Appendix

5. Proofs

Lemma 5.1. *If a random variable X is distributed according to p, under conditions on the kernel*

$$0 = \oint_{\partial \mathcal{X}} k(x,x')p(x)n(x)dS(x'),$$
$$0 = \oint_{\partial \mathcal{X}} \nabla_x k(x,x')^\top n(x')p(x')dS(x'),$$

then for all $f \in \mathcal{F}$, the expected value of T is zero, i.e. $\mathbb{E}(Tf)(X) = 0$.

Proof. This result was proved on bounded domains $\mathcal{X} \subset \mathbb{R}^d$ by Oates et al. (2015, Lemma 1), where $n(x)$ is the unit vector normal to the boundary at x, and $\oint_{\partial\mathcal{X}}$ is the surface integral over the boundary $\partial\mathcal{X}$. The case of unbounded domains was discussed by Oates et al. (2015, Remark 2). Here we provide an alternative, elementary proof for the latter case. First we show that the functions $g_i = p \cdot f_i$ vanish at infinity, by which we mean that for all dimensions j

$$\lim_{x_j \to \infty} g_i(x_1, \cdots, x_d) = 0.$$

The density function p vanishes at infinity. The function f is bounded, which is implied by Cauchy-Schwarz inequality – $|f(x)| \leq \|f\| \sqrt{k(x,x)}$. This implies that the function g vanishes at infinity. To show the expected value $\mathbb{E}(T_p)f(X)$ is zero, it is sufficient to show that for all dimensions i, the expected value of $\frac{\partial \log p(X)}{\partial x_i} f_i(X) + \frac{\partial f_i(X)}{\partial x_i}$ is zero.

$$\begin{aligned}
&\mathbb{E}\left(\frac{\partial \log p(x)}{\partial x_i} f_i(x) + \frac{\partial f_i(x)}{\partial x_i}\right) \\
&= \int_{R_d} \left[\frac{\partial \log p(x)}{\partial x_i} f_i(x) + \frac{\partial f_i(x)}{\partial x_i}\right] q(x)dx \\
&= \int_{R_d} \left[\frac{1}{p(x)} \frac{\partial q(x)}{\partial x_i} f(x) + \frac{\partial f(x)}{\partial x_i}\right] q(x)dx \\
&= \int_{R_d} \left[\frac{\partial p(x)}{\partial x_i} f_i(x) + \frac{\partial f_i(x)}{\partial x_i} q(x)\right] dx \\
&\overset{(a)}{=} \int_{R_{d-1}} \left(\lim_{R\to\infty} p(x)f_i(x)\Big|_{x_i=-R}^{x_i=R}\right) dx_1 \cdots dx_{i-1} \cdots dx_{i+1} \cdots dx_d \\
&= \int_{R_{d-1}} 0 dx_1 \cdots dx_{i-1} \cdots dx_{i+1} \cdots dx_d \\
&= 0.
\end{aligned}$$

For the equation (a) we have used integration by parts, fact that g_i vanishes at infinity and Fubini-Toneli theorem to show that we can do iterated integration. The sufficient condition for the Fubini-Toneli theorem is that $\int |g_i f(x)| q(x) dx$ exists. This is implied by existence of $\mathbb{E}|\frac{\partial \log p(X)}{\partial x_i} f_i(X)|$ and $\mathbb{E}|\frac{\partial f_i(X)}{\partial x_i}|$. Since f_i is bounded and $\mathbb{E}|\frac{\partial \log p(X)}{\partial x_i}| \leq \mathbb{E}\|\nabla \log p(Z)\|^2$, condition ii) guarantees that $\mathbb{E}|\frac{\partial \log p(X)}{\partial x_i} f_i(X)|$ is finite. For the second term we have $\mathbb{E}|\frac{\partial f_i(X)}{\partial x_i}| = \mathbb{E}|\langle \frac{\partial k(X,cdot)}{\partial x_i}, f_i|\rangle \leq \|f_i\|_{\mathcal{F}} \mathbb{E}\sqrt{\frac{\partial^2 k(X,X)}{dx_i dx_{i+d}}}$, which is guaranteed by the condition iv). □

