Self-Normalised Concentration via Sequential Probability Assignment

1 Introduction

We give a new proof of the following result (Abbasi-Yadkori et al., 2011).

Theorem 1. *Let $\mathbb{F} = (\mathscr{F}_t)_{t\geq 0}$ be a filtration. Let $(\varepsilon_t)_{t\geq 1}$ be an $\mathbb{F}$-adapted sequence of random variables such that each ε_t is conditionally 1-sub-Gaussian, i.e.,*

$$\forall \lambda \in \mathbb{R}\,, \quad \mathbb{E}[\exp(\lambda\varepsilon_t)|\mathscr{F}_{t-1}] \leq \exp(\lambda^2/2)\,.$$

Let $(X_t)_{t\geq 1}$ be an $\mathbb{F}$-predictable sequence of random vectors (in $\mathbb{R}^d$). Let $U \in \mathbb{R}^{d\times d}$ be a positive-definite matrix. For every $n \geq 1$, define

$$S_t = \sum_{t=1}^{n} X_t \epsilon_t\,, \quad V_t = \sum_{t=1}^{n} X_t X_t^\top\,.$$

For any $\delta \in (0, 1]$,

$$\mathbb{P}\left(\exists n \geq 1, \|S_n\|^2_{(V_n+U)^{-1}} > \log\det(U^{-1}V_n + \mathrm{Id}) + \log\tfrac{1}{\delta}\right) \leq \delta\,.$$

This can be proved using the method of mixtures (Abbasi-Yadkori et al., 2011) or the variational approach to concentration (Chugg & Ramdas, 2025), which is also known as PAC-Bayes. We show that this result can be proved using regret bounds for a learning problem known as sequential probability assignment.

Section 2 is an introduction to the setting of sequential probability assignment, and repeats a number of definitions and results from Clerico et al. (2025). In Section 3, we prove Theorem 1.

Notation. For vectors $x, y \in \mathbb{R}^d$, we use $\langle x, y\rangle$ to denote their Euclidean inner product. For a positive (semi-)definite matrix $U \in \mathbb{R}^{d\times d}$, we define the (semi-)norm $\|\theta\|_U := \langle \theta, U\theta\rangle$.

2 Introduction (continued)

We consider a sequential linear regression problem, which is a special case of the setting studied in Clerico et al. (2025). We observe the sequences $(X_t)_{t\geq 1}$ and $(Y_t)_{t\geq 1}$, where each $X_t \in \mathbb{R}^d$ is a vector of covariates and each $Y_t \in \mathbb{R}$ is a real-valued response. The covariates can be generated in an arbitrary sequential fashion, meaning each X_t can depend on $X_1, Y_1, \ldots, X_{t-1}, Y_{t-1}$. Each response Y_t is generated according to the linear model

$$Y_t = \langle \theta^\star, X_t\rangle + \varepsilon_t\,.$$

For each $t \geq 0$, we define $\mathscr{F}_t = \sigma(X_1, Y_1, \ldots, X_t, Y_t, X_{t+1})$ to be the σ-algebra generated by $X_1, Y_1, \ldots, X_t, Y_t, X_{t+1}$. Each noise variable ε_t is conditionally 1-sub-Gaussian, meaning

$$\forall \lambda \in \mathbb{R}\,, \;\; \mathbb{E}[\exp(\lambda\varepsilon_t)|\mathscr{F}_{t-1}] \leq \exp(\lambda^2/2)\,.$$

In Clerico et al. (2025), it is shown that one can construct confidence sequences for the parameter $\theta^\star$ via a reduction to an online learning problem known as sequential probability assignment. We are not interested confidence sequences, but we are interested in sequential probability assignment. Let us describe a special case of this problem. We define the loss function

$$\ell_t(\theta) := \tfrac{1}{2}(Y_t - \langle \theta^\star, X_t \rangle)^2 \,.$$

In addition, for any probability distribution $Q \in \Delta(\mathbb{R}^d)$ with Lebesgue density q, we define the log loss as

$$\mathscr{L}_t(Q) := -\log \int_{\mathbb{R}^d} \exp(-\ell_t(\theta)) \mathrm{d}Q(\theta) = -\log \int_{\mathbb{R}^d} \exp(-\ell_t(\theta)) q(\theta) \mathrm{d}\theta \,.$$

We consider the following game played between a learner and an environment, with the following steps repeated in each round $t = 1, \ldots, n$:

1. the environment reveals X_t to the learner,
2. the learner chooses a distribution $Q_t \in \Delta(\mathbb{R}^d)$,
3. the environment reveals Y_t to the learner, and
4. the learner incurs the log loss $\mathscr{L}_t(Q_t)$.

