A Bayesian Nonparametric Conditional Two-sample Test with an Application to Local Causal Discovery (Supplementary material)

Philip A. Boeken[1] **Joris M. Mooij**[1]

[1]Korteweg-de Vries Institute for Mathematics, University of Amsterdam

1 HYPOTHESIS TESTING WITH PÓLYA TREE PRIORS

In general, our setup for independence testing will assume availability of independent samples $X_1, ..., X_n$ of a random variable X with continuous distribution P. We let $\mathcal{X}$ denote the domain of X, and let $\mathcal{M}$ be the space of continuous distributions on $\mathcal{X}$. Our hypotheses will be of the form

$$H_0 : X \sim P \text{ with } P \in \mathcal{M}_0, \quad H_1 : X \sim P \text{ with } P \in \mathcal{M}_1, \tag{1}$$

where $\mathcal{M}_0, \mathcal{M}_1 \subset \mathcal{M}$, and $\mathcal{M}_0 \cap \mathcal{M}_1 = \emptyset$. Since we wish to device a Bayesian test, we will define prior distributions Π_0 and Π_1 with support on $\mathcal{M}_0$ and $\mathcal{M}_1$ respectively. Then we compare the evidence of the models given the data via the Bayes factor, i.e.

$$\text{BF}_{01} = \frac{\mathbb{P}(H_0|X_{1:n})}{\mathbb{P}(H_1|X_{1:n})} = \frac{p(X_{1:n}|H_0)}{p(X_{1:n}|H_1)} \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \frac{\int_{\mathcal{M}} \prod_{i=1}^n p(X_i) d\Pi_0(P)}{\int_{\mathcal{M}} \prod_{i=1}^n p(X_i) d\Pi_1(P)} \tag{2}$$

where we have placed equal prior weights on H_0 and H_1, so $\mathbb{P}(H_0) = \mathbb{P}(H_1) = 1/2$.

A canonical choice for a prior on a space of probability distributions is the Dirichlet Process. However, samples from the Dirichlet process are almost surely discrete distributions, so the Dirichlet Process is not a suitable choice for our setup. The Pólya tree prior does not suffer from this characteristic [Ferguson, 1974], and can be parametrised to be a suitable prior on $\mathcal{M}$. Since the elements of $\mathcal{M}$ have support on $\mathcal{X}$, we will speak of a Pólya tree on $\mathcal{X}$. We will first construct a Pólya tree on $\mathcal{X} \subseteq \mathbb{R}$, and then extend this definition to a Pólya tree on $\mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^2$.

First we recall the construction of the one-dimensional Pólya tree as described in the main paper. In particular, we construct a Pólya tree on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, where $\mathcal{X} \subseteq \mathbb{R}$, and $\mathcal{B}(\mathcal{X})$ denotes the Borel sigma-algebra on $\mathcal{X}$. In order to construct a random measure on $\mathcal{B}(\mathcal{X})$, we will assign random probabilities to a family of subsets $\mathcal{T}$ of $\mathcal{X}$ which generates the Borel sets. The family of subsets that we consider are the dyadic partitions of $[0, 1]$, mapped under the inverse of some cumulative distribution function G on $\mathcal{X}$. This results in the *canonical* family of partitions of $\mathcal{X}$, where for level j we have $\mathcal{X} = \bigcup_{\kappa \in \{0,1\}^j} B_\kappa$, with

