Distributions of a general reduced-order dependence measure and conditional independence testing: supplemental material

Mariusz Kubkowski[0000−0002−1453−5589], Małgorzata Łazęcka[1,2][0000−0003−0975−4274], and Jan Mielniczuk[1,2][0000−0003−2621−2303]

[1] Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland
[2] Warsaw University of Technology, Faculty of Mathematics and Information Sciences, Warsaw, Poland
m.kubkowski,miel,malgorzata.lazecka@ipipan.waw.pl

1 Supplemental material

1.1 Proofs of an auxiliary lemma, Theorems and and Corollary 2

Theorem 1. *(i) We have*

$$n^{1/2}(\hat{J}^{\beta,\gamma}(X,Y|Z) - J^{\beta,\gamma}(X,Y|Z)) \xrightarrow{d} N(0,\sigma^2_{\hat{J}}), \tag{1}$$

where $\sigma^2_{\hat{J}} = Df(p)^T \Sigma Df(p) = \mathrm{Var}(Df(p)^T\hat{p})$ *and* $\Sigma = n\mathrm{Var}(\hat{p}-p)$.
(ii) If $\sigma^2_{\hat{J}} = 0$ *then*

$$2n(\hat{J}^{\beta,\gamma}(X,Y|Z) - J^{\beta,\gamma}(X,Y|Z)) \xrightarrow{d} V^T HV, \tag{2}$$

where V *follows* $N(0,\Sigma)$ *distribution,* $\Sigma^{x'y'z'}_{xyz} = p(x',y',z')(I(x=x',y=y',z=z') - p(x,y,z))/n$ *and* $H = D^2 f(p)$ *is a Hessian of* f.

Proof. Note that $f(p) = J^{\beta,\gamma}(X,Y|Z)$ equals

$$I(X,Y) - \sum_{s\in S}(\beta I(X,Z_s) - \gamma I(X,Z_s|Y)) = \sum_{x,y,z} p(x,y,z)\Bigg(\ln\left(\frac{p(x,y)}{p(x)p(y)}\right) - \sum_{s\in S}\left(\beta\ln\left(\frac{p(x,z_s)}{p(x)p(z_s)}\right) - \gamma\ln\left(\frac{p(x,y,z_s)p(y)}{p(x,y)p(y,z_s)}\right)\right)\Bigg).$$

Thus we have for $z=(z_1,\ldots,z_{|s|})$ that $\frac{\partial f(p)}{\partial p(x,y,z)}$ equals

$$\ln\left(\frac{p(x,y)}{p(x)p(y)}\right) - \beta\sum_{s\in S}\left(\ln\left(\frac{p(x,z_s)}{p(x)p(z_s)}\right) - 1\right) + \gamma\sum_{s\in S}\ln\left(\frac{p(x,y,z_s)p(y)}{p(x,y)p(y,z_s)}\right) - 1. \tag{3}$$

Let $\hat{p}(x,y,z) = n(x,y,z)/n$, $\hat{p} = (\hat{p}(x,y,z))_{x,y,z}$. Then $\hat{J}^{\beta,\gamma}(X,Y|Z) = f(\hat{p})$. In view of Taylor's formula we have $f(\hat{p}) - f(p)$ equals

$$Df(p)^T(\hat{p}-p) + \frac{1}{2}(\hat{p}-p)^T D^2 f(p)(\hat{p}-p) + O(||\hat{p}-p||_2^3). \tag{4}$$

From CLT we obtain $\sqrt{n}\Sigma^{-\frac{1}{2}}(\hat{p}-p) \xrightarrow{d} N(0,I)$. Hence:

$$\begin{aligned}\sqrt{n}(f(\hat{p}) - f(p)) &= (Df(p)^T \Sigma^{\frac{1}{2}}(\sqrt{n}\Sigma^{-\frac{1}{2}}(\hat{p}-p)) \\ &+ \frac{1}{2\sqrt{n}}(\sqrt{n}(\hat{p}-p)\Sigma^{-\frac{1}{2}})^T \Sigma^{\frac{1}{2}} D^2 f(p) \Sigma^{\frac{1}{2}}(\sqrt{n}\Sigma^{-\frac{1}{2}}(\hat{p}-p)) \\ &+ O(\sqrt{n}||\hat{p}-p||_2^3) \xrightarrow{d} N(0, Df(p)^T \Sigma Df(p)) = N(0, \sigma_{\hat{J}}^2),\end{aligned}$$

