Send Orders for Reprints to reprints@benthamscience.ae

BENTHAM OPEN

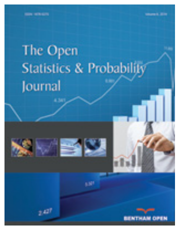
CrossMark

The Open Statistics & Probability Journal

Supplementary Material

Content list available at: www.benthamopen.com/TOSPJ/

DOI: 10.2174/1876527001607010020



Comparing Measures of Association in 2×2 Probability Tables

Dirk Hasenclever[1,*] and Markus Scholz[1,2]

[1]*Institute for Medical Informatics, Statistics and Epidemiology, Leipzig, Germany*
[2]*LIFE Research Center University of Leipzig, Leipzig, Germany*

In this section, we prove the propositions and theorems of our paper. In most situations it is sufficient to consider the case $x > 0$ since from symmetry conditions the case $x < 0$ follows analogously.

Proposition 6 (Margin weighting function for *r*): For *all* $x \in \mathbb{R} \setminus \{0\}$:

a) r_x *has exactly one extremum at the origin* $(y, z) = (0, 0)$, *corresponding to the diagonal symmetric table with the fixed odds-ratio*.

b) $\lim_{\|(y,z)\|\to\infty} r_x = 0$.

c) $\lim_{x\to\pm\infty} r_x = \pm 1$

d) r can be extended to $\bar{\mathbb{R}}^3$ *except for the lines* $(\pm,\pm, *)$ *and* $(\pm, *,\pm)$ *and the vertices V*.

e) r can be extended to $\bar{\mathbb{T}}$ *except for the vertices*.

Proof: a) We consider the maximum condition for r_x and $x > 0$:

$$\begin{aligned} r_x \quad \to \quad \max.! \quad &\Leftrightarrow \quad \frac{e^{y+z}}{\sqrt{\left(e^{x+y+z}+e^y\right)\left(e^{x+y+z}+e^z\right)\left(e^x+e^y\right)\left(e^x+e^z\right)}} \quad \to \quad \max.! \\ &\Leftrightarrow \quad \left(e^x+e^{-y}\right)\left(e^x+e^{-z}\right)\left(e^x+e^y\right)\left(e^x+e^z\right) \quad \to \quad \min.! \\ &\Leftrightarrow \quad \left(e^x+e^{-y}\right)\left(e^x+e^y\right) \quad \to \quad \min.! \quad \wedge \quad y=z \\ &\Leftrightarrow \quad y=z=0 \end{aligned}$$

b) and c) follow easily using equation (3). d) and e) are consequences of b) and c)

Proposition 7 (Margin weighting function for *D*′): For *all* $x \in \mathbb{R} \setminus \{0\}$:

a)D_x' *has a non-differentiable edge along the diagonal* $y = z$ *for* $D' > 0$ *and along the diagonal* $y = -z$ *for* $D' < 0$. *There is a non-smooth saddle point in the origin*.

b)

$$\begin{aligned} \lim_{y\to\pm\infty} D_x' \quad &= \quad \left(e^{2x}-1\right)\cdot\begin{cases} \left(e^{2x}+e^{x\pm z}\right)^{-1} & : \quad x>0 \\ \left(e^{x\mp z}+1\right)^{-1} & : \quad x<0 \end{cases} \\ \lim_{z\to\pm\infty} D_x' \quad &= \quad \left(e^{2x}-1\right)\cdot\begin{cases} \left(e^{2x}+e^{x\pm y}\right)^{-1} & : \quad x>0 \\ \left(e^{x\mp y}+1\right)^{-1} & : \quad x<0 \end{cases} \end{aligned}$$

Thus, limit functions have a range of $(0, 1 - e^{-2x})$ *for* $x > 0$ *and* $(e^{2x} - 1, 0)$ *for* $x < 0$, *where 0 is obtained for* $y \to \pm\infty$, $z \to \pm\infty$, $x > 0$ *and* $y \to \mp\infty$, $z \to \pm\infty$, $x < 0$.

c) $\lim_{x\to\pm\infty} D_x' = \pm 1$

d) D' *can be extended to* $\bar{\mathbb{R}}^3$ *except for the vertices* V_g.

e) D' *can be extended to* $\overline{\mathbb{T}}$ *except for the edges and vertices*.

