DZIAŁ NAUKOWY

MATEMATYKA STOSOWANA 8, 2007

Joanna Jarosz-Nowak (Wrocław)

Modele oceny stopnia zgody pomiędzy dwoma ekspertami z wykorzystaniem współczynników kappa

Streszczenie. Praca dotyczy modeli oszacowania stopnia zgody pomiędzy dwoma ekspertami oceniającymi te same obiekty. Rozważamy konstrukcję miar służących oszacowaniu poziomu zgodności oraz konteksty interpretacyjne otrzymanych wyników. W pracy rozpatrujemy również powiązania pomiędzy współczynnikami zdefiniowanymi w modelach o dwóch i więcej dopuszczalnych kategorii ocen.

Analiza poziomu zgody prowadzi do ilościowego ujęcia z wykorzystaniem współczynników kappa Cohena oraz Scotta. Pokażemy, że te heurystyczne propozycje w odpowiednio zdefiniowanych modelach dychotomicznych pokrywają się z estymatorami największej wiarogodności. Równocześnie wykażemy, że rozwinięcie metod dla modeli z większą niż dwa liczbą kategorii możliwe jest poprzez ważone uśrednienie współczynników kappa określonych w modelach binarnych wyznaczonych dla każdej kategorii oddzielnie.
Słowa kluczowe: zgodność, współczynnik kappa Cohena, współczynnik kappa Scotta.

1. Wprowadzenie. Zarówno w życiu codziennym, jak i w badaniach naukowych spotykamy sytuacje, w których dana osoba lub przedmiot poddane są ocenie kilku ekspertów. Grupą osób oceniających może być jury konkursowe, komisja egzaminacyjna, konsylium lekarzy itp. Niejednokrotnie, wystawione przez poszczególnych ekspertów oceny są w znacznym stopniu subiektywne i mogą się między sobą różnić. Interesuje nas jak bardzo te oceny są do siebie zbliżone, czyli jaki jest stopień zgody pomiędzy ekspertami. Oszacowanie poziomu zgodności oceniających pozwala określić regularność uzyskiwanych klasyfikacji. Oceniającym może być osoba (np. członek jury, egzaminator, lekarz wystawiający diagnozę itp.) jak również metoda lub narzędzie klasyfikujące (np. test diagnostyczny, laboratorium pomiarowe).

W pracy opiszemy modele i metody statystyki matematycznej służące określeniu stopnia zgody w przypadku klasyfikacji przez dwóch oceniających. Przyjmujemy, że skala pomiarowa, przy pomocy której dokonuje się oceny obiektu, jest skategoryzowana, tzn. interesujące obserwatora zdarzenia ujęte są w kategorie.

2000 Mathematics Subject Classification. Primary: 62P10; Secondary: 62H17.

Początkowo do oceny stopnia zgody używano statystyki testowej χ^2 lub współczynnika zgody obserwowanej P_o estymowanego jako frakcja obiektów o tych samych ocenach. Jednakże, otrzymywane rezultaty nie były satysfakcjonujące. Test χ^2 określa siłę związku pomiędzy ocenami, ale niekoniecznie jest to związek polegający na przypisaniu obserwacji do tej samej kategorii. Natomiast liczba przypadków, co do których oceniający zgodzili się w swojej ocenie, może być zawyżona o obserwacje, dla których eksperci wystawili oceny losowo i przypadkowo dali takie same. Fakt ten był powodem zaproponowania przez Scotta [21] oraz Cohena [5] metod poprawionych o zgodność przypadkową P_c określanych mianem współczynników typu kappa:

$$\kappa = \frac{P_o - P_c}{1 - P_c}.$$

Główna różnica pomiędzy tymi metodami polega na sposobie estymacji wartości prawdopodobieństwa P_c wystawienia tych samych ocen przypadkowo. Scott zakłada, że osoby oceniające nie umiejąc wystawić noty podają ocenę losową – obie z takim samym prawdopodobieństwem. Cohen argumentuje, że miara zgodności powinna uwzględniać pojedyncze preferencje każdego eksperta z osobna.

W pracy opiszemy modele matematyczne i metody estymacji w tych modelach, prowadzące do estymatorów identycznych z heurystycznie zaproponowanymi współczynnikami typu kappa. Okazuje się, że intuicyjne propozycje w przypadku modelu binarnego pokrywają się z estymatorami największej wiarogodności w odpowiednio zdefiniowanych modelach teoretycznych [1]. W pracy wykażemy hierarchiczne zależności pomiędzy modelami binarnymi a modelami skonstruowanymi dla danych politomicznych. Podamy analityczne uzasadnienie faktu, iż współczynniki kappa oraz odpowiadające im estymatory w modelach z wieloma kategoriami są średnimi ważonymi z wagami typu $(1 - P_c)$ współczynników uzyskanych dla modeli binarnych skonstruowanych dla każdej kategorii oddzielnie. Taka reprezentacja możliwa jest dla obu typów współczynników kappa.

Zaprezentujemy również własności uzyskanych estymatorów, w szczególności ich zależność od rozkładów brzegowych i rozpowszechnienia badanej cechy. Opiszemy problemy interpretacyjne [17, 7, 3, 14] oraz obliczeniowe dla tychże estymatorów. Okazuje się, że wykorzystanie standardowych metod zaimplementowanych w pakietach statystycznych nie gwarantuje poprawnych obliczeń [10, 6, 18]. W szczególnych przypadkach wyliczone wielkości mogą nie być poprawne, tj. współczynnik zgody nie zostaje obliczony lub jego wartość jest wyliczona błędnie.

2. Notacja. Przez N będziemy określać łączną liczbę obiektów poddanych ocenie. Indeks i, gdzie $i = 1, \ldots, N$, odpowiada i-temu obiektowi.

Przez R oznaczymy ogólną liczbę oceniających, czyli liczbę wykonanych pomiarów. Przy opisywanej klasyfikacji podwójnej przyjmujemy, że $R = 2$. Liczbę dopuszczalnych kategorii, do których można zaklasyfikować obiekt w trakcie pomiaru będziemy oznaczać przez C. Indeksy l oraz k będą oznaczać ustaloną kategorię ($l, k \in \{1, \dots, C\}$). Opisana w pracy klasyfikacja jest rozłączna, czyli wyborem oceniającego jest zawsze tylko jedna z dostępnych kategorii.

2.1. Sposoby prezentacji zgromadzonych danych. W literaturze przedmiotu szczególną uwagę, ze względu na szerokie zastosowania, poświęca się przypadkowi klasyfikacji binarnej dwóch ekspertów ($C = 2$; $R = 2$). Polega ona na przypisaniu ocenianemu obiektowi noty 0 (w przypadku gdy, obiekt nie wykazuje cech typowych dla grupy, nie należy do grupy) lub 1 (w przypadku gdy: obiekt posiada daną cechę, należy do grupy, wykazuje cechy typowe dla osobników z danej grupy). Ogólniejsze rozważania przeprowadza się dla $C > 2$. Dla takich eksperymentów dane można zgromadzić w postaci Tabeli 1, gdzie n_{lk} oznacza liczbę obiektów zaklasyfikowanych przez eksperta pierwszego (ozn. A) do kategorii l, a przez eksperta drugiego (B) do kategorii k ($l, k = 1, 2$).

		Oceniający B						
	Kategoria	1	2	$\dots$	k	$\dots$	C	Ogółem
	1	n_{11}	n_{12}	$\dots$	n_{1k}	$\dots$	n_{1C}	n_{1+}
	2	n_{21}	n_{22}	$\dots$	n_{2k}	$\dots$	n_{2C}	n_{2+}
Oceniajacy	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A	l	n_{l1}	n_{l2}	$\vdots$	n_{lk}	$\vdots$	n_{lC}	n_{l+}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	C	n_{C1}	n_{C2}	$\dots$	n_{Ck}	$\dots$	n_{CC}	n_{C+}
	Ogółem	n_{+1}	n_{+2}	$\dots$	n_{+k}	$\dots$	n_{+C}	N

Tablica 1. Format danych w przypadku dwóch oceniających z wykorzystaniem skali o wielu kategoriach.

Przez n_{l+} oznaczać będziemy liczebności brzegowe, czyli liczbę obiektów przypisanych do kategorii l-tej przez oceniającego A, a n_{+k} oznacza ogólną liczbę obiektów zaklasyfikowanych w trakcie eksperymentu do kategorii k-tej przez oceniającego B.

$$n_{l+} = \sum_{k=1}^{C} n_{lk} \qquad n_{+k} = \sum_{l=1}^{C} n_{lk}$$

Innym sposobem zapisu danych jest format zaprezentowany w Tabeli 2.

	Kategoria				
Obiekt	1	2	$\dots$	C	Razem
1	n_{11}	n_{12}	$\dots$	n_{1C}	R
2	n_{21}	n_{22}	$\dots$	n_{2C}	R
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
N	n_{N1}	n_{N2}	$\dots$	n_{NC}	R

TABLICA 2. Format danych w przypadku więcej niż dwóch oceniających z częściową utratą informacji.

W tym przypadku n_{il} jest liczbą oceniających, którzy zaklasyfikowali obiekt i-ty do l-tej kategorii, a n_{+l} jest liczbą obiektów przypisanych przez ekspertów do kategorii l-tej:

$$n_{+l} = \sum_{i=1}^{N} n_{il}.$$

Format ten może być wykorzystywany również w badanich z większą liczbą oceniających $R > 2$. Tak zgromadzone dane posiadają ograniczoną informację, ponieważ nie można na ich podstawie określić, który oceniający przypisał i-ty obiekt do l-tej kategorii, a tylko ile takich przypisań było w trakcie eksperymentu.

Format danych nie posiadający cech utraty informacji prezentuje Tabela 3, gdzie X_{ij} jest oceną wystawioną przez j-tego eksperta i-temu obiektowi ($i = 1, \dots, N$; $j = 1, \dots, R$).

	Oceniający			
Obiekt	1	2	$\dots$	R
1	X_{11}	X_{12}	$\dots$	X_{1R}
2	X_{21}	X_{22}	$\dots$	X_{2R}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N	X_{N1}	X_{N2}	$\dots$	X_{NR}

TABLICA 3. Ogólny format danych.

Dane zgromadzone w Tabeli 3 można przekodować za pomocą ślepych zmiennych (ang. *dummy variables*) wprowadzając dychotomizację. Przy takim kodowaniu tworzymy C tabel typu Tabeli 3, gdzie każda z nich odpowiada jednej kategorii. Elementy tabel X_{ijl} ($i = 1, \dots, N$; $j = 1, \dots, R$; $l = 1, \dots, C$) przyjmują wartości 0 lub 1. Wszystkie C tabel tworzy razem tablicę trójwymiarową o elementach zero-jedynkowych. Przykładową tablicę danych dla l-tej kategorii przedstawia Tabela 4.

	Oceniający			
Obiekt	1	2	...	R
1	X_{11l}	X_{12l}	...	X_{1Rl}
2	X_{21l}	X_{22l}	...	X_{2Rl}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N	X_{N1l}	X_{N2l}	...	X_{NRl}

Tablica 4. Ogólny format danych dla l-tej kategorii w postaci kodowania ślepego.

3. Współczynnik kappa Scotta κ_S. Pierwszą pracą, w której ujmuje się korektę o przypadkową zgodność w sposób zależny od próby, jest praca Scotta z roku 1955 [21]. Autor argumentuje, że miara oceny stopnia zgody powinna uwzględniać relatywną liczbę kategorii (tj. liczbę wykorzystanych/użytych kategorii) oraz sposób ich wykorzystania.

Jego propozycją jest współczynnik κ_S dany wzorem:

$$\kappa_S = \frac{P_o - P_c}{1 - P_c},$$

gdzie P_o jest obserwowaną zgodnością, a P_c określa stopień przypadkowej zgodności. Wielkość P_o estymowana jest następująco

$$\hat{P}_o = \frac{1}{N}\sum_{l=1}^{C} n_{ll}, \tag{3.1}$$

gdzie n_{ll} sa elementami na przekątnej Tabeli 1. Natomiast wielkość P_c estymowana jest przez

$$\hat{P}_c = \sum_{l=1}^{C} \hat{p}_l^2, \tag{3.2}$$

gdzie

$$\hat{p}_l = \frac{1}{2}\left(\frac{n_{+l}}{N} + \frac{n_{l+}}{N}\right). \tag{3.3}$$

Dla ustalonego l wyrażenie $\hat{p}_l$ może być interpretowane jako uśredniona tendencja przypisania obiektu do l-tej kategorii ($l = 1, \ldots, C$). Przyjmując, że oszacowaniem preferencji pierwszego oceniającego jest frakcja $\frac{n_{l+}}{N}$, a drugiego $\frac{n_{+l}}{N}$, można wyliczyć uśrednioną tendecję ekspertów jako średnią arytmetyczną tych frakcji zgodnie z wzorem (3.3). Przyjęcie zamiast uśrednionej tendencji, iloczynu empirycznych rozkładów brzegowych $\frac{n_{l+}}{N}\frac{n_{+l}}{N}$, prowadzi do propozycji Cohena opisanej w dalszej części pracy w rozdziale 4.

