Concentration inequalities for non-causal random fields [1]

Rémy Garnier[1], Raphaël Langhendries[2]

1 - AGM CY Cergy Paris Université, & Cdisccount

2 - SAMM - Université Panthéon Sorbonne & Safran Aircraft Engines

Abstract

Concentration inequalities are widely used for analyzing machine learning algorithms. However, the current concentration inequalities cannot be applied to some non-causal processes which appear for instance in Natural Language Processing (NLP). This is mainly due to the non-causal nature of such involved data, in the sense that each data point depends on other neighboring data points. In this paper, we establish a framework for modeling non-causal random fields and prove a Hoeffding-type concentration inequality. The proof of this result is based on a local approximation of the non-causal random field by a function of a finite number of i.i.d. random variables.

Random field

In this paper, we prove concentration inequalities on a κ-dimensional non-causal random field $(X_t)_{t\in\mathbb{Z}^\kappa}$ that is a stationary solution of the following equation.

$$\forall t \in \mathbb{Z}^\kappa, X_t = F((X_{t+s})_{s\in\mathcal{B}}, \varepsilon_t).$$

Such random fields have been introduced by [2]. We emphasize that for $s \in \mathcal{B}$, X_{t+s} depends on X_t, but X_t also depends on X_{t+s}. Therefore, it is no longer possible to describe $(X_t)_{t\in\mathbb{Z}^\kappa}$ as a result of a martingale process (this is why, we call $(X_t)_{t\in\mathbb{Z}^\kappa}$ a non-causal random field).

Notations

- $S_\mathcal{I} = \sum_{s\in\mathcal{I}} \Phi((X_{s+t})_{t\in\bar{\mathcal{B}}})$ for a given subset of index $\mathcal{I}$. $S_\mathcal{I}$ is the statistic for which we aim to prove concentration inequalities and Φ is a Lipschitz function.
- Let $n_B = \mathrm{Card}(\mathcal{B}) + 1$ and $n_{\bar{B}} = \mathrm{Card}(\bar{\mathcal{B}})$.

Contraction hypotheses

A common hypothesis used to prove concentration inequalities is the contraction hypothesis. This hypothesis can adapted to our non-causal framework.

Standard contraction hypothesis. It exists $(\lambda_t)_{t\in\mathcal{B}}$, such that $\rho := \sum_{t\in\mathcal{B}} \lambda_t < 1$, and for any $\mathcal{X}$-valued tuples $\mathcal{Y} = (y_t)_{t\in\mathcal{B}}$ and $\mathcal{Y}' = (y'_t)_{t\in\mathcal{B}}$ indexed by $\mathcal{B}$ and for all $\varepsilon \in E$.

$$\|F(\mathcal{Y}, \varepsilon) - F(\mathcal{Y}', \varepsilon)\| \le \sum_{s\in\mathcal{B}} \lambda_s \|y_s - y'_s\|.$$

In this paper, we introduced weaker versions of this hypothesis. Respectively, we used

- a *strong* contraction hypothesis that involves $\|F((Y_{t+s})_{s\in\mathcal{B}}, \varepsilon_t) - F((Y'_{t+s})_{s\in\mathcal{B}}, \varepsilon_t)\|_\infty$ (the uniform norm).
- a *weak* contraction hypothesis that involves $\|F((Y_{t+s})_{s\in\mathcal{B}}, \varepsilon_t) - F((Y'_{t+s})_{s\in\mathcal{B}}, \varepsilon_t)\|_m$ (the m norm).

Our results are based on these two hypotheses, the first one is more restrictive and therefore ensures stronger results.

Application to statistical learning

In the full paper, we also provide an application to learning theory (with a focus on model selection). We adapted the approach developed under the i.i.d setting ([3, 4]) to prove an oracle bound for the model selection problem.

References

[1] Rémy Garnier and Raphaël Langhendries.
Concentration inequalities for non-causal random fields.
Electronic Journal of Statistics, 16(1), January 2022.

[2] Paul Doukhan and Lionel Truquet.
A fixed point approach to model random fields.
Alea, 3:111–132, 2007.