Proof: a) Assume $x > 0$, consider the path $y = w + c$, $z = w - c$, w = const. Calculating the left-hand and right-hand derivative of D_{max} at $c = 0$ using equation (4) yields:

$$\lim_{c\to 0\pm 0} \frac{d}{dc} D_{max} = \mp e^w \left(e^x + 2e^w + e^{x+2w}\right)$$

Since the term in parentheses is positive, D_{max} has a wedge at $c = 0$. On the other hand, D_x' has a maximum at $y = z = 0$ along the path $y = z$ since

$$\begin{aligned} D_x'(y,y) \quad \to \quad \text{max.!} \quad &\Leftrightarrow \quad \frac{e^{2y}}{\left(e^{x+2y}+e^y\right)\left(e^x+e^y\right)} \quad \to \quad \text{max.!} \\ &\Leftrightarrow \quad \left(e^x + e^{-y}\right)\left(e^x+e^y\right) \quad \to \quad \text{min.!} \\ &\Leftrightarrow \quad y = 0 \end{aligned}$$

Hence, D' has a non-differentiable saddle point at $y = z = 0$. The case $x < 0$ follows analogously. The limit behaviour considered in b) to e) is easy to see using equation (4).

Proposition 8 (Margin weighting function for sMutInf): For all $x \in \mathbb{R} \setminus \{0\}$:

a) sMutInf_x *has exactly one maximum at the origin* $(y, z) = (0, 0)$.

b) $\lim_{\|(y,z)\|\to\infty} \text{sMutInf}_x = 0$.

c) $\lim_{x\to\pm\infty} \text{sMutInf}_x = \pm\left(\log_2\left(e^{y\pm z}+1\right) - \frac{e^{y\pm z}}{e^{y\pm z}+1}\log_2 e^{y\pm z}\right)$

Thus $\text{sMutInf}_x \to \pm 1$ *for* $y = \mp z$ *and* $x \to \pm\infty$ *respectively*.

d) sMutInf *can be extended to* $\bar{\mathbb{R}}^3$ *except for the vertices* V_b.

e) sMutInf *can be extended completely to* $\overline{\mathbb{T}}$.

Proof: a) We consider tables $t_\mu = \frac{1}{N_\mu}\begin{pmatrix} \mu p_{00} & p_{01} \\ p_{10} & p_{11}/\mu \end{pmatrix}$ and $t_\nu = \frac{1}{N_\nu}\begin{pmatrix} p_{00} & p_{01}/\nu \\ \nu p_{10} & p_{11} \end{pmatrix}$ of the same odds-ratio than t for $\mu, \nu > 0$ and N_μ and N_ν are the normalisation constants $N_\mu = \mu p_{00} + p_{01} + p_{10} + p_{11}/\mu$ and $N_\nu = p_{00} + \nu p_{01} + p_{10}/\nu + p_{11}$ respectively. We aim to proof that

$$\left.\frac{d}{d\mu}\text{sMutInf}_x\left(t_\mu\right)\right|_{\mu=1} = 0 \quad \Leftrightarrow \quad p_{00} = p_{11} \tag{A.1}$$

$$\left.\frac{d}{d\nu}\text{sMutInf}_x\left(t_\nu\right)\right|_{\nu=1} = 0 \quad \Leftrightarrow \quad p_{01} = p_{10} \tag{A.2}$$

