36-771 Fall 2018

Martingales 1 : Concentration inequalities

Improving Cramer-Chernoff & Freedman's, Hermitian dilation

Lecturer : Aaditya Ramdas

In the discrete-time, scalar setting, a simple sufficient condition for Assumption 1 is that

$$\mathbb{E}_{t-1} \exp\{\lambda \Delta S_t - \psi(\lambda)\Delta V_t\} \leq 1, \quad \forall t,$$

which is the standard assumption for a martingale-method Cramér-Chernoff inequality (McDiarmid 1998, Chung & Lu 2006, Boucheron et al. 2013). When V_t is deterministic, the fixed-time Cramér-Chernoff method gives, for fixed t and x,

$$\mathcal{P}(S_t \geq x) \leq \exp\{-V_t \psi^\star\left(\frac{x}{V_t}\right)\}, \tag{1}$$

so Theorem 1(b) is a uniform *extension* of the Cramér-Chernoff inequality, losing nothing at the fixed time t [B; C or D]. A stopping time argument due to Freedman (1975) extends this to the uniform bound

$$\mathcal{P}(\exists t \in \mathcal{T} : S_t \geq x \text{ and } V_t \leq m) \leq \exp\{-m\psi^\star\left(\frac{x}{m}\right)\}.$$

When V_t is deterministic, analogous uniform bounds follow from Doob's maximal inequality for submartingales, as in Hoeffding (1963, eq. 2.17). Theorem 1(b) strengthens this "Freedman-style" inequality [B; C or D], since it yields tighter bounds for all times t such that $V_t < m$, and also extends the inequality to hold for all times t with $V_t > m$, as illustrated by the figure.

Tropp (2011, 2012) extends the scalar Cramér-Chernoff approach to random matrices via control of the matrix moment-generating function, giving matrix analogues of Hoeffding's, Bennett's, Bernstein's and Freedman's inequalities. Following this approach, Theorem 1 gives corresponding strengthened versions of these inequalities for matrix-valued processes [B].

We summarize explicit results for special cases below. Recall the definitions of $\mathfrak{s}_P, \psi_P^\star, \mathfrak{s}_G, \psi_G^\star$ from earlier.

Corollary 1 *Let $\mathcal{T} = \mathbb{N}$ and $(Y_t)_{t\in\mathbb{N}}$ be an adapted, $\mathcal{H}^d$-valued martingale, or let $\mathcal{T} = (0, \infty)$ and $(Y_t)_{t\in(0,\infty)}$ be an adapted, real-valued local martingale. Let $S_t := \gamma_{\max}(Y_t)$.*

(a) When $\mathcal{T} = \mathbb{N}$, suppose $\Delta Y_t^2 \preceq A_t^2$ a.s. for all t for some $\mathcal{H}^d$-valued predictable sequence (A_t), and let either $V_t := \frac{1}{2}\gamma_{\max}\left(\langle Y\rangle_t + \sum_{i=1}^t A_i^2\right)$ or $V_t := \gamma_{\max}\left(\sum_{i=1}^t A_i^2\right)$. Then for