3.1. Model binarny. Model matematyczny pozwalający oszacować stopień zgody współczynnikiem Scotta w najprostszym przypadku dwóch kategorii można zapisać w następujący sposób. Każdy z dwóch oceniających klasyfikuje N obiektów niezależnie od siebie do dwóch rozłącznych kategorii.

Bez straty ogólności przyjmijmy, że określamy przynależność do danej grupy, posiadanie badanej cechy przez 1 a jej brak przez 0. W celu zdefiniowania modelu matematycznego oznaczymy ocenę j-tego oceniającego wystawioną i-temu obiektowi przez X_{ij}, gdzie $i = 1, \ldots, N$ oraz $j = 1, 2$. Plan eksperymentu w tym przypadku odpowiada formatowi danych określonemu Tabelą 1 przy $C = 2$.

Ponadto zakładamy, że zmienna losowa X_{ij} ma rozkład dwupunktowy z parametrem p $(X_{ij} \sim \mathcal{B}(1,p),\ i = 1, \ldots, N,\ j = 1, 2)$

$$\Pr\{X_{ij} = 1\} = p,$$

$$\Pr\{X_{ij} = 0\} = q,$$

gdzie $q = 1 - p$. Takie założenia oznaczają, że preferencje (tj. prawdopodobieństwa przypisania obiektowi danej cechy) oceniających przyjmujemy za jednakowo równe p. Niech ϱ oznacza korelację pomiędzy parą ocen. Zakładamy, że jest ona taka sama dla ocen każdego obiektu, tj. $\mathrm{Corr}(X_{i1}, X_{i2}) = \varrho$, $i = 1, \ldots, N$.

Stwierdzenie 1. *Dla ustalonego i $(i = 1, \ldots, N)$ rozkład łączny zmiennych losowych X_{i1} i X_{i2} w powyższym modelu dany jest następująco*:

$$\Pr\{X_{i1} = 1, X_{i2} = 1\} = p^2 + pq\varrho, \tag{3.4}$$

$$\Pr\{X_{i1} = 0, X_{i2} = 0\} = q^2 + pq\varrho, \tag{3.5}$$

$$\Pr\{X_{i1} = 0, X_{i2} = 1\} = \Pr\{X_{i1} = 1, X_{i2} = 0\} = pq(1 - \varrho). \tag{3.6}$$

Dowód. Ponieważ $\mathrm{Corr}(X_{i1}, X_{i2}) = \mathrm{Cov}(X_{i1}, X_{i2})/\sqrt{\mathrm{Var}(X_{i1})\,\mathrm{Var}(X_{i2})}$, to mamy $\mathbb{E}(X_{i1}X_{i2}) = \mathrm{Corr}(X_{i1}, X_{i2})\,\mathrm{Var}(X_{i1}) + \mathbb{E}(X_{i1})^2 = \varrho pq + p^2$. Z drugiej strony $\mathbb{E}(X_{i1}X_{i2}) = \Pr\{X_{i1}X_{i2} = 1\} = \Pr\{X_{i1} = 1, X_{i2} = 1\}$. Stąd otrzymujemy, że $\Pr\{X_{i1} = 1, X_{i2} = 1\} = p^2 + pq\varrho$. Przez symetrię można wykazać, że $\Pr\{X_{i1} = 0, X_{i2} = 0\} = q^2 + pq\varrho$. Wobec tego $\Pr\{X_{i1} = 0, X_{i2} = 1 \text{ lub } X_{i1} = 1, X_{i2} = 0\} = 1 - (p^2 + pq\varrho + q^2 + pq\varrho) = 2pq(1 - \varrho)$. Z rozkładu brzegowego otrzymujemy $\Pr\{X_{i1} = 1, X_{i2} = 0\} = \Pr\{X_{i1} = 1\} - \Pr\{X_{i1} = 1, X_{i2} = 1\}$ co daje, że $\Pr\{X_{i1} = 1, X_{i2} = 0\} = pq(1 - \varrho)$. Analogicznie rozumując otrzymujemy, że $\Pr\{X_{i1} = 0, X_{i2} = 1\} = pq(1 - \varrho)$. □

Oznaczmy przez P_o prawdopodobieństwo, że oceniający zgodzą się w swojej ocenie, czyli:

$$P_o = \Pr\{X_{i1} = X_{i2}\}.$$

Przez P_c oznaczmy prawdopodobieństwo, że oceniający dadzą te same oceny mimo, że nie są one skorelowane. Może się tak zdarzyć w przypadku, gdy oceny ekspertów były wystawiane losowo, bez szablonów postępowania i mimo to są identyczne.

Definicja 1. W modelu binarnym wartość P_o definiujemy następująco:

$$P_o = \Pr\{X_{i1} = 1, X_{i2} = 1\} + \Pr\{X_{i1} = 0, X_{i2} = 0\}. \tag{3.7}$$

Prawdopodobieństwo zgody przez przypadek P_c wyliczmy przyjmując, że oceny ekspertów są nieskorelowane ze sobą co oznacza jednocześnie, że P_c jest sumą iloczynów odpowiednich prawdopodobieństw rozkładów brzegowych.
W opisywanym modelu mamy

$$\begin{aligned} P_o &= \Pr\{X_{i1} = X_{i2} = 1\} + \Pr\{X_{i1} = X_{i2} = 0\} \\ &= p^2 + q^2 + 2pq\varrho = 1 - 2pq(1 - \varrho) \end{aligned} \tag{3.8}$$

oraz po podstawieniu $\varrho = 0$ do powyższego wzoru (3.8) otrzymujemy

$$P_c = 1 - 2pq = p^2 + q^2. \tag{3.9}$$

Twierdzenie 2. *W rozważanym modelu współczynnik zgodności typu kappa jest równy korelacji ϱ.*

Dowód. Przy P_o i P_c zdefiniowanych powyżej mamy:

$$\kappa_S = \frac{P_o - P_c}{1 - P_c} = \frac{1 - 2pq(1 - \varrho) - (1 - 2pq)}{1 - (1 - 2pq)} = \varrho.$$

□

3.2. Estymacja w modelu binarnym. Załóżmy, że dla próby N obiektów, obserwujemy liczebności par ocen każdego typu. Dane takie zgromadzimy w Tabeli 5 (uzyskanej z Tabeli 1 dla $C = 2$).

		Oceniający 2-gi		
	Kategoria	1	0	Ogółem
Oceniający	1	n_{11}	n_{12}	n_{1+}
1-szy	0	n_{21}	n_{22}	n_{2+}
	Ogółem	n_{+1}	n_{+2}	N

Tablica 5. Format danych w przypadku dwóch oceniających z wykorzystaniem skali o dwóch kategoriach.

Estymator współczynika kappa Scotta κ_S można wyznaczyć metodą największej wiarogodności [1]. Dla rozważanego modelu funkcja wiarogodności jest postaci

$$L(p,\varrho|n_{11},n_{22},n_{12},n_{21}) = (p^2+pq\varrho)^{n_{11}}(q^2+pq\varrho)^{n_{22}}(pq(1-\varrho))^{n_{12}}(pq(1-\varrho))^{n_{21}},$$

gdzie $q = 1-p$.
Wobec tego

$$\frac{\partial \ln L}{\partial \varrho} = n_{11}\frac{pq}{p^2+pq\varrho} + n_{22}\frac{pq}{q^2+pq\varrho} - (n_{12}+n_{21})\frac{1}{1-\varrho} \tag{3.10}$$

oraz

$$\frac{\partial \ln L}{\partial p} = \frac{n_{11}(2p(1-\varrho)+\varrho)}{p^2+pq\varrho} + \frac{(n_{12}+n_{21})(1-2p)}{pq} - \frac{n_{22}(2q(1-\varrho)+\varrho)}{q^2+pq\varrho}. \tag{3.11}$$

Przyrównując (3.10) oraz (3.11) do zera otrzymujemy następujące estymatory:

$$\hat{\kappa}_S = \hat{\varrho} = \frac{4n_{11}n_{22}-(n_{12}+n_{21})^2}{(2n_{11}+n_{21}+n_{12})(2n_{22}+n_{21}+n_{12})} \tag{3.12}$$

oraz

$$\hat{p} = \frac{1}{2}\left(\frac{n_{11}+n_{12}}{N}+\frac{n_{11}+n_{21}}{N}\right) = \frac{2n_{11}+n_{12}+n_{21}}{2N}. \tag{3.13}$$

Stąd

$$\hat{q} = 1-\hat{p} = \frac{2n_{22}+n_{12}+n_{21}}{2N},$$

oraz

$$\hat{P}_o = 1-2\hat{p}(1-\hat{p})(1-\hat{\varrho}) = \frac{n_{11}}{N}+\frac{n_{22}}{N},$$

$$\hat{P}_c = \hat{p}^2+(1-\hat{p})^2 = \left(\frac{2n_{11}+n_{21}+n_{12}}{2N}\right)^2 + \left(\frac{2n_{22}+n_{21}+n_{12}}{2N}\right)^2$$

co jest zgodne z wzorami (3.1) oraz (3.2) dla $C = 2$.

Zauważmy, że oba estymatory $\hat{p}$ oraz $\hat{\kappa}_S$ zależą w rzeczywistości od *trzech* liczebności n_{11}, n_{22} oraz sumy $(n_{12}+n_{21})$.

UWAGA 3. Należy zwrócić uwagę na dwa szczególne przypadki, gdy $p = 0$ lub $p = 1$. Dla $p = 0$ rozkład łączny zmiennych losowych X_{i1} i X_{i2} dany jest następująco:

$$\Pr\{X_{i1}=X_{i2}=1\} = \Pr\{X_{i1}=0, X_{i2}=1\} = \Pr\{X_{i1}=1, X_{i2}=0\} = 0,$$
$$\Pr\{X_{i1}=0, X_{i2}=0\} = 1.$$

Natomiast, dla $p = 1$ mamy

$$\Pr\{X_{i1} = X_{i2} = 0\} = \Pr\{X_{i1} = 0, X_{i2} = 1\} = \Pr\{X_{i1} = 1, X_{i2} = 0\} = 0,$$
$$\Pr\{X_{i1} = 1, X_{i2} = 1\} = 1.$$

Oznacza to, że *trzy* spośród obserwowanych liczebności wynoszą zero n_{12}, n_{21} oraz n_{11} lub n_{22}. W takich deterministycznych przypadkach, gdy obaj oceniający zgodzili się całkowicie, ale przypisali wszystkie obiekty do jednej kategorii estymator dany wzorem (3.12) nie jest dobrze określony. W zależności od intencji badaczy, albo można dookreślić model przyjmując, że $\hat{\kappa}_S = 1$, albo (jeśli mocniejsze są przesłanki, że dając wszystkim tę samą notę oceniający nie potrafią rozróżnić obiektów) nie podawać oszacowania stopnia zgody.

3.3. Model dla więcej niż dwóch kategorii. Rozważany do tej pory model określony był dla dwóch kategorii $C = 2$. Jednak swoją propozycję miary zgodności Scott wysunął dla dowolnej liczby $C \geq 2$. Wykażemy, że ten ogólniejszy estymator można skonstruować w oparciu o wartości współczynników kappa uzyskanych dla każdej z kategorii osobno bazując na zasadzie podstawiania. Okazuje się, że współczynnik kappa jest średnią ważoną binarnych współczynników kappa. Poprzez wstępną agregację obserwacji rozważania prowadzi się na poziomie przynależności lub nie do danej kategorii (bez względu na to jakie inne noty uzyskała dana obserwacja).

Niech X_{ij} (określająca ocenę i-tego obiektu wystawioną przez j-tego eksperta) będzie teraz zmienną losową z rozkładu wielomianowego z parametrami $(p_1, \ldots, p_C)$, $\sum_{l=1}^{C} p_l = 1$. Przykładowo zmienna X_{ij} przyjmuje wartości $1, \ldots, C$ z prawdopodobieństwami $p_1, \ldots, p_C$, odpowiednio.

Określmy na jej podstawie binarną ocenę przynależności X_{ijl} i-tego obiektu do kategorii l wystawioną przez j-tego oceniającego ($i = 1, \ldots, N$; $j = 1, 2$; $l = 1, \ldots, C$). Dla każdej kategorii l definiujemy:

$$X_{ijl} = \begin{cases} 1 & \text{gdy } X_{ij} = l, \\ 0 & \text{gdy } X_{ij} \neq l. \end{cases}$$

W następnym kroku dla każdej kategorii z osobna można zastosować model binarny opisany w poprzednich rozdziałach 3.1 oraz 3.2. Dla ustalonego l zmienna losowa X_{ijl} ma rozkład dwupunktowy z parametrem p_l $(X_{ijl} \sim \mathcal{B}(1, p_l),\ i = 1, \ldots, N;\ j = 1, 2)$

$$\Pr\{X_{ijl} = 1\} = p_l,$$
$$\Pr\{X_{ijl} = 0\} = q_l,$$

gdzie $q_l = 1 - p_l = \sum_{\substack{k=1 \\ k \neq l}}^{C} p_k$.