$$B_\kappa := [G^{-1}(\tfrac{k-1}{2^j}), G^{-1}(\tfrac{k}{2^j})), \tag{3}$$

and k is the natural number corresponding to the bit string $\kappa \in \{0, 1\}^j$. A schematic depiction of this binary tree of partitions is shown in Figure 1. We define the index set by $K := \{\{0, 1\}^j : j \in \mathbb{N}\}$, so the family of subsets of $\mathcal{X}$ that we consider is $\mathcal{T} := \{B_\kappa : \kappa \in K\}$. From basic measure theory we know that $\mathcal{T}$ indeed generates $\mathcal{B}(\mathcal{X})$. We assign random probabilities to the elements of $\mathcal{T}$ by first assigning random probabilities to B_0 and B_1, and randomly subdividing these masses among the children of B_0 and B_1. In particular, for the first level of the partition we assign the random probabilities $\mathcal{P}(B_0) = \theta_0$ and $\mathcal{P}(B_1) = \theta_1$ with $(\theta_0, \theta_1) \sim \text{Dir}(\alpha_0, \alpha_1)$, for some hyper-parameters α_0 and α_1. Then, for every $B_\kappa \in \mathcal{T}$ we split the mass that is assigned to B_κ by assigning a fraction $\theta_{\kappa 0}$ to $B_{\kappa 0}$ and a fraction $\theta_{\kappa 1}$ to $B_{\kappa 1}$, where we let $(\theta_{\kappa 0}, \theta_{\kappa 1}) \sim \text{Dir}(\alpha_{\kappa 0}, \alpha_{\kappa 1})$. This construction yields a Pólya tree on $\mathcal{X}$, which is a random measure on $\mathcal{B}(\mathcal{X})$:

Definition 1.1 (Lavine, 1992) *A random probability measure $\mathcal{P}$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is said to have a Pólya tree distribution with parameter $(\mathcal{T}, \mathcal{A})$, written $\mathcal{P} \sim \text{PT}(\mathcal{T}, \mathcal{A})$, if there exist nonnegative numbers $\mathcal{A} = \{(\alpha_{\kappa 0}, \alpha_{\kappa 1}) : \kappa \in K\}$ and random variables $\Theta = \{(\theta_{\kappa 0}, \theta_{\kappa 1}) : \kappa \in K\}$ such that the following hold:*

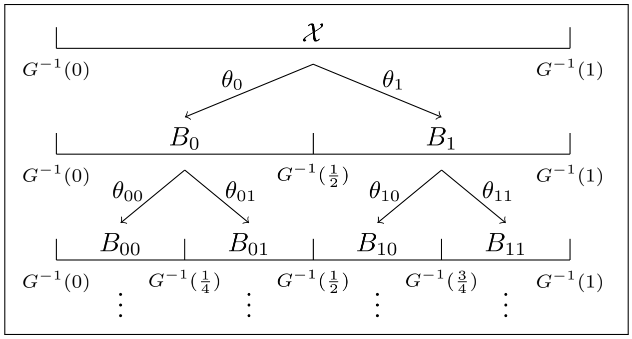


Figure 1: Construction of a one-dimensional Pólya tree based on canonical partitions.

1. *all the random variables in Θ are independent;*
2. *for every $\kappa \in K$, we have $(\theta_{\kappa 0}, \theta_{\kappa 1}) \sim \text{Dir}(\alpha_{\kappa 0}, \alpha_{\kappa 1})$;*
3. *for every $j \in \mathbb{N}$ and every $\kappa \in \{0,1\}^j$ we have $\mathcal{P}(B_\kappa|\Theta) = \prod_{i=1}^{j} \theta_{\kappa_1 \ldots \kappa_i}$.*

The support of the Pólya tree is determined by the choice of $\mathcal{T}$ and $\mathcal{A}$. In general, any separating binary tree of partitions of $\mathcal{X}$ can be considered. In this paper we only consider partitions of the type of equation (3). Ferguson [1974] shows that the Pólya tree is a Dirichlet process if $\alpha_\kappa = \alpha_{\kappa 0} + \alpha_{\kappa 1}$. The parameter of this Dirichlet process is the mean of the Pólya tree, i.e. the probability measure G_0 on $\mathcal{B}(\mathcal{X})$ defined by $G_0(B) := \mathbb{E}(\mathcal{P}(B))$ for all $B \in \mathcal{B}(\mathcal{X})$ [Lavine, 1994]. This implies that for this choice of $\mathcal{A}$, the support of the Pólya tree is contained in the space of discrete distributions. Sufficient conditions on $\mathcal{A}$ for samples of the Pólya tree to be continuous distributions are given by the following theorem:

Theorem 1.1 (Kraft, 1964) *Let $\bar{\sigma}_j := \sup\{Var(\theta_\kappa) : \kappa \in \{0,1\}^j\}$. If $\mathbb{E}(\theta_\kappa) = 1/2$ for all $\kappa \in K$ and $\sum_{j=1}^{\infty} \bar{\sigma}_j < \infty$, then with probability one, samples from $\mathcal{P}$ are absolutely continuous with respect to Lebesgue measure.*

