# Supplementary materials for "BET on Independence"

**Abstract**

In this supplementary file, we provide some additional numerical and graphical illustrations, as well as some key proofs. Some simple proofs are omitted.

*Keywords:* Nonparametric Inference; Multiple Testing; Binary Expansion; Contingency Table; Hadamard Transform

## 1 Additional Numerical Studies

### 1.1 Illustration of Power Loss in Detecting $BEX_4$

In this section we provide simulation results showing the possible power loss in detecting $BEX_4$. We consider two important tests in practice: the test based on the Hoeffding's D as a CDF based test and the test based on distance correlation as a kernel based test. We consider the sample size of $n = 5000, 10000, 15000, 20000$ and 25000. The data are generated by uniformly sampling from the uniform distribution over $BEX_4$. The level of the tests is set to be 0.1. The power is calculated based on 1000 simulated datasets and is presented in Table 1.

| | $n = 5000$ | $n = 10000$ | $n = 15000$ | $n = 20000$ | $n = 25000$ |
|---|---|---|---|---|---|
| Hoeffding's D | 0.114 | 0.149 | 0.172 | 0.181 | 0.216 |
| Distance Correlation | 0.096 | 0.121 | 0.109 | 0.114 | 0.115 |

Table 1: Power of Hoeffding's D and distance correlation for the uniform distribution over $BEX_4$.

We can see from Table 1 that the power of these two tests are low even when the sample size is as large as 25000. One intuition behind this power loss is that these tests are global and are not able to detect the local dependency in the uniform distribution over $BEX_4$. See the following illustration of a particular dataset of sample size 5000 as in Figure 1: Although data are locally collinear, globally they are not. Therefore, although the dependency in this particular dataset can be clearly observed by naked eyes, the $p$-value of Hoeffding's D test is 0.350 and the $p$-value of distance correlation test is 0.531. This power loss was quite large and was one main motivation of this paper.

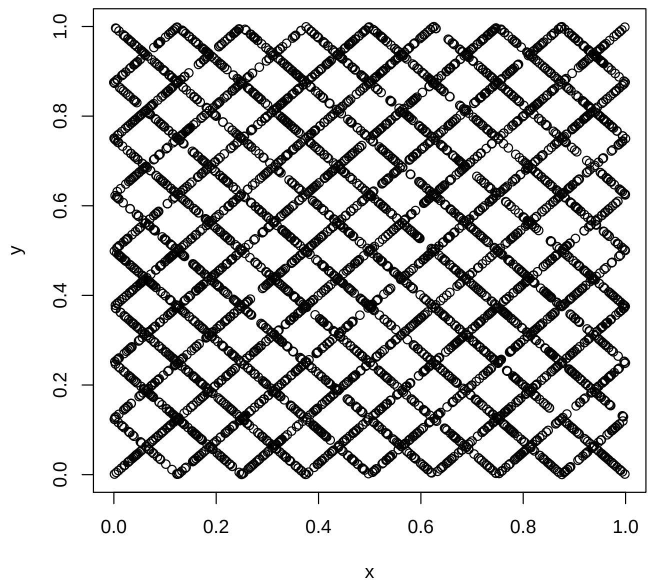


Figure 1: 5000 samples from the uniform distribution over the bisection expanding cross (BEX) at level $d = 4$. For this particular dataset, the $p$-value of Hoeffding's D test is 0.350 and the $p$-value of distance correlation test is 0.531.

For replicability purposes, we provide below our R code for generating i.i.d. samples from the BEX as in Figure 1.

```
library(energy)
library(Hmisc)
bex.sample <- function(n,depth=2){ # depth >=1
```

```
width <- 1/2^depth
which.be <- sample(1:2^(2*(depth-1)),n,replace=T)
bex.center <-  cbind(rep((1:2^(depth-1))/2^(depth-1),2^(depth-1)),
                  rep( (1:2^(depth-1))/2^(depth-1), rep(2^(depth-1),
                  2^(depth-1))   ))-1/2^(depth)
x_inc <- runif(n)*2*width-width
y_inc <- x_inc*sample(c(-1,1),n,replace=T)
xy <- bex.center[which.be,]+cbind(x_inc,y_inc)
colnames(xy) <- c("x","y")
xy
}
set.seed(1)
xy <- bex.sample(n=5000,depth=4);x <- xy[,1];y <- xy[,2]
plot(xy);dcor.ttest(x,y);hoeffd(x,y)$P
```

We remark here that the power of BET against this alternative distribution can be very high. See the discussions in Section 4.3 of the main paper.

## 1.2 Studies of the Power Loss Due to Multiplicity Control over the Depth

In this section we provide simulation results illustrating the adverse effect on power due to multiplicity control over depth discussed in Section 4.5 of the main paper. We consider the first four scenarios that are considered in Section 6 in the main paper since the dependency in these scenarios can be explained by one cross interaction, thus an "oracle" depth can be defined for each of these scenarios: As Figure 5 in the main paper shows, observations with linear dependency tend to fall in the positive region of $\widehat{\dot{A}}_1\widehat{\dot{B}}_1$, observations with the parabolic dependency tend to fall in the positive region of $\widehat{\dot{A}}_1\widehat{\dot{A}}_2\widehat{\dot{B}}_1$, observations with circular dependency tend to fall in the negative region of $\widehat{\dot{A}}_1\widehat{\dot{A}}_2\widehat{\dot{B}}_1\widehat{\dot{B}}_2$, observations with the sine dependency tend to fall in the negative region of $\widehat{\dot{A}}_2\widehat{\dot{B}}_1$. Therefore, we define the oracle depths for the linear, parabolic, circular and sine dependency to be 1, 2, 2 and 2 respectively. We compare the power of the Max BET with the above oracle depth (denoted by BETO in this subsection) and that

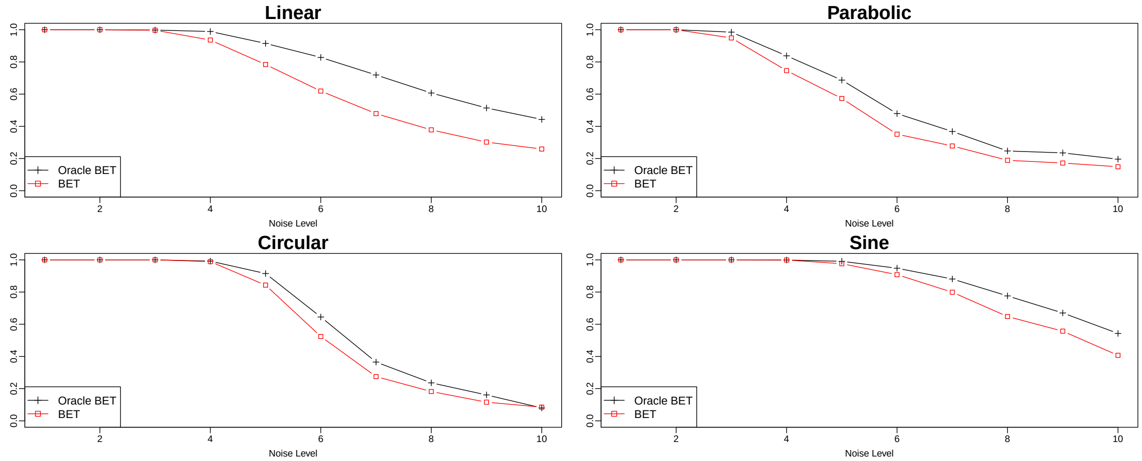


Figure 2: Comparison of the power of the Max BET with multiplicity control over depth (BETS) and that of the Max BET with an oracle depth (BETO). The loss of power is substantial in the linear case where the oracle depth is 1. The loss is less severe in other cases where the oracle depth is 2.

of the Max BET with the second stage multiplicity control (the procedure discussed in Section 4.5 and studied in Section 6 of the main paper, denoted by BETS in this subsection). See Figure 2. The loss of power is substantial in the linear case where the oracle depth is 1. However, the loss is much less in other cases where the oracle depth is 2.