as $\sqrt{n}||\hat{p}-p||_2 = O(1)$ in probability. Asymptotic variance of $\sqrt{n}(\hat{J}^{\beta,\gamma}(X,Y|Z) - J^{\beta,\gamma}(X,Y|Z))$ is thus equal to:

$$\sigma_{\hat{J}}^2 = Df(p)^T \Sigma Df(p) = \mathrm{Var}(Df(p)^T \hat{p}). \tag{5}$$

Moreover $\sigma_{\hat{J}}^2 = 0$ if and only if $Df(p)^T \hat{p} = C$ a.s. for some $C \in R$. This implies that for all x, y, z we have:

$$\frac{\partial f(p)}{\partial p(x,y,z)} = C, \tag{6}$$

because for all x, y, z $P(\hat{p}(x,y,z) = 1) = (p(x,y,z))^n > 0$ and thus under this event we obtain the desired equality.

This implies that $Df(p)^T(\hat{p}-p) = C - C = 0$ and $2n(f(\hat{p}) - f(p))$ equals:

$$\begin{aligned}&(\sqrt{n}\Sigma^{-\frac{1}{2}}(\hat{p}-p))^T \Sigma^{\frac{1}{2}} D^2 f(p) \Sigma^{\frac{1}{2}}(\sqrt{n}\Sigma^{-\frac{1}{2}}(\hat{p}-p)) \\ &\quad + O(2n||\hat{p}-p||_2^3) \xrightarrow{d} U^T \Sigma^{\frac{1}{2}} D^2 f(p) \Sigma^{\frac{1}{2}} U.\end{aligned}$$

Taking $V = \Sigma^{\frac{1}{2}} U$ and $H = D^2 f(p)$ ends the proof.

Before we prove Theorem 2 we state and prove instrumental Lemma. It is easy to see that if $p(x,y)/p(x)p(y) \equiv C$ then $C = 1$ and X and Y are independent. The same is not true for conditional independence. Lemma 1 which is used in theorem below specifies two cases when this holds.

Lemma 1. *Let $Y \in \{0,1\}$ be a binary random variable and $X, Z \in \mathbb{N}_+$ be discrete variables. If for all $y \in \{0,1\}$ and $x, z \in \mathbb{N}_+$ we have:*

$$\frac{\mathrm{P}(X = x, Y = y | Z = z)}{\mathrm{P}(X = x | Z = z)\,\mathrm{P}(Y = y | Z = z)} = a_{xy}, \tag{7}$$

where $a_{xy} > 0$ does not depend on z, then at least one of the following possibilities holds:

1. Y and Z are independent and Y and Z are conditionally independent given X, for all x, y:

$$a_{xy} = \frac{\mathrm{P}(X = x, Y = y)}{\mathrm{P}(X = x)\,\mathrm{P}(Y = y)},$$

where $a_{xy} \neq 1$ for some x, y (hence X and Y are not independent).

2. X and Y are conditionally independent given Z and $a_{xy} = 1$ for all x, y.

Conversely, if either of the above conditions is true then (7) holds.

Proof. First we observe that for all $x, z \in \mathbb{N}_+$ we have:

$$\begin{aligned}
&\sum_{y=0}^{1} a_{xy}\,\mathrm{P}(Y = y, Z = z) \\
&= \mathrm{P}(Z = z) \sum_{y=0}^{1} a_{xy}\,\mathrm{P}(Y = y | Z = z) \\
&= \mathrm{P}(Z = z) \sum_{y=0}^{1} \frac{\mathrm{P}(X = x, Y = y | Z = z)}{\mathrm{P}(X = x | Z = z)} = \mathrm{P}(Z = z).
\end{aligned} \tag{8}$$

This means that for all x we have:

$$\begin{aligned}
\sum_{y=0}^{1} a_{xy}\,\mathrm{P}(Y = y) &= \sum_{z\in\mathbb{N}_+} \sum_{y=0}^{1} a_{xy}\,\mathrm{P}(Y = y, Z = z) \\
&= \sum_{z\in\mathbb{N}_+} \mathrm{P}(Z = z) = 1.
\end{aligned} \tag{9}$$

