independent blocks technique

1. Stationarity and Mixing Sequences

Definition 1 (Stationarity)

Stationarity

A sequence of random variables

$$\mathbf{Z} = \{Z_t\}_{t=-\infty}^{\infty}$$

is said to be stationary if for any integer t and non-negative integers m and k, the random vectors

$$(Z_t, Z_{t+1}, \ldots, Z_{t+m})$$

and

$$(Z_{t+k}, Z_{t+k+1}, \ldots, Z_{t+m+k})$$

have the same distribution. In other words, the time index t does not affect the distribution of Z_t in a stationary sequence.

Note, however, that stationarity does not imply independence; for example, in general

$$\Pr[Z_j \mid Z_i] \quad \text{may not equal} \quad \Pr[Z_j],$$

or, equivalently, for $i < j < k$,

$$\Pr[Z_j \mid Z_i] \quad \text{may not equal} \quad \Pr[Z_k \mid Z_i].$$

Definition 2 (β-Mixing Coefficients)

β -Mixing Coefficients

Let $\mathbf{Z} = \{Z_t\}_{t=-\infty}^{\infty}$ be a stationary sequence and, for any integers $i \leq j$, let σ_i^j denote the σ-algebra generated by the random variables $Z_i, Z_{i+1}, \ldots, Z_j$. For any integer $m > 0$, the β-mixing coefficient is defined by

$$\beta(m) = \sup_{n \in \mathbb{Z}} \mathbb{E}_{A_1 \in \sigma_{-\infty}^n} \left[\sup_{A_2 \in \sigma_{n+m}^{+\infty}} |\Pr[A_2 \mid A_1] - \Pr[A_2]| \right].$$

The sequence $\mathbf{Z}$ is said to be β-mixing if $\beta(m) \to 0$ as $m \to \infty$; it is said to be algebraically β-mixing if there exist constants $\beta_0 > 0$ and $r > 0$ such that

$$\beta(m) \le \frac{\beta_0}{m^r} \quad \text{for all } m,$$

and exponentially β -mixing if there exist constants $\beta_0 > 0$, $\beta_1 > 0$, and $r > 0$ such that

$$\beta(m) \le \beta_0 \exp\left(-\beta_1 m^r\right) \quad \text{for all } m.$$

Both $\beta(m)$ and the corresponding $\varphi(m)$ (not detailed here) measure the dependence between events occurring more than m units of time apart. Although β -mixing is a weaker condition than φ -mixing, many previous studies on learning with dependent observations (e.g., [2,3,4,5]) adopt β -mixing as the standard assumption.

In the scenario of time series we studied (specifically, the assumptions about data in "l1-regularization"), when explaining the key assumptions of Satisfying DESIGN (3) and Satisfying WEIGHT (appendix A), the triplex inequality is used. The bound of the dependent term in the triplex inequality is assumed to satisfy the strong mixing condition (i.e., α-mixing) in the article. Here, β-mixing, as a weaker condition, clearly meets the scenario's requirements, and the independent blocks technique is also applicable.

A simple but useful observation is given by the following proposition.

Proposition 3.

Any i.i.d. sequence can be viewed as a special stationary β-mixing sequence with coefficients

$$\beta(m) = 0 \quad \text{for } m > 0.$$

For further technical development, we introduce a function $\tau(m)$ as follows.

Lemma

Lemma 4

If $\beta(m) \to 0$ as $m \to \infty$, then there exists a function $\tau(m) \le m$ such that

$$\tau(m) \to \infty \quad \text{and} \quad \tau(m)\, \beta\Big(\lfloor m/\tau(m) \rfloor\Big) \to 0 \quad \text{as } m \to \infty.$$

Proof:

If $\beta(m) = 0$ for all m then one may choose $\tau(m) = m$ and the result holds trivially. Now, assume that $\beta(m) \neq 0$ for $m > 0$ and that $\beta(m) \to 0$ as $m \to \infty$. Then, there exists a sequence $\{a_i\}$ such that $a_i \to \infty$ and

$$a_i\, \beta(i) \to 0 \quad \text{as } i \to \infty.$$

A possible choice is

$$a_i = \lfloor 1/\sqrt{\beta(i)} \rfloor.$$

For every $m \geq 1$, there exists an index $k \geq 1$ such that

$$(k-1)a_{k-1} \leq m \leq k\,a_k.$$

Set $\tau(m) = a_k$; then the lemma follows.

