Appendix

5. Proofs

Lemma 5.1. If a random variable $X$ is distributed according to $p$, under conditions on the kernel $k(x, x')$, we have

$$0 = \int_{\partial X} k(x, x') p(x) n(x) dS(x'),$$

where $n(x)$ is the vector normal to the boundary at $x$, and $f_{\partial X}$ is the surface integral over the boundary $\partial X$. 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 $j$

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

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.

$$\mathbb{E} \left( \frac{\partial \log p(x)}{\partial x_i} f_i(x) + \frac{\partial f_i(x)}{\partial x_i} \right)$$

$$= \int_{\mathbb{R}^d} \left[ \frac{\partial \log p(x)}{\partial x_i} f_i(x) + \frac{\partial f_i(x)}{\partial x_i} \right] q(x) dx$$

$$= \int_{\mathbb{R}^d} \left[ \frac{1}{p(x)} \frac{\partial q(x)}{\partial x_i} f_i(x) + \frac{\partial f_i(x)}{\partial x_i} \right] q(x) dx$$

$$= \int_{\mathbb{R}^d} \left[ \frac{\partial p(x)}{\partial x_i} f_i(x) + \frac{\partial f_i(x)}{\partial x_i} q(x) \right] dx$$

$$\overset{(a)}{=} \int_{\mathbb{R}^d} \left[ \lim_{R \to \infty} p(x) f_i(x) \right]_{x_i = R}^{x_i = -R} dx_1 \cdots dx_{i-1} \cdots dx_{i+1} \cdots dx_d$$

$$= \int_{\mathbb{R}^{d-1}} 0 dx_1 \cdots dx_{i-1} \cdots dx_{i+1} \cdots dx_d = 0.$$

For the equation (a) we have used integration by parts, fact that $g_i$ vanishes at infinity and Fubini-Tonelli theorem to show that we can do iterated integration. The sufficient condition for the Fubini-Tonelli 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 $\mathbb{E} |\frac{\partial f_i(X)}{\partial x_i}| = \mathbb{E} \|\frac{\partial h(x, \cdot)}{\partial x_i} f_i\| \leq \|f_i\|_x \mathbb{E} \sqrt{\frac{\partial^2 h(X)}{\partial x_i \partial x_j} f_i} dx_j dx_{i+d}$, which is guaranteed by the condition iv). \hfill \square

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 $\sum_{t=1}^{\infty} \sqrt{t^2 \sqrt{\tau(t)}} \leq \infty$ 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
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 $\sup_{1 \leq n \leq T} \mathbb{E}|W_{t,n}|^2 = 1 < \infty$ 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 $\lim_{n \to \infty} \rho(u) = 1$. 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 $\sum_{r=1}^{n-1} \rho(|r|/l_n) = \frac{1 - (1 - 2p_n)^n}{p_n} = O(w_n) = O(l_n)$. We show that, under the alternative hypothesis, $B_n$ converges to zero - we use (Chwialkowski et al., 2014, Theorem 3), the only assumption $\tau(r) = o(r^{-1})$ is satisfied since $\sum_{t=1}^{\infty} t^2 \sqrt{\tau(t)} \leq \infty$. We check the assumption

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

We have $\mathbb{E} h(Z, Z')^2 = \mathbb{E} (\xi(Z) \xi(Z'))^2 \leq \mathbb{E} \|\xi(Z)\|^2 \|\xi(Z')\|^2 \leq \sqrt{\mathbb{E} \|\xi(Z)\|^4} \mathbb{E} \|\xi(Z')\|^4 \leq \mathbb{E} h^2(Z, Z) < \infty$ and so for $i, j \in \mathbb{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 since $S(Z)^2 > 0$. We checked the assumption $\sup_{n} \sup_{i,j \in \mathbb{N}^2} \mathbb{E} h(Z_i) h(Z_j) < \infty$ 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

$$s(X, y) = \nabla \log p(X) g(X, y) - \nabla g(X, y).$$

For some number of random locations $Y_1, Y_2$ and a random variable $X$ define a random vector $Z_i$

$$Z_i = (s(X, Y_1), \cdots, s(X, Y_j)) \in \mathbb{R}^J.$$

Let $W_n$ be a mean of $Z_i$, $W_n = \frac{1}{n} \sum_{i=1}^{n} Z_i$, and $\Sigma_n$ its covariance matrix $\Sigma_n = \frac{1}{n} Z Z^T$. The test statistic is

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

The computation of $S_n$ requires inversion of a $J \times J$ matrix $\Sigma_n$, 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 $\mathbb{E} \xi(X, y) = 0$ for all $y$, then the statistic $S_n$ is a.s. asymptotically distributed as a $\chi^2$-random variable with $Jd$ degrees of freedom, where $d$ is $X$ dimensionality (as $n \to \infty$ with $d$ fixed). If $\mathbb{E} \xi(X, y) \neq 0$ for almost all $y$ then a.s. for any fixed $r$, $P(S_n > r) \to 1$ as $n \to \infty$.

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

A Kernel Test of Goodness of Fit
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 $Z$, define the test statistic is $\|S_n\|^2$, where $S_n = \frac{1}{n} \sum_{i=1}^{n} \xi_i$

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

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 $\tau$-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 ($\beta$-mixing) if $\beta(m) \to 0$, where

$$\beta(m) = \frac{1}{2n} \sup_{A,B} \sup_{i,j} \left| \sum_{i=1}^{J} \sum_{j=1}^{J} \left[ P(A_i \cap B_j) - P(A_i)P(B_j) \right] \right|$$

The second supremum in the $\beta(m)$ definition is taken over all pairs of finite partitions $\{A_1, \ldots, A_I\}$ and $\{B_1, \ldots, B_I\}$ of the sample space such that $A_i \in A_i^c$ and $B_j \in A^\infty_{n+m}$, and $A_i^c$ is a sigma field spanned by a subsequence, $A_i^c = \sigma(X_{ba}, X_{ba+1}, \ldots, X_c)$.

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

$$\alpha(m) = \sup_{A,B} \sup_{i,j} \left| \sum_{i=1}^{J} \sum_{j=1}^{J} \left[ P(B \cap A) - P(B)P(A) \right] \right|$$

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

$$\tilde{\alpha}(m) = \sup_{i \in \mathbb{N}} \frac{1}{\sum_{m \leq i_1 \leq \ldots \leq i_t} \tilde{\alpha}(\mathcal{F}_0, (X_i_1, \ldots, X_i_t))} \to 0$$

$$\tilde{\alpha}(\mathcal{M}, X) = \sup_{g \in \Lambda} \| \mathbb{E}(g(X)|\mathcal{M}) - \mathbb{E}g(X) \|_1$$

and $\Lambda$ 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 $\tau$-mixing and $\tilde{\alpha}$ mixing, as follows: if $Q_z$ is the generalized inverse of the tail function

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

then

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

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 $Q(u) = \frac{M}{t^p} \geq Q_z(u)$. As a result, we have the following inequality

$$\frac{\sqrt{\tilde{\alpha}(\mathcal{M}, X)}}{M} \geq C \tau(\mathcal{M}, X).$$
(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 $\tau$-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 $\beta$ coefficient. Interestingly (Bradley et al., 2005, Theorem 3.2) describes situations in which a non-stationary chain $\beta$-mixes exponentially.

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