Hence:

$$a_{x1} = \frac{1 - a_{x0}\,\mathrm{P}(Y = 0)}{\mathrm{P}(Y = 1)}. \tag{10}$$

From (8) it follows that for all x we have:

$$\begin{cases}
\mathrm{P}(Z = z) = \mathrm{P}(Y = 0, Z = z)a_{x0} + \mathrm{P}(Y = 1, Z = z)a_{x1}, \\
\mathrm{P}(Z = z) = \mathrm{P}(Y = 0, Z = z) + \mathrm{P}(Y = 1, Z = z).
\end{cases} \tag{11}$$

Subtracting second equation from the first and using (10) yields:

$$\begin{aligned}
0 &= \mathrm{P}(Y = 0, Z = z)(a_{x0} - 1) \\
&\quad + \mathrm{P}(Y = 1, Z = z)\left(\frac{1 - a_{x0}\,\mathrm{P}(Y = 0)}{\mathrm{P}(Y = 1)} - 1\right) \\
&= \mathrm{P}(Y = 0, Z = z)(a_{x0} - 1) \\
&\quad + \mathrm{P}(Y = 1, Z = z)(1 - a_{x0})\frac{\mathrm{P}(Y = 0)}{\mathrm{P}(Y = 1)}.
\end{aligned}$$

We have two cases:

1) If $a_{x0} \neq 1$ for some x (note that $a_{x0} = 1$ is equivalent to $a_{x1} = 1$ in view of (10)), then the above equation reduces to:

$$\mathrm{P}(Y = 0, Z = z) = \mathrm{P}(Y = 1, Z = z)\frac{\mathrm{P}(Y = 0)}{\mathrm{P}(Y = 1)}. \tag{12}$$

This yields:

$$\begin{aligned}\mathrm{P}(Z = z) &= \mathrm{P}(Y = 0, Z = z) + \mathrm{P}(Y = 1, Z = z)\\ &= \mathrm{P}(Y = 1, Z = z)\left(1 + \frac{\mathrm{P}(Y = 0)}{\mathrm{P}(Y = 1)}\right)\\ &= \frac{\mathrm{P}(Y = 1, Z = z)}{\mathrm{P}(Y = 1)}.\end{aligned}$$

Analogously, we obtain:

$$\mathrm{P}(Z = z) = \frac{\mathrm{P}(Y = 0, Z = z)}{\mathrm{P}(Y = 0)}. \tag{13}$$

Thus Y and Z are independent. This means that $\mathrm{P}(Y = y, Z = z) = \mathrm{P}(Y = y)\,\mathrm{P}(Z = z)$. Inserting this equation into (7) yields:

$$a_{xy} = \frac{\mathrm{P}(X = x, Y = y, Z = z)}{\mathrm{P}(X = x, Z = z)\,\mathrm{P}(Y = y)}. \tag{14}$$

Equivalently,

$$a_{xy}\,\mathrm{P}(X = x, Z = z) = \frac{\mathrm{P}(X = x, Y = y, Z = z)}{\mathrm{P}(Y = y)}.$$

Hence:

$$\begin{aligned}a_{xy}\,\mathrm{P}(X = x) &= \sum_z a_{xy}\,\mathrm{P}(X = x, Z = z)\\ &= \sum_z \tfrac{\mathrm{P}(X=x,Y=y,Z=z)}{\mathrm{P}(Y=y)} = \tfrac{\mathrm{P}(X=x,Y=y)}{\mathrm{P}(Y=y)}.\end{aligned}$$

It follows that:

$$a_{xy} = \frac{\mathrm{P}(X = x, Y = y)}{\mathrm{P}(X = x)\,\mathrm{P}(Y = y)}.$$

Thus, inserting this into (14), we obtain:

$$\frac{\mathrm{P}(X = x, Y = y, Z = z)}{\mathrm{P}(X = x, Z = z)\,\mathrm{P}(Y = y)} = \frac{\mathrm{P}(X = x, Y = y)}{\mathrm{P}(X = x)\,\mathrm{P}(Y = y)},$$

what is equivalent to conditional independence of Y and Z given X.