Przeprowadzając analogiczne rozumowanie jak w dowodzie Stwierdzenia 1 wyznaczymy rozkład łączny zmiennych losowych X_{i1l} i X_{i2l}.

FAKT 4. *Dla ustalonego i oraz l ($i = 1, \ldots, N; l = 1, \ldots, C$), przy założeniu, że $Corr(X_{i1l}, X_{i2l}) = \varrho_l$ rozkład łączny zmiennych losowych X_{i1l} i X_{i2l} dany jest następująco*:

$$\begin{aligned} \Pr\{X_{i1l} = 1, X_{i2l} = 1\} &= p_l^2 + p_l q_l \varrho_l, \\ \Pr\{X_{i1l} = 0, X_{i2l} = 0\} &= q_l^2 + p_l q_l \varrho_l, \\ \Pr\{X_{i1l} = 0, X_{i2l} = 1\} = \Pr\{X_{i1l} = 1, X_{i2l} = 0\} &= p_l q_l (1 - \varrho_l). \end{aligned} \tag{3.14}$$

Oznaczmy przez $P_{o,l}$ prawdopodobieństwo, że oceniający zgodzą się w swojej ocenie w modelu binarnym, a przez $P_{c,l}$ prawdopodobieństwo zgody losowej. W tym przypadku

$$\begin{aligned} P_{o,l} &= \Pr\{X_{i1l} = X_{i2l} = 1\} + \Pr\{X_{i1l} = X_{i2l} = 0\} \\ &= p_l^2 + (1 - p_l)^2 + 2p_l(1 - p_l)\varrho_l \end{aligned}$$

oraz

$$P_{c,l} = p_l^2 + (1 - p_l)^2.$$

Z Twierdzenia 2 wiadomo, że dla ustalonej kategorii współczynnik kappa Scotta wynosi:

$$\kappa_S^l = \frac{P_{o,l} - P_{c,l}}{1 - P_{c,l}} = \varrho_l.$$

DEFINICJA 2. Prawdopodobieństwo zgody P_o w modelu z wieloma kategoriami definiujemy w następujący sposób:

$$P_o = \sum_{l=1}^{C} \Pr\{X_{i1} = X_{i2} = l\} = \sum_{l=1}^{C} \Pr\{X_{i1l} = X_{i2l} = 1\}.$$

Wobec równości (3.14) w rozważanym modelu

$$P_o = \sum_{l=1}^{C} (p_l^2 + p_l(1 - p_l)\varrho_l)$$

natomiast prawdopodobieństwo zgody przez przypadek wynosi

$$P_c = \sum_{l=1}^{C} p_l^2.$$

DEFINICJA 3. Współczynnik kappa Scotta w modelu z wieloma kategoriami (uwzględniając powyższe równości) definiujemy standardowo

$$\kappa_S \stackrel{df}{=} \frac{P_o - P_c}{1 - P_c} = \frac{\sum_{l=1}^{C} p_l(1 - p_l)\varrho_l}{\sum_{l=1}^{C} p_l(1 - p_l)} \tag{3.15}$$

uzyskując zależność od wszystkich binarnych współczynników $\kappa_S^l = \varrho_l$.

Lemat 5. *Licznik i mianownik wyrażenia (3.15) można wyrazić w terminach $P_{o,l}$ oraz $P_{c,l}$ następująco:*

$$\sum_{l=1}^{C} \varrho_l p_l(1-p_l) = \frac{1}{2}\sum_{l=1}^{C}(P_{o,l}-P_{c,l}),$$
$$\sum_{l=1}^{C} p_l(1-p_l) = \frac{1}{2}\sum_{l=1}^{C}(1-P_{c,l}).$$

Dowód. Przypomnijmy, że dla modelu bianarnego $P_{o,l} = 1 - 2p_l(1 - p_l)(1-\varrho_l)$ oraz $P_{c,l} = 1-2p_l(1-p_l)$. W związku z tym $p_l(1-p_l) = \frac{1}{2}(1-P_{c,l})$. Ponadto $P_{o,l} - P_{c,l} = 2p_l(1-p_l)\varrho_l$, więc $\frac{1}{2}(P_{o,l} - P_{c,l}) = \varrho_l p_l(1-p_l)$. □

Stwierdzenie 6. *Współczynnik κ_S w terminach $P_{o,l}$, $P_{c,l}$ oraz ϱ_l ma postać:*

$$\begin{aligned}\kappa_S &= \frac{\sum_{l=1}^{C} p_l(1-p_l)\varrho_l}{\sum_{l=1}^{C} p_l(1-p_l)}\\ &= \frac{\sum_{l=1}^{C}(P_{o,l}-P_{c,l})}{\sum_{l=1}^{C}(1-P_{c,l})} \qquad (3.16)\\ &= \frac{\sum_{l=1}^{C} P_{o,l}-\sum_{l=1}^{C} P_{c,l}}{C-\sum_{l=1}^{C} P_{c,l}}\\ &= \frac{\sum_{l=1}^{C}(1-P_{c,l})\varrho_l}{\sum_{l=1}^{C}(1-P_{c,l})}. \qquad (3.17)\end{aligned}$$

Dowód. Na mocy Lematu 5 i definicji współczynnika κ_S danej wzorem (3.15) otrzymujemy równość (3.16). Mnożąc i dzieląc składniki sumy występującej w liczniku wyrażenia (3.16) przez $(1 - P_{c,l})$ otrzymujemy równość (3.17). □

Fakt 7. *Współczynnik κ_S jest średnią ważoną współczynników $\kappa_S^l = \varrho_l$ z wagami wynoszącymi $w_l = p_l(1-p_l) = \frac{1}{2}(1-P_{c,l})$.*

W dalszej części pracy podamy analityczne uzasadnienie, iż bazując na estymatorach dla modeli binarnych można uzyskać estymator współczynnika κ_S zaproponowany oryginalnie przez Scotta, dla którego $\hat{P}_c$ dane jest wzorem (3.2), a $\hat{P}_o$ standardowo wzorem (3.1).

3.3.1. Estymacja zgody obserwowanej $P_{o,l}$ w modelu binarnym wyznaczonym dla ustalonej kategorii l. Przypomnijmy, że w badaniach obserwujemy dane zgromadzone w Tabeli 1. Dwaj oceniający zgodzili się w swoich notach przypisując obiekt do ustalonej kategorii l dla n_{ll} obiektów poddanych ocenie. Kluczowym w rozumowaniu jest następujący fakt. Dla ustalonej kategorii l liczba obiektów, co do których obaj oceniający zgodzili się, że obiekt nie przynależy do kategorii l, jest *sumą* wszystkich par ocen różnych jednocześnie od l, tj. $X_{i1l} \neq l$ oraz $X_{i2l} \neq l$.

Zilustrujmy tę zależność na podstawie Tabeli 6, która może służyć gromadzeniu danych w eksperymencie z $C = 3$.

		Oceniający B			
	Kategoria	1	2	3	Ogółem
	1	n_{11}	n_{12}	n_{13}	n_{1+}
Oceniający	2	n_{21}	n_{22}	n_{23}	n_{2+}
A	3	n_{31}	n_{32}	n_{33}	n_{3+}
	Ogółem	n_{+1}	n_{+2}	n_{+3}	N

TABLICA 6. Tablica kontyngencji dla skali o trzech kategoriach.

Przykładowo ustalmy $l = 1$. Dla tej kategorii tabela binarna wygenerowana na podstawie Tabeli 6 będzie mieć postać Tabeli 7.

		Oceniający B		
	Kategoria	1	2+3	Ogółem
Oceniający	1	n_{11}	$n_{12} + n_{13}$	n_{1+}
A	2+3	$n_{21} + n_{31}$	$n_{22} + n_{23} + n_{33} + n_{32}$	$n_{(23)+}$
	Ogółem	n_{+1}	$n_{+(23)}$	N

TABLICA 7. Tablica kontyngencji modelu binarnego dla $l = 1$ w badaniu z pierwotną klasyfikacją do trzech kategorii.

Dla kategorii $l = 1$ liczba zgód „pozytywnych" wynosi n_{11}, a zgód „negatywnych" $n_{22} + n_{23} + n_{33} + n_{32}$.

Ogólnie

$$\hat{P}_{o,l} = \frac{n_{ll}}{N} + \frac{N - (n_{l+} + n_{+l}) + n_{ll}}{N}.$$

W dalszej części pracy opiszemy zależności pomiędzy estymatorami prawdopodobieństw zgody $P_{o,l}$ i P_o oraz estymatorami prawdopodobieństw zgody przypadkowej $P_{c,l}$ i P_c.

LEMAT 8. *Zachodzi związek*

$$\sum_{l=1}^{C} \hat{P}_{o,l} = 2\hat{P}_o + C - 2,$$

gdzie $\hat{P}_o$ *dane jest wzorem (3.1).*

Dowód.

$$\sum_{l=1}^{C} \hat{P}_{o,l} = \sum_{l=1}^{C} (N - (n_{l+} + n_{+l}) + 2n_{ll})/N$$
$$= (CN - 2N + 2\sum_{l=1}^{C} n_{ll})/N = C - 2 + 2\hat{P}_o.$$

□

Lemat 9. *Zachodzi związek*

$$\sum_{l=1}^{C} \hat{P}_{c,l} = 2\hat{P}_c + C - 2,$$

gdzie $\hat{P}_c$ *dane jest wzorem (3.2).*

Dowód.

$$\sum_{l=1}^{C} \hat{P}_{c,l} = \sum_{l=1}^{C} \left(\left(\frac{n_{l+} + n_{+l}}{2N} \right)^2 + \left(\frac{N - n_{l+} + N - n_{+l}}{2N} \right)^2 \right)$$
$$= \sum_{l=1}^{C} \left(2 \left(\frac{n_{l+} + n_{+l}}{2N} \right)^2 + 1 - 2 \frac{n_{l+} + n_{+l}}{2N} \right)$$
$$\overset{(3.2)}{=} 2\hat{P}_c + C - 2 \sum_{l=1}^{C} \frac{n_{l+} + n_{+l}}{2N} = 2\hat{P}_c + C - 2.$$

□

Twierdzenie 10. *Bazując na uzyskanych w modelach binarnych estymatorach* $\hat{P_{o,l}}$ *oraz* $\hat{P_{c,l}}$ *estymator współczynnika kappa dla modelu z wieloma kategoriami wynosi*

$$\hat{\kappa}_S = \frac{\sum_{l=1}^{C} \hat{P}_{o,l} - \sum_{l=1}^{C} \hat{P}_{c,l}}{C - \sum_{l=1}^{C} \hat{P}_{c,l}} = \frac{\hat{P}_o - \hat{P}_c}{1 - \hat{P}_c},$$

gdzie $\hat{P}_o$ *oraz* $\hat{P}_c$ *zostały zaproponowane przez Scotta wzorami (3.1) i (3.2).*

Dowód. Na mocy Lematów 8 i 9 otrzymujemy

$$\hat{\kappa_S} = \frac{\sum_{l=1}^{C} \hat{P}_{o,l} - \sum_{l=1}^{C} \hat{P}_{c,l}}{C - \sum_{l=1}^{C} \hat{P}_{c,l}} = \frac{C - 2 + 2\hat{P}_o - (C - 2 + 2\hat{P}_c)}{C - (C - 2) - 2\hat{P}_c} = \frac{\hat{P}_o - \hat{P}_c}{1 - \hat{P}_c}.$$

□

3.4. Wariancja estymatora, wartość maksymalna i minimalna. Wykorzystując metodę zaproponowaną przez Fishera, Bloch i Kraemer [1] przybliżyli wariancję asympotyczną estymatora współczynnika Scotta dla

$C = 2$ następująco:

$$\mathrm{Var}(\hat{\kappa}_S) = \frac{1-\kappa_S}{N}\left((1-\kappa_S)(1-2\kappa_S) + \frac{\kappa_S(2-\kappa_S)}{2pq}\right). \tag{3.18}$$

Metoda ta bazuje na rozwinięciu w szereg Taylora. Niech $T(n_1, n_2, \ldots, n_g)$ będzie funkcją zaobserwowanych częstości $n_1, n_2, \ldots, n_g$ dla próby o n elementach pochodzącej z rozkładu g-mianowego z prawdopodobieństwami $e_1, e_2, \ldots, e_g$ ($\sum n_h = n, \sum e_h = 1$). Wtedy asymptotycznie mamy

$$\frac{1}{n}\,\mathrm{Var}(T) = \sum_{h=1}^{g} e_h \left(\frac{\partial T}{\partial n_h}\right)^2 - \left(\frac{\partial T}{\partial n}\right)^2,$$

biorąc pochodne w punktach $n_h = e_h\, n$. Dla współczynnika kappa Scotta funkcja $T(n_1, n_2, n_3, n_4) = \hat{\kappa}_S(n_{11}, n_{12}, n_{21}, n_{22})$ dana wzorem (3.12), pochodna $\frac{\partial T}{\partial n} = 0$ a prawdopodobieństwa e_h określamy jak w modelu wzorami (3.4)–(3.6).