Assuming $\lambda > 1$, $p_{00} \le p_{11}$ without restriction of generality, we obtain after some calculations:

$$\begin{aligned}\left.\frac{d}{d\mu}\mathrm{sMutInf}_x\left(t_\mu\right)\right|_{\mu=1} &= \left(p_{11}-p_{00}\right)\mathrm{sMutInf}_x\left(t\right)\\ &\quad + p_{00}\log_2\frac{p_{00}}{\left(p_{00}+p_{01}\right)\left(p_{00}+p_{10}\right)} - p_{11}\log_2\frac{p_{11}}{\left(p_{11}+p_{01}\right)\left(p_{11}+p_{10}\right)}\\ &= \left(p_{11}-p_{00}\right)\mathrm{sMutInf}_x\left(t\right)\\ &\quad + p_{00}p_{11}\left(\frac{\log_2\left(1+p_{00}\left(\frac{1}{\lambda}-1\right)\right)}{p_{00}} - \frac{\log_2\left(1+p_{11}\left(\frac{1}{\lambda}-1\right)\right)}{p_{11}}\right)\end{aligned} \quad \textbf{(A.3)}$$

The first term is non-negative and equals zero iff $p_{00} = p_{11}$.

Consider the monotonicity of the term $\frac{\log_2\left(1+x\left(\frac{1}{\lambda}-1\right)\right)}{x}$ for $x \in (0,1)$. It holds that

$$\frac{d}{dx}\frac{\log_2\left(1+x\left(\frac{1}{\lambda}-1\right)\right)}{x} = \frac{1}{x^2\ln 2}\left(-\ln z+\frac{z-1}{z}\right) \quad \textbf{(A.4)}$$

with $z = x\left(\frac{1}{\lambda}-1\right)+1 \in (0,1)$. In this interval, (A.4) is negative since $-\ln z+\frac{z-1}{z}$ is monotonically increasing for $z \in (0,1)$, taking its maximum for $z \to 1$. Hence $\frac{\log_2\left(1+x\left(\frac{1}{\lambda}-1\right)\right)}{x}$ is monotonically decreasing for $x \in (0,1)$. In conclusion, the second term of (A.3) is non-negative too and equals zero iff $p_{00} = p_{11}$. This proves (A.1). Analogously, using t_ν instead of t_μ proves (A.2).

b)-e) are obvious exploiting the continuity of the functions involved, i.e. taking the limit of the tables first.

Theorem 1 (magic odds-ratio): *Define the "magic odds-ratio" by $L_{magic} = W(1/e)^{-2} \approx 12.89$. Let $L > 1$. The entropy H restricted to the submanifold of constant odds-ratio L in $\mathbb{T}$*

- *has a single maximum at the diagonal table of odds-ratio L if $1 < L \le L_{magic}$.*
- *has a saddle point at the diagonal table of odds-ratio L and two "L-shaped" tables as maxima which transpose with matrix transposition if $L_{magic} < L$.*

"Lshaped" means that for $L \to \infty$ one of the maxima approaches the table $\begin{pmatrix}1/3 & 1/3\\ 0 & 1/3\end{pmatrix}$. For the case $L < 1$ a similar result can be derived by transposing principal and secondary diagonals.

Proof: The constraint odds-ratio = L can be written in the form:

$$\ln(p_{00}) - \ln(p_{01}) - \ln(p_{10}) + \ln(p_{11}) = \ln(L) \quad \textbf{(A.5)}$$

We assume $L > 1$ in the following without restriction of generality since the case $L < 1$ can be studied analogously. A second constraint is given by:

$$p_{00}+p_{01}+p_{10}+p_{11} = 1 \quad \textbf{(A.6)}$$

In order to study the critical points of H, we now consider the extremal value problem of H given the constraints (A.5) and (A.6). For this purpose, we introduce Lagrange multipliers Λ_1 and Λ_2 and determine the first variation of the following function:

$$\begin{aligned}f\left(t,\Lambda_1,\Lambda_2\right) &= -\left(p_{00}\cdot\ln(p_{00})+p_{01}\cdot\ln(p_{01})+p_{10}\cdot\ln(p_{10})+p_{11}\cdot\ln(p_{11})\right)+\\ &+\Lambda_1\cdot\left(\ln(p_{00})-\ln(p_{01})-\ln(p_{10})+\ln(p_{11})-\ln(L)\right)+\Lambda_2\cdot\left(p_{00}+p_{01}+p_{10}+p_{11}-1\right)\end{aligned} \quad \textbf{(A.7)}$$