Below is the explanation of the above phenomenon: Suppose $d_{max} = 4$ and the overall FEWR is $\alpha$. When the oracle depth is $d_{oracle} = 1$, the level for the oracle cross interaction in BETO is $\alpha$, while that for BETS is $\frac{1}{d_{max}}\alpha = 0.25\alpha$. The ratio between them is 0.25. However, when $d_{oracle} = 2$, the level for BETO is $\frac{1}{(2^{d_{oracle}}-1)^2}\alpha = \frac{1}{9}\alpha$, while that for BETS is $\frac{1}{d_{max}((2^{d_{oracle}}-1)^2-(2^{d_{oracle}-1}-1)^2)}\alpha = \frac{1}{32}\alpha$. Note that the ratio in this case is $\frac{9}{32} = 0.281 > 0.25$. This means the difference in the levels is smaller, thus the loss of power is less severe.

In general, the ratio of levels is $\frac{(2^{d_{oracle}}-1)^2}{d_{max}((2^{d_{oracle}}-1)^2-(2^{d_{oracle}-1}-1)^2)}$, which is an increasing function in $d_{oracle}$ and is asymptotically equivalent to $\frac{4}{3d_{max}}$. Therefore, we conclude the studies about the power loss due to multiplicity control over depth with the following remarks:

1. The most severe power loss due to multiplicity control over depth occurs when $d_{oracle} = 1$. The loss with a larger $d_{oracle}$ is less serious.

2. By Hoeffding's inequality, the multiplicative effect on the critical value is $O(\sqrt{\log d_{max}})$, which is small when $d_{max}$ is not too large.

3. By Theorem 4.4 in the main paper, $d_{max} \leq \log_2 n$. Thus the multiplicative effect can be further bounded by $O(\sqrt{\log \log n})$.

One practical remedy to avoid the loss of power due to multiplicity control in BET when $d_{oracle} = 1$ could be considering both the analysis with BET and that of the distance correlation (which has the best power for linear dependency as shown in Section 6 of the main paper).

# 2 Proofs in Section 2 of the Main Paper

To facilitate the explanation in this section, we use the concept of dyadic squares of level $k \geq 0$, $E_{i,j}^k$, which are squares with four corners at

$$\{(\frac{i}{2^k}, \frac{j}{2^k}), (\frac{i+1}{2^k}, \frac{j}{2^k}), (\frac{i}{2^k}, \frac{j+1}{2^k}), (\frac{i+1}{2^k}, \frac{j+1}{2^k})\}_{i,j=0}^{2^k-1}.$$

We denote the uniform distribution over $[0,1]^2$ by $\mathbf{P}_0$. We also denote the probability measure of the uniform distribution over the BEX at level $d$ by $\mathbf{P}_d$. We state without proof below the following easy but useful lemma:

**Lemma 2.1.**

*(a) (Probability of a dyadic square) For $d \geq 1$, $\forall i, j = 0, \ldots, 2^{d-1} - 1$,*

$$\mathbf{P}_d(E_{i,j}^{d-1}) = \mathbf{P}_0(E_{i,j}^{d-1}) = \frac{1}{4^{d-1}}. \tag{2.1}$$

(b) *(Probability measure restricted to a dyadic square)* *For* $d \geq 1$, $\frac{i}{2^{d-1}} \leq x < \frac{i+1}{2^{d-1}}$, $i = 0, \ldots, 2^{d-1} - 1$,

$$\mathbf{P}_d\left(\frac{i}{2^{d-1}} \leq X_d \leq x \middle| E^{d-1}_{i+1,j}\right) = \mathbf{P}_1\left(0 \leq X_1 \leq 2^{d-1}\left(x - \frac{i}{2^{d-1}}\right)\right) = 2^{d-1}\left(x - \frac{i}{2^{d-1}}\right). \tag{2.2}$$

## 2.1 Proof of Proposition 2.1

*Proof.* (a) Without loss of generality, we show below the marginal distribution of $X$ is uniform. By Lemma 2.1, for any $d \geq 1$, suppose $x \in [\frac{i_0}{2^{d-1}}, \frac{i_0+1}{2^{d-1}}), i_0 = 0, \ldots, 2^{d-1} - 1$.

$$\begin{aligned}
&\mathbf{P}_d(X_d \leq x)\\
=&\mathbf{P}_d\left(X_d \leq \frac{i_0}{2^{d-1}}\right) + \mathbf{P}_d\left(\frac{i_0}{2^{d-1}} \leq X_d \leq x\right)\\
=&\begin{cases} \sum_{j=0}^{2^{d-1}-1} \mathbf{P}_d\left(\frac{i_0}{2^{d-1}} \leq X_d \leq x \middle| E^{d-1}_{i_0,j}\right)\mathbf{P}_d\left(E^{d-1}_{i_0,j}\right) & i_0 = 0\\ \sum_{i=0}^{i_0-1}\sum_{j=0}^{2^{d-1}-1} \mathbf{P}_d(E^{d-1}_{i,j}) + \sum_{j=0}^{2^{d-1}-1} \mathbf{P}_d\left(\frac{i_0}{2^{d-1}} \leq X_d \leq x \middle| E^{d-1}_{i_0,j}\right)\mathbf{P}_d\left(E^{d-1}_{i_0,j}\right) & i_0 \geq 1 \end{cases}\\
=&2^{d-1}\frac{i_0}{4^{d-1}} + 2^{d-1}2^{d-1}\left(x - \frac{i_0}{2^{d-1}}\right)\frac{1}{4^{d-1}}\\
=&x.
\end{aligned} \tag{2.3}$$

(b) The proof is immediate and is omitted.

(c) At level $d$, the coordinates of any point $(x, y) \in [0, 1]^2$ can be written as $x = \frac{2i_0+1\pm c_1}{2^d}, y = \frac{2j_0+1\pm c_2}{2^d}$ for some $0 < c_1, c_2 < 1$, and some $i_0, j_0 = 0, \ldots, 2^{d-1} - 1$. Without loss of generality, we give the proof for $x = \frac{2i_0+1-c_1}{2^d}, y = \frac{2j_0+1+c_2}{2^d}$, for $i_0 \geq 1$ and $j_0 \geq 1$, as other cases are similar and simpler. In this case,

$$xy = \frac{1}{4^d}(4i_0j_0 + 2i_0(1 + c_2) + 2j_0(1 - c_1)) + O(4^{-d}). \tag{2.4}$$

On the other hand, by Lemma 2.1,

$$
\begin{aligned}
&\mathbf{P}_d(X_d \le x, Y_d \le y)\\
=&\mathbf{P}_d\left(X_d \le \frac{i_0}{2^{d-1}}, Y_d \le \frac{j_0}{2^{d-1}}\right) + \mathbf{P}_d\left(X_d \le \frac{i_0}{2^{d-1}}, \frac{j_0}{2^{d-1}} \le Y_d \le y\right) +\\
&+\mathbf{P}_d\left(\frac{i_0}{2^{d-1}} \le X_d \le x, Y_d \le \frac{j_0}{2^{d-1}}\right) + \mathbf{P}_d\left(\frac{i_0}{2^{d-1}} \le X_d \le x, \frac{i_0}{2^{d-1}} \le Y_d \le y\right)\\
=&\sum_{i=0}^{i_0-1}\sum_{j=0}^{j_0}\mathbf{P}_d(E_{i,j}^{d-1}) + \sum_{i=0}^{i_0-1}\mathbf{P}_d\left(\frac{j_0}{2^{d-1}} \le Y_d \le y \middle| E_{i,j_0}^{d-1}\right)\mathbf{P}_d\left(E_{i,j_0}^{d-1}\right) +\\
&+\sum_{j=0}^{j_0-1}\mathbf{P}_d\left(\frac{i_0}{2^{d-1}} \le X_d \le x \middle| E_{i_0,j}^{d-1}\right)\mathbf{P}_d\left(E_{i_0,j}^{d-1}\right) + \mathbf{P}_d\left(\frac{i_0}{2^{d-1}} \le X_d \le x, \frac{j_0}{2^{d-1}} \le Y_d \le y\right)\\
=&\frac{i_0 j_0}{4^{d-1}} + \frac{2j_0(1-c_1)}{4^d} + \frac{2i_0(1+c_2)}{4^d} + O(4^{-d}).
\end{aligned}
\tag{2.5}
$$

By comparing (2.4) and (2.5), we see $\mathbf{P}_d(X_d \le x, Y_d \le y) = xy + O(4^{-d})$.

