Herleitung der ϕ_{max} Formel

Gegeben sei eine beliebige 2×2 Kontingenztabelle der folgenden Form

	x_1	x_2	Σ
y_1	n_{11}	n_{12}	$n_{1\bullet}$
y_2	n_{21}	n_{22}	$n_{2\bullet}$
Σ	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet\bullet}$

Wir haben gesehen, dass bei schiefen Randverteilungen der daraus berechnete ϕ Koeffizient nicht mehr den Wert 1 bzw. -1 erreichen kann. „Schief" bedeutet in Bezug auf die Randverteilungen dabei *nicht*, dass $n_{1\bullet} \neq n_{2\bullet}$ bzw. $n_{\bullet 1} \neq n_{\bullet 2}$, sondern dass $n_{1\bullet} \neq n_{\bullet 1}$ bzw. $n_{2\bullet} \neq n_{\bullet 2}$. Randverteilungen in einer 2×2 Kontingenztabelle sind also genau dann schief, wenn das Verhältnis von x_1 zu x_2 anders ist als das Verhältnis von y_1 zu y_2.

Für die Konstruktion der zugehörigen ϕ_{max} Tabelle müssen die Randhäufigkeiten erhalten bleiben, lediglich die Verbundhäufigkeiten ändern sich. Um die ϕ_{max} Tabelle mit positivem Vorzeichen zu erzeugen, kann nun auf der Nebendiagonale entweder anstelle von n_{12} oder n_{12} der Wert Null eingesetzt werden. Bei schiefen Randverteilungen ist diese Stelle keineswegs beliebig. Man betrachte dazu folgendes Beispiel.

Für die Tabelle

	x_1	x_2	Σ
y_1	*40*	*10*	*50*
y_2	*30*	*1*	*31*
Σ	*70*	*11*	*81*

sind die Randhäufigkeiten schief (70:11 versus 50:31). Es sind zwei mögliche ϕ_{max} Tabellen mit positiver Korrelation denkbar, je nachdem, ob n_{12} oder n_{12} gleich Null gesetzt wird. Diese lauten

	x_1	x_2	Σ
y_1	*50*	*0*	*50*
y_2	*20*	*11*	*31*
Σ	*70*	*11*	*81*

und

	x_1	x_2	Σ
y_1	*70*	*-20*	*50*
y_2	*0*	*31*	*31*
Σ	*70*	*11*	*81*

Man sieht sofort, dass die zweite Tabelle nicht zulässig ist, da eine Verbundhäufigkeit negativ wird. Ein solcher Widerspruch tritt bei schiefen Randhäufigkeiten zwangsläufig ein, so dass die ϕ_{max} Tabelle immer eindeutig zu bestimmen ist. Ob hierfür n_{12} oder n_{12} gleich Null gesetzt werden muss, wird allein durch die konkreten Randhäufigkeiten definiert.

In jedem Falle wird nach dem Nullsetzen der erlaubten Stelle der Nebendiagonale auf der Hauptdiagonale an der Stelle n_{11} die kleinere Zahl aus $n_{1\bullet}$ oder $n_{\bullet 1}$ stehen. Die größere Zahl würde zu einer unzulässigen Tabelle führen, da dann die kleinere der beiden Randhäufigkeiten überschritten wäre. Mit derselben Argumentation muss für n_{22} die kleinere Zahl $n_{2\bullet}$ oder $n_{\bullet 2}$ erscheinen. Es gilt also nach dem Nullsetzen

$$\begin{aligned} n_{11} &= min(n_{1\bullet}, n_{\bullet 1}) \\ n_{22} &= min(n_{2\bullet}, n_{\bullet 2}) \end{aligned}$$

Zusätzlich sieht man sofort, dass das Produkt der beiden Zahlen auf der Nebendiagonale Null sein muss, da nach dem Nullsetzen von entweder n_{12} oder n_{12} einer der beiden Multiplikanden Null ist. Setzt man diese Terme aus der ϕ_{max} Tabelle in die Formel für den ϕ Koeffizienten ein, so erhält man

$$\phi_{max} = \frac{min(n_{1\bullet}, n_{\bullet 1}) \cdot min(n_{2\bullet}, n_{\bullet 2}) - 0}{\sqrt{n_{1\bullet} \cdot n_{\bullet 1} \cdot n_{2\bullet} \cdot n_{\bullet 2}}}.$$