The performance of the learner is determined by the *regret* against a comparator strategy, which plays a fixed $\bar{\theta} \in \mathbb{R}^d$ in each round. The regret after n rounds is

$$\mathrm{regret}_n(\bar{\theta}) := \sum_{t=1}^n \mathscr{L}_t(Q_t) - \sum_{t=1}^n \ell_t(\bar{\theta}) \,.$$

Using the sub-Gaussian property of the noise variables, one can show that for any fixed $\bar{\theta}$, the difference between the total loss of $\theta^\star$ and the total loss of $\bar{\theta}$ is the logarithm of a non-negative supermartingale. As a result, Ville's inequality for non-negative supermartingales tells us that

$$\mathbb{P}\left(\exists n \in \mathbb{N}, \;\; \sum_{t=1}^n \ell_t(\theta^\star) - \sum_{t=1}^n \ell_t(\theta) > \log \tfrac{1}{\delta} \right) \leq \delta \,.$$

Theorem 2.2 of Clerico et al. (2025) states that if we add the regret of any strategy $(Q_t)_{t \geq 1}$ against $\bar{\theta}$ to the right-hand side, then this bound holds uniformly over all $\bar{\theta} \in \mathbb{R}^d$.

Theorem 2. *For any $\delta \in (0, 1)$,*

$$\mathbb{P}\left(\exists (n, \bar{\theta}) \in \mathbb{N} \times \mathbb{R}^d, \;\; \sum_{t=1}^n \ell_t(\theta^\star) - \sum_{t=1}^n \ell_t(\bar{\theta}) > \mathrm{regret}_n(\bar{\theta}) + \log \tfrac{1}{\delta} \right) \leq \delta \,.$$

Notice that if we replace the regret by any upper bound on the regret, then the statement of Theorem 2 remains true. Upper bounds on the regret are provided in Section 3 of Clerico et al. (2025). To state (one of) them, we must first introduce a few more things. Let $\rho : \mathbb{R}^d \to \mathbb{R}$ be a convex and differentiable function such that $\int_{\mathbb{R}^d} \exp(-\rho(\theta)) \mathrm{d}\theta < \infty$. Let us define the total regularised loss as

$$Z_n(\theta) := \sum_{t=1}^n \ell_t(\theta) + \rho(\theta) \,.$$

For any convex and differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, the *Bregman divergence* induced by f is

$$\mathscr{B}_f(\theta, \theta') := f(\theta) - f(\theta') - \langle \theta - \theta', \nabla f(\theta') \rangle \,.$$

Notice that this is the difference between $f(\theta)$ and the first-order Taylor polynomial of f around θ'. Let us define the *Bregman information gain* as

$$\gamma_n := -\log\left(\frac{\int \exp(-\mathscr{B}_{Z_n}(\theta, \widehat{\theta}_n))\mathrm{d}\theta}{\int \exp(-\rho(\theta))\mathrm{d}\theta}\right),$$

Proposition 3.1 in Clerico et al. (2025) states that the regret of the exponentially weighted average (EWA) forecaster (a.k.a. Vovk's aggregating algorithm) can be bounded in terms of the Bregman information gain.

Proposition 3. *There exists a strategy $(Q_t)_{t\geq 1}$ such that for all comparators $\bar{\theta}$ that satisfy $\rho(\bar{\theta}) < \infty$,*

$$\mathrm{regret}_n(\bar{\theta}) \leq \rho(\bar{\theta}) + \gamma_n\,.$$

Substituting this regret bound into Theorem 2 gives the following corollary.