This condition is satisfied if for each $\kappa \in \{0,1\}^j$ we take $\alpha_{\kappa 0} = \alpha_{\kappa 1} = j^2$, which is promoted as a 'sensible canonical choice' by Lavine [1992]. In this case we indeed have $\mathbb{E}(\theta_\kappa) = 1/2$, and thus for every $j \in \mathbb{N}$, the mass is (in expectation) split uniformly over the B_κ for all $\kappa \in \{0,1\}^j$. As a consequence the Pólya tree is centred on the base distribution with cumulative distribution function G, i.e. $\mathbb{E}(\mathcal{P}(B_\kappa)) = \int_{B_\kappa} G'(x)dx$. As mentioned in the main paper we only consider partitions up to a pre-determined level $J(n)$.

Let X be a continuous random variable with a distribution that lies in the support of the Pólya tree $\mathcal{P} \sim \text{PT}(\mathcal{T}, \mathcal{A})$. Drawing a distribution from $\mathcal{P}$ is done by drawing from each of the random variables in Θ. If we let $X_1, ..., X_n$ be a sample from X, then the likelihood of that sample with respect to a sampled distribution Θ from the Pólya tree $\text{PT}(\mathcal{T}, \mathcal{A})$ is

$$p(X_{1:n}|\Theta, \mathcal{T}, \mathcal{A}) = \prod_{\kappa \in K} \theta_{\kappa 0}^{n_{\kappa 0}} \theta_{\kappa 1}^{n_{\kappa 1}}, \tag{4}$$

where n_κ denotes the number of observations lying in B_κ, i.e. $n_\kappa := |X_{1:n} \cap B_\kappa|$. If we integrate over all possible values of all θ_κ, we obtain the marginal likelihood

$$p(X_{1:n}|\mathcal{T}, \mathcal{A}) = \prod_{\kappa \in K} \frac{\text{B}(\alpha_{\kappa 0} + n_{\kappa 0}, \alpha_{\kappa 1} + n_{\kappa 1})}{\text{B}(\alpha_{\kappa 0}, \alpha_{\kappa 1})}, \tag{5}$$

where $\text{B}(\cdot)$ denotes the Beta function. Note that this quantity corresponds to the marginal likelihood $\int_{\mathcal{M}} \prod_{i=1}^{n} p(X_i) d\Pi(P)$, a version of which occurs in the numerator and denominator of the right-hand side of equation (2). This marginal likelihood will therefore be a fundamental quantity in the Bayesian tests that we consider.

1.1 A NONPARAMETRIC TWO-SAMPLE TEST

In order to use the Pólya tree prior for Bayesian testing, we have to formulate our hypotheses H_0 and H_1 in terms of the relevant spaces of distributions $\mathcal{M}_0$ and $\mathcal{M}_1$, as suggested by equation (1). This is done by picking Pólya tree prior $\mathcal{P}_i \sim \Pi_i$

under H_i, and defining $\mathcal{M}_i$ to be the support of Π_i, for $i = 0, 1$. Given data to test our hypothesis with, we calculate marginal likelihoods via equation (5) for both Pólya trees $\mathcal{P}_0$ and $\mathcal{P}_1$, which are in turn used for calculating the Bayes factor via (2).

We first use this procedure to describe the nonparametric two-sample test, as proposed by Holmes et al. [2015]. Given a sample $\{(X_1, B_1), ..., (X_n, B_n)\}$ from binary variable C and continuous variable X, define $X^{(0)} := \{X_i : C_i = 0, i = 1, .., n\}$ and $X^{(1)} := \{X_i : C_i = 1, i = 1, .., n\}$. Let F denote the distribution of X, and let $F^{(0)}$ and $F^{(1)}$ denote the distributions of $X^{(0)}$ and $X^{(1)}$. We formulate the independence between X and C as a two-sample test, i.e.

$$H_0 : X \perp\!\!\!\perp C \iff F^{(0)} = F^{(1)} = F \tag{6}$$

$$H_1 : X \not\perp\!\!\!\perp C \iff F^{(0)} \neq F^{(1)}. \tag{7}$$

Under H_0 we standardise the sample $X_{1:n}$, and compute its marginal likelihood using equation (5). Under H_1, we model $X^{(0)}$ and $X^{(1)}$ as being samples from independent random variables, having different distributions. Since separately normalising $X^{(0)}$ and $X^{(1)}$ may erase distinctive features between the samples, we first standardise X, and then subdivide X into $X^{(0)}$ and $X^{(1)}$.

