36-771 Fall 2018

Martingales 1 : Concentration inequalities

**The big reference table**

Lecturer : Aaditya Ramdas

# 1 What conditions imply sub-$\psi$?

In what follows, the matrix conditional variance is $\mathrm{Var}_t X := \mathbb{E}_t X^2 - (E_t X)^2$. We let $I_d$ denote the $d \times d$ identity matrix. For a process $(Y_t)_{t\in\mathcal{T}}$, let $[Y]_t$ denote the quadratic variation and $\langle Y \rangle_t$ the conditional quadratic variation; in discrete time, $[Y]_t := \sum_{i=1}^t \Delta Y_i^2$ and $\langle Y \rangle_t := \sum_{i=1}^t \mathbb{E}_{i-1}\Delta Y_i^2$. In the discrete time case, we have the following known results.

**Fact 1** *Let $(Y_t)_{t\in\mathcal{N}}$ be any $\mathcal{H}^d$-valued martingale.*

1. *(Scalar parametric) If $d = 1$ and $Y_t$ is a cumulative sum of i.i.d., real-valued random variables, each of which is mean zero with known cumulant generating function $\psi(\lambda)$ that is finite on $\lambda \in [0, \lambda_{\max})$, then $(Y_t)$ is sub-$\psi$ with variance process $W_t = t$.*

2. *(Bernoulli) If $-gI_d \preceq \Delta Y_t \preceq hI_d$ a.s. for all $t \in \mathcal{N}$, then $(Y_t)$ is sub-Bernoulli with variance process $W_t = tI_d$ and range parameters $g, h$ (Hoeffding 1963, Tropp 2012).*

3. *(Bennett) If $\Delta Y_t \preceq cI_d$ a.s. for all $t \in \mathcal{N}$ for some $c > 0$, then $(Y_t)$ is sub-Poisson with variance process $W_t = \langle Y \rangle_t$ and scale parameter $c$ (Bennett 1962, Hoeffding 1963, Tropp 2012).*

4. *(Bernstein) If $\mathbb{E}_{t-1}(\Delta Y_t)^k \preceq (k!/2)c^{k-2}\mathrm{Var}_{t-1}(\Delta Y_t)$ for all $t \in \mathcal{N}$ and $k = 2, 3, \ldots$, then $(Y_t)$ is sub-gamma with variance process $W_t = \langle Y \rangle_t$ and scale parameter $c$ (Bernstein 1927, Tropp 2012, Boucheron et al. 2013).*

5. *(Heavy on left) Let $T_a(y) := (y \wedge a) \vee -a$ for $a > 0$ denote the truncation of $y$. If $d = 1$ and*

$$\mathbb{E}_{t-1}T_a(\Delta Y_t) \leq 0 \quad \text{for all } a > 0, t \in \mathcal{N}, \tag{1}$$

*then $(Y_t)$ is sub-Gaussian with self-normalizing process $U_t = [Y]_t$. A random variable satisfying (1) is called* heavy on left*, and $(Y_t)$ need not be a martingale in this case (Bercu & Touati 2008, Delyon 2015, Bercu et al. 2015). When $-\Delta Y_t$ satisfies (1) we say $\Delta Y_t$ is* heavy on right.

| | Condition | $\psi$ | $U_t$ | $W_t$ |
|---|---|---|---|---|
| *Discrete time* | | | | |
| Parametric ($d = 1$) | $\Delta Y_t \overset{\text{i.i.d.}}{\sim} F$ | $\log \mathbb{E} e^{\lambda \Delta Y_1}$ | | $t$ |
| Bernoulli | $-gI_d \preceq \Delta Y_t \preceq hI_d$ | $\psi_B$ | | $tI_d$ |
| Bennett | $\Delta Y_t \preceq cI_d$ | $\psi_P$ | | $\langle Y \rangle_t$ |
| Bernstein | $\mathbb{E}_{t-1}(\Delta Y_t)^k \preceq \frac{k!}{2} c^{k-2} \mathbb{E}_{t-1} \Delta Y_t^2$ | $\psi_G$ | | $\langle Y \rangle_t$ |
| Heavy on left | $\mathbb{E}_{t-1} T_a(\Delta Y_t) \leq 0$ for all $a > 0$ | $\psi_N$ | $[Y]_t$ | |
| Hoeffding I | $-G_t I_d \preceq \Delta Y_t \preceq H_t I_d$ | $\psi_N$ | | $\sum_{i=1}^t \left( \frac{G_i + H_i}{2} \right)^2 I_d$ |
| Symmetric | $\Delta Y_t \sim -\Delta Y_t \mid \mathcal{F}_{t-1}$ | $\psi_N$ | $[Y]_t$ | |
| Bounded below | $\Delta Y_t \succeq -cI_d$ | $\psi_E$ | $[Y]_t$ | |
| Self-normalized I | $\mathbb{E}_{t-1} \Delta Y_t^2 < \infty$ | $\psi_N$ | $[Y]_t/3$ | $2 \langle Y \rangle_t / 3$ |
| Self-normalized II | $\mathbb{E}_{t-1} \Delta Y_t^2 < \infty$ | $\psi_N$ | $[Y_+]_t/2$ | $\langle Y_- \rangle_t / 2$ |
| Hoeffding II | $\Delta Y_t^2 \preceq A_t^2$ | $\psi_N$ | | $\sum_{i=1}^t A_i^2$ |
| Cubic self-normalized | $\mathbb{E}_{t-1} \lvert \Delta Y_t \rvert^3 < \infty$ | $\psi_G$ | $[Y]_t$ | $\sum_{i=1}^t \mathbb{E}_{i-1} \lvert \Delta Y_i \rvert^3$ |
| *Continuous time ($d = 1$)* | | | | |
| Lévy | $\mathbb{E} e^{\lambda Y_1} < \infty$ | $\log \mathbb{E} e^{\lambda Y_1}$ | | $t$ |
| Bennett | $\Delta Y_t \leq c$ | $\psi_P$ | | $\langle Y \rangle_t$ |
| Bernstein | $V_{m,t} \leq \frac{m!}{2} c^{m-2} W_t$ | $\psi_G$ | | $W_t$ |
| Continuous paths | $\Delta Y_t \equiv 0$ | $\psi_N$ | | $\langle Y \rangle_t$ |