Nun muss ein kleiner Trick angewandt werden. Betrachtet man die beiden Zahlen $n_{1\bullet}$ und $n_{\bullet 1}$, so ist unmittelbar evident, dass eine der beiden Zahlen $min(n_{1\bullet}, n_{\bullet 1})$ und die andere der beiden Zahlen $max(n_{1\bullet}, n_{\bullet 1})$ sein muss, da bei nur zwei Zahlen notwendigerweise die eine der beiden die größere und die andere die kleinere ist. Sind die beiden Zahlen gleich, bleibt diese Betrachtung prinzipiell korrekt, nur ist die Zuordnung dann beliebig. Dieselbe Betrachtung gilt für die Zahlen $n_{2\bullet}$ und $n_{\bullet 2}$. Auch hier muss eine der beiden Zahlen gleich $min(n_{2\bullet}, n_{\bullet 2})$ und die andere gleich $max(n_{2\bullet}, n_{\bullet 2})$ sein.

Der Nenner des ϕ_{max}–Koeffizienten kann dementsprechend ersetzt werden als

$$\phi_{max} = \frac{min(n_{1\bullet}, n_{\bullet 1}) \cdot min(n_{2\bullet}, n_{\bullet 2})}{\sqrt{min(n_{1\bullet}, n_{\bullet 1}) \cdot max(n_{1\bullet}, n_{\bullet 1}) \cdot min(n_{2\bullet}, n_{\bullet 2}) \cdot max(n_{2\bullet}, n_{\bullet 2})}}$$

Nun gilt aber wegen $x = \sqrt{x} \cdot \sqrt{x}$

$$\begin{aligned} min(n_{1\bullet}, n_{\bullet 1}) &= \sqrt{min(n_{1\bullet}, n_{\bullet 1})} \cdot \sqrt{min(n_{1\bullet}, n_{\bullet 1})} \\ min(n_{2\bullet}, n_{\bullet 2}) &= \sqrt{min(n_{2\bullet}, n_{\bullet 2})} \cdot \sqrt{min(n_{2\bullet}, n_{\bullet 2})}, \end{aligned}$$

so dass wir den ϕ_{max}–Koeffizienten schreiben können als

$$\phi_{max} = \frac{\sqrt{min(n_{1\bullet}, n_{\bullet 1})} \cdot \sqrt{min(n_{1\bullet}, n_{\bullet 1})} \cdot \sqrt{min(n_{2\bullet}, n_{\bullet 2})} \cdot \sqrt{min(n_{2\bullet}, n_{\bullet 2})}}{\sqrt{min(n_{1\bullet}, n_{\bullet 1}) \cdot max(n_{1\bullet}, n_{\bullet 1}) \cdot min(n_{2\bullet}, n_{\bullet 2}) \cdot max(n_{2\bullet}, n_{\bullet 2})}}.$$

Da $\sqrt{x \cdot y} = \sqrt{x} \cdot \sqrt{y}$, kann der Nenner nun ebenfalls aufgelöst werden, womit man erhält

$$\phi_{max} = \frac{\sqrt{min(n_{1\bullet}, n_{\bullet 1})} \cdot \sqrt{min(n_{1\bullet}, n_{\bullet 1})} \cdot \sqrt{min(n_{2\bullet}, n_{\bullet 2})} \cdot \sqrt{min(n_{2\bullet}, n_{\bullet 2})}}{\sqrt{min(n_{1\bullet}, n_{\bullet 1})} \cdot \sqrt{max(n_{1\bullet}, n_{\bullet 1})} \cdot \sqrt{min(n_{2\bullet}, n_{\bullet 2})} \cdot \sqrt{max(n_{2\bullet}, n_{\bullet 2})}}.$$

Dies kürzt sich schließlich zu

$$\phi_{max} = \frac{\sqrt{min(n_{1\bullet}, n_{\bullet 1})} \cdot \sqrt{min(n_{1\bullet}, n_{\bullet 1})}}{\sqrt{min(n_{1\bullet}, n_{\bullet 1})} \cdot \sqrt{max(n_{1\bullet}, n_{\bullet 1})}}.$$