2) If $a_{x0} = 1$ for all x, then in view of (10) we obtain $a_{x1} = 1$ for all x. This implies conditional independence of (X, Y) given Z. To see the converse note that a_{xy} in (7) equals 1 when 2) is true and $a_{xy} = p(x, y)/(p(x)p(y))$ when 1) holds.

Now we state and prove Theorem 2. Recall that W is defined as

$$W = \Big\{ s \in S : \exists_{x,y,z_s} \frac{p(x,y,z_s)p(z_s)}{p(x,z_s)p(y,z_s)} \neq 1 \Big\}. \tag{15}$$

Theorem 2. *Assume that $\sigma^2_{\hat{J}} = 0$ and $\beta = \gamma \neq 0$. Then we have:*
(i) If $|S| > 1$ and $\beta^{-1} \in \{1, 2, \ldots, |S|-1\}$ then one of the above scenarios holds with W defined in 15.
(ii) If $\beta^{-1} = |S|$ or $\beta^{-1} \notin \{1, 2, \ldots, |S|-1\}$ then Scenario 1 is valid.

Proof. As $\sigma^2_{\hat{J}} = 0$, then in view of (6) and (3) we obtain:

$$\begin{aligned} \ln\left(\tfrac{p(x,y)}{p(x)p(y)}\right) - 1 - \beta \sum_{s\in S} \left(\ln\left(\tfrac{p(x,z_s)}{p(x)p(z_s)}\right) - 1\right) \\ +\gamma \sum_{s\in S} \ln\left(\tfrac{p(x,y,z_s)p(y)}{p(x,y)p(y,z_s)}\right) \equiv C. \end{aligned} \tag{16}$$

Let $s_0 \in S$. We observe that:

$$\begin{aligned} &\beta \left(\ln\left(\frac{p(x,z_{s_0})}{p(x)p(z_{s_0})}\right) - 1 \right) - \gamma \ln\left(\frac{p(x,y,z_{s_0})p(y)}{p(x,y)p(y,z_{s_0})}\right) \\ &\quad \equiv \ln\left(\frac{p(x,y)}{p(x)p(y)}\right) - \beta \sum_{s\in S\setminus\{s_0\}} \left(\ln\left(\frac{p(x,z_s)}{p(x)p(z_s)}\right) - 1\right) \\ &\qquad + \gamma \sum_{s\in S\setminus\{s_0\}} \ln\left(\frac{p(x,y,z_s)p(y)}{p(x,y)p(y,z_s)}\right) - 1 - C. \end{aligned} \tag{17}$$

Thus the left side of above equality does not depend on z_{s_0}, as the right side does not depend on it. Thus we have for all x, y, z_{s_0}:

$$\beta \left(\ln\left(\frac{p(x,z_{s_0})}{p(x)p(z_{s_0})}\right) - 1 \right) - \gamma \ln\left(\frac{p(x,y,z_{s_0})p(y)}{p(x,y)p(y,z_{s_0})}\right) := a_{xy}. \tag{18}$$

Rearranging terms, we obtain:

$$\begin{aligned} \beta \ln\left(\tfrac{p(x,z_{s_0})}{p(z_{s_0})}\right) - \gamma \ln\left(\tfrac{p(x,y,z_{s_0})}{p(y,z_{s_0})}\right) = a_{xy} + \beta \ln(p(x)) \\ +\beta + \gamma \ln(p(y)) - \gamma \ln(p(x,y)) := b_{xy}. \end{aligned} \tag{19}$$

For $\gamma = \beta \neq 0$ equation (24) takes the form:

$$\ln\left(\frac{p(x,z_{s_0})p(y,z_{s_0})}{p(z_{s_0})p(x,y,z_{s_0})}\right) = \frac{b_{xy}}{\beta} := c_{xy}. \tag{20}$$

If $s_0 \in W$, then in view of Lemma 1

$$c_{xy} = \ln\left(\frac{p(x)p(y)}{p(x,y)}\right). \tag{21}$$

If $s_0 \in W^c$, then $c_{xy} = 0$, and (X, Y) are conditionally independent given Z_{s_0}. Thus (27) can be written as:

$$\begin{aligned} &\ln(\tfrac{p(x,y)}{p(x)p(y)}) - 1 + \beta|S| \\ &+\beta \sum_{s\in S} \ln\left\{\tfrac{p(x,y,z_s)p(x)p(y)p(z_s)}{p(x,y)p(y,z_s)p(x,z_s)}\right\} \\ &= (1-\beta|W^c|)\ln\left(\tfrac{p(x,y)}{p(x)p(y)}\right) - 1 + \beta|S| = C. \end{aligned} \tag{22}$$