Proof of proposition 3.1. We check assumptions of the Theorem 2.1 (Leucht, 2012). The condition A1, $\sum_{t=1}^{\infty} \sqrt{\tau(t)} \leq \infty$, is implied by assumption $\sum_{t=1}^{\infty} t^2 \sqrt{\tau(t)} \leq \infty$ in Section 3. Condition A2 (iv), Lipschitz continuity of h is assumed. Conditions A2 i), ii) positive definiteness, symmetry and degeneracy of h follow from the proof of Theorem (2.2). Indeed

$$h(x, y) = \langle \xi(x), \xi(y) \rangle_{\mathcal{F}^d}$$

so the statistic is an inner product and hence positive definite. Degeneracy under the null follows from the fact that, by Theorem 2.1,$\mathbb{E}\xi(Z) = 0$. Finally, condition A2 (iii), $\mathbb{E}h(X, X) \leq \infty$ is assumed. □

Proof of proposition 3.2. We use Theorem 2.1 (Leucht, 2012) to see that, under the null hypothesis, $f(Z_{1,n}, \cdots, Z_{t,n})$ converges to zero in probability. The condition A1, $\sum_{t=1}^{\infty} \sqrt{\tau(t)} \leq \infty$, is implied by assumption $\sum_{t=1}^{\infty} t^2 \sqrt{\tau(t)} \leq \infty$ in Section 3. Condition A2 (iv), Lipschitz continuity of h is assumed.. Assumption B1 is identical to our assumption $\sum_{t=1}^{\infty} t^2 \sqrt{\tau(t)} \leq \infty$ from Section 3. Finally we check assumption B2 (bootstrap assumption): $\{W_{t,n}\}_{1 \leq t \leq n}$ is a row-wise strictly stationary triangular array independent of all Z_t such that $\mathbb{E}W_{t,n} = 0$ and $\sup_n \mathbb{E}|W_{t,n}^{2+\sigma}| = 1 < \infty$ for some $\sigma > 0$. The auto-covariance of the process is given by $\mathbb{E}W_{s,n}W_{t,n} = (1 - 2p_n)^{-|s-t|}$, so the function $\rho(x) = \exp(-x)$, and $l_n = \log(1 - 2p_n)^{-1}$. We verify that $\lim_{u \to 0} \rho(u) = 1$. If we set $p_n = w_n^{-1}$, such that $w_n = o(n)$ and $\lim_{n \to \infty} w_n = \infty$, then $l_n = O(w_n)$ and $\sum_{r=1}^{n-1} \rho(|r|/l_n) = \frac{1-(1-2p_n)^{n+1}}{p_n} = O(w_n) = O(l_n)$. We show that,under the alternative hypothesis, B_n converges to zero - we use (Chwialkowski et al., 2014, Theorem 3), the only assumption $\tau(r) = o(r^{-4})$ is satisfied since $\sum_{t=1}^{\infty} t^2 \sqrt{\tau(t)} \leq \infty$. We check the assumption

$$\sup_N \sup_{i,j \in N^2} \mathbb{E}h(Z_i)h(Z_j) < \infty.$$

We have $\mathbb{E}h(Z, Z')^2 = \mathbb{E}\langle \xi(Z, \xi(Z') \rangle^2 \leq \mathbb{E}\|\xi(Z)\|^2 \|\xi(Z')\|^2 \leq \sqrt{\mathbb{E}\|\xi(Z)\|^4 \mathbb{E}\|\xi(Z')\|^4} \leq \mathbb{E}h^2(Z, Z) < \infty$ and so for $i, j \in N^2$

$$\mathbb{E}h(Z_i)h(Z_j) \leq \mathbb{E}h(Z_i)^2 \mathbb{E}h(Z_j)^2 \leq (\mathbb{E}h^2(Z, Z))^2 < \infty.$$

We show that, under the alternative hypothesis, V_n converges to a positive constant - we use (Chwialkowski et al., 2014, Theorem 3). The zero comportment of h is positive sine $S(Z)^2 > 0$. We checked the assumption $\sup_N \sup_{i,j \in N^2} \mathbb{E}h(Z_i)h(Z_j) < \infty$ above.

