36-771 Fall 2018

Martingales 1 : Concentration inequalities

Improving de la Peña's self-normalized inequalities

Lecturer : Aaditya Ramdas

1 Self-normalized uniform bounds

De la Peña (1999) and de la Peña et al. (2004, 2007, 2009) give a variety of sufficient conditions for Assumption 1 to hold with equality in the scalar case in both discrete- and continuous-time settings. They formulate their bounds for ratios involving S_t in the numerator and V_t in the denominator, as in Theorem 1(c), and they often specify initial-time conditions, as in Theorem 1(d). In this section we draw some direct comparisons between Theorem 1 and their results. As a first example, consider the boundary of Theorem 1(c) for the ratio S_t/V_t, strictly decreasing towards the asymptotic level $\mathfrak{s}(x)$. In particular, at time $V_t = m$ the boundary equals x, so Theorem 1(c) strengthens various theorems of de la Peña (1999) and de la Peña et al. (2007) which use a constant boundary after time $V_t = m$ [B; C or D]. The figure below illustrates the relationship between the boundary of Theorem 1(c) and those of de la Peña et al. As before, we give explicit results for special cases.

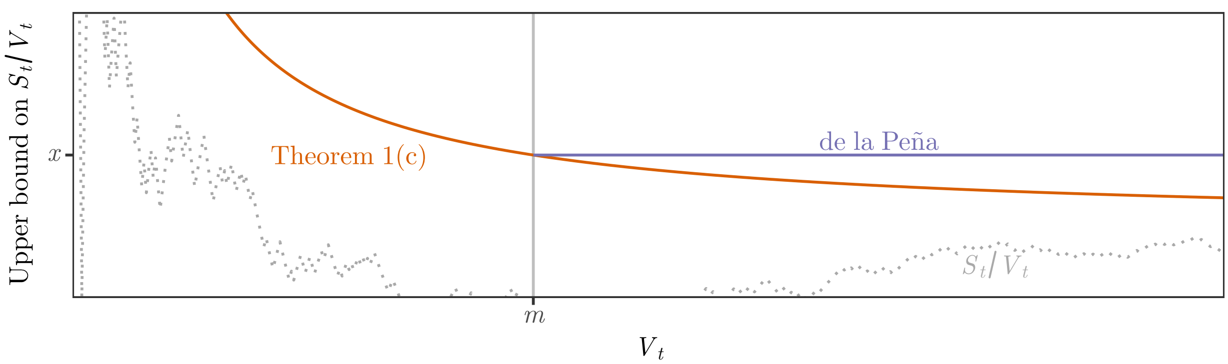

Figure 1: Comparing our decreasing boundary from Theorem 1(c) to the constant boundaries of de la Peña (1999).

Corollary 1 *Let $\mathcal{T} = \mathcal{N}$ and $(Y_t)_{t \in \mathcal{N}}$ be an adapted, $\mathcal{H}^d$-valued process, or let $\mathcal{T} = (0, \infty)$ and $(Y_t)_{t \in (0,\infty)}$ be an adapted, real-valued process. Suppose (Y_t) is sub-gamma with self-normalizing process (U_t), variance process (W_t) and scale parameter c, and let $S_t := \gamma_{\max}(Y_t)$, $V_t := \gamma_{\max}(U_t + W_t)$. Then for any $x, m \geq 0$, we have*

$$\mathcal{P}\left(\exists t \in \mathcal{T} : \frac{S_t}{V_t} \geq \mathfrak{s}_G(x)\left(1 + \frac{m\sqrt{1+2cx}}{V_t}\right)\right) \leq d\exp\{-m\psi_G^{\star}(x)\} \leq d\exp\{-\frac{mx^2}{2(1+cx)}\}.$$

This strengthens the final statement of Theorem 1.2B of de la Peña (1999) [B; C or D]. In the sub-Gaussian case (obtained as $c \to 0$), the above bound simplifies to:

$$\mathcal{P}\left(\exists t \in \mathcal{T} : \frac{S_t}{V_t + m} \geq x\right) \leq d \exp\{-2mx^2\}.$$

This strengthens Theorem 2.1 of de la Peña et al. (2007) and Theorem 6.1 of de la Peña (1999) [B, C or D].

More generally, when we normalize by $\alpha + \beta V_t$ and include an initial time condition $V_t \geq m$, Theorem 1(d) becomes the following:

Corollary 2 *If Assumption 1 holds for some real-valued processes $(S_t)_{t\in\mathcal{T}}$ and $(V_t)_{t\in\mathcal{T}}$, then*

$$\mathcal{P}\left(\exists t \in \mathcal{T} : V_t \geq m \text{ and } \frac{S_t}{\alpha + \beta V_t} \geq x\right) \leq \begin{cases} (\mathbb{E}L_0) \exp\{-\alpha x D(\beta x)\}, & \beta x \leq \mathfrak{s}\left(\frac{x(\alpha+\beta m)}{m}\right) \\ (\mathbb{E}L_0) \exp\{-m\psi^\star\left(\frac{x(\alpha+\beta m)}{m}\right)\}, & \beta x \geq \mathfrak{s}\left(\frac{x(\alpha+\beta m)}{m}\right) \end{cases}$$
$$\leq (\mathbb{E}L_0) \exp\{-m\psi^\star(\beta x) - \alpha x \psi^{\star\prime}(\beta x)\}.$$

For the sub-Gaussian case, let $\mathcal{T} = \mathcal{N}$ and $(Y_t)_{t\in\mathcal{N}}$ be an adapted, $\mathcal{H}^d$-valued process, or let $\mathcal{T} = (0,\infty)$ and $(Y_t)_{t\in(0,\infty)}$ be an adapted, real-valued process. Suppose (Y_t) is sub-Gaussian with self-normalizing process (U_t) and variance process (W_t), and let $S_t := \gamma_{\max}(Y_t)$, $V_t := \gamma_{\max}(U_t + W_t)$. Then for any $\alpha, \beta, m \geq 0$, we have

$$\mathcal{P}\left(\exists t \in \mathcal{T} : V_t \geq m \text{ and } \frac{S_t}{\alpha + \beta V_t} \geq x\right) \leq \exp\{-x^2\left(2\alpha\beta + \frac{(\beta m - \alpha)^2}{2m}\mathbf{1}\{\alpha \leq \beta m\}\right)\}.$$

This improves the final statement in Theorem 6.2 of de la Peña (1999) [B; C or D; E].