Corollary 4. *For any $\delta \in (0, 1)$,*

$$\mathbb{P}\left(\exists (n, \bar{\theta}) \in \mathbb{N} \times \mathbb{R}^d,\ \ \sum_{t=1}^{n} \ell_t(\theta^\star) - \sum_{t=1}^{n} \ell_t(\bar{\theta}) > \rho(\bar{\theta}) + \gamma_n + \log\tfrac{1}{\delta}\right) \leq \delta\,.$$

3 Self-normalised concentration via sequential probability assignment

Using the results from the previous section, we can now prove Theorem 1. We choose $\rho(\theta) = \frac{1}{2}\|\theta - \theta^\star\|_U^2$. With this choice, one can verify that the minimiser of Z_n is

$$\widehat{\theta}_n := \underset{\theta\in\mathbb{R}^d}{\operatorname{argmin}}\{Z_n(\theta)\} = (V_n + U)^{-1}\left(\sum_{t=1}^{n} X_t Y_t + U\theta^\star\right).$$

We begin by finding an expression for the Bregman divergence induced by Z_n.

Lemma 5. *If $\rho(\theta) = \frac{1}{2}\|\theta - \theta^\star\|_U^2$, then*

$$\mathscr{B}_{Z_n}(\theta, \theta') = \frac{1}{2}\|\theta - \theta'\|_{V_n+U}^2\,.$$

Proof. Since Z_n is quadratic, it is equal to its second-order Taylor polynomial (around any point). The Bregman divergence $\mathscr{B}_{Z_n}(\theta, \theta')$ is equal to the difference between $Z_n(\theta)$ and the first-order Taylor polynomial of Z_n around θ'. Thus $\mathscr{B}_{Z_n}(\theta, \theta')$ is equal to the quadratic term in the Taylor series of Z_n around θ'. Therefore,

$$\mathscr{B}_{Z_n}(\theta, \theta') = \frac{1}{2}\|\theta - \theta'\|_{\nabla Z_n(\theta')}^2 = \frac{1}{2}\|\theta - \theta'\|_{V_n+U}^2\,.$$

This concludes the proof. □

With this expression for $\mathscr{B}_{Z_n}(\theta, \theta')$, we can evaluate the Bregman information gain.

Lemma 6. *If $\rho(\theta) = \frac{1}{2}\|\theta - \theta^\star\|_U^2$, then*

$$\gamma_n = \frac{1}{2}\log\det(U^{-1}V_n + \mathrm{Id})\,.$$

Proof. We will use that fact that for any $\mu \in \mathbb{R}^d$ and any positive-definite matrix $\Sigma \in \mathbb{R}^{d\times d}$,

$$\int_{\mathbb{R}^d} \exp(-\tfrac{1}{2}\|\theta - \mu\|^2_{\Sigma^{-1}})\mathrm{d}\theta = \sqrt{(2\pi)^d \det(\Sigma)}\,.$$

Using the expression for $\mathscr{B}_{Z_n}(\theta, \theta')$ in Lemma 5, we obtain

$$\gamma_n = -\log\left(\frac{\int_{\mathbb{R}^d} \exp(-\frac{1}{2}\|\theta - \widehat{\theta}_n\|^2_{V_n+U})\mathrm{d}\theta}{\int_{\mathbb{R}^d} \exp(-\frac{1}{2}\|\theta - \theta^\star\|^2_{U})\mathrm{d}\theta}\right) = \frac{1}{2}\log(\det(U^{-1})\det(V_n + U)) = \frac{1}{2}\log\det(U^{-1}V_n + \mathrm{Id})\,.$$

This concludes the proof. ☐

We're now ready to prove Theorem 1.

Proof of Theorem 1. First, we notice that

$$\widehat{\theta}_n - \theta^\star = (V_n + U)^{-1}\left(\sum_{t=1}^n X_t Y_t + U\theta^\star - (V_n + U)\theta^\star\right) = (V_n + U)^{-1}\sum_{t=1}^n X_t \varepsilon_t\,.$$

This allows us to write

$$\tfrac{1}{2}\|S_n\|^2_{(V_n+U)^{-1}} = \tfrac{1}{2}\|\theta^\star - \widehat{\theta}_n\|^2_{V_n+U}\,.$$

Since $\widehat{\theta}_n$ minimises Z_n,

$$\tfrac{1}{2}\|\theta^\star - \widehat{\theta}_n\|^2_{V_n+U} = \mathscr{B}_{Z_n}(\theta^\star, \widehat{\theta}_n) = \sum_{t=1}^n \ell_t(\theta^\star) + \rho(\theta^\star) - \sum_{t=1}^n \ell_t(\widehat{\theta}_n) - \rho(\widehat{\theta}_n)\,.$$