We formulate the Bayes factor as

$$\mathrm{BF}_{01} = \frac{p(X_{1:n}|\mathcal{T}, \mathcal{A})}{p(X^{(0)}|\mathcal{T}, \mathcal{A})p(X^{(1)}|\mathcal{T}, \mathcal{A})}. \tag{8}$$

Upon inspection of equation (5) we see that the Bayes factor can be written as an infinite product of fractions, being

$$\mathrm{BF}_{01} = \prod_{\kappa \in K} \frac{\mathrm{B}(\alpha_{\kappa 0} + n_{X|\kappa 0}, \alpha_{\kappa 1} + n_{X|\kappa 1})\mathrm{B}(\alpha_{\kappa 0}, \alpha_{\kappa 1})}{\mathrm{B}(\alpha_{\kappa 0} + n_{X^{(0)}|\kappa 0}, \alpha_{\kappa 1} + n_{X^{(0)}|\kappa 1})\mathrm{B}(\alpha_{\kappa 0} + n_{X^{(1)}|\kappa 0}, \alpha_{\kappa 1} + n_{X^{(1)}|\kappa 1})}, \tag{9}$$

where $n_{X|\kappa} := |X_{1:n} \cap B_\kappa|$, and $n_{X^{(0)}|\kappa}, n_{X^{(1)}|\kappa}$ are defined similarly. We note that whenever $n_{X|\kappa} \leq 1$ the fraction has a value of 1, so we calculate the marginal likelihoods until we either reach the maximum partitioning depth $J(n)$, or until $n_{X|\kappa} \leq 1$.

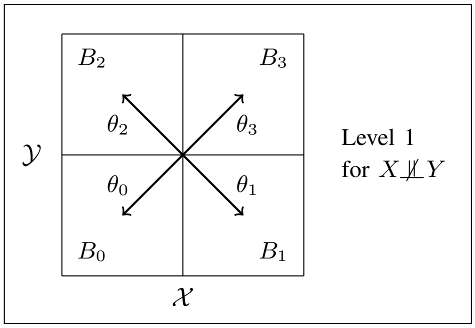
1.2 TWO-DIMENSIONAL PÓLYA TREES

Now that we have defined a Pólya tree on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with $\mathcal{X} \subseteq \mathbb{R}$, we extend this definition to a Pólya tree on $(\mathcal{X} \times \mathcal{Y}, \mathcal{B}(\mathcal{X} \times \mathcal{Y}))$ with $\mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^2$. This construction is done similarly to the construction on $\mathcal{X}$. We consider a base measure with cumulative distribution function G on $\mathcal{X} \cup \mathcal{Y}$, and partition $\mathcal{X} \times \mathcal{Y}$ into the four quadrants B_0, B_1, B_2 and B_3, where the boundaries of the B_i are determined by G^{-1}. We assign random probability θ_i to quadrant B_i with $(\theta_0, ..., \theta_3) \sim \mathrm{Dir}(\alpha_0, ..., \alpha_3)$. Then we recursively partition B_κ into quadrants $B_{\kappa 0}, ..., B_{\kappa 3}$, and split the mass assigned to B_κ according to $(\theta_{\kappa 0}, ..., \theta_{\kappa 3}) \sim \mathrm{Dir}(\alpha_{\kappa 0}, ..., \alpha_{\kappa 3})$. This partitioning scheme is shown in Figure 2a. We will denote this two-dimensional *canonical* family of partitions with $\mathcal{T}_2$, the set of parameters α_κ with $\mathcal{A}_2$, and the set of splitting variables θ_κ with Θ_2, where the subscript $_2$ emphasises the dimension of the space $\mathcal{X} \times \mathcal{Y}$. This leads to the following definition of the two-dimensional Pólya tree:

