) \leq \left(\frac{q}{2q-1}\right)^{\frac{q}{2q-1}} x^{\frac{-q}{2q-1}} exp\{-x^2/2\}, x > 0 \tag{12}$$

Remark

If we assume A is integrable, or square-integrable, the bound in the RHS of (12) becomes $\frac{2}{3}^{2/3} x^{-2/3} \exp(-x^2/2)$ or $2^{-1/2} x^{-1/2} \exp(-x^2/2)$.

We establish the following identity: for any $C > 0$,

$$\mathbb{E}\left[\frac{C}{\sqrt{B^2+C}}\exp\left(\frac{A^2}{2(B^2+C)}\right)\right] \leq 1,$$

which is exactly the lemma of the one-dimensional version (where $C = y^2 > 0$). You are suggested to recall how to obtain this inequality with Gaussian mixture. Now let $G \in \mathcal{F}$ be any measurable set. Then, by MARKOV's inequality,

$$\begin{aligned}
\mathbb{P}\left(\frac{|A|}{\sqrt{B^2+C}} \geq x, G\right) &= \mathbb{P}\left(\frac{|A|^2}{4(B^2+C)} \geq \frac{x^2}{4}, G\right) \\
&\leq \mathbb{P}\left(\frac{|A|^{1/2}}{(B^2+C)^{1/4}}\exp\left(\frac{|A|^2}{4(B^2+C)}\right) \geq x^{1/2}\exp(x^2/4), G\right) \\
&\leq x^{-1/2}\exp(-x^2/4)\mathbb{E}\left[\left(\frac{|A|^2}{(B^2+C)}\right)^{1/4}\exp\left(\frac{|A|^2}{4(B^2+C)}\right) I_G\right].
\end{aligned}$$

Now, by HÖLDER's inequality,

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{|A|^2}{(B^2+C)}\right)^{1/4}\exp\left(\frac{|A|^2}{4(B^2+C)}\right)I_G\right] \\
&=\mathbb{E}\left[\left(\frac{C^{1/4}}{|A|^{1/2}}\frac{|A|^2}{(B^2+C)}\right)^{1/4}\exp\left(\frac{|A|^2}{4(B^2+C)}\right)\frac{|A|^{1/2}}{C^{1/4}}I_G\right] \\
&\le\left(\mathbb{E}\left(\frac{C}{B^2+C}\right)^{1/2}\exp\left(\frac{|A|^2}{2(B^2+C)}\right)\right)^{1/2}\left(\mathbb{E}\left[\frac{|A|^{1/2}}{C^{1/2}}I_G\right]\right)^{1/2} \\
&\le\left(\mathbb{E}\left[\frac{|A|}{C^{1/2}}I_G\right]\right)^{1/2}.
\end{aligned}$$

If $A \in \mathbb{L}^p$ for $p > 1$, we can choose $C = (\mathbb{E}|A|^p)^{2/p}$ so that for p, q satisfying $1/p + 1/q = 1$, again by HÖLDER's inequality,

$$\mathbb{E}\left[\frac{|A|}{C^{1/2}} I_G\right] \leq \left(\mathbb{E}\left[\frac{|A|^p}{C^{p/2}} I_G\right]\right)^{1/p} \mathbb{P}(G)^{1/q} \leq \mathbb{P}(G)^{1/q}.$$

This implies that

$$\mathbb{P}\left(\frac{|A|}{\sqrt{B^2 + (\mathbb{E}|A|^p)^{2/p}}} \geq x, G\right) \leq x^{1/2} \exp\left(\frac{-x^2}{4}\right) \mathbb{P}(G)^{1/2q}.$$

Now letting $G = \{\frac{|A|}{\sqrt{B^2+(\mathbb{E}|A|^p)^{2/p}}} \geq x\}$, we obtain

$$\mathbb{P}\left(\frac{|A|}{\sqrt{B^2 + (\mathbb{E}|A|^p)^{2/p}}} \geq x\right) \leq x^{-\frac{q}{2q-1}} \exp\left(\frac{-q}{2(2q-1)} x^2\right),$$

as we claimed.

Martingale Inequality

Now let us focus on the case of martingale. Bercu and Touati [1], provides the following lemma, which satisfies the canonical assumption.

Lemma

Let $\{X_i : i \geq 1\}$ *be a martingale difference sequence w.r.t. the filtration* $\mathbb{F} = \{\mathcal{F}_n : n \geq 1\}$ *and suppose that* $\mathbb{E}X_i^2 < \infty$ *for all* $i \geq 1$*. Then, for all* $\lambda \in \mathbb{R}$,

$$\mathbb{E}\left[\exp\left(\sum_{i=1}^{n} X_i - \frac{\lambda^2}{2}\left(\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} \mathbb{E}[X_i^2 | \mathcal{F}_{i-1}]\right)\right)\right] \leq 1. \tag{13}$$

This lemma and the above theorem lead to the following concentration inequality.