□

5.1. Linear time test

We may use similar reasoning for the quadratic time test to define a linear time test, based on the two-sample test of Chwialkowski et al. (2015). For some fixed location y and a random variable X, define a random variable $s(X, y)$ as

$$s(X, y) = \nabla \log p(X) g(X, y) - \nabla g(X, y). \tag{3}$$

For some number of random locations Y_1, Y_J and a random variable X define a random vector Z_i

$$Z_i = (s(X_i, Y_1), \cdots, s(X_i, Y_J)) \in \mathbf{R}^J. \tag{4}$$

Let W_n be a mean of Z_i's $W_n = \frac{1}{n} \sum_{i=1}^{n} Z_i$, and Σ_n its covariance matrix $\Sigma_n = \frac{1}{n} Z Z^T$. The test statistic is

$$S_n = n W_n \Sigma_n^{-1} W_n. \tag{5}$$

The computation of S_n requires inversion of a $J \times J$ matrix Σ_n, but this is fast and numerically stable: J will typically be small, and is less than 10 in our experiments. The next proposition demonstrates the use of S_n as a one-sample test.

Proposition 5.2 (Asymptotic behavior of S_n)**.** *If $\mathbb{E}s(X, y) = 0$ for all y, then the statistic S_n is a.s. asymptotically distributed as a χ^2-random variable with Jd degrees of freedom, where d is X dimensionality (as $n \to \infty$ with d fixed). If $\mathbb{E}s(X, y) \neq 0$ for almost all y then a.s. for any fixed r, $\mathbb{P}(S_n > r) \to 1$ as $n \to \infty$.*

One sample test Calculate S_n. Choose a threshold r_α corresponding to the $1 - \alpha$ quantile of a χ^2 distribution with J degrees of freedom, and reject the null hypothesis whenever S_n is larger than r_α.

6. MCMC convergence testing

Convergence phase. We only can show consistency during the convergence phase on the empirical data. During convergence phase the null hypothesis is not true, but can be 'almost' true. For a sequence Z_i define the test statistic is $\|S_n\|^2$, where $S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i$

$$\mathbb{E}\|S_n\|^2 \geq \|\mathbb{E}S_n\|^2 > \frac{1}{n}\sum_{i,j}^{n}\langle \mathbb{E}\xi_i, \mathbb{E}\xi_i\rangle$$

If it is reasonable to assume that $\mathbb{E}\xi_i \simeq \mathbb{E}\xi_j$ the sum diverges, which makes the test consistent. One way to achieve $\mathbb{E}\xi_i = \mathbb{E}\xi_j$ it is to introduce an extra level of randomization i.e. sample the sequence without replacement and used the re-sampled sequence in the test. This may however change the structure of temporal dependence.

Stationary phase. In the stationary phase there are number of results which might be used to show that the chain is τ-mixing.

Strong mixing coefficients. Strong mixing is historically the most studied type of temporal dependence – a lot of models, example being Markov Chains, are proved to be strongly mixing, therefore it's useful to relate weak mixing to strong mixing.