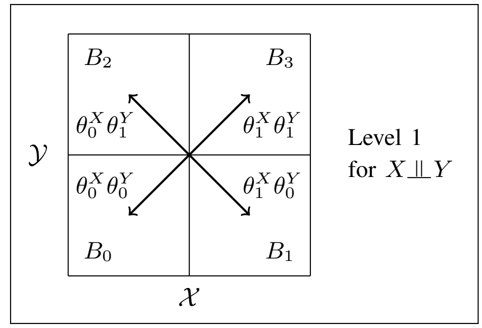
Definition 1.2 (Hanson [2006]) *A random probability measure $\mathcal{P}$ on $(\mathcal{X} \times \mathcal{Y}, \mathcal{B}(\mathcal{X} \times \mathcal{Y}))$ is said to have a Pólya tree distribution with parameter $(\mathcal{T}_2, \mathcal{A}_2)$, written $\mathcal{P} \sim \mathrm{PT}(\mathcal{T}_2, \mathcal{A}_2)$, if there exist nonnegative numbers $\mathcal{A}_2 = \{(\alpha_{\kappa 0}, \alpha_{\kappa 1}, \alpha_{\kappa 2}, \alpha_{\kappa 3}) : \kappa \in K_2\}$ and random variables $\Theta_2 = \{(\theta_{\kappa 0}, \theta_{\kappa 1}, \theta_{\kappa 2}, \theta_{\kappa 3}) : \kappa \in K_2\}$ such that the following hold:*

1. *all the random variables in Θ_2 are independent;*
2. *for every $\kappa \in K_2$ we have $(\theta_{\kappa 0}, \theta_{\kappa 1}, \theta_{\kappa 2}, \theta_{\kappa 3}) \sim \mathrm{Dir}(\alpha_{\kappa 0}, \alpha_{\kappa 1}, \alpha_{\kappa 2}, \alpha_{\kappa 3})$;*
3. *for every $j \in \mathbb{N}$ and every $\kappa \in \{0, 1, 2, 3\}^j$ we have $\mathcal{P}(B_\kappa|\Theta_2) = \prod_{i=1}^{j} \theta_{\kappa_1 \ldots \kappa_i}$.*

Similarly to the one-dimensional case, samples from the Pólya tree $\mathcal{P} \sim \mathrm{PT}(\mathcal{T}_2, \mathcal{A}_2)$ are continuous with respect to the two-dimensional Lebesgue measure if we take $\alpha_{\kappa 0} = \alpha_{\kappa 1} = \alpha_{\kappa 2} = \alpha_{\kappa 3} = (j+1)^2$, where j denotes the length in the string $\kappa \in K_2$ [Walker and Mallick, 1999]. Similar to the one-dimensional case, we only consider partitions up to a pre-specified depth $J(n)$.



(a) Partitioning scheme for $X\not\perp\!\!\!\perp Y$.



(b) Partitioning scheme for $X\perp\!\!\!\perp Y$.

When observing a sample $(X_1, Y_1), ..., (X_n, Y_n)$ from continuous random variables X and Y of which the joint distribution lies in the support of the two-dimensional Pólya tree $\mathcal{P}$, we have that the marginal likelihood of that sample is

$$p((X,Y)_{1:n}|\Theta_2, \mathcal{T}_2, \mathcal{A}_2) = \prod_{\kappa \in K} \theta_{\kappa 0}^{n_{\kappa 0}} \theta_{\kappa 1}^{n_{\kappa 1}} \theta_{\kappa 2}^{n_{\kappa 2}} \theta_{\kappa 3}^{n_{\kappa 3}}. \tag{10}$$

If we integrate over all possible values of all θ_κ, we obtain the marginal likelihood

$$p((X,Y)_{1:n}|\mathcal{T}_2, \mathcal{A}_2) = \prod_{\kappa \in K} \frac{\tilde{\mathrm{B}}(n_{\kappa 0} + \alpha_{\kappa 0}, n_{\kappa 1} + \alpha_{\kappa 1}, n_{\kappa 2} + \alpha_{\kappa 2}, n_{\kappa 3} + \alpha_{\kappa 3})}{\tilde{\mathrm{B}}(\alpha_{\kappa 0}, \alpha_{\kappa 1}, \alpha_{\kappa 2}, \alpha_{\kappa 3})}, \tag{11}$$

where $\tilde{\mathrm{B}}$ denotes the multivariate Beta function.[1]