Theorem (de la Peña & Pang, [2])

With the same setting in Lemma 9, let τ be any stopping time with respect to the filtration $\mathbb{F}$ and assume $\tau < \infty$ a.s.. then equation (13) holds even when n is replaced by τ, and for $x > 0$,

$$\mathbb{P}\left(\frac{|\sum_{i=1}^{\tau} X_i|}{\sqrt{\frac{3}{2}\left(\sum_{i=1}^{\tau} X_i^2 + \sum_{i=1}^{\tau}\mathbb{E}[X_i^2|\mathcal{F}_{i-1}] + \mathbb{E}[\sum_{i=1}^{\tau} X_i^2]\right)}} \geq x\right) \leq \left(\frac{2}{3}\right)^{\frac{2}{3}} x^{-\frac{2}{3}} e^{-\frac{x^2}{2}}.$$

Fan et al. [7] applied those theorems by de la Peña and Pang to construct self-normalized Cramér type moderate deviations for self-normalized martingales.

Topics Preview

1. Background on Pseudo-maximization (Method of Mixtures)
2. Gaussian Bounds for $\frac{A}{\sqrt{B^2+\mathbb{E}^2 B}}$
3. Matrix-Normalized Processes
4. Gaussian Bounds for $\frac{A}{\sqrt{(B^2+(\mathbb{E}|A|^p)^{2/p})}}$
5. Boundary-crossing Problem

In the next section, we will use the method of mixtures to do an analysis of the boundary-crossing problems. This method was first introduced by ROBBINS and SIEGMUND [10] and later refined by LAI [8], at Columbia.

Theorem (Law of the Iterated Logarithm)

Let Y_n be independent, identically distributed random variables with means zero and variances σ^2. Let $S_n = Y_1 + ... + Y_n$. Then

$$\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{2n\sigma^2 loglogn}} = 1 \ a.s.. \tag{14}$$

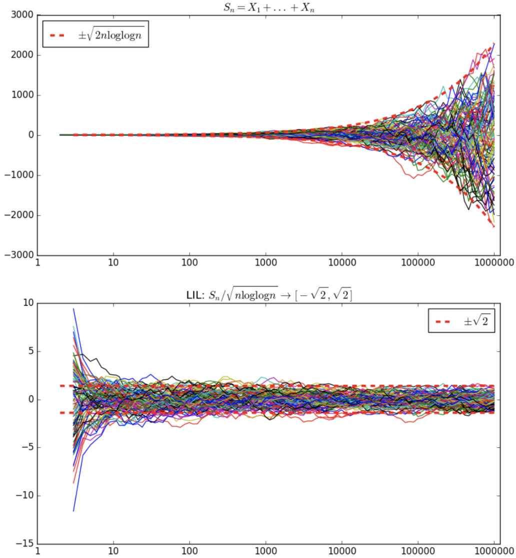

Figure: Law of the Iterated Logarithm: Simple Symmetric Random Walk, obtained from Wikipedia

Canonical Assumption

We plan to extend this limiting theorem to self-normalization $\frac{A_t}{B_t}$ and we make the following refinement of the canonical assumption:

$$\{\exp(\lambda A_t - \lambda^2 B_t^2/2), t \geq 0\} \text{ is a supermartingale with mean } \leq 1, \quad (15)$$

for $0 \leq \lambda < \lambda_0$, with $A_0 = 0$.

Instead of constructing a probability measure, we let F be a finite positive measure on $(0, \lambda_0)$ and assume that $F(0, \lambda_0) > 0$. Without assuming the exact density (or mass) of F, let us first consider the following function $\Psi : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^+$, such that

$$\Psi(u, v) = \int_{(0,\lambda_0)} \exp\left(\lambda u - \frac{\lambda^2 v}{2}\right) dF(\lambda).$$

When $v > 0$ is fixed, $\Psi(\cdot, v)$, strictly increasing, map onto $\mathbb{R}^+$, and this implies that when we further fix $c > 0$, the equation

$$\Psi(u, v) = c$$

has a unique solution $u = \beta_F(v, c)$.
Moreover, you can verify that $\beta_F(v, c)$ is a concave function of v and

$$\lim_{v \to \infty} \frac{\beta_F(v, c)}{v} = \frac{b}{2},$$

where

$$b := \sup\{y > 0 : \int_0^y F(d\lambda) = 0\},$$

with sup over the empty set equal to zero.

The ROBBINS-SIEGMUND (R-S) boundaries $\beta_F(v, c)$ can be extended to analyse the random boundary crossing probability

$$\mathbb{P}(A_t \geq g(B_t) \text{ for some } t \geq 0),$$

when $g(B_t) = \beta_F(B_t^2, c)$ for some F and $c > 0$. This probability equals

$$\mathbb{P}(A_t \geq g(B_t) \text{ for some } t \geq 0) = \mathbb{P}(\Psi(A_t, B_t^2) \geq c \text{ for some } t \geq 0)$$

$$= \lim_{T\uparrow\infty} \mathbb{P}(\sup_{t\leq T} \int_{(0,\lambda_0)} e^{\lambda A_t - \frac{\lambda^2 B_t^2}{2}} dF(\lambda) \geq c)$$

$$\leq \lim_{T\uparrow\infty} \frac{1}{c}\mathbb{E}[\int_{(0,\lambda_0)} e^{\lambda A_T - \frac{\lambda^2 B_T^2}{2}} dF(\lambda)] \qquad \text{[by DOOB's Inequality]}$$

$$= \lim_{T\uparrow\infty} \frac{1}{c}\int_{(0,\lambda_0)} \mathbb{E}\left[e^{\lambda A_T - \frac{\lambda^2 B_T^2}{2}}\right] dF(\lambda) \qquad \text{[using pseudo-maximization]}$$

$$\leq \frac{1}{c}\int_{(0,\lambda_0)} dF(\lambda) = \frac{F((0,\lambda_0))}{c}.$$

