Estimating beta mixing coefficients

Daniel J. McDonald
Department of Statistics
Indiana University, Bloomington
mypage.iu.edu/~dajmcdon

1 April 2017

Motivation

- Theoretical statistics assumes independent observations.
- The reason is often mis-understood.
- It is **not** to guarantee that large samples are representative of the underlying population.
- That needs only the ergodic theorem to work for dependent sources equally well.
- Independence lets theorists discuss the **rate** at which growing samples approximate the truth.

CONVERGENCE RATES

- Independence $\Rightarrow$ every observation is "unpredictable" from every other.
- Each datum provides a new piece of information about the source.
- Thus, information is proportional to the number of observations.
- Under dependence, later events are more or less **predictable** from earlier ones.
- Assuming ergodicity alone, the convergence of samples on the source can be arbitrarily slow.
- This leads to an $n = 1$ situation no matter how many observations one has.
- Need to assume something

Quantifying dependence

- For time series, the natural replacement for independence is to require asymptotic independence of events.
- Known as **mixing**.
- Quantifies the decay in dependence as the future moves farther from the past.
- Many types with matching dependence coefficients (e.g. Doukhan, 1994; Bradley, 2005; Dedecker et al., 2007).
- We focus on β-mixing or absolute regularity.

Notation

- $\mathbf{X} = \{X_t\}_{t=-\infty}^{\infty}$ is a sequence of random variables.
- $X_t : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^q$.
- $\mathbf{X}_{i:j} \equiv \{X_t\}_{t=i}^{j}$.
- Similar notation for the sigma fields generated by these blocks.
- The joint distribution of $\mathbf{X}_{i:j}$ is $\mathbb{P}_{i:j}$.
- Products of marginal distributions are, e.g., $\mathbb{P}_{i:j} \otimes \mathbb{P}_{k:l}$.

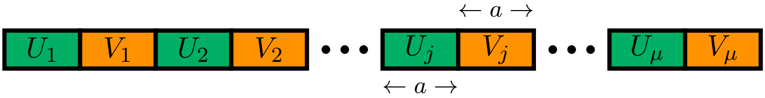
For each $a \in \mathbb{N}$, the coefficient of absolute regularity, or β-dependence coefficient, is

$$\beta(a) := \|\mathbb{P}_{-\infty:0} \otimes \mathbb{P}_{a:\infty} - \mathbb{P}_{-\infty:0,a:\infty}\|_{TV}$$

X is said to be β-mixing if $\beta(a) \to 0$ as $a \to \infty$.

WHY IS THIS USEFUL?

- Observe $\mathbf{X}_{1:n}$. Divide this into μ blocks, each containing a points.



Theorem: (Yu, 1994) Let Z be an event with respect to the $\mathbf{U}$ blocks. Then

$$|\mathbb{P}_\infty(Z) - \widetilde{\mathbb{P}}(Z)| \leq \beta(a)(\mu - 1).$$

- $\mathbb{P}_\infty$ in the joint of $\mathbf{X}$
- $\widetilde{\mathbb{P}}$ is the μ-fold product of $\mathbb{P}_{1:a}$.

Implication: n points from a mixing sequence is like μ independent blocks.

Further steps

- Reuse concentration of measure inequalities with only minor changes.
- Hoeffding's inequality for n i.i.d. RV's $X_i : \Omega \to [0,1]$

$$\mathbb{P}_\infty\left(\left|\overline{\mathbf{X}}_{1:n} - \mathbb{E}[X_1]\right| > t\right) \leq 2e^{-nt^2}.$$

- Hoeffding's modified for dependent $\mathbf{X}_{1:n}$

$$\mathbb{P}_\infty\left(\left|\overline{\mathbf{U}} - \mathbb{E}[\overline{U}_1]\right| > t\right) \leq 2e^{-\mu t^2} + \beta(a)(\mu - 1).$$

Problem

- For some $\mathbf{X}$, we know $\beta(a)$ (usually up to unknown constants).
- If you give me data, I can't tell you if the data is mixing, let alone tell you $\beta(a)$.
- We solve (part of) this problem.
- Basic tools: use the many equivalent definitions of mixing (see Bradley, 2005; Dedecker et al., 2007).

Idea

- $\beta(a)$ depends on an infinite process.
- Can't estimate it with only finite data.
- Define

$$\beta^{d+1}(a) := \left\| \mathbb{P}_{[d+1]} \otimes \mathbb{P}_{[d+1]} - \mathbb{P}_{-d:0,a:(a+d)} \right\|_{TV}.$$

IDEA

- Estimate $\beta^d(a)$.
- Use the triangle inequality:

$$|\widehat{\beta}^d(a) - \beta(a)| \leq |\widehat{\beta}^d(a) - \beta^d(a)| + |\beta^d(a) - \beta(a)|.$$

- Bound the first term for fixed d.
- As $n \to \infty$, let $d \to \infty$.
- Hope the second half (doesn't depend on the data) goes to zero.
- To get $\widehat{\beta}^d(a)$, rewrite Total Variation in terms of densities, and plug in density estimators.

Main results

Theorem 1: Let $\mathbf{X}_{1:n}$ be a sample from a Markov process of order no larger than d. Then,

1. $$\mathbb{P}\left(|\widehat{\beta}^d(a) - \beta^d(a)| > \epsilon\right) \leq f(\epsilon, \mu).$$

2. $$\mathbb{E}[|\widehat{\beta}^d(a) - \beta^d(a)|] = O\left(\sqrt{\frac{W(n)}{n}}\right).$$

Proof sketch

- Show that density estimates converge in L_1 if $d \to \infty$.
- Propogate the convergence rate to mixing data using methods like above (see also Tran, 1989, 1994).
- Find the appropriate bandwidth for the histograms and the rate at which d can grow.