Under the assumption $X \perp\!\!\!\perp Y$, we construct a prior similar to the two-dimensional Pólya tree. First we note that the two-dimensional family of partitions $\mathcal{T}_2$ can be regarded as the per-level Cartesian product of the partitions, i.e.

$$\mathcal{T}_2 = \left\{ \{B_\kappa \times B_\ell : B_\kappa \in \mathcal{T}_X, B_\ell \in \mathcal{T}_Y, \kappa, \ell \in \{0,1\}^j\} : j \in \mathbb{N} \right\} \tag{12}$$

where $\mathcal{T}_X$ and $\mathcal{T}_Y$ are one-dimensional canonical partitions $\mathcal{X}$ and $\mathcal{Y}$ respectively. For every level κ, we first split the mass over the elements of $\mathcal{T}_X$ according to $(\theta_{\kappa 0}^X, \theta_{\kappa 1}^X) \sim \mathrm{Dir}(\alpha_{\kappa 0}^X, \alpha_{\kappa 1}^X)$, and then independently split the mass over the elements of $\mathcal{T}_Y$ according to $(\theta_{\kappa 0}^Y, \theta_{\kappa 1}^Y) \sim \mathrm{Dir}(\alpha_{\kappa 0}^Y, \alpha_{\kappa 1}^Y)$. We denote the set of parameters α_κ^X with $\mathcal{A}_X$, and the parameters α_κ^Y with $\mathcal{A}_Y$. This prior yields a marginal likelihood of

$$\begin{aligned} p((X,Y)_{1:n}|\mathcal{T}_2, \mathcal{A}_X, \mathcal{A}_Y) = \prod_{\kappa \in K} & \frac{\mathrm{B}(n_{\kappa 0} + n_{\kappa 2} + \alpha_{\kappa 0}^X, n_{\kappa 1} + n_{\kappa 3} + \alpha_{\kappa 1}^X)}{\mathrm{B}(\alpha_{\kappa 0}^X, \alpha_{\kappa 1}^X)} \\ & \times \frac{\mathrm{B}(n_{\kappa 0} + n_{\kappa 1} + \alpha_{\kappa 0}^Y, n_{\kappa 2} + n_{\kappa 3} + \alpha_{\kappa 1}^Y)}{\mathrm{B}(\alpha_{\kappa 0}^Y, \alpha_{\kappa 1}^Y)}, \end{aligned} \tag{13}$$

as shown by Filippi and Holmes [2017]. We notice that this equals the product of the marginal likelihoods of X and Y according to independent one-dimensional Pólya tree priors $\mathcal{P}_X \sim \mathrm{PT}(\mathcal{T}_X, \mathcal{A}_X)$ on $\mathcal{X}$ and $\mathcal{P}_Y \sim \mathrm{PT}(\mathcal{T}_Y, \mathcal{A}_Y)$ on $\mathcal{Y}$, i.e.

$$p((X,Y)_{1:n}|\mathcal{T}_2, \mathcal{A}_X, \mathcal{A}_Y) = p(X_{1:n}|\mathcal{T}_X, \mathcal{A}_X) p(Y_{1:n}|\mathcal{T}_Y, \mathcal{A}_Y), \tag{14}$$

where the univariate marginal likelihoods are computed according to equation (5). To ensure that this prior is not biased when considered in conjunction with the two-dimensional Pólya tree, we consider parameters $\alpha_{\kappa 0}^X = \alpha_{\kappa 0} + \alpha_{\kappa 2}$, $\alpha_{\kappa 1}^X = \alpha_{\kappa 1} + \alpha_{\kappa 3}$, $\alpha_{\kappa 0}^Y = \alpha_{\kappa 0} + \alpha_{\kappa 1}$ and $\alpha_{\kappa 1}^Y = \alpha_{\kappa 2} + \alpha_{\kappa 3}$ [Filippi and Holmes, 2017]. When using the set of standard parameters $\mathcal{A}_2$ for the two-dimensional Pólya tree, we have $\mathcal{A}' := \mathcal{A}_X = \mathcal{A}_Y = \{2j^2 : j \in \mathbb{N}\}$.

