Concentration: Negative association

Master 2 Mathematics and Computer Science

Guillaume Aubrun, Aurélien Garivier, Rémi Gribonval

2020-2021

UNIVERSITÉ DE LYON
ENS DE LYON

Table of contents

References:

Negative Association - Definition, Properties, and Applications, *by David Wajc* https://www.cs.cmu.edu/~dwajc/notes/Negative%20Association.pdf

Balls and Bins:A Study in Negative Dependence, *by Balls and Bins:A Study in Negative Dependence*, https://www.brics.dk/RS/96/25/BRICS-RS-96-25.pdf

Definition

Intuitively: $X_1, \ldots, X_n$ are negatively associated when, if a subset I a variables is "high", a disjoint subset J has to be "low".

Definition

A set of real-valued random variables $X_1, X_2, ..., X_n$ is said to be negatively associated (NA) if for any two disjoint index sets $I, J \subset [n]$ and two functions f, g both monotone increasing or both monotone decreasing, it holds

$$\mathbb{E}\Big[f(X_i : i \in I)\, g\big(X_j : j \in J\big)\Big] \leq \mathbb{E}\big[f(X_i : i \in I)\big]\, \mathbb{E}\Big[g\big(X_j : j \in J\big)\Big]$$

NB: f is *monotone increasing* if $\forall i \in I, x_i \leq x'_i$ implies $f(x) \leq f(x')$.

First properties

Let $X_1, X_2, ..., X_n$ be NA.

- For all $i \neq j$, $\mathbb{E}[X_i X_j] \leq \mathbb{E}[X_i]\,\mathbb{E}[X_j]$ i.e. $\mathrm{Cov}(X_i, X_j) \leq 0$.
- For any disjoints subsets $I, J \subset [n]$ and all $x_1, \ldots, x_n$,

$$\mathbb{P}\big(X_i \geq x_i : i \in I \cup J\big) \leq \mathbb{P}\big(X_i \geq x_i : i \in I\big)\,\mathbb{P}\big(X_j \geq x_j : j \in J\big) \quad \text{and}$$
$$\mathbb{P}\big(X_i \leq x_i : i \in I \cup J\big) \leq \mathbb{P}\big(X_i \leq x_i : i \in I\big)\,\mathbb{P}\big(X_j \leq x_j : j \in J\big)$$

- For all $x_1, \ldots, x_n$,

$$\mathbb{P}\left(\bigcap_i \{X_i \geq x_i\right) \leq \prod_i \mathbb{P}\big(X_i \geq x_i\big) \quad \text{and}$$
$$\mathbb{P}\left(\bigcap_i \{X_i \leq x_i\}\right) \leq \prod_i \mathbb{P}\big(X_i \leq x_i\big)$$

- For all monotone increasing functions $f_1, \ldots, f_k$ depending on disjoint subsets of the $(X_i)_i$,

$$\mathbb{E}\Big[\prod_j f_j(X)\Big] \leq \prod_j \mathbb{E}\big[f_j(X)\big]$$

Consequence: NA concentrates better than independent

Chernoff-Hoeffding bound

Let $X_1, \ldots, X_n$ be NA random variables with $X_i \in [a_i, b_i]$ a.s. Then $S = X_1 + \cdots + X_n$ satifies Hoeffding's tail bound: for all $t \geq 0$,

$$\mathbb{P}\Big[\big|S - E[S]\big| \geq t\Big] \leq 2\exp\left(-\frac{2t^2}{\sum_i (b_i - a_i)^2}\right)$$

Examples of NA variables

- Independent variables...
- **0-1 principle** If $X_1, \dots X_n$ are Bernoulli variables and $\sum_i X_i \leq 1$ a.s., then they are NA.
- **Permutation distributions** If $x_1 \leq \cdots \leq x_n$ and if $X_1, \dots, X_n$ are random variables such that $\left\{X_1, \dots, X_n\right\} = \left\{x_1, \dots, x_n\right\}$ a.s., then they are NA.
- **Sampling without replacement** If $X_1, \dots, X_n$ are sample without replacement from $\{x_1, \dots, x_N\}$ (with $N \geq n$), then they are NA.

Closure properties

Union

If the $\{X_i : i \in I\}$ are NA, if $\{Y_j : j \in J\}$ are NA, and if the $\{X_i\}$ are independent from the $\{Y_j\}$, then the $\{X_i, Y_j : i \in I, j \in J\}$ are NA.

Concordant monotone

If the $\{X_i : i \in I\}$ are NA, if $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ are all monotonically increasing and depend on different subsets of $[n]$, then $\left\{f_j(X) : 1 \leq j \leq k\right\}$ are NA.
The same holds if $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ are all monotonically decreasing.

Bins and balls

The standard bins and balls process consists of m balls and n bins. Each ball b is independently placed in bin i with probability $p_{b,i}$. The *occupancy number* B_i be the number of balls placed in bin i. In particular $\sum_i B_i = m$.

Prop: The B_i are NA.

Consequence: Concentration of the number $N = \sum_i \mathbb{1}\{B_i = 0\}$ of empty bins.

Proof: just apply the previous ideas to the $\left\{Z_{b,i} = \mathbb{1}\{\text{ball b fell into box i}\} : 1 \leq b \leq m, 1 \leq i \leq n\right\}$.