□

## 2.2 Proof of Theorem 2.2

*Proof.* Consider the case when $\delta = 1$. Since $n$ is fixed and $\mathbf{P}_0(\partial C_n) = 0$ ($C_n$ is referred to as Jordan measurable in $\mathbb{R}^n$), $\forall \epsilon > 0$, there exists a finite collection of $N$ almost disjoint dyadic hypercubes $\{\boldsymbol{G}_i\}_{i=1}^N$, which are direct products of dyadic squares, such that

$$C_n \subset \cup_{i=1}^N \boldsymbol{G}_i \tag{2.6}$$

and

$$\mathbf{P}_0(\cup_{i=1}^N \boldsymbol{G}_i) \le \mathbf{P}_0(C_n) + \epsilon \le \alpha + \epsilon. \tag{2.7}$$

Suppose the smallest dyadic square in forming $\{\boldsymbol{G}_i\}_{i=1}^N$ is at level $k$. Now consider the BEX at level $k+1$, we see from Lemma 2.1 that

$$\mathbf{P}_{k+1}(\cup_{i=1}^N \boldsymbol{G}_i) = \mathbf{P}_0(\cup_{i=1}^N \boldsymbol{G}_i). \tag{2.8}$$

Therefore,

$$\mathbf{P}_{k+1}(C_n) \le \mathbf{P}_{k+1}(\cup_{i=1}^N \boldsymbol{G}_i) = \mathbf{P}_0(\cup_{i=1}^N \boldsymbol{G}_i) \le \mathbf{P}_0(C_n) + \epsilon \le \alpha + \epsilon. \tag{2.9}$$

This implies that with $n$ samples, the test with critical region $C_n$ is powerless with $F_n$ being the uniform distribution over the BEX at level $k+1$. $\square$

# 3 Proofs in Section 3 of the Main Paper

## 3.1 Proof of Theorem 3.4

*Proof.* We prove part (a) below by induction. The proof of part (b) is similar. For $d_1 = d_2 = 1$, we consider calculation of expectations by conditioning on $\dot{A}_1$ and $\dot{B}_1$. For example,

$$\begin{aligned}
&\mathbf{E}[\dot{B}_1] \\
=&\mathbf{E}[\dot{B}_1|A_1=1,B_1=1]\mathbf{P}(A_1=1,B_1=1)+\mathbf{E}[\dot{B}_1|A_1=1,B_1=0]\mathbf{P}(A_1=1,B_1=0)+ \\
&+\mathbf{E}[\dot{B}_1|A_1=0,B_1=1]\mathbf{P}(A_1=0,B_1=1)+\mathbf{E}[\dot{B}_1|A_1=0,B_1=0]\mathbf{P}(A_1=0,B_1=0) \\
=&p_{(11)}-p_{(10)}+p_{(01)}-p_{(00)}.
\end{aligned} \tag{3.1}$$

In general, we have the following four equations:

$$\begin{aligned}
1 = E_{(00)} = \mathbf{E}[1] =& p_{(11)}+p_{(10)}+p_{(01)}+p_{(00)} \\
E_{(01)} = \mathbf{E}[\dot{B}_1] =& p_{(11)}-p_{(10)}+p_{(01)}-p_{(00)} \\
E_{(10)} = \mathbf{E}[\dot{A}_1] =& p_{(11)}+p_{(10)}-p_{(01)}-p_{(00)} \\
E_{(11)} = \mathbf{E}[\dot{A}_1\dot{B}_1] =& p_{(11)}-p_{(10)}-p_{(01)}+p_{(00)}.
\end{aligned} \tag{3.2}$$

Note that

$$\mathbf{H}_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}. \tag{3.3}$$

Thus, by considering the matrix form of (3.2) with (3.3), we see $\boldsymbol{E} = \mathbf{H}\boldsymbol{p}$ for $d_1 = d_2 = 1$.

Suppose the result is true for $d_1$ and $d_2$. Without loss of generality, consider the case with depths $d_1$ and $d_2+1$ with a new binary variable $B_{d_2+1}$. Note that by the induction assumption, we have a similar identity as in (3.2) that for any $(d_1+d_2)$-dimensional binary integer indices $(\boldsymbol{ab})$ and $(\boldsymbol{a'b'})$, the corresponding entry in $\mathbf{H}_{2^{d_1+d_2}}$ is

$$\mathbf{H}_{(\boldsymbol{ab}),\overline{(\boldsymbol{a'b'})}} = \mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'}] = (-1)^{(\boldsymbol{ab})^T\overline{(\boldsymbol{a'b'})}}. \tag{3.4}$$

To consider the probabilities in the $\sigma$-field at depths $d_1$ and $d_2+1$, we write probabilities in the $\sigma$-field at depths $d_1$ and $d_2$ by $p_{(\boldsymbol{a'b'}\cdot)} = p_{(\boldsymbol{a'b'}1)} + p_{(\boldsymbol{a'b'}0)}$ where two latter probabilities are in the $\sigma$-field at depths $d_1$ and $d_2+1$.

With $B_{d_2+1}$, we can expand the following two quantities through condition expectation:

$$\begin{aligned}
&\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'}]p_{(\boldsymbol{a'b'}\cdot)}\\
=&\left(\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=1]\frac{p_{(\boldsymbol{a'b'}1)}}{p_{(\boldsymbol{a'b'}\cdot)}} + \mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=0]\frac{p_{(\boldsymbol{a'b'}0)}}{p_{(\boldsymbol{a'b'}\cdot)}}\right)p_{(\boldsymbol{a'b'}\cdot)}\\
=&\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=1]p_{(\boldsymbol{a'b'}1)} + \mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=0]p_{(\boldsymbol{a'b'}0)}
\end{aligned} \tag{3.5}$$

and

$$\begin{aligned}
&\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}\dot{B}_{d_2+1}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'}]p_{(\boldsymbol{a'b'}\cdot)}\\
=&\left(\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}\dot{B}_{d_2+1}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=1]\frac{p_{(\boldsymbol{a'b'}1)}}{p_{(\boldsymbol{a'b'}\cdot)}} + \mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}\dot{B}_{d_2+1}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=0]\frac{p_{(\boldsymbol{a'b'}0)}}{p_{(\boldsymbol{a'b'}\cdot)}}\right)p_{(\boldsymbol{a'b'}\cdot)}\\
=&\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}\dot{B}_{d_2+1}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=1]p_{(\boldsymbol{a'b'}1)} + \mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}\dot{B}_{d_2+1}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=0]p_{(\boldsymbol{a'b'}0)}.
\end{aligned} \tag{3.6}$$

Now note that

$$\begin{aligned}
\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=1] =& \mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'}] = (-1)^{(\boldsymbol{ab}0)^T\overline{(\boldsymbol{a'b'}1)}}\\
\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=0] =& \mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'}] = (-1)^{(\boldsymbol{ab}0)^T\overline{(\boldsymbol{a'b'}0)}}\\
\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}\dot{B}_{d_2+1}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=1] =& \mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'}] = (-1)^{(\boldsymbol{ab}1)^T\overline{(\boldsymbol{a'b'}1)}}\\
\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}\dot{B}_{d_2+1}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'},B_{d_2+1}=0] =& -\mathbf{E}[\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}|\boldsymbol{A}=\boldsymbol{a'},\boldsymbol{B}=\boldsymbol{b'}] = (-1)^{(\boldsymbol{ab}1)^T\overline{(\boldsymbol{a'b'}0)}}.
\end{aligned} \tag{3.7}$$

By using part (c) in Proposition 3.5 that $\mathbf{H}_{2^{d_1+d_2+1}} = \mathbf{H}_{2^{d_1+d_2}} \otimes \mathbf{H}_2$, putting (3.7) in (3.5) and (3.6) and combining over $(\boldsymbol{ab})$ and $(\boldsymbol{a'b'})$, the equation $\boldsymbol{E} = \mathbf{H}\boldsymbol{p}$ is shown at the depths $d_1$ and $d_2+1$, and the induction proof is complete. □

## 3.2 Proof of Theorem 3.7

*Proof.* This proof is seen by using part (a) in Proposition 3.5. □

## 3.3 Proof of Theorem 3.8

*Proof.* The "only if" part is easy. To prove the "if" part of Theorem 3.8, we first show the following lemma, which connects independence to the odds ratio between each cell and corresponding cells in the last row and column. Note that the same result is shown in Cornfield (1956) for the case when marginal counts are fixed, but we prove the general case in Section 3.4.

