Dependent Wild Bootstrap for Degenerate U- and V-Statistics

A. Leucht, M. Neumann

Arthur Gretton's notes

March 28, 2014

Why it's interesting

The MMD is a distance between probabilities P and Q, based on i.i.d. samples from each.
What if we want to compute the distance when P and Q are random processes?
Applications:

1. Most real data (eg music, text),
2. Comparison of MCMC methods (convergence diagnostics, benchmarking of MCMC)

Why it's interesting

The MMD is a distance between probabilities P and Q, based on i.i.d. samples from each.
What if we want to compute the distance when P and Q are random processes?
Applications:

1. Most real data (eg music, text),
2. Comparison of MCMC methods (convergence diagnostics, benchmarking of MCMC)

The challenges:

1. What is the asymptotic distribution?
2. What is a good empirical construction (e.g. bootstrap) for the asymptotic distribution?

Assumptions on statistic

We want an estimate of the asymptotic distribution of an empirical quantity

$$V_n = \frac{1}{n} \sum_{s,t=1}^{n} h(x_s, x_t)$$

(which is a V-statistic - very close to a U-statistic, easier to analyse). This statistic is:

1. Degenerate: $E_X h(x, X) = 0$.
2. Symmetric, positive definite (MMD is the sum of several of these, so more complicated)
3. Lipschitz continuous:

$$\sup_{x,x',y,x' \neq x} \left| h(x,y) - h(x',y) \right| / \left\| x - x' \right\|_1 = \mathrm{Lip}(h) < \infty.$$

Assumptions on time series

The time series $(X_t)_{t\in\mathbb{Z}}$ is:

1. Strictly stationary,
2. τ-dependent, with $\sum_{r=1}^{\infty}\sqrt{\tau(r)} < \infty$.

The formal definition of the second condition is complicated. But the property which makes the proofs work is

$$E\left\|X_r - \tilde{X}_r\right\|_1 < \tau(r),$$

where X_r is dependent on X_0, whereas $\tilde{X}_r$ is a copy of X_r independent of X_0.

Asymptotics of V_n under the assumption

By a generalization of Mercer's theorem, we can write:

$$h(x, y) = \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \phi_k(y),$$

where λ_k, ϕ_k are solutions of the eigenvalue equation

$$E_{P_0}\left[h(x, X_0)\phi(X_0)\right] = \lambda \phi(x),$$

and P_0 is the stationary distribution of $(X_t)_{t \in \mathbb{Z}}$.

Asymptotics of V_n under the assumption

By a generalization of Mercer's theorem, we can write:

$$h(x,y) = \sum_{k=1}^{\infty} \lambda_k \phi_k(x)\phi_k(y),$$

where λ_k, ϕ_k are solutions of the eigenvalue equation

$$E_{P_0}\left[h(x, X_0)\phi(X_0)\right] = \lambda\phi(x),$$

and P_0 is the stationary distribution of $(X_t)_{t\in\mathbb{Z}}$.

$$\text{Asymptotic distribution :} \qquad V_n \xrightarrow{d} Z := \sum_{k=1}^{\infty} \lambda_k Z_k^2,$$

where

$$\mathrm{cov}(Z_j, Z_k) := \sum_{r=-\infty}^{\infty} \mathrm{cov}(\phi_j(X_0), \phi_k(X_r)).$$

Asymptotics of V_n: proof idea (1)

First define a truncation of $h(x, y)$:

$$h^{(K)}(x, y) = \sum_{k=1}^{K} \lambda_k \phi_k(x) \phi_k(y).$$

We can prove that as $K \to \infty$ the asymptotics of the truncation approach those of h.

The associated V-statistic is:

$$V_n^{(K)} = \frac{1}{n} \sum_{s,t=1}^{n} \underbrace{\left(\sum_{k=1}^{K} \lambda_k \phi_k(X_s) \phi_k(X_t) \right)}_{h^{(K)}(X_s, X_t)}$$

$$= \sum_{k=1}^{K} \lambda_k \left(n^{-1/2} \sum_{t=1}^{n} \phi_k(X_t) \right)^2$$

Asymptotics of V_n: proof idea (2)

Under the assumptions on x_t, we can apply a central limit theorem for weakly dependent random variables on the inner sum:

$$n^{-1/2}\sum_{t=1}^{n}\begin{bmatrix} \phi_1(X_t) & \ldots & \phi_K(X_t) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} Z_1 & \ldots & Z_K \end{bmatrix}$$

Bootstrap estimate of the asymptotic distribution

Define a new time series W_t^* with the property

$$\mathrm{cov}(W_s^*, W_t^*) = \rho\left(|s-t|/\ell_n\right),$$

where ℓ_n is a width parameter growing with n, and ρ is a window, e.g.

$$\mathrm{cov}(W_s^*, W_t^*) = \exp\left(-|s-t|/\ell_n\right).$$

Make a stronger assumption on X_t: it is τ-dependent with

$$\sum_{r=1}^{\infty} r^2\sqrt{\tau(r)} < \infty.$$

Then

$$V_n^* := \frac{1}{n}\sum_{s,t=1}^{n} h(X_s, X_t) W_s^* W_t^* \xrightarrow{d} Z \qquad \text{in probability}$$

Meaning: as measured via Prokhorov metric d_p,

$$d_p(V_n^*, Z) \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty.$$

How the proof works (1)

Again define a finite approximation,

$$V_n^{(K)*} = \frac{1}{n}\sum_{s,t=1}^{n} h^{(K)}(X_s, X_t) W_s^* W_t^*$$
$$= \sum_{k=1}^{K} \lambda_k \left(n^{-1/2} \sum_{t=1}^{n} \phi_k(X_t) W_t^* \right)^2$$

which can be shown to converge as $K \to \infty$. Define

$$Y_t^* := \left[\begin{array}{ccc} \phi_1(X_t) & \ldots & \phi_K(X_t) \end{array} \right] W_t^*$$

We need that in probability (as $n \to \infty$),

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t^* \xrightarrow{d} \mathcal{N}(0, \Sigma_K)$$

How the proof works (2)

$$\mathrm{cov}\left(n^{-1/2}\sum_{s=1}^{n}\phi_j(X_s)W_s^*, n^{-1/2}\sum_{t=1}^{n}\phi_k(X_t)W_t^*\right)$$

$$= \frac{1}{n}\sum_{s,t=1}^{n}\phi_j(X_s)\phi_k(X_t)\rho(|s-t|\,\ell_n)$$

$$= \underbrace{\frac{1}{n}\sum_{s,t=1}^{n}\left(\phi_j(X_s)\phi_k(X_t) - E\left[\phi_j(X_s)\phi_k(X_t)\right]\right)\rho(|s-t|\,\ell_n)}_{\text{converges to } 0}$$

$$+ \underbrace{\sum_{r=-\infty}^{\infty} E\left(\phi_j(X_0)\phi_k(X_r)\right)\rho(|r|/\ell_n)\max\left\{1-|r|/n, 0\right\}}_{\text{converges to } (\Sigma_K)_{j,k}}$$

Appendix: More detail on window assumptions

Requirements on ℓ_n:

$$\lim_{n\to\infty} \ell_n = \infty \qquad \lim_{n\to\infty} \ell_n/n = 0$$

Requirements on ρ:

$$\sum_{r=1}^{n-1} \rho(r/\ell_n) = O(l_n).$$

This being the case, W_t^* are weakly τ-dependent.