Maksymalną wielkością osiąganą przez estymator (przy pełnej zgodności) jest wartość 1, a minimalną -1.

4. Współczynnik kappa Cohena κ_C. W nurcie rozważań na temat oceny stopnia zgody dwóch oceniających pozostaje propozycja Cohena z roku 1960 [5], w której autor krytykuje wcześniejsze podejście argumentując, iż dla oceniających nie można przyjąć, że mają takie same preferencje. Uznaje, że tendencje wyboru przypisania obiektu do badanej klasy należy określać osobno dla każdego z ekspertów i dopiero na tej podstawie wyliczać wielkość P_c odpowiadającą szansie zgodności przez przypadek. Propozycja Cohena oparta ponownie na intuicyjnej idei wyraża się wzorem:

$$\kappa_C = \frac{P_o - P_c}{1 - P_c},$$

gdzie P_o estymowane jest standardowo zgodnie ze wzorem (3.1), natomiast P_c tym razem przybliżamy następująco:

$$\hat{P}_c = \sum_{l=1}^{C}\left(\frac{n_{l+}}{N}\frac{n_{+l}}{N}\right). \tag{4.19}$$

Wielkości $\frac{n_{l+}}{N}$ oraz $\frac{n_{+l}}{N}$ występujące w powyższym wyrażeniu interpretujemy jako preferencje przypisania ocenianego obiektu do kategorii l-tej przez każdego z ekspertów. We wcześniej podanej formule (3.2) preferencje obu oceniających były przyjęte jako jednakowe i wyestymowane jako średnia z pojedynczych preferencji.

W tym rozdziale opiszemy teoretyczny model matematyczny, dla którego uzyskane estymatory pokrywają się z heurystyczną propozycją Cohena.

4.1. Model binarny. W rozdziale 3 rozważaliśmy przypadek, gdy obaj oceniający z tym samym prawdopodobieństwem klasyfikują obiekty. Zakładamy teraz, że te prawdopodobieństwa są różne.

Dla ustalonych i oraz j $(i = 1, \ldots, N;\ j = 1, 2)$ mamy

$$\begin{aligned}\Pr\{X_{ij} = 1\} &= p_j,\\ \Pr\{X_{ij} = 0\} &= q_j = 1 - p_j.\end{aligned}$$

Przez ϱ oznaczmy korelację pomiędzy parą ocen. Załóżmy jak poprzednio, że jest ona jednakowa dla każdego obiektu tj. $\operatorname{Corr}(X_{i1}, X_{i2}) = \varrho, i = 1, \ldots, N$.

STWIERDZENIE 11. *Rozkład łączny zmiennych losowych* X_{i1} *i* X_{i2} *w powyższym modelu (dla ustalonego* i*) dany jest następująco*:

$$\Pr\{X_{i1} = 1, X_{i2} = 1\} = p_1p_2 + \varrho\sqrt{p_1p_2q_1q_2}, \tag{4.20}$$

$$\Pr\{X_{i1} = 0, X_{i2} = 0\} = q_1q_2 + \varrho\sqrt{p_1p_2q_1q_2}, \tag{4.21}$$

$$\begin{aligned}\Pr\{X_{i1} = 0, X_{i2} = 1\} &= q_1p_2 - \varrho\sqrt{p_1p_2q_1q_2},\\ \Pr\{X_{i1} = 1, X_{i2} = 0\} &= p_1q_2 - \varrho\sqrt{p_1p_2q_1q_2}.\end{aligned}$$

Dowód. Z faktu $\operatorname{Corr}(X_{i1}, X_{i2}) = \operatorname{Cov}(X_{i1}, X_{i2})/\sqrt{\operatorname{Var}(X_{i1})\operatorname{Var}(X_{i2})}$ wynika, że $\mathbb{E}(X_{i1}X_{i2}) = \operatorname{Corr}(X_{i1}, X_{i2})\sqrt{\operatorname{Var}(X_{i1})\operatorname{Var}(X_{i2})} + \mathbb{E}(X_{i1})\mathbb{E}(X_{i2}) = \varrho\sqrt{p_1q_1p_2q_2} + p_1p_2$. Z drugiej strony $\mathbb{E}(X_{i1}X_{i2}) = \Pr\{X_{i1}X_{i2} = 1\} = \Pr\{X_{i1} = 1, X_{i2} = 1\}$. Stąd otrzymujemy, że $\Pr\{X_{i1} = 1, X_{i2} = 1\} = p_1p_2 + \varrho\sqrt{p_1q_1p_2q_2}$. Przez symetrię można wykazać, że $\Pr\{X_{i1} = 0, X_{i2} = 0\} = q_1q_2 + \varrho\sqrt{p_1q_1p_2q_2}$. Wobec powyższego $\Pr\{X_{i1} = 0, X_{i2} = 1 \text{ lub } X_{i1} = 1, X_{i2} = 0\} = 1 - (p_1p_2 + \varrho\sqrt{p_1q_1p_2q_2} + q_1q_2 + \varrho\sqrt{p_1q_1p_2q_2}) = p_1 + q_1 - (p_1p_2 + q_1q_2 + 2\varrho\sqrt{p_1q_1p_2q_2}) = p_1q_2 + q_1p_2 - 2\varrho\sqrt{p_1q_1p_2q_2}$. Biorąc pod uwagę rozkład brzegowy otrzymujemy $\Pr\{X_{i1} = 1, X_{i2} = 0\} = \Pr\{X_{i1} = 1\} - \Pr\{X_{i1} = 1, X_{i2} = 1\}$ co daje, że $\Pr\{X_{i1} = 1, X_{i2} = 0\} = p_1q_2 - \sqrt{p_1p_2q_1q_2}\varrho$. Analogicznie rozumując otrzymujemy, że $\Pr\{X_{i1} = 0, X_{i2} = 1\} = q_1p_2 - \sqrt{p_1p_2q_1q_2}\varrho$. □

Wobec Definicji 1 oraz równości (4.20) i (4.21) prawdopodobieństwo P_o, że oceniający wystawią te same oceny, wynosi

$$\begin{aligned}P_o &= \Pr\{X_{i1} = 1, X_{i2} = 1\} + \Pr\{X_{i1} = 0, X_{i2} = 0\}\\ &= p_1p_2 + q_1q_2 + 2\varrho\sqrt{p_1p_2q_1q_2}.\end{aligned}$$

Przez P_c oznaczmy standardowo prawdopodobieństwo, że oceniający wystawią te same oceny mimo, że nie są one skorelowane. Prawdopodobieństwo zgody przez przypadek P_c wyliczamy podstawiając do P_o wartość $\varrho = 0$. Otrzymujemy, że $P_c = p_1p_2 + q_1q_2$.

TWIERDZENIE 12. *Dla opisywanego modelu współczynnik oceny stopnia zgody typu kappa równa się*

$$\kappa_C = 2\varrho\frac{\sqrt{p_1p_2q_1q_2}}{p_1q_2 + p_2q_1}. \tag{4.22}$$

Dowód. Przy P_o i P_c zdefiniowanych powyżej mamy:

$$\kappa_C = \frac{P_o - P_c}{1 - P_c} = \frac{p_1p_2 + q_1q_2 + 2\varrho\sqrt{p_1p_2q_1q_2} - (p_1p_2 + q_1q_2)}{1 - (p_1p_2 + q_1q_2)}$$
$$= \frac{2\varrho\sqrt{p_1p_2q_1q_2}}{p_1(1-p_2) + q_1(1-q_2)} = 2\varrho\frac{\sqrt{p_1p_2q_1q_2}}{p_1q_2 + q_1p_2}.$$

□

WNIOSEK 13. *Opisany model zakłada różne preferencje oceniających. Okazuje się, że przy takim założeniu współczynnik zgodności poprawiony o zgodę przez przypadek jest współczynnikiem korelacji pomnożonym przez pewien współczynnik zależny od rozkładów brzegowych.*

4.2. Estymacja w modelu binarnym. Dla próby N obiektów, obserwujemy liczebności par ocen każdego typu. Dane takie można zgromadzić jak poprzednio w Tabeli 5.

Wyznaczmy estymator współczynnika kappa κ_C oraz prawdopodobieństw p_1, p_2 metodą największej wiarogodności. Dla rozważanego modelu funkcja wiarogodności jest postaci

$$\begin{aligned} L(p_1, p_2, \varrho | n_{11}, n_{22}, n_{12}, n_{21}) &= \\ &(p_1p_2 + \varrho\sqrt{p_1p_2q_1q_2})^{n_{11}}(q_1q_2 + \varrho\sqrt{p_1p_2q_1q_2})^{n_{22}} \\ &\times(p_1q_2 - \varrho\sqrt{p_1p_2q_1q_2})^{n_{12}}(q_1p_2 - \varrho\sqrt{p_1p_2q_1q_2})^{n_{21}}. \end{aligned}$$

Szukamy wartości maksymalizujących powyższą funkcję. W celu obliczenia estymatora współczynnika κ_C dokonujemy podstawienia zgodnie z wzorem (4.22) otrzymując, że

$$\varrho\sqrt{p_1p_2q_1q_2} = \kappa_C(p_1q_2 + p_2q_1)/2.$$

Następnie przyrównujemy pochodne $\frac{\partial \ln L}{\partial p_1}$, $\frac{\partial \ln L}{\partial p_2}$ oraz $\frac{\partial \ln L}{\partial \kappa_C}$ do zera. Bloch i Kreamer [1] stwierdzili, że estymatory spełniające taki układ równań są postaci:

$$\hat{p}_1 = \frac{n_{11} + n_{12}}{N} = \frac{n_{1+}}{N},$$
$$\hat{p}_2 = \frac{n_{11} + n_{21}}{N} = \frac{n_{+1}}{N},$$
$$\hat{\kappa}_C = \frac{2(n_{11}n_{22} - n_{12}n_{21})}{n_{1+}n_{+2} + n_{+1}n_{2+}}, \tag{4.23}$$

gdzie dla uproszczenia wprowadzono notację adekwatną z Tabelą 5. Stąd

$$\hat{P}_o = \frac{n_{11} + n_{22}}{N}$$

oraz

$$\hat{P}_c = \frac{n_{1+}n_{+1}}{N^2} + \frac{n_{2+}n_{+2}}{N^2}$$

co jest zgodne z wzorami (3.1) oraz (4.19) dla $C = 2$.

UWAGA 14. Należy zwrócić uwagę na szczególne przypadki, gdy $p_1 = 0$ lub $p_2 = 0$. Dla $p_1 = 0$ i $p_2 \in (0, 1)$ rozkład łączny zmiennych losowych X_{i1} i X_{i2} dany jest następująco:

$$\begin{aligned} \Pr\{X_{i1} = 1, X_{i2} = 1\} = \Pr\{X_{i1} = 1, X_{i2} = 0\} &= 0, \\ \Pr\{X_{i1} = 0, X_{i2} = 0\} &= q_2, \\ \Pr\{X_{i1} = 0, X_{i2} = 1\} &= p_2. \end{aligned}$$

Dla $p_2 = 0$ i $p_1 \in (0, 1)$ rozkład łączny zmiennych losowych X_{i1} i X_{i2}dany jest następująco:

$$\begin{aligned} \Pr\{X_{i1} = 1, X_{i2} = 1\} = \Pr\{X_{i1} = 0, X_{i2} = 1\} &= 0, \\ \Pr\{X_{i1} = 0, X_{i2} = 0\} &= q_1, \\ \Pr\{X_{i1} = 1, X_{i2} = 0\} &= p_1. \end{aligned}$$

Oznacza to, że *dwie* spośród obserwowanych liczebności wynoszą zero n_{12} lub n_{21} oraz n_{11} lub n_{22}. W tych przypadkach estymator dany wzorem (4.23) wynosi 0. Jeśli natomiast $p_1 = p_2 = 0$ lub $p_1 = p_2 = 1$, to estymator nie jest dobrze określony. W praktyce oznacza to, że trzy liczebności n_{12}, n_{21} oraz n_{11} lub n_{22} wynoszą zero. Takie deterministyczne przypadki należy rozważać osobno. W zależności od planu eksperymentu można dookreślić estymator lub w ogóle nie określać stopnia zgody. Jesli $p_1 = 1 - p_2 = 1$ lub $p_2 = 1 - p_1 = 1$, to estymator wynosi -1, co dobrze odzwierciedla pełną niezgodę.

4.3. Model dla więcej niż dwóch kategorii. Model opisywany w poprzednich podrozdziałach określony był dla dwóch kategorii $C = 2$. Miara Cohena oceny stopnia zgodności określona jest dla dowolnej liczby $C \geq 2$. Podobnie jak dla współczynnnika kappa Scotta wykażemy, że estymator w modelu z wieloma kategoriami można uzyskać wykorzystując wartości współczynników kappa uzyskanych dla każdej z kategorii osobno, bazując na zasadzie podstawiania. Ponownie okazuje się, że współczynnik kappa jest średnią ważoną współczynników kappa uzyskanych dla modeli binarnych ze wstępną agregacją danych.