**Lemma 3.1.** *$U_{d_1}$ and $V_{d_2}$ are independent if and only if for any $\boldsymbol{a} \neq \mathbf{1}_{d_1}$ and $\boldsymbol{b} \neq \mathbf{1}_{d_2}$,*

$$\theta_{(\boldsymbol{ab})} = \frac{p_{(\mathbf{11})}p_{(\boldsymbol{ab})}}{p_{(\boldsymbol{a}\mathbf{1})}p_{(\mathbf{1}\boldsymbol{b})}} = 1. \tag{3.8}$$

With this lemma, we show below that $\lambda_{(\boldsymbol{ab})} = 1$ for all $\boldsymbol{a} \neq \mathbf{0}$ and $\boldsymbol{b} \neq \mathbf{0}$ implies $\theta_{(\boldsymbol{a}'\boldsymbol{b}')} = 1$ and $\log \theta_{(\boldsymbol{a}'\boldsymbol{b}')} = 0$ for any $\boldsymbol{a}' \neq \mathbf{1}_{d_1}$ and $\boldsymbol{b}' \neq \mathbf{1}_{d_2}$.

Note first that $\log \theta_{(\boldsymbol{a}')(\boldsymbol{b}')}$ can be written as a linear combination of log cell probabilities, i.e., $\log \theta_{(\boldsymbol{a}'\boldsymbol{b}')} = \ell^T_{(\boldsymbol{a}'\boldsymbol{b}')}\boldsymbol{p}_l$. Moreover, the binary integer indices of the only four non-zero entries in $\ell_{(\boldsymbol{a}'\boldsymbol{b}')}$ are 1 for $\overline{(\mathbf{11})}$, $-1$ for $\overline{(\boldsymbol{a}'\mathbf{1})}$, $-1$ for $\overline{(\mathbf{1}\boldsymbol{b}')}$ and 1 for $\overline{(\boldsymbol{a}'\boldsymbol{b}')}$. Denote the $(2^{d_1}-1)(2^{d_2}-1) \times 2^{d_1+d_2}$ matrix with rows of $\ell^T_{(\boldsymbol{a}'\boldsymbol{b}')}$ by $\mathbf{L}$.

Now note that by Theorem 3.7, $\boldsymbol{p}_l = \frac{1}{2^{d_1+d_2}}\mathbf{H}\boldsymbol{\lambda}_l$. Therefore, the event $\log \theta_{(\boldsymbol{a}'\boldsymbol{b}')} = 0$ for any $\boldsymbol{a}' \neq \mathbf{1}_{d_1}$ and $\boldsymbol{b}' \neq \mathbf{1}_{d_2}$ can be summarized by the following equation in $\boldsymbol{\lambda}_l$:

$$\mathbf{LH}\boldsymbol{\lambda}_l = \mathbf{0}. \tag{3.9}$$

To solve (3.9), we consider the inner product $\ell^T_{(\boldsymbol{a}'\boldsymbol{b}')}\mathbf{H}_{(\boldsymbol{ab})}$. Because all entries in $\ell_{(\boldsymbol{a}'\boldsymbol{b}')}$ are 0's except four of them, we have

$$\ell^T_{(\boldsymbol{a}'\boldsymbol{b}')}\mathbf{H}_{(\boldsymbol{ab})} = 1\cdot(-1)^{\overline{(\mathbf{11})}^T(\boldsymbol{ab})} + (-1)\cdot(-1)^{\overline{(\mathbf{1}\boldsymbol{b}')}^T(\boldsymbol{ab})} + (-1)\cdot(-1)^{\overline{(\boldsymbol{a}'\mathbf{1})}^T(\boldsymbol{ab})} + 1\cdot(-1)^{\overline{(\boldsymbol{a}'\boldsymbol{b}')}^T(\boldsymbol{ab})}. \tag{3.10}$$

Now note that for $\boldsymbol{a} = \mathbf{0}$,

$$\ell_{(\boldsymbol{a}'\boldsymbol{b}')}^{T} \mathbf{H}_{(\mathbf{0}\boldsymbol{b})} = 1 - (-1)^{\overline{(\mathbf{1}\boldsymbol{b}')}^{T}(\mathbf{0}\boldsymbol{b})} - 1 + (-1)^{\overline{(\boldsymbol{a}'\boldsymbol{b}')}^{T}(\mathbf{0}\boldsymbol{b})} = 0. \tag{3.11}$$

Similarly, for $\boldsymbol{b} = \mathbf{0}$,

$$\ell_{(\boldsymbol{a}'\boldsymbol{b}')}^{T} \mathbf{H}_{(\boldsymbol{a}\mathbf{0})} = 0. \tag{3.12}$$

With (3.11) and (3.12), we see that in (3.9), the coefficients for $\log\lambda_{(\mathbf{00})}$, $\log\lambda_{(\boldsymbol{a}\mathbf{0})}$'s and $\log\lambda_{(\mathbf{0}\boldsymbol{b})}$'s are all 0's. Therefore, (3.9) is a system of $(2^{d_1}-1)(2^{d_2}-1)$ linear equations with $(2^{d_1}-1)(2^{d_2}-1)$ unknowns of $\log\lambda_{(\boldsymbol{ab})}$ with $\boldsymbol{a} \neq \mathbf{0}$ and $\boldsymbol{b} \neq \mathbf{0}$. Because $\mathbf{L}$ is full rank, the unique solution to (3.9) is $\log\lambda_{(\boldsymbol{ab})} = 0$ and $\lambda_{(\boldsymbol{ab})} = 1$ for all $\boldsymbol{a} \neq \mathbf{0}$ and $\boldsymbol{b} \neq \mathbf{0}$.

$\square$

## 3.4 Proof of Lemma 3.1

*Proof.* The "only if" part is immediate. To see the "if" part, note that (3.8) implies that $p_{(\boldsymbol{ab})} = \frac{p_{(\boldsymbol{a}\mathbf{1})}p_{(\mathbf{1}\boldsymbol{b})}}{p_{(\mathbf{11})}}$. Denote the row total probability and column total probabilities by $p_{(\boldsymbol{a}\cdot)}$ and $p_{(\cdot\boldsymbol{b})}$ respectively. By summing over rows and columns, we see that

$$\frac{\sum_{\boldsymbol{a}\neq\mathbf{1}_{d_1}} p_{(\boldsymbol{a}\mathbf{1})} \sum_{\boldsymbol{b}\neq\mathbf{1}_{d_2}} p_{(\mathbf{1}\boldsymbol{b})}}{p_{(\mathbf{11})}} + \sum_{\boldsymbol{a}\neq\mathbf{1}_{d_1}} p_{(\boldsymbol{a}\mathbf{1})} + \sum_{\boldsymbol{b}\neq\mathbf{1}_{d_2}} p_{(\mathbf{1}\boldsymbol{b})} + p_{(\mathbf{11})} = 1, \tag{3.13}$$

which in turn gives

$$\left(\sum_{\boldsymbol{a}\neq\mathbf{1}_{d_1}} p_{(\boldsymbol{a}\mathbf{1})} + p_{(\mathbf{11})}\right)\left(\sum_{\boldsymbol{b}\neq\mathbf{1}_{d_2}} p_{(\mathbf{1}\boldsymbol{b})} + p_{(\mathbf{11})}\right) = p_{(\mathbf{1}\cdot)}p_{(\cdot\mathbf{1})} = p_{(\mathbf{11})}. \tag{3.14}$$