Finally, for any $\delta \in (0, 1)$, the combination of Corollary 4 and Lemma 6 tells us that, with probability at least δ, for all $n \geq 1$,

$$\sum_{t=1}^n \ell_t(\theta^\star) + \rho(\theta^\star) - \sum_{t=1}^n \ell_t(\widehat{\theta}_n) - \rho(\widehat{\theta}_n) \leq \gamma_n + \log\tfrac{1}{\delta} = \frac{1}{2}\log\det(U^{-1}V_n + \mathrm{Id}) + \log\tfrac{1}{\delta}\,.$$

Putting everything together proves Theorem 1. ☐

4 Conclusion

We have seen a new proof of the self-normalised concentration inequality in Abbasi-Yadkori et al. (2011). We conclude by exploring some connections between this proof, the method of mixtures, and the proof in Abbasi-Yadkori et al. (2011). Using a standard telescoping sum argument (cf. Lemma B.1 in Clerico et al. (2025)), the regret of the EWA forecaster can be re-written as

$$\mathrm{regret}_n(\bar{\theta}) = -\log\int \exp\left(\sum_{t=1}^n \ell_t(\bar{\theta}) - \ell_t(\theta)\right)\mathrm{d}Q_1(\theta)\,.$$

Here, Q_1 is the distribution with density $q_1(\theta) \propto \exp(-\rho(\theta))$. Combining this with Theorem 2, one has

$$\mathbb{P}\left(\exists n \in \mathbb{N},\ \ \log\int \exp\left(\sum_{t=1}^n \ell_t(\theta^\star) - \sum_{t=1}^n \ell_t(\theta)\right)\mathrm{d}Q_1(\theta) > \log\tfrac{1}{\delta}\right) \leq \delta\,.$$

This is what we would get by applying the method of mixtures (with the mixture distribution Q_1) to the collection of non-negative supermartingales $(M_n(\theta))_{\theta\in\mathbb{R}^d}$, where

$$M_n(\theta) := \exp\left(\sum_{t=1}^n \ell_t(\theta^\star) - \sum_{t=1}^n \ell_t(\theta)\right).$$

Since we already know that Theorem 1 can be proved using the method of mixtures (Abbasi-Yadkori et al., 2011), it is not so surprising that the approach taken here is closely related to the method of mixtures. Note however, that the mixing distribution and the collection of supermartingales that appear here are not the same as those used in the proof of Theorem 1 in Abbasi-Yadkori et al. (2011). There, the mixture distribution has density $h(x) \propto \exp(-\frac{1}{2}\|x\|_U^2)$ and the collection of supermartingales is $(M_n(x))_{x\in\mathbb{R}^d}$, where

$$M_n(x) := \exp\left(\langle x, S_n\rangle - \tfrac{1}{2}\|x\|_{V_t}^2\right).$$

Nevertheless, the resulting mixture supermartingale $\int M_n(x)h(x)\mathrm{d}x$ is the same. In particular,

$$\begin{aligned}
\int \exp\left(\sum_{t=1}^n \ell_t(\theta^\star) - \sum_{t=1}^n \ell_t(\theta)\right)\mathrm{d}Q_1(\theta) &= \int \exp\left(\langle \theta-\theta^\star, S_n\rangle - \tfrac{1}{2}\|\theta-\theta^\star\|_{V_n}^2\right)\frac{\exp(-\frac{1}{2}\|\theta-\theta^\star\|_U^2)}{\int \exp(-\frac{1}{2}\|\theta-\theta^\star\|_U^2)\mathrm{d}\theta}\mathrm{d}\theta \\
&= \int \exp\left(\langle x, S_n\rangle - \tfrac{1}{2}\|x\|_{V_n}^2\right)\frac{\exp(-\frac{1}{2}\|x\|_U^2)}{\int \exp(-\frac{1}{2}\|x\|_U^2)\mathrm{d}x}\mathrm{d}x \\
&= \int M_n(x)h(x)\mathrm{d}x\,.
\end{aligned}$$

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Chugg, B. and Ramdas, A. A variational approach to dimension-free self-normalized concentration. *arXiv preprint arXiv:2508.06483*, 2025.

Clerico, E., Flynn, H., Kotłowski, W., and Neu, G. Confidence sequences for generalized linear models via regret analysis. *arXiv preprint arXiv:2504.16555*, 2025.