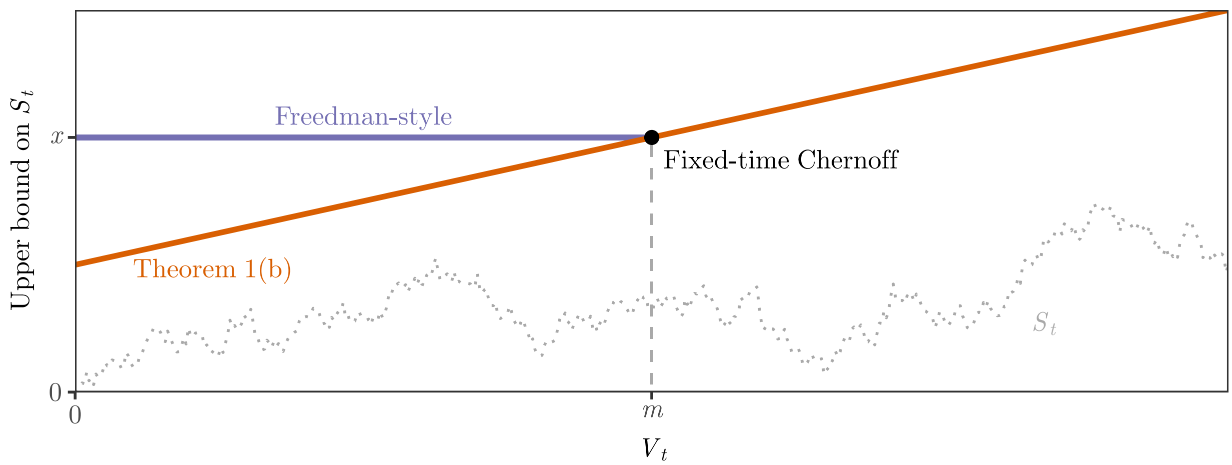

Figure 1: Comparison of (i) fixed-time Cramér-Chernoff bound, which bounds the deviations of S_m at a fixed time m; (ii) "Freedman-style" constant uniform bound, which bounds the deviations of S_t for all t such that $V_t \leq m$, with a constant boundary equal in value to the fixed-time Cramér-Chernoff bound; and (iii) linear uniform bound from Theorem 1, which bounds the deviations of S_t for all $1 \leq n < \infty$, with a boundary growing linearly in V_t. Each bound gives the same tail probability and thus implies the preceding one.

any $x, m > 0$, we have

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

This strengthens Hoeffding's inequality (Hoeffding 1963) [A,B,D] and its matrix analogues in Tropp (2012, Theorem 7.1) [B,E] and Mackey et al. (2014, Corollary 4.2) [A,B].

(b) *Suppose $\gamma_{\max}(\Delta Y_t) \leq c$ a.s. for all t for some constant c, and let $V_t := \gamma_{\max}(\langle Y \rangle_t)$. Then for any $x, m > 0$, we have*

$$\mathcal{P}\left(\exists t \in \mathcal{T} : S_t \geq x + \mathfrak{s}_P\left(\frac{x}{m}\right) \cdot (V_t - m)\right) \leq d \exp\{-m\psi_P^{\star}\left(\frac{x}{m}\right)\} \leq d \exp\{-\frac{x^2}{2(m + cx/3)}\}.$$

This strengthens Bennett's and Freedman's inequalities (Bennett 1962, Freedman 1975) [B; C or D] for scalars and the corresponding matrix bounds from Tropp (2011, 2012) [B].

(c) *Suppose (Y_t) is sub-gamma with self-normalizing process (U_t), variance process (W_t) and scale parameter c, and let $V_t := \gamma_{\max}(U_t + W_t)$. Then for any $x, m > 0$, we have*

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

This strengthens Bernstein's inequality (Bernstein 1927) [B; C or D], along with the matrix Bernstein inequality (Tropp 2012) [B].

The first setting of V_t in case (a) follows from the bound $[Y_+]_t \preceq \sum_{i=1}^t A_i^2$, and further upper bounding $\langle Y_- \rangle_t \preceq \sum_{i=1}^t A_i^2$ yields the second setting of V_t. As is well known, the Hoeffding-style bound in part (a) and the Bennett-style bound in part (b) are not directly comparable: V_t may be smaller in part (b), but $\psi_P^\star \leq \psi_N^\star$, so neither subsumes the other. Additionally, the Hoeffding-style bound requires two-sided boundedness of increments while the Bennett-style bound requires only an upper bound on the deviations of increments above their expectations. It is also worth remarking that $\psi_P^\star(u) \geq \frac{u}{2c} arcsinh\left(\frac{cu}{2}\right)$, so the Bennett-style inequality in part (b) is an improvement on the inequality of Prokhorov (1959) for sums of independent random variables, as noted by Hoeffding (1963), as well as its extension to martingales in de la Peña (1999).