[3] Peter L. Bartlett, Stephane Boucheron, and Gábor Lugosi.
Model Selection and Error Estimation.
SSRN Electronic Journal, 2000.

[4] G. Lugosi.
Pattern Classification and Learning Theory, pages 1–56.
Springer Vienna, Vienna, 2002.

Main results

Theorem (Concentration inequalities)

- *If the strong hypothesis is verified, there is a constant A such that, for all $\varepsilon > 2n_{\bar{B}}\mathbb{V}_\infty$,*

$$\mathbb{P}\left(|S_\mathcal{I} - \mathbb{E}[S_\mathcal{I}]| \ge \varepsilon\right) \le 2\exp\left(\frac{-2\left(\varepsilon - 2n_{\bar{B}}\mathbb{V}_\infty\right)^2}{\left(n_{\bar{B}}\mathbb{V}_\infty\right)^2\left(1 + An_{\bar{B}}n_B^3\kappa!^2\lceil\ln(n)\rceil^\kappa n\right)}\right).$$

With A such that $\Upsilon(d)^2 d^\kappa \le A(\kappa!)^2\lceil\ln(n)\rceil^\kappa$, where Υ is a function that can be bounded independently of $n_{\bar{B}}$, m or n.

- *If the weak hypothesis is verified, there are constants A,B,C,D,E,F,H such that, for all $\varepsilon \ge 2F\left(n_{\bar{B}}n_B\right)^3\kappa!\lceil\ln(n)\rceil^{2\kappa} n^{\frac{2}{m}}$.*

$$\mathbb{P}\left(|S_\mathcal{I} - \mathbb{E}[S_\mathcal{I}]| \ge \varepsilon\right) \le 2\exp\left(\frac{-2\left(\frac{\varepsilon}{2} - F\left(n_{\bar{B}}n_B\right)^3\kappa!\lceil\ln(n)\rceil^{2\kappa} n^{\frac{2}{m}}\right)^2}{\left(Hn_{\bar{B}}n^{\frac{2}{m}}\right)^2\left(1 + En_{\bar{B}}n_B^3(\kappa!)^2\lceil\ln(n)\rceil^\kappa n\right)}\right.$$
$$\left. + \frac{\rho^m}{n}\left(2n_Bn_{\bar{B}}C\lceil\ln(n)\rceil^\kappa + \left(\frac{D}{n_B^3 n_{\bar{B}}^2\Upsilon(d)\ln(n)^{2\kappa}}\right)^m\right)\right).$$

Constants involved are explicits and do not increase with n.

Moment inequalities

Corollary (Moments inequalities: asymptotic equivalents)

- *If the strong hypothesis is verified.*

$$\mathbb{E}[|S_\mathcal{I} - \mathbb{E}[S_\mathcal{I}]|] \le G_1(\kappa, \rho, \mathbb{V}_\infty, n_{\bar{B}}, n_B, n) \underset{n\to\infty}{\sim} n_Bn_{\bar{B}}\mathbb{V}_\infty \upsilon\sqrt{\frac{\pi}{2}n_Bn_{\bar{B}}(K_\rho(\infty))^\kappa n\ln(n)^\kappa},$$
$$\underset{n\to\infty}{=} \mathcal{O}\left(\sqrt{n\ln(n)^\kappa}\right),$$

- *If the weak hypothesis is verified.*

$$\mathbb{E}[|S_\mathcal{I} - \mathbb{E}[S_\mathcal{I}]|] \le G_2(\kappa, \rho, \mathbb{V}_m, n_{\bar{B}}, n_B, n) \underset{n\to\infty}{\sim} \frac{n_{\bar{B}}n_B\mathbb{V}_m\upsilon n^{\frac{2}{m}}}{\rho}\sqrt{2\pi n_{\bar{B}}n_BK_\rho(m)^\kappa\ln(n)^\kappa n}$$
$$\underset{n\to\infty}{=} \mathcal{O}\left(n^{\frac{2}{m}}\sqrt{n\ln(n)^\kappa}\right),$$

with $K_\rho(\infty) = \frac{1}{\ln(\rho^{-1})}$, $K_\rho(m) = \frac{1-\frac{1}{m}}{\ln(\rho^{-1})}$ and ν a constants depending on κ and ρ.