Let us denote by $\log_2(\cdot) = \log\log(\cdot)$ and $\log_3(\cdot) = \log\log\log(\cdot)$. For $\delta > 0$, $\lambda \in (0, e^{-e})$ we assume $F \ll$ *Lebesgue*, such that

$$dF(\lambda) = \frac{1}{\lambda(\log\frac{1}{\lambda}(\log_2\frac{1}{\lambda})^{1+\delta})}d\lambda$$

As shown in example 4 of Robbins and Siegmund [10], for fixed c,

$$\beta_F(v,c) = \sqrt{2v[\log_2 v + (\frac{3}{2}+\delta)\log_3 v + \log(\frac{c}{2\sqrt{\pi}}) + o(1)]},$$

as $v \to \infty$. With this choice of F, the probability $\mathbb{P}(A_t \geq g(B_t)$ for some $t \geq 0)$ is bounded by $F(0, e^{-e})/c$ for all $c > 0$. Given $\epsilon > 0$, take c large enough so that $F(0, e^{-e})/c < \epsilon$. Since ϵ can be arbitrarily small and since for fixed c, $\beta_F(v,c) \sim \sqrt{2v\log\log v}$ as $v \to \infty$,

$$\limsup \frac{A_t}{2B_t^2 \log\log B_t^2} \leq 1,$$

on the event set $\{\lim_{t\to\infty} B_t = \infty\}$.

Theorem (de la Peña, Klass & Lai, [6])

Assume that

$$\{\exp\left(\lambda A_t - \lambda \frac{B_t^2}{2}\right), t \geq 0\}$$

is a supermartingale with mean ≤ 1. *Then on the set* $\{\lim_{t\to\infty} B_t^2 = \infty\}$,

$$\limsup_{t\to\infty} \frac{A_t}{\sqrt{2B_t^2 \log\log B_t^2}} \leq 1.$$

Let us recall the following example introduced on Monday, where we made no integrability assumption.

Lemma (de la Peña [3])

Let $\{d_i\}$ be a sequence of variables adapted to an increasing sequence of $\sigma-$fields $\{\mathcal{F}_i\}$. Assume that the d_i's are conditionally symmetric (i.e., $\mathcal{L}(d_i|\mathcal{F}_{i-1}) = \mathcal{L}(-d_i|\mathcal{F}_{i-1})$). Then $\exp(\lambda \sum_{i=1}^n d_i - \lambda^2 \sum_{i=1}^n d_i^2/2)$, $n \geq 1$, is a supermartingale with mean ≤ 1, for all $\lambda \in \mathbb{R}$.

We can have, on the set $\{\lim_{n\to\infty} \sum_{i=1}^n d_i^2 = \infty\}$, that

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^n d_i}{\sqrt{2(\sum_{i=1}^n d_i^2) \log\log(\sum_{i=1}^n d_i^2)}} \leq 1,$$

a sharp extension of Kolmogorov's LIL without moment assumptions, which is also valid for i.i.d. centered Cauchy variables.

References I

[1] Bernard Bercu and Abderrahmen Touati. "Exponential inequalities for self-normalized martingales with applications". In: (2008).

[2] V.H. de la Pena and G. Pang. "Exponential inequalities for self-normalized processes with applications". In: (2009).

[3] V. H. de la Peña. "A general class of exponential inequalities for martingales and ratios". In: *The Annals of Probability* 27.1 (1999), pp. 537–564.

[4] V. H. de la Peña, M. J. Klass, and T. L. Lai. "Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws". In: *The Annals of Probability* (2004).

[5] V. H. de la Peña, T. L. Lai, and Q. Shao. *Self-normalized processes: Limit theory and Statistical Applications*. Springer, 2009.

[6] V.H. de la Peña, M.J. Klass, and T.L. Lai. "Pseudo-maximization and self-normalized processes". In: (2007).

References II

[7] Xiequan Fan et al. "Self-normalized Cramér type moderate deviations for martingales". In: (2019).

[8] Tze Leung Lai. "Boundary crossing probabilities for sample sums and confidence sequences". In: *The Annals of Probability* 4.2 (1976), pp. 299–312.

[9] Victor H de la Pena, Michael J Klass, and Tze Leung Lai. "Theory and applications of multivariate self-normalized processes". In: *Stochastic Processes and their Applications* 119.12 (2009), pp. 4210–4227.

[10] Herbert Robbins and David Siegmund. "Boundary crossing probabilities for the Wiener process and sample sums". In: *The Annals of Mathematical Statistics* (1970), pp. 1410–1429.