[1] which is defined as $\tilde{\mathrm{B}}(\alpha_1, \alpha_2, \alpha_3, \alpha_4) := \prod_{i=1}^4 \Gamma(\alpha_i) / \Gamma(\sum_{i=1}^4 \alpha_i)$.

1.3 A NONPARAMETRIC INDEPENDENCE TEST

A Bayesian independence test that utilises two-dimensional Pólya trees is proposed by Filippi and Holmes [2017]. Considering one-dimensional continuous random variables X and Y, we test the hypotheses

$$H_0 : X \perp\!\!\!\perp Y, \quad H_1 : X \not\perp\!\!\!\perp Y \tag{15}$$

using the Bayes factor

$$\mathrm{BF}_{01} = \frac{p(X_{1:n}|\mathcal{T}, \mathcal{A}')p(Y_{1:n}|\mathcal{T}, \mathcal{A}')}{p((X,Y)_{1:n}|\mathcal{T}_2, \mathcal{A}_2)}, \tag{16}$$

where the marginal likelihoods are computed according to equations (5) and (13).

Using similar arguments as for the two-sample test, the Bayes factor can be denoted as an infinite product, of which the terms are equal to one when $n_{XY|\kappa} \leq 1$. Therefore we compute the marginal likelihoods up to level $J(n)$, or until all elements of the partition contain at most one observation.

2 PROTEIN EXPRESSION DATA

In the main paper we apply the LCD algorithm, implemented with the Bayesian ensemble of independence tests, to protein expression data [Sachs et al., 2005]. The data set consists of measurements of 11 phosphorylated proteins and phospholipids (Raf, Erk, p38, JNK, Akt, Mek, PKA, PLCg, PKC, PIP2 and PIP3) and 8 indicators of different interventions, performed by adding reagents to the cellular system, which are depicted in Table 1.[2] The biological details of these proteins, phospholipids, and reagents are described in Sachs et al. [2005]. Using flow cytometry, the activity of the 11 proteins and phospholipids are measured from a single human immune system cell. Flow cytometry allows for simultaneous, independent observation of hundreds of cells, producing a statistically large sample, and thus allowing for the application of causal inference algorithms [Sachs et al., 2005]. The 'expert network' from Sachs et al. [2005] is depicted in Figure 3. We note that, as argued in the main paper, we do not accept this network as the true causal graph, but merely display it suggestively.

Table 1: Interventions from the data set of Sachs et al. [2005].

	Description	Nr. of observations
1	CD3, CD28	853
2	CD3, CD28, Akt-inhibitor	911
3	CD3, CD28, G0076	723
4	CD3, CD28, Psitectorigenin	810
5	CD3, CD28, U0126	799
6	CD3, CD28, LY294002	848
7	PMA	913
8	β2CAMP	707

We assume that adding the reagents is not caused by the activity of the proteins and phospholipids, which justifies the application of the LCD algorithm to this dataset, as per Proposition 3.1 of the main paper. When performing a statistical test we always use the entire set of observations. As is common when analysing flow cytometry data, we preprocessed the data by taking the log of the raw measurement values.

[2]Similarly to most analyses of this data, we restrict our attention to 8 out of 14 experimental conditions, namely those in which no ICAM was added.