For example, in the case of an algebraically β-mixing sequence with coefficients $\beta(m) \leq \beta_0/m^r$, one may select

$$\tau(m) = m^{r_0} \quad \text{with } 0 < r_0 < \frac{r}{1+r};$$

and for an exponentially β-mixing sequence with

$$\beta(m) = \beta_0 \exp\left(-\beta_1 m^r\right),$$

one may choose

$$\tau(m) = m^{r_0} \quad \text{with } 0 < r_0 < 1.$$

It is worth noticing that the choice of $\tau(m)$ is not unique, and different choices lead to different convergence rates for $\tau(m)\beta(\lfloor m/\tau(m)\rfloor) \to 0$. Throughout this note, we assume, without loss of generality, that the sequence $\tau(m)\beta(\lfloor m/\tau(m)\rfloor)$ is non-increasing.

2. Independent Blocks Technique

When dealing with dependent observations from a stationary β -mixing sequence, one powerful method is to transfer the original problem (which is based on dependent points) to one based on a sequence of independent blocks. This technique permits the application of standard i.i.d. results—such as concentration inequalities and other important theorems and tools —that originally required independence.

A General Lemma for Dependent versus Independent Blocks

Let $\mu \geq 1$ and suppose that h is a measurable function defined on a product probability space

$$\left(\prod_{j=1}^{\mu} \Omega_j, \prod_{j=1}^{\mu} \sigma_{r_j}^{s_j}\right)$$

and that $|h| \leq M$. Here, the σ-algebras $\sigma_{r_j}^{s_j}$ are generated by the random variables in the j th block, with

$$r_j \leq s_j \leq r_{j+1} \quad \text{for all } j.$$

Let Q be a probability measure on the product space with marginal measures Q_j on $\left(\Omega_j, \sigma_{r_j}^{s_j}\right)$, and define

$$P = \prod_{j=1}^{\mu} Q_j.$$

If we set

$$k_j = r_{j+1} - s_j \quad \text{and} \quad \beta(Q) = \sup_{1 \le j \le \mu - 1} \beta(k_j),$$

then the following holds (see Yu [1, Corollary 2.7]):

Lemma 5 (Yu, Corollary 2.7) (original)

$$|\mathbb{E}_Q[h] - \mathbb{E}_P[h]| \le (\mu - 1) M \, \beta(Q).$$

This lemma quantifies the difference between the expectation of h when the blocks are dependent (with joint distribution Q) versus when the blocks are independent (with product distribution P). In particular, if the function $\beta(\cdot)$ is monotonically decreasing, then

$$\beta(Q) = \beta(k^*) \quad \text{with} \quad k^* = \min_{1 \le j \le \mu - 1} k_j,$$

where k^* is the smallest gap between consecutive blocks.

Construction of Independent Blocks

Consider a stationary β-mixing sequence

$$S = (z_1, z_2, \ldots, z_m)$$

with

$$m = 2a\mu,$$

where a and μ are positive integers.