Niech X_{ij} (określająca ocenę i-tego obiektu wystawioną przez j-tego eksperta) będzie zmienną losową z rozkładu wielomianowego z parametrami

$(p_1, \ldots, p_C)$, $\sum_{l=1}^{C} p_l = 1$. Wyznaczmy na jej podstawie binarną ocenę przynależności do danej kategorii. Niech X_{ijl} $(i = 1, \ldots, N; j = 1, 2; l = 1, \ldots, C)$ będzie oceną przynależności i-tego obiektu do kategorii l wystawioną przez j-tego oceniającego. Dla każdej l-tej kategorii, wartości X_{ijl} zależą od ocen X_{ij} w następujący sposób:

$$X_{ijl} = \begin{cases} 1 & \text{gdy } X_{ij} = l, \\ 0 & \text{gdy } X_{ij} \neq l. \end{cases}$$

W następnym kroku dla każdej kategorii z osobna można zastosować model binarny opisany w poprzednich rozdziałach 4.1 oraz 4.2. Niech dla ustalonego l zmienna losowa X_{ijl} ma rozkład dwupunktowy z parametrem $p_{j,l}$ $(X_{ijl} \sim \mathcal{B}(1, p_{j,l}),\ i = 1, \ldots, N;\ j = 1, 2)$:

$$\Pr\{X_{ijl} = 1\} = p_{j,l}$$

oraz

$$\Pr\{X_{ijl} = 0\} = q_{j,l}\,.$$

Z założenia

$$q_{j,l} = 1 - p_{j,l} = \sum_{\substack{k=1 \\ k \neq l}}^{C} p_{j,k}\,.$$

Przeprowadzając analogiczne rozumowanie jak w dowodzie Stwierdzenia 11 wyznaczymy rozkład łączny zmiennych losowych X_{i1l} i X_{i2l}.

FAKT 15. *Niech dla ustalonego i oraz l $(i = 1, \ldots, N;\ l = 1, \ldots, C)$ korelacja $Corr(X_{i1l}, X_{i2l})$ wynosi ϱ_l. Rozkład łączny zmiennych losowych X_{i1l} i X_{i2l} dany jest następująco*:

$$\begin{aligned} \Pr\{X_{i1l} = 1, X_{i2l} = 1\} &= p_{1,l}p_{2,l} + \varrho_l\sqrt{p_{1,l}p_{2,l}q_{1,l}q_{2,l}}, \\ \Pr\{X_{i1l} = 0, X_{i2l} = 0\} &= q_{1,l}q_{2,l} + \varrho_l\sqrt{p_{1,l}p_{2,l}q_{1,l}q_{2,l}}, \\ \Pr\{X_{i1l} = 0, X_{i2l} = 1\} &= q_{1,l}p_{2,l} - \varrho_l\sqrt{p_{1,l}p_{2,l}q_{1,l}q_{2,l}}, \\ \Pr\{X_{i1l} = 1, X_{i2l} = 0\} &= p_{1,l}q_{2,l} - \varrho_l\sqrt{p_{1,l}p_{2,l}q_{1,l}q_{2,l}}\,. \end{aligned} \tag{4.24}$$

Zgodnie z Definicją 2, wobec równości (4.24),

$$P_o = \sum_{l=1}^{C} \Pr\{X_{i1l} = X_{i2l} = 1\} = \sum_{l=1}^{C} \left(p_{1,l}p_{2,l} + \varrho_l\sqrt{p_{1,l}p_{2,l}q_{1,l}q_{2,l}}\right).$$

Natomiast prawdopodobieństwo zgody przez przypadek wynosi

$$P_c = \sum_{l=1}^{C} p_{1,l}p_{2,l}\,.$$

Definicja 4. Współczynnik kappa Cohena w modelu z wieloma kategoriami definiujemy standardowo:

$$\kappa_C \overset{df}{=} \frac{P_o - P_c}{1 - P_c} = \frac{\sum_{l=1}^{C} \varrho_l \sqrt{p_{1,l} p_{2,l} q_{1,l} q_{2,l}}}{1 - \sum_{l=1}^{C} p_{1,l} p_{2,l}}. \tag{4.25}$$

Przypomnijmy, że $P_{o,l}$ oznacza prawdopodobieństwo, że oceniający zgodzą się w swojej ocenie w modelu binarnym

$$\begin{aligned} P_{o,l} &= \Pr\{X_{i1l} = X_{i2l} = 1\} + \Pr\{X_{i1l} = X_{i2l} = 0\} \\ &= p_{1,l} p_{2,l} + q_{1,l} q_{2,l} + 2\varrho_l \sqrt{p_{1,l} p_{2,l} q_{1,l} q_{2,l}} \end{aligned}$$

oraz

$$P_{c,l} = p_{1,l} p_{2,l} + q_{1,l} q_{2,l} = p_{1,l} p_{2,l} + (1 - p_{1,l})(1 - p_{2,l}). \tag{4.26}$$

Dla ustalonej l-tej kategorii współczynnik kappa Cohena (zgodnie z tezą Twierdzenia 12) wynosi:

$$\kappa_C^l = \frac{P_{o,l} - P_{c,l}}{1 - P_{c,l}} = 2\varrho_l \frac{\sqrt{p_{1,l} p_{2,l} q_{1,l} q_{2,l}}}{p_{1,l} q_{2,l} + p_{2,l} q_{1,l}}.$$

W następnym kroku chcemy wyrazić współczynnik κ_C dany wzorem (4.25) w terminach $P_{o,l}$ i $P_{c,l}$ oraz pokazać, że takie podejście jest równoważne z wyrażeniem współczynnika κ_C w zależności od κ_C^l oraz $P_{c,l}$.

Lemat 16. *Mamy*

$$P_c = \frac{1}{2} \left(\sum_{l=1}^{C} P_{c,l} - C + 2 \right).$$

Dowód. Z równości (4.26) wynika, że

$$p_{1,l} p_{2,l} = \frac{1}{2} \left(P_{c,l} + p_{1,l} + p_{2,l} - 1 \right).$$

Więc

$$P_c = \sum_{l=1}^{C} p_{1,l} p_{2,l} = \frac{1}{2} \sum_{l=1}^{C} (P_{c,l} + p_{1,l} + p_{2,l} - 1) = \frac{1}{2} \left(\sum_{l=1}^{C} P_{c,l} + 2 - C \right).$$

□

Lemat 17. *Mamy*

$$P_o = \frac{1}{2} \left(\sum_{l=1}^{C} P_{o,l} - C + 2 \right).$$

Dowód. Ponieważ

$$P_{o,l} = p_{1,l}p_{2,l} + (1-p_{1,l})(1-p_{2,l}) + 2\varrho_l\sqrt{p_{1,l}p_{2,l}(1-p_{1,l})(1-p_{2,l})}$$
$$= 2\left(p_{1,l}p_{2,l} + \varrho_l\sqrt{p_{1,l}p_{2,l}(1-p_{1,l})(1-p_{2,l})}\right) + 1 - p_{1,l} - p_{2,l},$$

dostajemy

$$p_{1,l}p_{2,l} + \varrho_l\sqrt{p_{1,l}p_{2,l}(1-p_{1,l})(1-p_{2,l})} = \frac{1}{2}(P_{o,l} - 1 + p_{1,l} + p_{2,l}).$$

To daje równość

$$P_o = \frac{1}{2}\sum_{l=1}^{C}(P_{o,l} - 1 + p_{1,l} + p_{2,l}) = \frac{1}{2}\left(\sum_{l=1}^{C}P_{o,l} - C + 2\right).$$

□

STWIERDZENIE 18. *Współczynnik κ_C w terminach $P_{o,l}$, $P_{c,l}$ oraz κ_C^l ma postać*:

$$\kappa_C = \frac{\sum_{l=1}^{C}(P_{o,l} - P_{c,l})}{\sum_{l=1}^{C}(1 - P_{c,l})} \tag{4.27}$$

$$= \frac{\sum_{l=1}^{C}P_{o,l} - \sum_{l=1}^{C}P_{c,l}}{C - \sum_{l=1}^{C}P_{c,l}} \tag{4.28}$$

$$= \frac{\sum_{l=1}^{C}(1 - P_{c,l})\kappa_C^l}{\sum_{l=1}^{C}(1 - P_{c,l})} \tag{4.29}$$

Dowód. Korzystając z Lematów 16 oraz 17 możemy uzasadnić równość (4.28) w następujący sposób

$$\kappa_C = \frac{P_o - P_c}{1 - P_c} = \frac{\frac{1}{2}(\sum_{l=1}^{C}P_{o,l} - C + 2) - \frac{1}{2}(\sum_{l=1}^{C}P_{c,l} - C + 2)}{1 - \frac{1}{2}(\sum_{l=1}^{C}P_{c,l} - C + 2)}$$
$$= \frac{\sum_{l=1}^{C}P_{o,l} - \sum_{l=1}^{C}P_{c,l}}{C - \sum_{l=1}^{C}P_{c,l}}.$$

Wymnażając kolejne składniki sumy występującej w liczniku wyrażenia (4.27) przez $\frac{1-P_{c,l}}{1-P_{c,l}}$ wobec faktu, że $\kappa_C^l = \frac{P_{o,l}-P_{c,l}}{1-P_{c,l}}$ otrzymujemy równość daną wzorem (4.29). □

FAKT 19. *Współczynnik κ_C jest średnią ważoną współczynników κ_C^l z wagami wynoszącymi $w_l = 1 - P_{c,l}$.*

Własność opisana w Fakcie 19 była ogólnie znana [8, str. 606], jednakże podane uzasadnienie bazowało na numerycznych przykładach. Wykażemy, że podstawiając do wzoru (4.28) estymatory wielkości $P_{o,l}$ i $P_{c,l}$ uzyskane dla modeli binarnych otrzymujemy oryginalny estymator zaproponowany przez Cohena dla wielu kategorii, dla którego $\hat{P}_c$ dane jest wzorem (4.19), a $\hat{P}_o$ standardowo wzorem (3.1).

Liczebności n_{lk} $(l, k = 1, \ldots, C)$ zgromadzone w Tabeli 1 przekształcamy dokonując dychotomizacji względem każdej z kategorii z osobna, jak to zostało opisane w rozdziale 3.3.1. Przypomnijmy, że

$$\hat{P}_{o,l} = \frac{n_{ll}}{N} + \frac{N - (n_{l+} + n_{+l}) + n_{ll}}{N}.$$

LEMAT 20. *Mamy*

$$\sum_{l=1}^{C} \hat{P}_{o,l} = 2\hat{P}_o + C - 2 \tag{4.30}$$

oraz

$$\sum_{l=1}^{C} \hat{P}_{c,l} = 2\hat{P}_c + C - 2, \tag{4.31}$$

gdzie $\hat{P}_c$ dane jest wzorem (4.19), a $\hat{P}_o$ wzorem (3.1).

Dowód. Równość (4.30) została wykazana w Lemacie 8. Pozostaje wykazać równość (4.31). Mamy

$$\begin{aligned}\sum_{l=1}^{C} \hat{P}_{c,l} &= \sum_{l=1}^{C} \left(\frac{n_{l+}}{N} \frac{n_{+l}}{N} + \frac{N - n_{l+}}{N} \frac{N - n_{+l}}{N} \right) \\ &= \sum_{l=1}^{C} \left(2 \frac{n_{l+}}{N} \frac{n_{+l}}{N} + 1 - \frac{n_{l+}}{N} - \frac{n_{+l}}{N} \right) \\ &= 2 \sum_{l=1}^{C} \left(\frac{n_{l+}}{N} \frac{n_{+l}}{N} \right) + C - 1 - 1 = 2\hat{P}_c + C - 2.\end{aligned}$$

□

TWIERDZENIE 21. *Bazując na uzyskanych w modelach binarnych estymatorach $\hat{P}_{o,l}$ oraz $\hat{P}_{c,l}$ estymator współczynnika kappa dla modelu z wieloma kategoriami wynosi*

$$\hat{\kappa}_C = \frac{\sum_{l=1}^{C} \hat{P}_{o,l} - \sum_{l=1}^{C} \hat{P}_{c,l}}{C - \sum_{l=1}^{C} \hat{P}_{c,l}} = \frac{\hat{P}_o - \hat{P}_c}{1 - \hat{P}_c},$$

gdzie $\hat{P}_o$ oraz $\hat{P}_c$ zostały zaproponowane przez Cohena wzorami (4.19) i (3.1).