Calculating the partial derivatives $\frac{\partial}{\partial p_{ij}}$ gives four equations:

$$\begin{aligned}\ln(p_{00})+1-\frac{\Lambda_1}{p_{00}}-\Lambda_2 &= 0\\ \ln(p_{11})+1-\frac{\Lambda_1}{p_{11}}-\Lambda_2 &= 0\\ \ln(p_{10})+1+\frac{\Lambda_1}{p_{10}}-\Lambda_2 &= 0\\ \ln(p_{01})+1+\frac{\Lambda_1}{p_{01}}-\Lambda_2 &= 0\end{aligned}$$

In order to solve this system explicitly, we recall that Lambert's W function is defined as the inverse function to $x\exp(x)$. Hence, it holds that

$$p_{ij}=\frac{\pm\Lambda_1}{W(\pm\Lambda_1\exp(1-\Lambda_2))}$$

where the upper sign holds for p_{00} and p_{11} and the lower sign for p_{01} and p_{10} respectively. At the first look it seems as that the only solution is the diagonal-symmetric table. But W is a multi-branch function since $y = x\exp(x)$ has two solutions for $y \in (-1/e,0)$. The two real-valued braches are traditionally called W_0 when $x \in (-1/e,0)$ and W_{-1} when $x \in (-\infty, -1/e)$. Note that $W_{-1}(-1/e) = W_0(-1/e) = -1$. Assume $\Lambda_1 > 0$. Inserting the solutions for p_{ij} in the condition on $\ln(L)$ we get three possible solutions.

a) $\ln(L) = \ln(\frac{W_0(-z)^2}{W(z)^2})$

This solution exists only for $L \in (1, W(1/e)^{-2}]$. $W(1/e)^{-2} \approx 12.89615....$

b) $\ln(L) = \ln(\frac{W_{-1}(-z)^2}{W(z)^2})$

This solution exists only for $L \in [W(1/e)^{-2}, \infty)$.

c) $\ln(L) = \ln(\frac{W_{-1}(-z)W_0(-z)}{W(z)^2})$

These solutions exist only for $L \in [W(1/e)^{-2}, \infty)$.

Hence, we have a single critical point for $L \in \left(1,\ W(1/e)^{-2}\right)$ but three critical points for $L \in \left(W(1/e)^{-2}, \infty\right)$. The next lemma characterises these critical points.

Lemma S1: (Characterisation of the critical points of H for given odds-ratio)

a) For $L \in (1, W(1/e)^{-2})$, *H has a maximum at the diagonal table of odds-ratio L.*

b) For $L \in (W(1/e)^{-2}, \infty)$, *H has a saddle-point at the diagonal table and two maxima at the other two critical points. If* $L \to \infty$ *these maxima tend to the tables* $\begin{pmatrix}1/3 & 1/3\\ 0 & 1/3\end{pmatrix}$ *and* $\begin{pmatrix}1/3 & 0\\ 1/3 & 1/3\end{pmatrix}$ *respectively.*

Proof: We study the following tables: $t_\mu = \frac{1}{N_\mu}\begin{pmatrix}\mu a & c\\ c & a/\mu\end{pmatrix}$ with $N_\mu = \mu a + \frac{a}{\mu} + 2c$ and $t_\nu = \frac{1}{N_\nu}\begin{pmatrix}a & \nu c\\ c/\nu & a\end{pmatrix}$ with $N_\nu = 2a + \mu c + {}^{c}/_{\nu}$, $a, c > 0$, $a + c = 1/2$. We calculate the second derivative of $H(t_\mu)$ and $H(t_\nu)$ at $\mu = 1$ and $\nu = 1$ respectively. After some calculations one obtains

$$\left.\frac{d^2H}{d\mu^2}(t_\mu)\right|_{\mu=1} = -\frac{1}{\ln 2}\frac{\sqrt{\lambda}}{1+\sqrt{\lambda}}\left(1+\frac{1}{1+\sqrt{\lambda}}\ln\sqrt{\lambda}\right) < 0$$