A process is called absolutely regular (β-mixing) if $\beta(m) \to 0$, where

$$\beta(m) = \frac{1}{2}\sup_{n}\sup\sum_{i=1}^{I}\sum_{j=1}^{J}|P(A_i \cap B_j) - P(A_i)P(B_j)|.$$

The second supremum in the $\beta(m)$ definition is taken over all pairs of finite partitions $\{A_1, \cdots, A_I\}$ and $\{B_1, \cdots, B_J\}$ of the sample space such that $A_i \in \mathcal{A}_1^n$ and $B_j \in \mathcal{A}_{n+m}^{\infty}$, and $\mathcal{A}_b^c$ is a sigma field spanned by a subsequence, $\mathcal{A}_b^c = \sigma(X_b, X_{b+1}, ..., X_c)$.

A process is called strongly mixing (α-mixing) if $\alpha(m) \to 0$, where

$$\alpha(m) = \sup_{n}\sup_{A\in\mathcal{A}_1^n}\sup_{B\in\mathcal{A}_{n+m}^{\infty}}|P(B \cap A) - P(B)P(A)|.$$

By (Bradley et al., 2005) we have $\alpha(m) \leq \beta(m)$.

Using another weak mixing coefficient $\tilde{\alpha}(m)$ we can relate strong mixing to weak mixing. The process is called $\tilde{\alpha}$-mixing if $\tilde{\alpha}(m) \to 0$, where

$$\tilde{\alpha}(m) = \sup_{l\in\mathbb{N}}\frac{1}{l}\sup_{m\leq i_1\leq\ldots\leq i_l}\tilde{\alpha}(\mathcal{F}_0, (X_{i_1}, ..., X_{i_l})) \xrightarrow{r\to\infty} 0, \text{ where}$$
$$\tilde{\alpha}(\mathcal{M}, X) = \sup_{g\in\Lambda}\|\ \mathbb{E}(g(X)|\mathcal{M}) - \mathbb{E}g(X)\ \|_1$$

and Λ is the set of all one-Lipschitz continuous real-valued functions on the domain of X. (Dedecker et al., 2007, Remark 2.4) show that $\tilde{\alpha}(m) \leq 2\alpha(m)$. (Dedecker & Prieur, 2005, Proposition 2) relates τ-mixing and $\tilde{\alpha}$ mixing, as follows: if Q_x is the generalized inverse of the tail function

$$Q_x(u) = \inf_{t\in R}\{P(|X| > t) \leq u\},$$

then

$$\tau(\mathcal{M}, X) \leq 2\int_0^{\tilde{\alpha}(\mathcal{M},X)} Q_x(u)du.$$

While this definition can be hard to interpret, it can be simplified in the case $E|X|^p = M$ for some $p > 1$, since via Markov's inequality $P(|X| > t) \leq \frac{M}{t^p}$, and thus $\frac{M}{t^p} \leq u$ implies $P(|X| > t) \leq u$. Therefore $Q'(u) = \frac{M}{\sqrt[p]{u}} \geq Q_x(u)$. As a result, we have the following inequality

$$\frac{\sqrt[p]{\tilde{\alpha}(\mathcal{M}, X)}}{M} \geq C\tau(\mathcal{M}, X).$$

(Dedecker & Prieur, 2005) provides examples of systems that are tau-mixing. In particular, given that certain assumptions are satisfied causal functions of stationary sequences, iterated random functions, Markov chains, expanding maps are all τ-mixing.

Of particular interest to this work are Markov chains. The assumptions provided by (Dedecker & Prieur, 2005), under which Markov chains are tau-mixing are somehow difficult to check but we can use classical theorems about the absolute regularity (beta mixing). In particular (Bradley et al., 2005, Corollary 3.6) states that a Harris recurrent and aperiodic Markov chain satisfies absolute regularity and (Bradley et al., 2005, Theorem 3.7) states that geometric ergodicity implies geometric decay of the β coefficient. Interestingly (Bradley et al., 2005, Theorem 3.2) describes situations in which a non-stationary chain β-mixes exponentially.

Using inequalities between τ-mixing coefficient and strong mixing coefficients one can use those classical theorems show that e.g Markov chains are τ-mixing sequences.