Dowód. Na mocy Lematu 20 otrzymujemy

$$\hat{\kappa}_C = \frac{\sum_{l=1}^{C}\hat{P}_{o,l} - \sum_{l=1}^{C}\hat{P}_{c,l}}{C - \sum_{l=1}^{C}\hat{P}_{c,l}} = \frac{C - 2 + 2\hat{P}_o - (C - 2 + 2\hat{P}_c)}{C - (C-2) - 2\hat{P}_c} = \frac{\hat{P}_o - \hat{P}_c}{1 - \hat{P}_c}.$$

□

4.4. Wariancja, wartośc maksymalna i minimalna. Wykorzystując metodę opisaną wcześniej w rozdziale 3.4, można przybliżyć asymptotyczną wariancję estymatora współczynnika kappa Cohena dla $C = 2$ następująco [1]:

$$\mathrm{Var}(\hat{\kappa}_C) = \frac{4(p_1p_2q_1q_2)}{(p_1q_2 + p_2q_1)^2}U,$$

gdzie

$$U = 1 + 4U_{X_1}U_{X_2}\varrho - (1 + 3U_{X_1}^2 + 3U_{X_2}^2)\varrho^2 + 2U_{X_1}U_{X_2}\varrho^3,$$

$$U_{X_1} = \frac{\frac{1}{2} - p_1}{\sqrt{p_1q_1}},$$

$$U_{X_2} = \frac{\frac{1}{2} - p_2}{\sqrt{p_2q_2}},$$

oraz współczynnik ϱ dany jest w modelu przez

$$\varrho = \frac{\kappa_C(p_1q_2 + p_2q_1)}{2\sqrt{p_1p_2q_1q_2}}.$$

Estymator wariancji otrzymamy wstawiając do powyższego wzoru uzyskane wcześniej metodą największej wiarogodności estymatory wielkości $\hat{p}_1$, $\hat{p}_2$ oraz $\hat{\kappa}_C$. Wartość największą jaką przyjmuje estymator jest równa 1, gdy $P_o = 1$. Wartość minimalna jest z przedziału $[-1{,}0]$ i zależy od rozkładów brzegowych oraz liczby kategorii [5]. Jednakże, jak argumentuje Cohen, jeśli wartość współczynnika jest mniejsza od zera, oznacza to, że obserwowana zgodność jest mniejsza niż ta oczekiwana losowo, co z punktu widzenia praktyka już samo świadczy o zasadniczym braku zgodności.

5. Własności analizowanych miar. Rozdział ten zawiera opis zagadnień będących swoistą charakterystyką współczynników kappa:
(1) sposobów interpretacji uzyskanych wartości [16, 8, 4],
(2) zależności kappy Cohena od rozpowszechnienia badanej cechy w populacji oraz jednorodności rozkładów brzegowych [9, 13, 2, 17, 7, 3, 11, 20, 12, 22, 19, 15, 14],
(3) problemów technicznych pojawiających się przy wykorzystaniu do obliczeń pakietów statystycznych [14, 10, 6, 18].

5.1. Interpretacje. Testowanie hipotezy o istotności współczynnika kappa z punktu widzenia zastosowań ma znikomą wartość poznawczą. Powszechnie stosowaną przez praktyków skalą porównawczą dla współczynnika Cohena są wartości progowe podane przez Landisa i Kocha [16] przedstawione w Tabeli 8. Jak piszą autorzy, progi te są arbitralne, jednakże mogą być przydatne przy porównaniach. Fleiss i inni [8] proponują uproszczenia, które znajdują się w Tabeli 9. W Tabeli 10 zebrano nieco odmienne uproszczone wartości progowe zaproponowane przez Cicchetti i innych [4].

Kappa	Interpretacja
$< 0{,}00$	brak zgodności
$0{,}00 - 0{,}20$	słaba
$0{,}21 - 0{,}40$	średnia
$0{,}41 - 0{,}60$	umiarkowana
$0{,}61 - 0{,}80$	pokaźna
$0{,}81 - 1{,}00$	prawie perfekcyjna

Tablica 8. Interpretacja według Landisa i Kocha wartości przyjmowanych przez współczynnik κ_C.

Kappa	Interpretacja
$< 0{,}40$	słaba zgodność
$0{,}40 - 0{,}74$	umiarkowana lub dobra
$0{,}75 - 1{,}00$	perfekcyjna

Tablica 9. Interpretacja wartości współczynnika κ_C według Fleissa.

Kappa	Interpretacja
$< 0{,}40$	słaba
$0{,}40 - 0{,}59$	umiarkowana
$0{,}60 - 0{,}74$	dobra
$0{,}75 - 1{,}00$	wyśmienita

Tablica 10. Interpretacja według Cicchetti i innych wartości przyjmowanych przez współczynnik κ_C.

5.2. Zależność od rozkładów brzegowych i rozpowszechnienia. Problemy interpretacji wartości współczynnika kappa, jego wad, zależności od rozpowszechnienia oraz prawdopodobieństw brzegowych, były tematem niejednokrotnie podejmowanym w literaturze.

Przykładem zachowania się wartości współczynników κ_S oraz κ_C są dane zgromadzone w Tabeli 11, gdzie przedstawiono zmienność wartości kapp mimo, iż obserwowana zgodność jest taka sama we wszystkich przypadkach i wynosi $\hat{P}_o = 0,80$. Zmianom podlega tylko rozkład wartości na poszczególne komórki tabeli. Przy różnych układach wartości współczynnika stopnia zgody wahają się od 0,22 do 0,62.

Kategoria	1	2	κ_S	κ_C	Poziom zgody wg Landisa i Kocha
1	75	10			
2	10	5	0,22	0,22	średni
Kategoria	1	2	κ_S	κ_C	
1	70	10			
2	10	10	0,38	0,38	średni
Kategoria	1	2	κ_S	κ_C	
1	40	10			
2	10	40	0,60	0,60	umiarkowany
Kategoria	1	2	κ_S	κ_C	
1	75	0			
2	20	5	0,22	0,27	średni
Kategoria	1	2	κ_S	κ_C	
1	70	0			
2	20	10	0,38	0,41	umiarkoway
Kategoria	1	2	κ_S	κ_C	
1	40	0			
2	20	40	0,60	0,62	pokaźny

TABLICA 11. Przykłady zachowania współczynników κ_C oraz κ_S.

Zwróćmy uwagę, że im większa koncentracja ocen w jednej z komórek na przekątnej, która odzwierciedla obserwacje zgodne, tym mniejsza wartość współczynników kappa. „Przesuwanie" niezgód tj. liczby przypadków, dla których oceniający dali odmienne oceny, nie prowadzi do istotnej zmiany w wartości współczynnika zgody.

Cicchetti i Feinstein [3] podkreślają fakt, że w badanich nad zgodnością można rozpatrywać zgodność na każdej z kategorii osobno, jako bardziej efektywny sposób analizy danych. Definicje modeli tego typu były już rozważane w pracy w kontekście zdefiniowania i określenia współczynników dla wielu kategorii.

W świetle przykładów zawartych w Tabeli 11 wszelkie odniesienia do wartości progowych (zawartych w Tabelach 8, 9, 10) mogą być mylące dla praktyków i mogą powodować niewłaściwą interpretację. Wydaje się, iż najwłaściwszym postępowaniem obok podania miar typu kappa jest prezentacja bazowych tabel kontyngencji, pokazujących dokładną istotę badanych zjawisk.

5.2.1. Nieokreśloność estymatorów współczynników kappa. W tym miejscu należy również zwrócić uwagę na problem nieokreśloności estymatorów współczynników kappa Scotta oraz kappa Cohena sygnalizowany już wcześniej w Uwagach 3 oraz 14.

W przypadku, gdy obu oceniających zaklasyfikuje wszystkie obiekty zgodnie tylko do jednej i tej samej kategorii estymatory współczynników κ_S oraz κ_C nie są dobrze określone. Naturalnym rozwiązaniem problemu może się wydawać dookreślenie tychże estymatorów współczynników poprzez przyjęcie wartości 1 w takich sytuacjach. Jednakże, niektórzy autorzy krytykują takie podejście argumentując, że dane tego typu oznaczają, iż oceniający nie potrafią rozróżnić obiektów i takie eksperymenty nie powinny być oceniane pod kątem zgodności. W naszym odczuciu wybór podejścia zależy od intencji badacza i typu eksperymentu. Bez podjęcia jednoznacznej decyzji problem ten może być istotnym źródłem błędu wyników numerycznych.

5.3. Implementacje w pakietach statystycznych. Większość popularnych pakietów statystycznych takich jak SAS, SPSS posiada zaimplementowane metody obliczeń współczynnika kappa Cohena. Jednakże, wielkości te mogą być błędnie wyliczone w przypadku otrzymania niezrównoważonych tablic kontyngencji (ang. *unbalanced tables*). Takie przypadki to:

(1) niekwadratowa tablica; gdy liczba kategorii wykorzystanych do oceny N obiektów przez oceniającego A jest różna od tej wykorzystanej przez oceniającego B. W tym przypadku wyliczona tablica kontyngencji nie jest kwadratowa i współczynnik nie jest podany. W Tabeli 12 przedstawiono przykładową pełną tablicę kontyngencji otrzymaną w wyniku oceny grupy obiektów. Natomiast Tabela 13 jest jej obrazem uzykanym w pakiecie statystycznym.

		Oceniający B			
	Kategoria	1	2	3	Ogółem
	1	n_{11}	n_{12}	0	n_{1+}
Oceniający	2	n_{21}	n_{22}	0	n_{2+}
A	3	n_{31}	n_{32}	0	n_{3+}
	Ogółem	n_{+1}	n_{+2}	$n_{+3} = 0$	N

TABLICA 12. Przykład pełnej tablicy kontyngencji jako wyniku oceny grupy obiektów.

	Kategoria	Oceniający B 1	2	Ogółem
	1	n_{11}	n_{12}	n_{1+}
Oceniający	2	n_{21}	n_{22}	n_{2+}
A	3	n_{31}	n_{32}	n_{3+}
	Ogółem	n_{+1}	n_{+2}	N

Tablica 13. Przykład niekwadratowej tablicy kontyngencji.

(2) nieregularna kwadratowa tablica; gdy liczba kategorii wykorzystanych do oceny N obiektów jest taka sama dla oceniającego A i oceniającego B, ale nie jest to ten sam zbiór ocen. Prowadzi to do wyliczenia współczynnika κ_C na podstawie kwadratowej tabeli kontyngencji, gdzie na przekątnej *nie leżą* przypadki zgodne (patrz Tabela 14). W tej sytuacji współczynnik kappa zostaje wyliczony, ale błędnie. Wielkości leżące na przekątnej są traktowane jako obserwacje zgodne, co prowadzi do niepoprawnych obliczeń.

	Kategoria	Oceniający B 1	2	3	Ogółem
	1	n_{11}	n_{12}	n_{13}	n_{1+}
Oceniający	2	n_{21}	n_{22}	n_{23}	n_{2+}
A	4	n_{41}	n_{42}	n_{43}	n_{4+}
	Ogółem	n_{+1}	n_{+2}	n_{+3}	N

Tablica 14. Przykład kwadratowej tablicy kontyngencji o niejednakowych zbiorach użytych kategorii.

Oba typy nieregularności powstają z powodu usuwania przez pakiety komputerowe z tablicy kontyngencji o maksymalnym rozmiarze kolumn lub wierszy o elementach równych zero. Problem ten potrafi być istotnym źródłem błędu wyników w badaniach empirycznych. Został on opisany w pracach Liu i Hays [18], Crewsona [6] oraz Gweta [10]. W publikacjach tych autorzy podają również różne sposoby uniknięcia problemu (dla pakietu SAS), w celu otrzymania poprawnych wyników.

Jedną z metod jest dodanie sztucznych obserwacji w ten sposób, aby w każdej kategorii znalazł się przynajmniej jeden obiekt dla każdego oceniającego. Następnie sztucznie dodanym obserwacjom przypisuje się znikome wagi, tak aby ich wprowadzenie nie miało istotnego wpływu na wynik numeryczny. Drugą metodą jest uzupełnienie tabeli kontyngencji wierszami lub kolumnami złożonymi z zer, w celu otrzymania tablicy zrównoważonej (ang. *balanced*). Inny pomysł polega na przypisaniu oryginalnych danych do jednej warstwy, a do drugiej przypisanie wszystkich kombinacji zero-jedynkowych możliwych par ocen [6].

6. Przykład. Zastosowanie omawianych w pracy współczynników oceny stopnia zgody typu kappa zilustrujemy na przykładzie danych pochodzących z treningów wykorzystania narzędzi oceny psychiatrycznej stanu zdrowia pacjenta przeprowadzonych w ramach międzynarodowego badania EDEN([1]).

Ocenie poddano $N = 68$ pacjentów. Każdy z nich był oceniany przy pomocy skali GSDS (Groningen Social Disability Scale). Narzędzie to pozwala ocenić stopień niesprawności w zakresie 8 ról społecznych. Dla każdej roli oceniamy najpierw jej różne wymiary, a następnie na ich podstawie wystawiamy ocenę ogólną. Zastosowanie ma skala o 5 poziomach:

[0] – brak niesprawności w zakresie roli,
[1] – niewielka niesprawność,
[2] – średnia niesprawność,
[3] – znacząca/duża niesprawność,
[8] – nie potrafię określić, nie wiem, mam za mało danych.

W Tabeli 15 zgromadzono wyniki uzyskane przez dwóch lekarzy dla ogólnej oceny wybranej jednej roli – „Dbanie o siebie".

		Oceniający B					
	Kategoria	0	1	2	3	8	Ogółem
	0	25	2	1	0	3	31
	1	4	12	6	3	1	26
Oceniajacy A	2	0	2	4	3	0	9
	3	0	0	1	0	0	1
	8	0	0	0	0	1	1
	Ogółem	29	16	12	6	5	$N = 68$

Tablica 15. Przykładowe dane pochodzące z badania EDEN.