Now for any $\boldsymbol{a} \neq \mathbf{1}_{d_1}$, note that the row total probability is

$$p_{(\boldsymbol{a}\cdot)} = \sum_{\boldsymbol{b}'\neq\mathbf{1}_{d_2}} p_{(\boldsymbol{ab}')} + p_{(\boldsymbol{a}\mathbf{1})} = \frac{p_{(\boldsymbol{a}\mathbf{1})}p_{(\cdot\mathbf{1})}}{p_{(\mathbf{11})}} \tag{3.15}$$

and similarly, for any $\boldsymbol{b} \neq \mathbf{1}_{d_2}$, the column total probability is

$$p_{(\cdot\boldsymbol{b})} = \frac{p_{(\mathbf{1}\boldsymbol{b})}p_{(\mathbf{1}\cdot)}}{p_{(\mathbf{11})}}. \tag{3.16}$$

Therefore, by (3.14), for any $\boldsymbol{a} \neq \mathbf{1}_{d_1}$ and $\boldsymbol{b} \neq \mathbf{1}_{d_2}$,

$$p_{(\boldsymbol{a}\cdot)}p_{(\cdot\boldsymbol{b})} = \frac{p_{(\boldsymbol{a}\mathbf{1})}p_{(\cdot\mathbf{1})}}{p_{(\mathbf{1}\mathbf{1})}}\frac{p_{(\mathbf{1}\boldsymbol{b})}p_{(\mathbf{1}\cdot)}}{p_{(\mathbf{1}\mathbf{1})}} = \frac{p_{(\boldsymbol{a}\mathbf{1})}p_{(\mathbf{1}\boldsymbol{b})}}{p_{(\mathbf{1}\mathbf{1})}} = p_{(\boldsymbol{a}\boldsymbol{b})}. \tag{3.17}$$

Now for any $\boldsymbol{a} \neq \mathbf{1}_{d_1}$, by (3.17),

$$p_{(\boldsymbol{a}\mathbf{1})} = p_{(\boldsymbol{a}\cdot)} - \sum_{\boldsymbol{b}'\neq\mathbf{1}_{d_2}} p_{(\boldsymbol{a}\boldsymbol{b}')} = p_{(\boldsymbol{a}\cdot)} - \sum_{\boldsymbol{b}'\neq\mathbf{1}_{d_2}} p_{(\boldsymbol{a}\cdot)}p_{(\cdot\boldsymbol{b}')} = p_{(\boldsymbol{a}\cdot)}(1 - \sum_{\boldsymbol{b}'\neq\mathbf{1}_{d_2}} p_{(\cdot\boldsymbol{b}')}) = p_{(\boldsymbol{a}\cdot)}p_{(\cdot\mathbf{1})}. \tag{3.18}$$

Similarly,

$$p_{(\mathbf{1}\boldsymbol{b})} = p_{(\cdot\boldsymbol{b})}p_{(\mathbf{1}\cdot)}. \tag{3.19}$$

By (3.14),(3.17), (3.18) and (3.19), $U_{d_1}$ and $V_{d_2}$ are independent. □

# 4 Proofs in Section 4 of the Main Paper

## 4.1 Proof of Theorem 4.1

*Proof.* This is a consequence of Theorem 3.4 with $\boldsymbol{p} = \frac{1}{2^{d_1+d_2}}\mathbf{1}$. □

## 4.2 Proof of Theorem 4.2

*Proof.* To see how part (b) and (c) imply (a), note the following general facts:

1. $(\widehat{S}_{(\boldsymbol{a}\boldsymbol{b})} + n)/4 \sim \mathit{Hypergeometric}(n, n/2, n/2, w_{\boldsymbol{a}\boldsymbol{b}})$, where the noncentral parameter $\omega_{(\boldsymbol{a}\boldsymbol{b})}$ is given by

$$\omega_{(\boldsymbol{a}\boldsymbol{b})} = \frac{\mathbf{P}(\widehat{\dot{A}}_{\boldsymbol{a}} = 1, \widehat{\dot{B}}_{\boldsymbol{b}} = 1)\mathbf{P}(\widehat{\dot{A}}_{\boldsymbol{a}} = -1, \widehat{\dot{B}}_{\boldsymbol{b}} = -1)}{\mathbf{P}(\widehat{\dot{A}}_{\boldsymbol{a}} = 1, \widehat{\dot{B}}_{\boldsymbol{b}} = -1)\mathbf{P}(\widehat{\dot{A}}_{\boldsymbol{a}} = -1, \widehat{\dot{B}}_{\boldsymbol{b}} = 1)}. \tag{4.1}$$

2. The inclusion of events $\{\widehat{\boldsymbol{A}}_{\boldsymbol{a}} = \boldsymbol{a}', \widehat{\boldsymbol{B}}_{\boldsymbol{b}} = \boldsymbol{b}'\} \subseteq \{\widehat{\dot{A}}_{\boldsymbol{a}} = a, \widehat{\dot{B}}_{\boldsymbol{b}} = b\}$ for $a = \pm 1$ and $b = \pm 1$ if and only if $(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})} = a$ and $(-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})} = b$. In particular, $\{\widehat{\boldsymbol{A}}_{\boldsymbol{a}} = \mathbf{1}, \widehat{\boldsymbol{B}}_{\boldsymbol{b}} = \mathbf{1}\} \subseteq \{\widehat{\dot{A}}_{\boldsymbol{a}} = 1, \widehat{\dot{B}}_{\boldsymbol{b}} = 1\}$ for any $\boldsymbol{a}$ and $\boldsymbol{b}$ since $(-1)^{\overline{(\mathbf{1})}^T(\boldsymbol{a})} = (-1)^{\overline{(\mathbf{1})}^T(\boldsymbol{b})} = 1$.

It can be helpful to understand the above two facts with Figure 2 in the main paper. To see fact 1, note that for interaction $\widehat{\dot{A}_1}\widehat{A_2}\widehat{\dot{B}_1}$, the event $\{\widehat{\dot{A}}_{\boldsymbol{a},i} = 1, \widehat{\dot{B}}_{\boldsymbol{b},i} = 1\}$ corresponding to the event that observation $i$ falling into either the upper right white region or the upper left white region. To see fact 2, note that the upper right region is always white for any variables in the filtration.

With the above facts, to show that parts (b) and (c) imply (a) is to show that $\omega_{(\boldsymbol{ab})} = 1$ for any $\boldsymbol{a} \neq \boldsymbol{0}$ and $\boldsymbol{b} \neq \boldsymbol{0}$ implies independence, which is equivalent to, by Lemma 3.1, $\theta_{(\boldsymbol{a}'\boldsymbol{b}')} = 1$ for any $\boldsymbol{a}' \neq \boldsymbol{1}$ and $\boldsymbol{b}' \neq \boldsymbol{1}$.