THE NON-RANDOM PART

Theorem 2: $\lim_{d \to \infty} \beta^d(a) = \beta(a)$

- Don't know how fast

PROOF SKETCH

- Easy to show that, $\forall d$, $\beta^d(a) \leq \beta(a)$.
- Rewrite the definition of $\beta^d(a)$ as a supremum over measurable sets.
- Note that the σ-field generated by $\mathbf{X}_{1:d_1}$ is nested in that generated by $\mathbf{X}_{1:d_2}$ for $d_1 < d_2$.
- Conclude that $\beta^d(a)$ is a bounded monotone increasing sequence.
- Now show that the limit is $\beta(a)$.

FINDING THE LIMIT

- Write

$$R = \mathbb{P}_{-\infty:0} \otimes \mathbb{P}_{a:\infty} - \mathbb{P}_{-\infty:0,a:\infty},$$

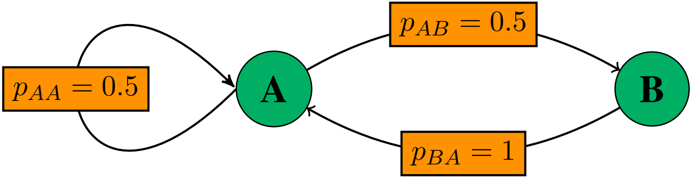
  a signed measure on σ_∞.
- Write

$$R^d = \mathbb{P}_{[d]} \otimes \mathbb{P}_{[d]} - \mathbb{P}_{-(d+1):0,a:(a+d-1)},$$

  a signed measure on the σ field generated by 2 d-blocks with a separation of length a.
- (We call the above $\sigma_{[d],a}$ in terribly overloaded notation.)
- Decompose R^d into positive and negative parts: Q^+ and Q^-.
- Set $\sigma_f = \bigcup_d \sigma_{[d],a}$.
- Can show that, [omit many steps] for any $\epsilon > 0$, $\exists A \in \sigma_f$ and $\exists d \geq 1$ such that

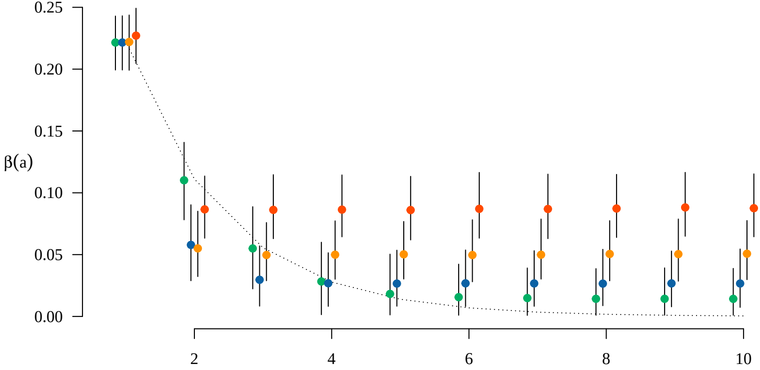
$$\beta^{d_1}(a) \geq \beta^d(a) \geq Q^+(A) \geq \beta(a) - \epsilon.$$

Markov process



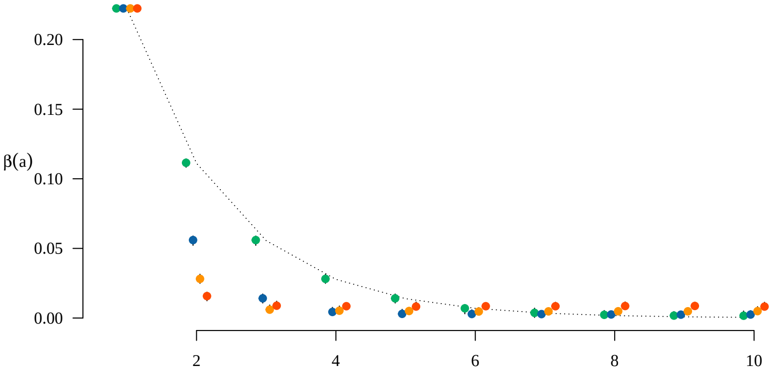
$$\beta(a) = \frac{4}{9}\left(\frac{1}{2}\right)^a$$

Results



- $n = 1000$

Bigger sample



- $n = 100,000$

Conclusions

- We did some other simulations.
- Seems to work pretty well.
- Upward bias because estimator is a.e. > 0, but $\beta(a) \to 0$.
- Can we extend to other mixing coefficients (or dependence coefficients).
- What if we want the curve $\beta(1), \ldots, \beta(a), \ldots$?
- Can we characterize $\beta(a) - \beta^d(a)$? Presumably depends on $\beta(a)$.
- Thanks to NSF and INET for support.

References

Bradley, R.C. (2005) Basic properties of strong mixing conditions. A survey and some open questions. Probability Surveys, **2**, 107–144.

Dedecker, J., Doukhan, P., Lang, G., Leon R., J.R., Louhichi, S. and Prieur, C. (2007) Weak Dependence: With Examples and Applications. Springer Verlag, New York.

Doukhan, P. (1994) Mixing: Properties and Examples. Springer Verlag, New York.

Tran, L. (1989) The L_1 convergence of kernel density estimates under dependence. The Canadian Journal of Statistics/La Revue Canadienne de Statistique, **17**, 197–208.

Tran, L. (1994) Density estimation for time series by histograms. Journal of Statistical Planning and Inference, **40**, 61–79.

Yu, B. (1994) Rates of convergence for empirical processes of stationary mixing sequences. The Annals of Probability, **22**, 94–116.