We have two cases:
a) If $1-\beta|S|+\beta|W| \neq 0$, then $|W| \neq |S| - \beta^{-1}$. In this case from (22) it follows that (X, Y) are independent. This in view of definition of W and (21) we have that $W = \emptyset$ and $|S| = \beta^{-1}$. But if $|S| = \beta^{-1}$, and $|W| > 0$ we obtain from (22) that X and Y are independent and thus $W = \emptyset$. This proves (ii).
b) If $1-\beta|S|+\beta|W| = 0$, then $|W| = |S| - \beta^{-1}$. The case $\beta^{-1} \notin \{1, 2, \ldots, |S|\}$ is excluded as then $|W| \notin \mathbb{N}$ or $|W| < 0$ or $|W| > |S|$. If $|S| > 1$ and $\beta^{-1} \in \{1, 2, \ldots, |S| - 1\}$, then we obtain (i) from the Lemma 1 and above remarks. Note that it follows from the proof that W is a proper subset of S.

Theorem 3. *If X, Y, Z are discrete, $\sigma^2_{\hat{J}} = 0$, then:*
1) If $\gamma = 0$ and $\beta = 0$, then X and Y are independent.
2) If $\beta = 0$ and $\gamma \neq 0$, then for all $s \in S$ (X, Z_s) are independent and (X, Y) are independent.
3) If $\gamma = 0$ and $\beta \neq 0$, then for all $s \in S$ (Y, Z_s) and X are independent.
Moreover if $\gamma = 0$ or $\beta = 0$, then $J^{\beta,\gamma}(X, Y|Z) = 0$.

Proof. 1) If $\gamma = 0$ and $\beta = 0$, then in view of (27) we have:

$$\ln\left(\frac{p(x,y)}{p(x)p(y)}\right) = C + 1. \tag{23}$$

From this equation independence of X and Y follows.
2) If $\gamma = 0$ and $\beta \neq 0$, then in view of

$$\begin{aligned} &\beta\ln\left(\tfrac{p(x,z_{s_0})}{p(z_{s_0})}\right) - \gamma\ln\left(\tfrac{p(x,y,z_{s_0})}{p(y,z_{s_0})}\right) = a_{xy} + \beta\ln(p(x)) \\ &\quad +\beta + \gamma\ln(p(y)) - \gamma\ln(p(x,y)) := b_{xy} \end{aligned} \tag{24}$$

(cf proof of Theorem 2) we have:

$$p(x, z_{s_0}) = p(z_{s_0})\exp\left(\frac{b_{xy}}{\beta}\right). \tag{25}$$

Summing both sides over z_{s_0} yields:

$$p(x) = \exp\left(\frac{b_{xy}}{\beta}\right). \tag{26}$$

Hence $p(x, z_{s_0}) = p(z_{s_0})p(x)$. This means that X and Z_{s_0} are independent for all $s_0 \in S$. Substituting this into

$$\ln\left(\tfrac{p(x,y)}{p(x)p(y)}\right) - 1 - \beta\sum_{s\in S}\left(\ln\left(\tfrac{p(x,z_s)}{p(x)p(z_s)}\right) - 1\right) +$$

$$\gamma \sum_{s\in S} \ln\left(\tfrac{p(x,y,z_s)p(y)}{p(x,y)p(y,z_s)}\right) \equiv C \tag{27}$$

(cf. proof of Theorem 2) gives:

$$\ln\left(\frac{p(x,y)}{p(x)p(y)}\right) = C + 1 + \beta|\mathbf{S}|. \tag{28}$$

Hence X and Y are independent.
3) If $\gamma \neq 0$ and $\beta = 0$, then in view of (24) we have:

$$p(x,y,z_{s_0}) = p(y,z_{s_0}) \exp\left(-\frac{b_{xy}}{\gamma}\right). \tag{29}$$