Table 1: Summary of sufficient conditions for a martingale $(Y_t)$ to be sub-$\psi$ with the given self-normalizing and variance processes. See text for details of each case.

In addition, we give the following novel results for matrices by extending the corresponding scalar results. Here $[Y_+]_t := \sum_{i=1}^t \max(0, \Delta Y_i)^2$ and $\langle Y_- \rangle_t := \sum_{i=1}^t \mathbb{E}_{i-1} \min(0, \Delta Y_i)^2$, where the functions $\max(0, \cdot)$ and $\min(0, \cdot)$ extend to $^d$ by truncating the eigenvalues.

**Lemma 2** *Let $(Y_t)_{t \in \mathcal{N}}$ be any $\mathcal{H}^d$-valued martingale.*

1. *(Hoeffding I) If $-G_t I_d \preceq \Delta Y_t \preceq H_t I_d$ a.s. for all $t \in \mathcal{N}$ for some real-valued, pre-*

*dictable sequences* $(G_t)$ *and* $(H_t)$*, then* $(Y_t)$ *is sub-Gaussian with variance process* $W_t = \left[\sum_{i=1}^t (G_i + H_i)^2/4\right] I_d$.

2. *(Conditionally symmetric) If* $\Delta Y_t$ *and* $-\Delta Y_t$ *have the same distribution conditional on* $\mathcal{F}_{t-1}$ *for all* $t \in \mathcal{N}$*, then* $(Y_t)$ *is sub-Gaussian with self-normalizing process* $U_t = [Y]_t$*. In this case,* $(Y_t)$ *need not be a martingale, i.e., it need not be integrable.*

3. *(Bounded from below) If* $\Delta Y_t \succeq -cI_d$ *a.s. for all* $t \in \mathcal{N}$ *for some* $c > 0$*, then* $(Y_t)$ *is sub-exponential with self-normalizing process* $U_t = [Y]_t$ *and scale parameter* $c$*.*

4. *(General self-normalized I) If* $\mathbb{E}_{t-1}\Delta Y_t^2$ *is finite for all* $t \in \mathcal{N}$*, then* $(Y_t)$ *is sub-Gaussian with self-normalizing process* $U_t = [Y]_t/3$ *and variance process* $W_t = 2\langle Y \rangle_t/3$*.*

5. *(General self-normalized II) If* $\mathbb{E}_{t-1}\Delta Y_t^2$ *is finite for all* $t \in \mathcal{N}$*, then* $(Y_t)$ *is sub-Gaussian with self-normalizing process* $U_t = [Y_+]_t/2$ *and variance process* $W_t = \langle Y_- \rangle_t/2$*.*

6. *(Hoeffding II) If* $\Delta Y_t^2 \preceq A_t^2$ *a.s. for all* $t \in \mathcal{N}$ *for some* $\mathcal{H}^d$*-valued predictable sequence* $(A_t)$*, then* $(Y_t)$ *is sub-Gaussian with* $W_t = \sum_{i=1}^t A_i^2$*.*

7. *(Cubic self-normalized) If* $\mathbb{E}_{t-1}|\Delta Y_t|^3$ *is finite for all* $t \in \mathcal{N}$*, then* $(Y_t)$ *is sub-gamma with self-normalizing process* $U_t = [Y]_t$*, variance process* $W_t = \sum_{i=1}^t \mathbb{E}_{i-1}|\Delta Y_i|^3$*, and scale parameter* $c = 1/6$*.*

# References


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

Bercu, B., Delyon, B. & Rio, E. (2015), *Concentration Inequalities for Sums and Martingales*, Springer International Publishing, Cham.

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

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.

Delyon, B. (2015), Exponential inequalities for dependent processes, Technical report.

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

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