To show that $\theta_{(\boldsymbol{a}'\boldsymbol{b}')} = 1$ for any $\boldsymbol{a}' \neq \boldsymbol{1}$ and $\boldsymbol{b}' \neq \boldsymbol{1}$, note that by (4.1) and fact 2 above,

$$\omega_{(\boldsymbol{ab})} = \frac{\sum_{\substack{(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})}=1 \\ (-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})}=1}} p_{(\boldsymbol{a}'\boldsymbol{b}')} \sum_{\substack{(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})}=-1 \\ (-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})}=-1}} p_{(\boldsymbol{a}'\boldsymbol{b}')}}{\sum_{\substack{(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})}=-1 \\ (-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})}=1}} p_{(\boldsymbol{a}'\boldsymbol{b}')} \sum_{\substack{(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})}=1 \\ (-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})}=-1}} p_{(\boldsymbol{a}'\boldsymbol{b}')}}. \tag{4.2}$$

Therefore, setting $\omega_{(\boldsymbol{ab})} = 1$ for $\boldsymbol{a} \neq \boldsymbol{0}$ and $\boldsymbol{b} \neq \boldsymbol{0}$ is to set

$$\sum_{\substack{(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})}=1 \\ (-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})}=1}} p_{(\boldsymbol{a}'\boldsymbol{b}')} \sum_{\substack{(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})}=-1 \\ (-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})}=-1}} p_{(\boldsymbol{a}'\boldsymbol{b}')} = \sum_{\substack{(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})}=-1 \\ (-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})}=1}} p_{(\boldsymbol{a}'\boldsymbol{b}')} \sum_{\substack{(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})}=1 \\ (-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})}=-1}} p_{(\boldsymbol{a}'\boldsymbol{b}')}. \tag{4.3}$$

Consider (4.3) as a system of $(2^{d_1}-1)(2^{d_2}-1)$ linear equations of $p_{(\boldsymbol{11})}p_{(\boldsymbol{a}'\boldsymbol{b}')}$ for $\boldsymbol{a}' \neq \boldsymbol{1}$ and $\boldsymbol{b}' \neq \boldsymbol{1}$. Denote the coefficient matrix by $\mathbf{M}_{d_1,d_2}$, with the rows corresponding to $\omega_{(\boldsymbol{ab})}$ and columns corresponding to $p_{(\boldsymbol{11})}p_{(\boldsymbol{a}'\boldsymbol{b}')}$. Note that if $\mathbf{M}_{d_1,d_2}$ is invertible, then (4.3) has a unique solution. Note also that Lemma 3.1 implies $\theta_{(\boldsymbol{a}'\boldsymbol{b}')} = 1$, or $p_{(\boldsymbol{11})}p_{(\boldsymbol{a}'\boldsymbol{b}')} = p_{(\boldsymbol{a}'\boldsymbol{1})}p_{(\boldsymbol{1b}')}$ is a solution to (4.3). Thus, if $\mathbf{M}_{d_1,d_2}$ is invertible, $p_{(\boldsymbol{11})}p_{(\boldsymbol{a}'\boldsymbol{b}')} = p_{(\boldsymbol{a}'\boldsymbol{1})}p_{(\boldsymbol{1b}')}$ is the unique solution to (4.3). This means that $\omega_{(\boldsymbol{ab})} = 1$ for any $\boldsymbol{a} \neq \boldsymbol{0}$ and $\boldsymbol{b} \neq \boldsymbol{0}$ implies $\theta_{(\boldsymbol{a}'\boldsymbol{b}')} = 1$ for any $\boldsymbol{a}' \neq \boldsymbol{1}$ and $\boldsymbol{b}' \neq \boldsymbol{1}$, i.e., independence.

To show the invertibility of $\mathbf{M}_{d_1,d_2}$, note that by fact 2, $\mathbf{M}_{d_1,d_2}$ is a binary matrix with entries

$$\mathbf{M}_{d_1,d_2,(\boldsymbol{ab}),(\boldsymbol{a}'\boldsymbol{b}')} = \begin{cases} 1 & (-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})} = -1 \text{ and } (-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})} = -1 \\ 0 & \text{otherwise} \end{cases}, \tag{4.4}$$

which can be simplified to

$$\mathbf{M}_{d_1,d_2,(\boldsymbol{a}\boldsymbol{b}),(\boldsymbol{a}'\boldsymbol{b}')} = \frac{1}{4}(1-(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})})(1-(-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})}). \tag{4.5}$$

Consider the special case when $d_2 = 1$. In this case, the only possible values for $\boldsymbol{b}$ and $\boldsymbol{b}'$ are $\boldsymbol{b} = 1$ and $\boldsymbol{b}' = 0$. Therefore,

$$\mathbf{M}_{d_1,1,(\boldsymbol{a}1),(\boldsymbol{a}'0)} = \frac{1}{2}(1-(-1)^{\overline{(\boldsymbol{a}')}^T(\boldsymbol{a})}). \tag{4.6}$$

Note that (4.6) implies that one can write the matrix $\mathbf{H}_{2^{d_1}}$ in block form as

$$\mathbf{H}_{2^{d_1}} = \begin{pmatrix} 1 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{1}\mathbf{1}^T - 2\mathbf{M}_{d_1,1} \end{pmatrix} \tag{4.7}$$

Now note that $\mathbf{H}_{2^{d_1}}$ is invertible with $\mathbf{H}_{2^{d_1}}^{-1} = \frac{1}{2^{d_1}}\mathbf{H}_{2^{d_1}}$. Therefore, by the inversion formula of a partitioned matrix,

$$\frac{1}{2^{d_1}}(\mathbf{1}\mathbf{1}^T - 2\mathbf{M}_{d_1,1}) = (\mathbf{1}\mathbf{1}^T - 2\mathbf{M}_{d_1,1} - \mathbf{1}\mathbf{1}^T)^{-1} = -\frac{1}{2}\mathbf{M}_{d_1,1}^{-1}, \tag{4.8}$$

i.e., $\mathbf{M}_{d_1,1}$ is invertible and $\mathbf{M}_{d_1,1}^{-1} = -\frac{1}{2^{d_1-1}}(\mathbf{1}\mathbf{1}^T - 2\mathbf{M}_{d_1,1})$.

Similarly, when $d_1 = 1$, $\mathbf{M}_{1,d_2}$ with

$$\mathbf{M}_{1,d_2,(1\boldsymbol{b}),(0\boldsymbol{b}')} = \frac{1}{2}(1-(-1)^{\overline{(\boldsymbol{b}')}^T(\boldsymbol{b})}) \tag{4.9}$$

is invertible.

Now note that by (4.5), (4.6) and (4.9), we have $\mathbf{M}_{d_1,d_2} = \mathbf{M}_{d_1,1} \otimes \mathbf{M}_{1,d_2}$. Thus, $\mathbf{M}_{d_1,d_2}$ is invertible. □

## 4.3 Proof of Theorem 4.3

*Proof.* (a) Consider any two cross interaction variables $\dot{A}_{\boldsymbol{a}_1}\dot{B}_{\boldsymbol{b}_1}$ and $\dot{A}_{\boldsymbol{a}_2}\dot{B}_{\boldsymbol{b}_2}$ such that $(\boldsymbol{a}_1\boldsymbol{b}_1) \neq (\boldsymbol{a}_2\boldsymbol{b}_2)$. Under independence of $U_{d_1}$ and $V_{d_2}$, any $\dot{A}_i$ and $\dot{B}_j$ are independent. Note that their marginal expectations are also all 0's. Thus, we have

$$\mathbf{E}[\dot{A}_{\boldsymbol{a}_1}\dot{B}_{\boldsymbol{b}_1}\dot{A}_{\boldsymbol{a}_2}\dot{B}_{\boldsymbol{b}_2}] = 0 = \mathbf{E}[\dot{A}_{\boldsymbol{a}_1}\dot{B}_{\boldsymbol{b}_1}]\mathbf{E}[\dot{A}_{\boldsymbol{a}_2}\dot{B}_{\boldsymbol{b}_2}]. \tag{4.10}$$

Therefore, $\dot{A}_{\boldsymbol{a}_1}\dot{B}_{\boldsymbol{b}_1}$ and $\dot{A}_{\boldsymbol{a}_2}\dot{B}_{\boldsymbol{b}_2}$ are uncorrelated binary variables. This means they are independent. The pairwise independence of $S_{(\boldsymbol{ab})}$'s is a consequence of the pairwise independence of cross interaction variables.