Dla zaprezentowanych danych wyliczone współczynniki typu kappa wynoszą:
(1) $\hat{\kappa}_S = 0{,}4390$ przy $\hat{P}_o = 0{,}6176$ oraz $\hat{P}_c = 0{,}3184$,
(2) $\hat{\kappa}_C = 0{,}4458$ przy $\hat{P}_o = 0{,}6176$ oraz $\hat{P}_c = 0{,}3101$.

Rezultaty otrzymane w modelach binarnych uzyskanych dla każdej z kategorii z osobna zostały ujęte w Tabeli 16.

Zauważmy, że emipryczne wyniki potwierdzają zależności wykazane w Twierdzeniach 10 oraz 21.

([1]) Akronim EDEN pochodzi od nazwy badania „**E**uropean **D**ay hospital **E**valuatio**N**". Szczegółowy opis badania znajduje się pod adresem internetowym: `www.edenstudy.com`

0 – brak	1	0	$\hat{\kappa}_S$	$\hat{\kappa}_C$
1	25	6		
0	4	33	0,7018	0,7020
1 – niewielka	1	0	κ_S	κ_C
1	12	14		
0	4	38	0,3799	0,3953
2 – średnia	1	0	κ_S	κ_C
1	4	5		
0	8	51	0,2679	0,2706
3 – znacząca	1	0	κ_S	κ_C
1	0	1		
0	6	61	-0,0543	-0,0259
8 – brak oceny	1	0	κ_S	κ_C
1	1	0		
0	4	63	0,3026	0,3166

TABLICA 16. Wyniki dla tabel uzyskanych dla każdej z kategorii osobno.

Literatura

[1] D.A. Bloch and H.Ch. Kraemer, *2x2 Kappa Coefficients: Measures of Agreement or Association* Biometrics, 45:269 – 287, 1989.

[2] T. Byrt, J. Bishop, and J.B. Carlin, *Bias, prevalence and kappa*, Journal of Clinical Epidemiology, 46(5):423 – 429, 1993.

[3] D.V. Cicchetti and A.R. Feinstein, *High Agreement but low kappa: II. Resolving the paradoxes*, Journal of Clinical Epidemiology, 43(6):551 – 558, 1990.

[4] D.V. Cicchetti, F. Volkmar, S.S. Sparrow, and D. Cohen, *Assessing the Reliability of Clinical Scales When the Data Have Both Nominal and Ordinal Features: Proposed guidelines for neuropsychological assessments*, Journal of Clinical and Experimantal Neuropsychology, 14(5):673 – 686, 1992.

[5] J. Cohen. *A Coefficient of Agreement for Nominal Scales*, Educational and Psychological Measurement, 20(1):37 – 46, 1960.

[6] P.E. Crewson, *A Correction for Unbalanced Kappa Tables SAS 6.12*, Proceedings of the 26-th Annual SAS Users Group International Conference, 194:1 – 3, 2001.

[7] A.R. Feinstein and D.V. Cicchetti, *High Agreement but low kappa: I. The problems of two paradoxes*, Journal of Clinical Epidemiology, 43(6):543 – 549, 1990.

[8] J.L. Fleiss, B. Levin, and M.C. Paik, *Statistical Methods for Raters and Proportions.* Wiley and Sons, Third edition, 2003.

[9] I. Guggenmoos-Holzman, *How reliable are chance-corrected measures of agreement?*, Statistics in Medicine, 12:2191 – 2205, 1993.

[10] K. Gwet, *Computing Inter-rater reliability with the SAS system*, Statistical Methods For Inter-Rater Reliability Assessment, 3:1 – 16, 2002.

[11] K. Gwet, *Inter-Rater Reliability: Dependency on Trait Prevalence and Marginal Homogenity*, Statistical Methods For Inter-Rater Reliability Assessment, 2:1 – 9, 2002.

[12] K. Gwet, *Kappa Statistic is not Satisfactory for Assessing the Extent of Agreement Between Raters*, Statistical Methods For Inter-Rater Reliability Assessment, 1:1 – 6, 2002.
[13] F.K. Hoehler, *Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity*, Journal of Clinical Epidemiology, 53:499 – 503, 2000.
[14] J. Jarosz-Nowak, *On same methods of assessment of interrater reliability for nominal and ordinal scale*, Prace naukowe Instytutu Matematyki Politechniki Wrocławskiej „II Konferencja dla Młodych Matematyków Zastosowania Matematyki, Lądek Zdrój 2001", 25(4):41 – 49, 2003.
[15] J. Jarosz-Nowak, *Modified kappa as a measure of interrater agreement*, Proceedings of the Tenth National Conference on Application of Mathematics in Biology and Medicine, Święty Krzyż, Poland, pages 79 – 83, 2004.
[16] J.R. Landis and G.G. Koch, *The Measurement of Observer Agreement for Categorical Data* Biometrics, 33:159 – 174, 1977.
[17] Ch.A. Lantz and E. Nebenzahl, *Behavior and Interpretation of the κ Statistics: Resolution of the Two Paradoxes*, Journal of Clinical Epidemiology, 49(4):431 – 434, 1996.
[18] H. Liu and R.D. Hays, *Measurement of Interrater Agreement A SAS/IML Macro Kappa Procedure for Handling Incomplete Data*, Proceedings of the 24-th Annual SAS Users Group International Conference, 280:1620 – 1625, 1999.
[19] M. Maclure and W.C. Willett, *Misinterpretation and misuse of the kappa statistic*, American Journal of Epidemiology, 126(2):161 – 169, 1987.
[20] J.M. Sargeant and S.W. Martin, *The dependence of kappa on attribute pevalence when assessing the repeatability of questionnaire data*, Preventive Veterinary Medicine, 34:115 – 123, 1998.
[21] W.A. Scott, *Reliability of Content Analysis: The Case of Nominal Scale Coding*, The Public Opinion Quarterly, 19(3):321 – 325, 1955.
[22] W.D. Thompson and S.D. Walter, *A reappraisal of the kappa coefficient*, Journal of Clinical Epidemiology, 41(10):949 – 958, 1988.

Instytut Matematyki i Informatyki, Politechnika Wrocławska
Wyb. Wyspiańskiego 27, 50–370 Wrocław
E-mail: Joanna.Jarosz-Nowak@pwr.wroc.pl

Models of assessing the extent of agreement beteween raters using kappa coefficients

Abstract. In medical studies quality of assessment is of great importance. Typically it is characterized by reliability. A fundamental sense of interobserver reliability is to evaluate a degree of agreement between independent judges examining the same objects. The paper is an evaluation of some interrater reliability measures and their interpretation. Cohen's kappa and Scott's coefficient are considered. We describe models and connections between coefficients defined for dichotomous and politomous data. We show that abovementioned estimators of kappa for classification into more than 2 categories are weighted averages of kappas in binary models defined for each category separately.
Key words: Agreement, Cohen's kappa, Scott's coefficient

(wpłynęło 3 czerwca 2007 r.)

MATEMATYKA STOSOWANA 8, 2007

Ilona Kopocińska (Wrocław)
Bolesław Kopociński (Wrocław)

Zagadnienie Steinhausa o szacowaniu strat wojennych na podstawie analizy nekrologów prasowych

Józefowi Łukaszewiczowi
w osiemdziesiątą rocznicę Urodzin

Streszczenie. Przedmiotem noty jest próba rekonstrukcji, wzmiankowanej we wspomnieniach Hugona Steinhausa, metody oszacowania strat wojennych w armii Niemieckiej Rzeszy podczas drugiej wojny światowej przy wykorzystaniu ówczesnych nekrologów prasowych. Zakładając dla wielkości rodziny niemieckiej rozdęty rozkład geometryczny i wprowadzając pewne przekształcenia tej zmiennej losowej oraz biorąc pod uwagę rozkład treści nekrologów, wskazujemy na możliwość oszacowania frakcji poległych w armii.
Słowa kluczowe: metoda klepsydr Steinhausa, rozdęty rozkład geometryczny, rozrzedzenie dwumianowe, estymacja parametrów.

1. Wprowadzenie. W przekazach o dziełach Profesora Hugona Steinhausa znajdujemy wzmiankę o sugerowanej przez Niego możliwości oszacowania strat wojennych Niemieckiej Rzeszy poniesionych podczas drugiej wojny światowej, na podstawie analizy treści nekrologów prasowych. Problem był wielokrotnie przywoływany przez Profesora na seminarium zastosowań matematyki we Wrocławiu na przełomie lat pięćdziesiątych i sześćdziesiątych ubiegłego stulecia. Jest o nim wzmianka we *Wspomnieniach i zapiskach* (patrz [4], s. 520): *23 IX 1963. Parę dni temu zjawił się u mnie pan Gdula, któremu dałem jako temat pracy magisterskiej zbadanie strat niemieckich w zabitych, chciałem przekonać się, jak moja metoda klepsydr (Osiczyna, styczeń 1942) wygląda, gdy ma się tych klepsydr więcej. Pan Gdula znalazł we Wrocławiu komplety „Völkischer Beobachter" obejmujące kilkaset klepsydr. W dziele Schirera (The Rise and Fall of the Third Reich) autor podaje liczby z diariusza feldmarszałka Holdera, który objął po Brauchitschu komendę nad całym rosyjskim frontem; liczby te obejmują czas od*

22 VI 1941 do 28 II 1942 roku i tylko front rosyjski (nie licząc Włochów i Rumunów czy też Węgrów). Okazuje się niekonsekwencja tych dat, tak że można raczej kontrolować je moim sposobem, niż przeciwnie.

Nie są nam znane informacje o rozwiązaniu wspomnianego zagadnienia przez Hugona Steinhausa, Edmunda Gdulę czy też kogoś innego. Niniejsza nota jest więc próbą rekonstrukcji matematycznego modelu prowadzącego do pewnego pozytywnego rezultatu, jednakże bez weryfikacji wyników na podstawie danych liczbowych.

Idąc tropem Steinhausa, przyjmując założenia modelu, zobowiązani jesteśmy uwzględnić ograniczone możliwości dostępu do danych demograficznych Niemiec w warunkach okupowanej Polski, co zmusza do daleko idących uproszczeń. Przyjmiemy zatem, naszym zdaniem, rozsądne założenia dotyczące rodziny i reprezentatywność próby nekrologów. Dalej przy opisie probabilistycznym komentujemy zakładaną równość prawdopodobieństw i niezależność rozważanych zdarzeń losowych.

Podstawowym obiektem naszych rozważań jest rodzina niemiecka. Przyjmujemy, że ma ona D dzieci, w tym S synów, spośród których W wcielono do wojska, a z nich Z poległo. Ogólna liczba żołnierzy N może być nieznana – przedmiotem estymacji jest liczba M żołnierzy poległych do ustalonej daty albo lepiej frakcja $z = M/N$ poległych. W tym celu, zgodnie z ideą Steinhausa, mamy wykorzystać nekrologi prasowe. Nekrologi zawierają następujące treści: A_0 – nasz jedyny syn poległ, A_j – nasz j-ty ($j \geq 1$) syn poległ. Odnotujmy, że zdarzenie A_1, czyli „nasz pierwszy syn poległ" zawiera A_0, czyli „nasz jedyny syn poległ". Wyróżnienie A_0 i wprowadzenie $A_1^* = A_1 \setminus A_0$, co pozostało w naszej pamięci, pochodzi od Steinhausa. Zwiększa ono liczbę zdarzeń, które w praktyce mogą być wzięte pod uwagę.

2. Model. Główną rolę w naszych rozważaniach gra zmienna losowa $G = G_{\alpha,p}$ mająca rozkład rozdęty geometryczny:

$$P(G_{\alpha,p} = 0) = 1 - \frac{\alpha p}{q},$$
$$P(G_{\alpha,p} = n) = \alpha p^n, \quad 0 < p < 1, \quad q = 1 - p, \quad 0 < \alpha < q/p, \quad n \geq 1.$$

Niech X będzie zmienną losową o rozkładzie $P(X = n) = p_n$, $n \geq 0$. Weźmy pod uwagę operator B z parametrem b, który działa na X dając zmienną $B_b X$ o rozkładzie $P(B_b X = k) = \sum_{n=k}^{\infty} p_n b(k; n, b)$, gdzie $b(k; n, b) = \binom{n}{k} b^k (1-b)^{n-k}$. Łatwo sprawdzić, że $B_a B_b X = B_{ab} X$, $B_b G_{\alpha,p} =^d G_{\alpha',p'}$, gdzie $\alpha' = \frac{\alpha}{q+bp}$, $p' = \frac{bp}{q+bp}$. Odnotujmy, że

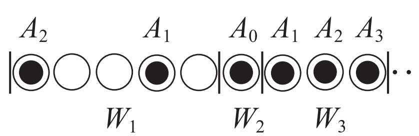
$$P(B_b G_{\alpha,p} = 0) = 1 - \frac{\alpha b p}{q(q + bp)},$$

$$P(B_b G_{\alpha,p} = n) = \frac{\alpha}{q+bp}\Big(\frac{bp}{q+bp}\Big)^n, \quad n \geq 1.$$

Przechodzimy do analizy zagadnienia Steinhausa. Założenia modelu są następujące: za Lotką [1] przyjmujemy, że $D = G_{\alpha,p}$, $S = B_s D$, $W = B_w S =^d B_{sw} D$, $Z = B_z W =^d B_{swz} D$, gdzie $0 < s < 1$, $0 < w < 1$, $0 < z < 1$. Operator B (patrz [3]) zwany rozrzedzeniem dwumianowym, zawiera *implicite* założenie prób Bernoulliego. Zakładamy więc w szczególności, że śmierć spotyka żołnierza niezależnie od liczebności rodziny i liczby żołnierzy służących w armii, należących do owej rodziny. Są to być może daleko idące uproszczenia, ale ich zmiana wymagałaby znajomości nowych, trudnych do wyestymowania parametrów. W szczególności przyjęcie jednakowego w dla braci wynika z braku wiedzy o strukturze wieku w rodzinach.