As an example of the Hermitian dilation technique, we give a bound for rectangular matrix Gaussian and Rademacher series, following Tropp (2012); here $\|A\|_{op}$ denotes the largest singular value of A. The proof will be given later.

Corollary 2 *Let $\mathcal{T} = \mathbb{N}$, consider a sequence $(B_t)_{t\in\mathbb{N}}$ of fixed matrices with dimension $d_1 \times d_2$, and let $(\epsilon_t)_{t\in\mathbb{N}}$ be a sequence of independent standard normal or Rademacher variables. Let $S_t := \| \sum_{i=1}^t \epsilon_i B_i \|_{op}$ and $V_t := \max\{\| \sum_{i=1}^t B_i B_i^\star \|_{op}, \| \sum_{i=1}^t B_i^\star B_i \|_{op}\}$. Then for any $x, m > 0$, we have*

$$\mathcal{P}\left(\exists t \in \mathbb{N} : S_t \geq x + \frac{x}{2m}(V_t - m)\right) \leq (d_1 + d_2) \exp\{-\frac{x^2}{2m}\}.$$

This strengthens Corollary 4.2 of Tropp (2012) [B].

Proof: Define the $\mathcal{H}^{d_1+d_2}$-valued process (Y_t) using the dilation of B_t:

$$\Delta Y_t := \epsilon_t \begin{pmatrix} 0 & B_t \\ B_t^\star & 0 \end{pmatrix}.$$

Since the dilation operation is linear and preserves spectral information, $\gamma_{\max}(Y_t) = \| \sum_{i=1}^t \epsilon_i B_i \|_{op}$ (Tropp 2012, Eq. 2.12). Furthermore, since each B_i is fixed and ϵ_i is 1-sub-Gaussian, (Y_t) is sub-Gaussian with variance process

$$W_t = \sum_{i=1}^t \begin{pmatrix} B_i B_i^\star & 0 \\ 0 & B_i^\star B_i \end{pmatrix},$$

which has $\|W_t\|_{op} = \max\{\| \sum_{i=1}^t B_i B_i^\star \|_{op}, \| \sum_{i=1}^t B_i^\star B_i \|_{op}\}$ (Tropp 2012, Lemma 4.3). The result now follows the connector lemma and Theorem 1(b) applied to (Y_t) and (W_t). ■

References

Bennett, G. (1962), 'Probability Inequalities for the Sum of Independent Random Variables', *Journal of the American Statistical Association* **57**(297), 33–45.

Bernstein, S. (1927), *Theory of probability*, Gastehizdat Publishing House, Moscow.

Boucheron, S., Lugosi, G. & Massart, P. (2013), *Concentration inequalities: a nonasymptotic theory of independence*, 1st edn, Oxford University Press, Oxford.

Chung, F. & Lu, L. (2006), 'Concentration inequalities and martingale inequalities: a survey', *Internet Mathematics* **3**(1), 79–127.

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

Freedman, D. A. (1975), 'On Tail Probabilities for Martingales', *The Annals of Probability* **3**(1), 100–118.

Hoeffding, W. (1963), 'Probability Inequalities for Sums of Bounded Random Variables', *Journal of the American Statistical Association* **58**(301), 13–30.

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.

McDiarmid, C. (1998), Concentration, *in* M. Habib, C. McDiarmid, J. Ramirez-Alfonsin & B. Reed, eds, 'Probabilistic Methods for Algorithmic Discrete Mathematics', Spinger, New York, pp. 195–248.

Prokhorov, Y. V. (1959), 'An Extremal Problem in Probability Theory', *Theory of Probability & Its Applications* **4**(2), 201–203.

Tropp, J. A. (2011), 'Freedman's inequality for matrix martingales', *Electronic Communications in Probability* **16**, 262–270.

Tropp, J. A. (2012), 'User-friendly tail bounds for sums of random matrices', *Foundations of Computational Mathematics* **12**(4), 389–434.