A defining feature of self-normalized bounds is that they involve an intrinsic time process (V_t) constructed with the squared observations themselves rather than just conditional variances or constants. Such normalization can be found in common statistical procedures such as the t-test. Furthermore, it allows for Gaussian-like concentration while reducing or eliminating moment conditions.

Corollary 3 *Suppose $\mathcal{T} = \mathcal{N}$ and $(Y_t)_{t\in\mathcal{N}}$ is an $\mathcal{H}^d$-valued martingale with $\mathbb{E}Y_t^2 < \infty$ for all $t \in \mathcal{N}$, and let $S_t := \gamma_{\max}(Y_t)$ and either $V_t := \frac{1}{2}\gamma_{\max}([Y_+]_t + \langle Y_-\rangle_t)$ or $V_t := \frac{1}{3}\gamma_{\max}([Y]_t + 2\langle Y\rangle_t)$. Then for any $x, m \geq 0$, we have*

$$\mathcal{P}\left(\exists t \in \mathcal{N} : \frac{S_t}{V_t + m} \geq x\right) \leq d \exp\{-2mx^2\}.$$

This strengthens the third statement in Theorem 4 of Delyon (2009) [B,D], Theorem 2.1 of Bercu & Touati (2008) [B,D,E], and an implicit self-normalized bound of Mackey et al. (2014, Corollary 4.2) [B].

The above corollary is remarkable for the fact that it gives Gaussian-like concentration with only the existence of second moments for the increments. If the increments have conditionally symmetric distributions, one may instead achieve Gaussian-like concentration without existence of any moments, as illustrated by the following example.

Example 4 (Cauchy increments) *Let $(\Delta S_t)_{t\in\mathcal{N}}$ be i.i.d. standard Cauchy random variables. Since the distribution of ΔS_t is symmetric about zero, we earlier proved that (S_t) is sub-Gaussian with variance process $V_t = [S]_t$. Hence our corollary yields for any $m, x \geq 0$,*

$$\mathcal{P}\left(\exists t \in \mathcal{N} : \frac{S_t}{[S]_t + m} \geq x\right) \leq \exp\{-2mx^2\}.$$

The above result is new to the best of our knowledge, and we are not aware of other ways to prove it. For another example, we give a self-normalized bound involving third rather than second moments:

Corollary 5 *Suppose $\mathcal{T} = \mathcal{N}$ and $(Y_t)_{t\in\mathcal{N}}$ is an $\mathcal{H}^d$-valued martingale with $\mathbb{E}|Y_t|^3 < \infty$ for all $t \in \mathcal{N}$, and let $S_t := \gamma_{\max}(Y_t)$ and $V_t := \gamma_{\max}([Y]_t + \sum_{i=1}^t \mathbb{E}_{i-1}(\Delta Y_i)_-^3)$. Then for any $x, m \geq 0$, we have*

$$\mathcal{P}\left(\exists t \in \mathcal{N} : S_t \geq x + \mathfrak{s}_G\left(\frac{x}{m}\right) \cdot (V_t - m)\right) \leq d\exp\{-m\psi_G^\star\left(\frac{x}{m}\right)\} \leq d\exp\{-\frac{x^2}{2(m + x/6)}\}, \tag{1}$$

where $\mathfrak{s}_G$ and $\psi_G^\star$ use $c = 1/6$. This is a uniform alternative to Corollary 2.2 of Fan et al. (2015) [B,D].

Note the exponent in (1) is different from that in Fan et al. (2015), and neither strictly dominates the other. Also note that, unlike the classical Bernstein bound, neither of the above two bounds assume existence of moments of all orders.

References

Bercu, B. & Touati, A. (2008), 'Exponential inequalities for self-normalized martingales with applications', *The Annals of Applied Probability* **18**(5), 1848–1869.

de la Peña, V. H. (1999), 'A General Class of Exponential Inequalities for Martingales and Ratios', *The Annals of Probability* **27**(1), 537–564.

de la Peña, V. H., Klass, M. J. & Lai, T. L. (2004), 'Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws', *The Annals of Probability* **32**(3), 1902–1933.

de la Peña, V. H., Klass, M. J. & Lai, T. L. (2007), 'Pseudo-maximization and self-normalized processes', *Probability Surveys* **4**, 172–192.

de la Peña, V. H., Lai, T. L. & Shao, Q.-M. (2009), *Self-normalized processes: limit theory and statistical applications*, Springer, Berlin.

Delyon, B. (2009), 'Exponential inequalities for sums of weakly dependent variables', *Electronic Journal of Probability* **14**, 752–779.

Fan, X., Grama, I. & Liu, Q. (2015), 'Exponential inequalities for martingales with applications', *Electronic Journal of Probability* **20**(1), 1–22.

Mackey, L., Jordan, M. I., Chen, R. Y., Farrell, B. & Tropp, J. A. (2014), 'Matrix concentration inequalities via the method of exchangeable pairs', *The Annals of Probability* **42**(3), 906–945.