(b) Under independence of $\widehat{U}_{d_1,i}$ and $\widehat{V}_{d_2,i}$ for each $i$, the joint distribution of cell counts $\widehat{\boldsymbol{N}}$ is central multivariate hypergeometric. This implies that the cell counts $\widehat{N}_{(\boldsymbol{ab})}$'s are exchangeable. In particular, the covariance matrix of $\widehat{\boldsymbol{N}}$ takes the form

$$\mathrm{Var}[\widehat{\boldsymbol{N}}] = c_1\mathbf{I}_{2^{d_1+d_2}} + c_2\mathbf{1}\mathbf{1}^T \tag{4.11}$$

for some constants $c_1$ and $c_2$.

Now consider $\widehat{S}_{(\boldsymbol{a}_1\boldsymbol{b}_1)}$ and $\widehat{S}_{(\boldsymbol{a}_2\boldsymbol{b}_2)}$ with $(\boldsymbol{a}_1\boldsymbol{b}_1) \neq (\boldsymbol{a}_2\boldsymbol{b}_2)$. Note that by Theorem 3.4, $\widehat{S}_{(\boldsymbol{a}_1\boldsymbol{b}_1)} = \mathbf{H}^T_{(\boldsymbol{a}_1\boldsymbol{b}_1)}\widehat{\boldsymbol{N}}$ and $\widehat{S}_{(\boldsymbol{a}_2\boldsymbol{b}_2)} = \mathbf{H}^T_{(\boldsymbol{a}_2\boldsymbol{b}_2)}\widehat{\boldsymbol{N}}$. Therefore,

$$\mathbf{E}[\widehat{S}_{(\boldsymbol{a}_1\boldsymbol{b}_1)}\widehat{S}_{(\boldsymbol{a}_2\boldsymbol{b}_2)}] = \mathbf{H}^T_{(\boldsymbol{a}_1\boldsymbol{b}_1)}(c_1\mathbf{I}_{2^{d_1+d_2}} + c_2\mathbf{1}\mathbf{1}^T)\mathbf{H}_{(\boldsymbol{a}_2\boldsymbol{b}_2)} = 0. \tag{4.12}$$

The last equation is because of the orthogonality of $\mathbf{H}$:

$$\mathbf{H}^T_{(\boldsymbol{a}_1\boldsymbol{b}_1)}\mathbf{H}_{(\boldsymbol{a}_2\boldsymbol{b}_2)} = \mathbf{H}^T_{(\boldsymbol{a}_1\boldsymbol{b}_1)}\mathbf{1} = \mathbf{1}^T\mathbf{H}_{(\boldsymbol{a}_2\boldsymbol{b}_2)} = 0. \tag{4.13}$$

(c) Note that the expected count in each cell is $\frac{n}{2^{d_1+d_2}}$. Therefore, by Theorem 3.4, we have

$$\begin{aligned}
C =& \frac{2^{d_1+d_2}}{n}\sum_{\boldsymbol{a},\boldsymbol{b}}\big(\widehat{N}_{(\boldsymbol{ab})} - \frac{n}{2^{d_1+d_2}}\big)^2 \\
=& \frac{2^{d_1+d_2}}{n}\big(\widehat{\boldsymbol{N}} - \frac{n}{2^{d_1+d_2}}\mathbf{1}\big)^T\big(\widehat{\boldsymbol{N}} - \frac{n}{2^{d_1+d_2}}\mathbf{1}\big) \\
=& \frac{2^{d_1+d_2}}{n}\big(\frac{1}{2^{d_1+d_2}}\mathbf{H}\widehat{\boldsymbol{S}} - \frac{n}{2^{d_1+d_2}}\mathbf{1}\big)^T\big(\frac{1}{2^{d_1+d_2}}\mathbf{H}\widehat{\boldsymbol{S}} - \frac{n}{2^{d_1+d_2}}\mathbf{1}\big) \\
=& \frac{2^{d_1+d_2}}{n}\big(\frac{1}{2^{d_1+d_2}}\sum_{\boldsymbol{a},\boldsymbol{b}}\widehat{S}^2_{(\boldsymbol{ab})} - \frac{n^2}{2^{d_1+d_2}}\big) \\
=& \frac{1}{n}\sum_{\boldsymbol{a}\neq\mathbf{0},\boldsymbol{b}\neq\mathbf{0}}\widehat{S}^2_{(\boldsymbol{ab})}.
\end{aligned} \tag{4.14}$$

The last equation comes from the fact that $\widehat{S}_{(\boldsymbol{a}\mathbf{0})} = \widehat{S}_{(\mathbf{0}\boldsymbol{b})} = 0$ except $\widehat{S}_{(\mathbf{00})} = n$.

$\square$

## 4.4 Proof of Theorem 4.4

*Proof.*

1. Note that the joint distribution of $(U_{d_1}, V_{d_2})$, $\mathbf{P}_{(U_{d_1},V_{d_2})}$, can be represented by the vector of cell probabilities $\boldsymbol{p}$ as in Theorem 3.4. Note also that the (independent) bivariate uniform distribution $\mathbf{P}_{0,d_1,d_2}$ can be represented as the vector $\boldsymbol{p}_{0,d_1,d_2} = \frac{1}{2^{d_1+d_2}}\mathbf{1}_{2^{d_1+d_2}}$. Denote the difference of these two vectors by $\boldsymbol{q} = \boldsymbol{p} - \frac{1}{2^{d_1+d_2}}\mathbf{1}_{2^{d_1+d_2}}$. Then the total variation distance between $\mathbf{P}_{(U_{d_1},V_{d_2})}$ and $\mathbf{P}_{0,d_1,d_2}$ can be written as the $\ell_1$ norm of $\boldsymbol{q}$:

$$\|\boldsymbol{q}\|_1 = \sum_{\boldsymbol{ab}} \left|p_{(\boldsymbol{ab})} - \frac{1}{2^{d_1+d_2}}\right| = 2TV(\mathbf{P}_{(U_{d_1},V_{d_2})}, \mathbf{P}_{0,d_1,d_2}) \geq 2\delta. \tag{4.15}$$

   We shall show next that (4.15) implies that for the expectation of some cross interaction is greater than $\frac{2\sqrt{d_1+d_2}\delta}{2^{(d_1+d_2)/4}}$, in the proof of which we utilize again Theorem 3.4. To see this, note that

$$\mathbf{H}_{2^{d_1+d_2}}\boldsymbol{q} = \mathbf{H}_{2^{d_1+d_2}}\boldsymbol{p} - \frac{1}{2^{d_1+d_2}}\mathbf{H}_{2^{d_1+d_2}}\mathbf{1}_{2^{d_1+d_2}} = \boldsymbol{E} - \boldsymbol{e}_{(\mathbf{00})}. \tag{4.16}$$

   Note that $E_{(\mathbf{00})} = 1$ and $E_{(\boldsymbol{a}\mathbf{0})} = E_{(\mathbf{0}\boldsymbol{b})} = 0$ for any $\boldsymbol{a} \neq \mathbf{0}$ and $\boldsymbol{b} \neq \mathbf{0}$. Therefore, all entries in $\mathbf{H}_{2^{d_1+d_2}}\boldsymbol{q}$ not corresponding to cross interactions are 0's. Now the maximal absolute entry in $\mathbf{H}_{2^{d_1+d_2}}\boldsymbol{q}$ corresponds to the maximal absolute expectation among cross interactions:

$$\|\mathbf{H}_{2^{d_1+d_2}}\boldsymbol{q}\|_\infty = \max_{\boldsymbol{a}\neq\mathbf{0},\boldsymbol{b}\neq\mathbf{0}} |E_{(\boldsymbol{ab})}|. \tag{4.17}$$

Now note that by the condition $\|\boldsymbol{E}-\boldsymbol{e}_{(\mathbf{00})}\|_\infty \geq \sqrt{d_1+d_2}2^{-(d_1+d_2)/4}\|\boldsymbol{E}-\boldsymbol{e}_{(\mathbf{00})}\|_2$,

$$\begin{aligned}
&\max_{\boldsymbol{a}\neq\mathbf{0},\boldsymbol{b}\neq\mathbf{0}} |E_{(\boldsymbol{ab})}| \\
=&\|\mathbf{H}_{2^{d_1+d_2}}\boldsymbol{q}\|_\infty \\
\geq&\sqrt{d_1+d_2}2^{-(d_1+d_2)/4}\|\mathbf{H}_{2^{d_1+d_2}}\boldsymbol{q}\|_2 \\
=&\sqrt{d_1+d_2}2^{(d_1+d_2)/4}\|\boldsymbol{q}\|_2 \\
\geq&\sqrt{d_1+d_2}2^{-(d_1+d_2)/4}\|\boldsymbol{q}\|_1 \\
\geq&\frac{2\sqrt{d_1+d_2}\delta}{2^{(d_1+d_2)/4}}.
\end{aligned} \tag{4.18}$$

By (4.18), for some $(\boldsymbol{ab})$, the cross interaction $\dot{A}_{\boldsymbol{a}}\dot{B}_{\boldsymbol{b}}$ has an expectation that is at least $\frac{2\sqrt{d_1+d_2}\delta}{2^{(d_1+d_2)/4}}$ away from 0.