Summing both sides by z_{s_0} yields:

$$p(x,y) = p(y) \exp\left(-\frac{b_{xy}}{\gamma}\right). \tag{30}$$

Hence using two last equations we have:

$$p(x,y,z_{s_0}) = p(y,z_{s_0}) \frac{p(x,y)}{p(y)}. \tag{31}$$

This means that (X, Z_{s_0}) are conditionally independent given Y for all $s_0 \in S$. Substituting this into (27) gives:

$$\ln\left(\frac{p(x,y)}{p(x)p(y)}\right) = C + 1. \tag{32}$$

Hence again it follows easily that X and Y are independent. Thus equation (31) takes the form: $p(x,y,z_{s_0}) = p(y,z_{s_0})p(x)$. This means that (Y, Z_{s_0}) and X are independent for all $s_0 \in S$.

We now state the asymptotic result for CIFE which is analogous to that of JMI. Let

$$\sigma^2_{\widehat{CIFE}} = \sum_{x,y,z} p(x,y,z)\Bigg(\ln\Bigg[\left(\frac{p(x,y)}{p(x)p(y)}\right)^{1-|S|} \cdot \prod_{s\in S} \frac{p(x,y,z_s)p(z_s)}{p(x,z_s)p(y,z_s)}\Bigg]\Bigg)^2 - (CIFE)^2. \tag{33}$$

We have

Corollary 1. *Let Y be binary. (i) If $\sigma^2_{\widehat{CIFE}} \neq 0$ then*

$$n^{1/2}(\widehat{CIFE} - CIFE) \xrightarrow{d} N(0, \sigma^2_{\widehat{CIFE}}).$$

(ii) If $\sigma^2_{\widehat{CIFE}} = 0$ then $CIFE = 0$ and

$$2n\widehat{CIFE} \xrightarrow{d} V^T H V,$$

where V and H are defined in Theorem 1. Moreover in this case either Scenario 1 holds or Scenario 2 holds with W such that $|W| = 1$.

1.2 Analysis of real data sets

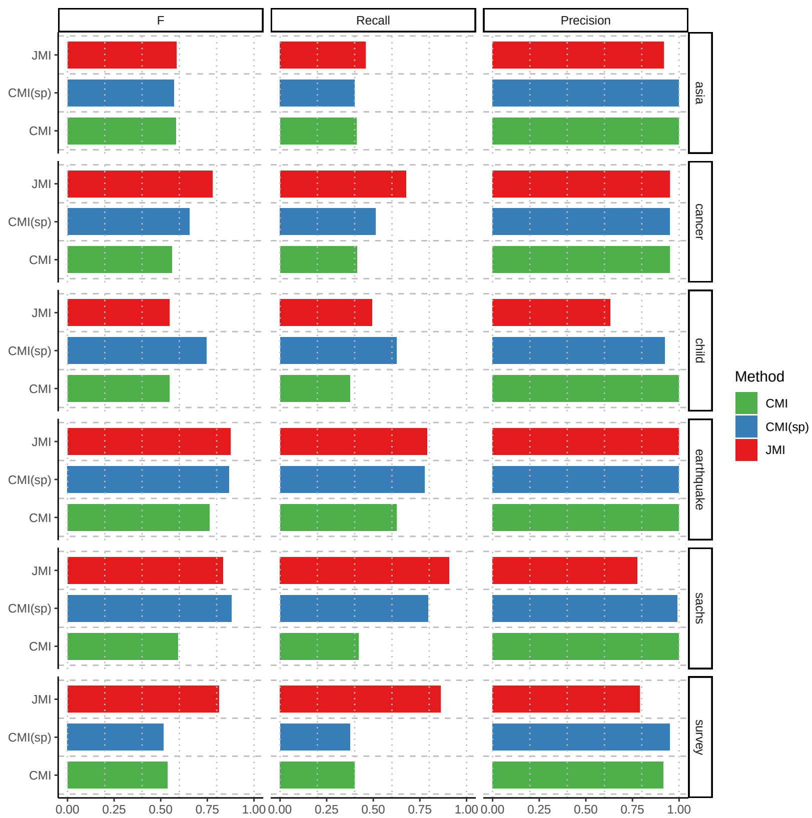

Fig. 1. F measure, precision and recall for GS algorithm using JMI, CMI and CMI(sp) tests