In contrast

$$\left.\frac{d^2H}{d\nu^2}(t_\nu)\right|_{\nu=1} = -\frac{1}{\ln 2}\frac{1}{1+\sqrt{\lambda}}\left(1-\frac{\sqrt{\lambda}}{1+\sqrt{\lambda}}\ln\sqrt{\lambda}\right)$$

Which is greater than 0 for $\lambda > W(1/e)^{-2}$. Thus $\left(\begin{smallmatrix} a & c \\ c & a \end{smallmatrix}\right)$ becames a saddle point for $\lambda > W(1/e)^{-2}$ but is a maximum for $\lambda \le W(1/e)^{-2}$. The other suppositions of the lemma and theorem 1 are then easy to see.

Lemma 1 (Monotony of the entropy difference): *Let H be the entropy of t and H_{diag} be the entropy of the corresponding diagonal table of the same odds-ratio λ. Then, H_{diag}–H is monotonically decreasing for increasing $\lambda > 1$ and constant margins.*

Proof: Let ε > 0 and $t_\varepsilon = \begin{pmatrix} p_{00}+\varepsilon & p_{01}-\varepsilon \\ p_{10}-\varepsilon & p_{11}+\varepsilon \end{pmatrix}$ a table with increased odds-ratio but same margins compared to t. We show that $\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} H_{diag}(t_\varepsilon) - H(t_\varepsilon) \le 0$, where equality holds iff t is diagonal. After some calculations we obtain

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} H_{diag}(t_\varepsilon) = -\frac{\sqrt{\lambda}\log_2\lambda}{4\left(1+\sqrt{\lambda}\right)^2}\sum_{i,j=0}^{1}\frac{1}{p_{ij}} \quad \textbf{(A.8)}$$

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} H(t_\varepsilon) = -\log_2\lambda \quad \textbf{(A.9)}$$

Thus

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \left(H_{diag}(t_\varepsilon) - H(t_\varepsilon)\right) = \frac{\log_2\lambda}{4}\left(4-\frac{\sqrt{\lambda}}{\left(1+\sqrt{\lambda}\right)^2}\sum_{i,j=0}^{1}\frac{1}{p_{ij}}\right) \quad \textbf{(A.10)}$$

Now consider the tables $t_\mu = \frac{1}{N_\mu}\begin{pmatrix} \mu p_{00} & p_{01} \\ p_{10} & p_{11}/\mu \end{pmatrix}$ and $t_\nu = \frac{1}{N_\nu}\begin{pmatrix} p_{00} & p_{01}/\nu \\ \nu p_{10} & p_{11} \end{pmatrix}$ of the same odds-ratio than t for $\mu, \nu \ge 1$ and the normalisation constants $N_\mu = \mu p_{00} + p_{01} + p_{10} + p_{11}/\mu$ and $N_\nu = p_{00} + \nu p_{01} + p_{10}/\nu + p_{11}$ respectively.

Assume $p_{00} \le p_{11}$ and $p_{01} \le p_{10}$ without restriction of generality, we see that for $f(t) = \sum_{i,j=0}^{1}\frac{1}{p_{ij}}$ it holds that

$$\left.\frac{d}{d\mu}\right|_{\mu=1} f(t_\mu) = (p_{00}-p_{11})\left(\sum_{i,j=0}^{1}\frac{1}{p_{ij}}+\frac{1}{p_{00}p_{11}}\right) \le 0$$

$$\left.\frac{d}{d\nu}\right|_{\nu=1} f(t_\nu) = (p_{10}-p_{01})\left(\sum_{i,j=0}^{1}\frac{1}{p_{ij}}+\frac{1}{p_{01}p_{10}}\right) \le 0$$

were equality holds iff t is diagonal. Hence the maximum of the term in parenthesis of (A.10) is obtained iff t is diagonal. On the other hand, for t diagonal it hold that

$$\frac{\sqrt{\lambda}}{\left(1+\sqrt{\lambda}\right)^2} \sum_{i,j=0}^{1} \frac{1}{p_{ij}} - 4 \quad = \quad 0 \tag{A.11}$$