Now for $\alpha>0$, consider the following test for $(S_{(\boldsymbol{ab})}+n)/2 \sim Binomial(n,q)$:

$$H_0: q=\frac{1}{2} \quad v.s. \quad H_1: |q-1/2| \geq \frac{2\sqrt{d_1+d_2}\delta}{2^{(d_1+d_2)/4+1}}. \tag{4.19}$$

For the test statistic $\frac{1}{\sqrt{n}}S_{(\boldsymbol{ab})}$ and the level $\frac{\alpha}{(2^{d_1}-1)(2^{d_2}-1)}$, by Hoeffding's inequality under $H_0$, the critical value $c_\alpha$ is such that $c_\alpha \asymp \sqrt{d_1+d_2}$. Note that under $H_1$, $\mathbf{E}\big[\big|\frac{1}{\sqrt{n}}S_{(\boldsymbol{ab})}\big|\big] \geq \sqrt{n}\frac{2\sqrt{d_1+d_2}\delta}{2^{(d_1+d_2)/4}}$. Since the variance of $\frac{1}{\sqrt{n}}S_{(\boldsymbol{ab})}$ is bounded, the proof follows by comparing the magnitude of $\mathbf{E}\big[\big|\frac{1}{\sqrt{n}}S_{(\boldsymbol{ab})}\big|\big]$ and $c_\alpha$.

2. The proof follows from the general method in Le Cam (2012) and is similar to the approach in Paninski (2008). For $0<\delta<1/2$, we consider a collection of alternative distributions $\mathcal{P}$ in $H_{1,d_1,d_2}$ such that

$$\mathcal{P}=\{\mathbf{P}_{(\boldsymbol{ab}),c}: \boldsymbol{E}_{\mathbf{P}_{(\boldsymbol{ab})}}=\boldsymbol{e}_{(\mathbf{00})}+2c\delta\boldsymbol{e}_{(\boldsymbol{ab})}, c=\pm1\}. \tag{4.20}$$

By Theorem 3.4, for each distribution in $\mathcal{P}$, the vector of probability $\boldsymbol{p}_{(\boldsymbol{ab}),c}=(p_{(\boldsymbol{a}'\boldsymbol{b}')})$ of $\mathbf{P}_{(\boldsymbol{ab}),c}$ satisfies that

$$p_{(\boldsymbol{a}'\boldsymbol{b}')}=\begin{cases} \frac{1}{2^{d_1+d_2}}(1+2\delta) & \text{if } (-1)^{\overline{(\boldsymbol{a}'\boldsymbol{b}')}^T(\boldsymbol{ab})}=c \\ \frac{1}{2^{d_1+d_2}}(1-2\delta) & \text{if } (-1)^{\overline{(\boldsymbol{a}'\boldsymbol{b}')}^T(\boldsymbol{ab})}=-c \end{cases} \tag{4.21}$$

Therefore, $TV(\mathbf{P}_{(\boldsymbol{ab}),c}, \mathbf{P}_{0,d_1,d_2}) = \delta$. Moreover, $\|\boldsymbol{E} - \boldsymbol{e}_{(\mathbf{00})}\|_\infty = \|\boldsymbol{E} - \boldsymbol{e}_{(\mathbf{00})}\|_2 = 2\delta$ so the conditions in the theorem are satisfied.

Let $m = (2^{d_1} - 1)(2^{d_2} - 1)$ for simplicity in notation. We shall show that no test can be consistent for the $2m$ distributions in $\mathcal{P}$ if $n = o(\frac{\sqrt{m}}{\delta^2})$. To this end, consider the uniform measure $\mu_{\mathcal{P}}$ over $\mathcal{P}$. Consider the likelihood ratio between $\mu_{\mathcal{P}}$ and $\mathbf{P}_{0,d_1,d_2}$ based on $n$ samples,

$$L_{\mu_{\mathcal{P}}} = \frac{(1-4\delta^2)^{\frac{n}{2}}}{2m} \sum_{\substack{\boldsymbol{a}\neq\mathbf{0}\\ \boldsymbol{b}\neq\mathbf{0}}} \left( \left(\frac{1+2\delta}{1-2\delta}\right)^{\frac{s_{(\boldsymbol{ab})}}{2}} + \left(\frac{1-2\delta}{1+2\delta}\right)^{\frac{s_{(\boldsymbol{ab})}}{2}} \right). \tag{4.22}$$

By Le Cam (2012), it suffices to show that (see for example the proof of Theorem 5 in Han et al. (2017)) if $n = o(\frac{\sqrt{m}}{\delta^2})$,

$$\mathbf{E}_{\mathbf{P}_{0,d_1,d_2}}[L^2_{\mu_{\mathcal{P}}}] = 1 + o(1). \tag{4.23}$$

To see this, we utilize the pairwise independence of $S_{(\boldsymbol{a}_1\boldsymbol{b}_2)}$ and $S_{(\boldsymbol{a}_2\boldsymbol{b}_2)}$ for $(\boldsymbol{a}_1\boldsymbol{b}_1) \neq (\boldsymbol{a}_2\boldsymbol{b}_2)$ as in Theorem 4.3. Let $S$ denote the sum of $n$ independent Rademacher variables, and note that $\mathbf{E}[e^{tS}] = \frac{1}{2^n}(e^{-t} + e^t)^n$. We have

$$\begin{aligned}
&\mathbf{E}_{\mathbf{P}_{0,d_1,d_2}}[L^2_{\mu_{\mathcal{P}}}] \\
=&\frac{(1-4\delta^2)^n}{4m^2}\left( m\left(\mathbf{E}\left[\left(\frac{1+2\delta}{1-2\delta}\right)^S\right] + \mathbf{E}\left[\left(\frac{1-2\delta}{1+2\delta}\right)^S\right] + 2\right) + \right. \\
&\left. m(m-1)\left(\mathbf{E}\left[\left(\frac{1+2\delta}{1-2\delta}\right)^{\frac{S}{2}}\right] + \mathbf{E}\left[\left(\frac{1-2\delta}{1+2\delta}\right)^{\frac{S}{2}}\right]\right)^2\right) \\
=&\frac{(1-4\delta^2)^n}{4m^2}\left(2m\left(\left(\frac{1+4\delta^2}{1-4\delta^2}\right)^n + 1\right) + m(m-1)\left(\frac{2}{(1-4\delta^2)^{\frac{n}{2}}}\right)^2\right) \\
=&1 + \frac{1}{2m}[(1+4\delta^2)^n + (1-4\delta^2)^n - 2] \\
=&1 + O\left(\frac{n^2\delta^4}{m}\right).
\end{aligned} \tag{4.24}$$

□

# References


Cornfield, J. (1956). A statistical problem arising from retrospective studies. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, Volume 4, pp. 135–148. University of California Press Berkeley.

Han, F., S. Chen, and H. Liu (2017). Distribution-free tests of independence in high dimensions. *Biometrika 104*(4), 813–828.

Le Cam, L. (2012). *Asymptotic methods in statistical decision theory*. Springer Science & Business Media.

Paninski, L. (2008). A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory 54*(10), 4750–4755.