Following previous non-i.i.d. analyses (e.g., [2,3,4,5]), we transfer the problem based on the original dependent points to one based on a sequence of independent blocks by splitting S into two subsequences S_0 and S_1. Each of these subsequences consists of μ blocks of a consecutive points. They are defined as follows:

$$\begin{aligned} S_0 &= (Z_1, Z_2, \ldots, Z_\mu), && \text{where } Z_i = \left(z_{(2i-1)+1}, \ldots, z_{(2i-1)+a}\right), \\ S_1 &= \left(Z_1^{(1)}, Z_2^{(1)}, \ldots, Z_\mu^{(1)}\right), && \text{where } Z_i^{(1)} = \left(z_{2i+1}, \ldots, z_{2i+a}\right). \end{aligned}$$

Instead of working with the original sequence of (say) odd blocks S_0, we construct a new sequence $\widetilde{S}_0$ of independent blocks of equal size a. That is,

$$\widetilde{S}_0 = \left(\widetilde{Z}_1, \widetilde{Z}_2, \ldots, \widetilde{Z}_\mu\right),$$

where the blocks $\widetilde{Z}_k$ are mutually independent and, for each k, the points within $\widetilde{Z}_k$ follow the same distribution as those in Z_k.

The key result (see Yu [1, Corollary 2.7]) is that, for a sufficiently large spacing a between blocks and under a sufficiently fast decay of the mixing coefficients, the expectation of any bounded measurable function h defined on the blocks is essentially unchanged when computed with respect to S_0 or $\widetilde{S}_0$. More precisely, we have:

Corollary 1 (a changed version)

Let h be a measurable function bounded by $M \geq 0$ defined over the blocks Z_k. Then,

$$\left|\mathbb{E}_{S_0}[h] - \mathbb{E}_{\widetilde{S}_0}[h]\right| \leq (\mu - 1)M\,\beta(a),$$

where $\mathbb{E}_{S_0}$ denotes the expectation with respect to the dependent blocks S_0, and $\mathbb{E}_{\widetilde{S}_0}$ the expectation with respect to the independent blocks $\widetilde{S}_0$.

We denote by $\widetilde{D}$ the distribution corresponding to the independent blocks $\widetilde{Z}_k$.

Extensions to Block-Based Hypotheses

To facilitate working with block sequences, it is useful to extend hypotheses defined on single data points to those defined on blocks. For any hypothesis $h \in H$, define its block-extension $h_a : Z^a \to \mathbb{R}$ by

$$h_a(B) = \frac{1}{a}\sum_{i=1}^{a} h(z_i) \quad \text{for any block } B = (z_1, \ldots, z_a) \in Z^a.$$

Let H_a denote the set of all block-based hypotheses h_a generated from $h \in H$.

It is also useful to define a subsequence S_μ, which consists of μ singleton points separated by a gap of $2a - 1$ points. Such a subsequence can be constructed from either S_0 or S_1 by selecting only the j th point from each block, for any fixed $j \in \{1, \ldots, a\}$.

In summary, the independent blocks technique allows us to transform a problem involving a dependent, stationary β -mixing sequence (that's what we are facing as a important difficulty in time series, or explicitly, time-dependent) into one involving independent blocks.

This transformation—quantified by Lemma 5 and Corollary 1—permits the application of i.i.d. methods to data that are originally non-i.i.d. and dependent.

Reference:
[0] S. N. Bernstein. Sur l'extension du th´eor eme limite du calcul des probabilit´es aux sommes de quantit´es d´ependantes. Math. Ann., 97:1–59, 1927.
[1] B. Yu. Rates of convergence for empirical processes of stationary mixing sequences. Annals Probability, 22(1):94–116, 1994.
[2] A. Lozano, S. Kulkarni, and R. Schapire. Convergence and consistency of regularized boosting algorithms with stationary β-mixing observations. Advances in Neural Information Processing Systems, 18, 2006.
[3] R. Meir. Nonparametric time series prediction through adaptive model selection. Machine Learning, 39(1):5–34, 2000.

[4] M. Mohri and A. Rostamizadeh. Stability bounds for non-iid processes. Advances in Neural Information Processing Systems, 2007.
[5] B. Yu. Rates of convergence for empirical processes of stationary mixing sequences. Annals Probability, 22(1):94–116, 1994.