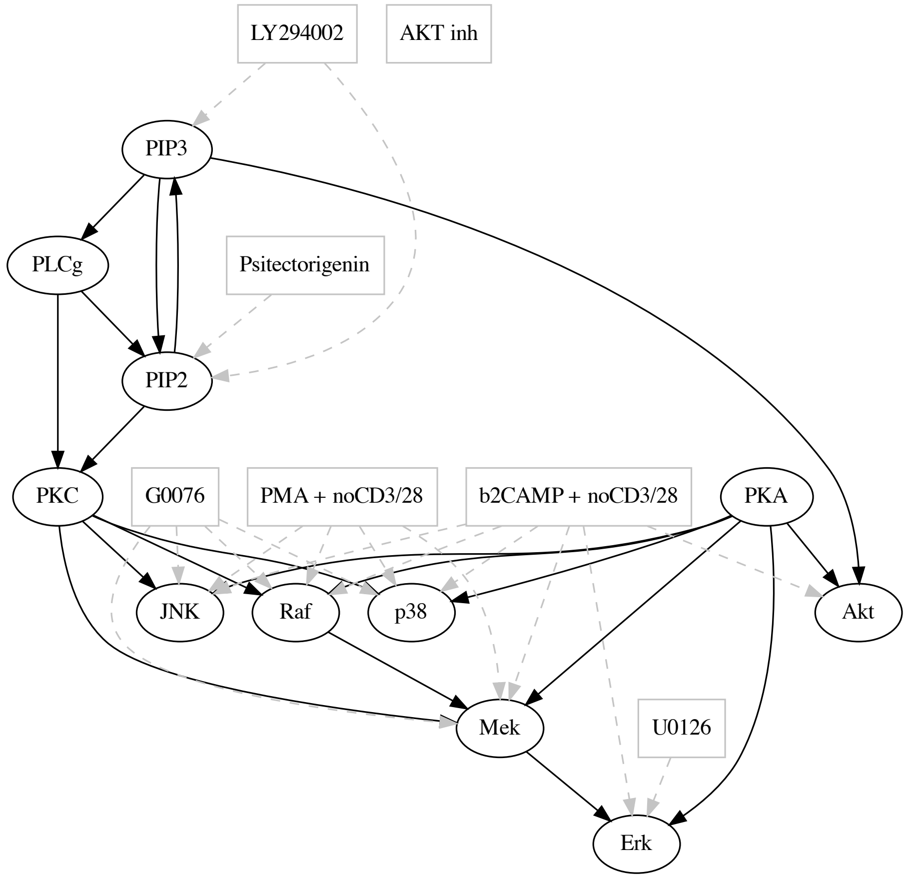


Figure 3: The ‘expert network’ as provided by Sachs et al. [2005]. Edges indicate direct causal effects between the nodes. Interventions and their direct causal effects are indicated with light-coloured and dashed nodes and edges.

References

Thomas S. Ferguson. Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2(4):615–629, 07 1974. URL `https://doi.org/10.1214/aos/1176342752`.

Sarah Filippi and Chris C. Holmes. A Bayesian nonparametric approach to testing for dependence between random variables. *Bayesian Analysis*, 12(4):919–938, 12 2017. URL `https://doi.org/10.1214/16-BA1027`.

Timothy E. Hanson. Inference for mixtures of finite Pólya tree models. *Journal of the American Statistical Association*, 101 (476):1548–1565, 2006. URL `http://www.jstor.org/stable/27639772`.

Chris C. Holmes, François Caron, Jim E. Griffin, and David A. Stephens. Two-sample Bayesian nonparametric hypothesis testing. *Bayesian Analysis*, 10(2):297–320, 06 2015. URL `https://doi.org/10.1214/14-BA914`.

Charles H. Kraft. A class of distribution function processes which have derivatives. *Journal of Applied Probability*, 1(2): 385–388, 1964. URL `http://www.jstor.org/stable/3211867`.

Michael Lavine. Some aspects of Pólya tree distributions for statistical modelling. *The Annals of Statistics*, 20(3):1222–1235, 09 1992. URL `https://doi.org/10.1214/aos/1176348767`.

Michael Lavine. More aspects of Pólya tree distributions for statistical modelling. *The Annals of Statistics*, 22(3):1161–1176, 1994. URL `http://www.jstor.org/stable/2242220`.

Karen Sachs, Omar Perez, Dana Pe’er, Douglas A. Lauffenburger, and Garry P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):529–528, 2005. URL `http://www.jstor.org/stable/3841298`.

Stephen Walker and Bani K. Mallick. A Bayesian semiparametric accelerated failure time model. *Biometrics*, 55(2): 477–483, 1999. URL `http://www.jstor.org/stable/2533795`.