Ogół żołnierzy numerujemy liczbami $1, 2, \ldots, N$; braciom w wojsku przypisujemy numery kolejne. Zatem żołnierze będący w grupach rodzinnych utworzą dyskretny proces odnowy $1, 2, \ldots, W_1, W_1+1, \ldots, W_1+W_2, W_1+W_2+1, \ldots$, gdzie cykle procesu odnowy $W_1, W_2, \ldots$ są niezależnymi zmiennymi losowymi o rozkładzie G z odpowiednimi parametrami. Każdemu żołnierzowi zabitemu przypisujemy jako markę nekrolog jemu poświęcony. Cykl charakteryzujemy jego długością W oraz markami: liczbą synów S, liczbą zabitych Z i treścią ewentualnego nekrologu A od tej rodziny, losowanego ze wszystkich nekrologów możliwych w tym cyklu. Mamy więc do czynienia z markowanym procesem odnowy (patrz [2]). Marka S jest użyteczna przy analizowaniu nekrologów treści A_0. Zauważmy, że odczytany nekrolog umieszczamy w kontekście stanu jego grupy rodzinnej.

Rys. 1 przedstawia fragment omawianego procesu odnowy. Kółko oznacza żołnierza – poległych zaczerniono. Są tam trzy cykle $W_1 = 5$, $W_2 = 1$, $W_3 = 3$, dla nich $Z_1 = 2$, $Z_2 = 1$, $Z_3 = 3$. Na rysunku umieszczono ewentualne nekrologi. Nekrolog, który do nas dociera, jest losowany spośród tych, które są w cyklu.



Rys. 1. Fragment sekwencji żołnierzy.

Niech $H_j(N)$ oznacza oczekiwaną liczbę nekrologów o treści A_j w zbiorze żołnierzy $\{1, 2, \ldots, N\}$. W [2] znajdujemy asymptotykę tych wielkości. Nam wystarczy pierwszy wyraz podanego tam rozwinięcia asymptotycznego:

$$H_j(N) = \frac{N}{\mathrm{E}(W)} P(A = A_j) + O(1), \quad N \to \infty, \quad j \geq 0.$$

Pozostają więc do obliczenia prawdopodobieństwa $P(A = A_j) = a_j$, $j \geq 0$. Niech $a_1^* = P(A_1^*) = a_1 - a_0$. Oznaczmy przez π prawdopodobieństwo zamieszczenia w prasie nekrologu przez rodziców i odnalezienie jego przez nas. Nie będziemy identyfikowali rodzin, gdyby powtarzały się one w odczytywanych nekrologach. Można sprawdzić (patrz Dodatek 1), że

$$a_0 = \frac{\pi\alpha swzp}{(q+sp)^2}, \quad a_j = \pi\alpha_z \sum_{k=j}^{\infty} \frac{1}{k} p_z^k, \quad j \ge 1, \tag{1}$$

gdzie $\alpha_z = \dfrac{\alpha}{q+swzp}$, $p_z = \dfrac{swzp}{q+swzp}$.

3. Estymacja parametrów modelu. Zbieranie nekrologów daje możliwość oszacowania frakcji zdań określonej treści. Empirycznie możemy bowiem oszacować frakcje $\hat{a}_j = H_j(N)/\sum_{j=1}^{\infty} H_j(N) = a_j/a$, $j \ge 0$, gdzie $a = \sum_{j=1}^{\infty} a_j$. Można sprawdzić (patrz Dodatek 2), że

$$a = \frac{\pi\alpha_z p_z}{q_z}, \text{ gdzie } q_z = 1 - p_z. \tag{2}$$

Odnotujmy, że wyrażenia $\hat{a}_j$, $j \ge 0$, nie zależą od α, π; parametry $s, p/q$ oraz w, z zawsze pozostają w iloczynie, zatem nie da się ich estymować oddzielnie. W analizie naszego problemu, zakładając niewielką dokładność obliczeń, można wziąć $s = 0.5$. Należy przypuścić, że Steinhaus poznał w w jakiś inny sposób. My tutaj przyjęliśmy $w = 0.5$ całkowicie arbitralnie, a następnie w swoich doświadczeniach numerycznych estymowaliśmy parametry p, z przyjmując jako kryterium minimum kwadratów odchyleń prawdopodobieństw $\hat{a}_j$ od ich wartości obserwowanych.

Wiadomo, że Hugon Steinhaus w realnych zastosowaniach matematyki cenił prostotę formuł, nawet kosztem ich optymalności. Przy znajomości $s = 0.5$, p, w, Steinhausowskiego typu estymatorem z może być

$$\tilde{z} = \frac{\tilde{p}_z}{w\tilde{r}(1-\tilde{p}_z)}, \text{ gdzie } \tilde{p}_z = 2\frac{\hat{a}_2 - \hat{a}_3}{\hat{a}_1 - \hat{a}_2} \text{ jest estymatorem } p_z,$$

natomiast estymator $\tilde{r}$ parametru $p/2q$ spełnia równanie

$$\hat{a}_0 = \frac{1 + w\tilde{z}\tilde{r}}{(1+\tilde{r})^2}.$$

Uogólnienie. Dotychczas zakładaliśmy, że nekrologi obserwujemy od początku wojny do określonego dnia. Można rozważać również badania okresowe, dotyczące zbierania nekrologów ogłoszonych w pewnym przedziale czasu. Problem badań okresowych poruszamy w Dodatku 2.

4. Próby numeryczne. Aby zbadać praktyczną użyteczność obliczeniową przytoczonych wzorów, przyjęliśmy pewien (nieuzasadniony jednak realnymi przesłankami) zestaw wartości parametrów: $p = 0.75$, $\alpha = 0.25$, $s = w = 0.5$, $z = 0.3$, $\pi = 1$. Obliczyliśmy: $a_0 = 0.0360$, $a_1^* = 0.1297$,

$a_2 = 0.0157$, $a_3 = 0.0020$, $a_4 = 0.0003$, przy czym $a = 0.1837$. Widzimy, że a_4 jest empirycznie raczej niedostępne. Namiastkę danych empirycznych utworzyliśmy modyfikując nieco te liczby. Sugerując się możliwością emocjonalnej reakcji rodziców w ekstremalnych dla nich przypadkach, zwiększyliśmy liczby a_0 i a_3 biorąc po 5% od sąsiednich prawdopodobieństw, a_4 zaniedbaliśmy. Przyjęliśmy więc dalej jako empiryczne prawdopodobieństwa $a\hat{a}_0 = 0.0378$, $a\hat{a}_1 = 0.1279$, $a\hat{a}_2 = 0.0156$, $a\hat{a}_3 = 0.0021$. Ustalmy $s = w = 0.5$. Estymując p, z otrzymaliśmy: $p = 0.744$, $z = 0.320$, przy czym teraz jest $a_0 = 0.0377$, $a_1^* = 0.1279$, $a_2 = 0.0162$, $a_3 = 0.0021$, a więc metoda najmniejszych kwadratów estymacji działa sprawnie. Trzeba dodać, że zadanie numeryczne nie jest trudne, albowiem w procedurze obliczeniowej mamy dobre przybliżenie wstępne. Zniekształcenie parametrów po zniekształceniu danych okazało się raczej duże, co daje pewną wskazówkę o skuteczności metody nekrologów w praktyce. Wykorzystaliśmy cztery dane liczby (sumujące się do jedności) do estymowania dwóch parametrów. Ponieważ nie ma praktycznie szans na sensowne oszacowanie a_4, więc możliwości uogólnienia modelu i szacowania większej liczby parametrów są nikłe.

Dodatek 1. *Dowód wzoru* (1). Weźmy pod uwagę zmienną losową S oraz W i Z odpowiednio warunkowane. Prawdopodobieństwo a_j znajdujemy biorąc pod uwagę zdarzenia $\{S = m\}$, $\{W = n|S = m\}$, $\{Z = k|W = n\}$, faktu napisania nekrologu i jego odnalezieniu przez nas. Obliczamy:

$$a_0 = P(S=1)P(W=1|S=1)P(Z=1|W=1)\pi = \frac{\alpha sp}{(q+sp)^2} wz\pi,$$

$$a_j = \sum_{m=j}^{\infty} P(S=m) \sum_{n=j}^{m} P(W=n|S=m) \sum_{k=j}^{n} P(Z=k|W=n) \frac{1}{k}\pi,$$

gdzie $P(W = n|S = m) = b(n; m, w)$, $P(Z = k|W = n) = b(k; n, z)$.

Dalej, wykorzystując operator B, obliczamy

$$\begin{aligned} a_j &= \pi \sum_{n=j}^{\infty} \sum_{m=n}^{\infty} P(S=m) b(n; m, w) \sum_{k=j}^{n} \frac{1}{k} b(k; n, z) \\ &= \pi \sum_{n=j}^{\infty} P(W=n) \sum_{k=j}^{n} \frac{1}{k} b(k; n, z) = \pi \sum_{k=j}^{\infty} \frac{1}{k} \sum_{n=k}^{\infty} P(W=n) b(k; n, z) \\ &= \pi \sum_{k=j}^{\infty} \frac{1}{k} P(Z=k) = \pi \alpha_z \sum_{k=j}^{\infty} \frac{1}{k} p_z^k. \end{aligned}$$

Dodatek 2. *Dowód wzoru* (2). Obliczamy

$$a = a_0 + a_1^* + \sum_{j=2}^{\infty} a_j = \sum_{j=1}^{\infty} a_j = \pi\alpha_z \sum_{j=1}^{\infty}\sum_{k=j}^{\infty} \frac{1}{k} p_z^k$$
$$= \pi\alpha_z \sum_{k=1}^{\infty} \frac{1}{k} \sum_{j=1}^{k} p_z^k = \pi\alpha_z \sum_{k=1}^{\infty} p_z^k = \frac{\pi\alpha_z p_z}{1-p_z}.$$

Dodatek 3. *Badanie okresowe.* Przypuśćmy, że zbieramy nekrologi w pewnym przedziale czasu i niech M_0 będzie liczbą poległych do początku i M_1 będzie liczbą poległych potem do końca tego przedziału. Teraz mamy do wyestymowania dwa interesujące parametry: $z_0 = M_0/N$ i $z_1 = M_1/N$. Nietrudno zauważyć, że wzór (1) przyjmie teraz postaci

$$(3) \qquad \begin{aligned} a_0 &= \pi P(S=1)P(W=1|S=1)P(Z_0=0|W=1)\times \\ &\times P(Z_1=1|W=1,Z_0=0) = \frac{\alpha sp}{(q+sp)^2} w(1-z_0)z_1, \end{aligned}$$

$$\begin{aligned} a_j &= \pi \sum_{m=j}^{\infty} P(S=m) \sum_{n=j}^{m} P(W=n|S=m) \sum_{i=0}^{n-j} P(Z_0=i|W=n)\times \\ &\times \sum_{k=j}^{n-i} \frac{1}{k} P(Z_1=k|W=n,Z_0=i),\ j \ge 1, \end{aligned}$$

gdzie

$$P(Z_0=i|W=n) = b(i;n,z_0),\ P(Z_1=k|W=n,Z_0=i) = b(k;n-i,z_1).$$

Stąd, po przekształceniu otrzymujemy

$$(4) \qquad a_j = \pi \frac{\alpha}{q+zp} \sum_{k=j}^{\infty} \frac{1}{k}\Big(\frac{zp}{q+zp}\Big)^k,\ j \ge 1,\ \text{gdzie } z = sw(1-z_0)z_1.$$

W badaniach okresowych mamy więc do czynienia z sytuacją podobną do badań ciągłych, ale będących tu parametrów w iloczynie $w(1-z_0)z_1$ nie da się rozdzielić.

Prace cytowane

[1] A. J. Lotka, *Théorie analytique des associations biologiques II*, Actualités scientifique et industrielles, **780**, Paris 1939 [cytowane za W. Fellerem, *Wstęp do rachunku prawdopodobieństwa*, wyd. II, PWN, Warszawa 1966, s. 127].

[2] I. Kopocińska, *Estimation of multivariate cumulative processes*, Complex Systems, **13**, 2 (2001), 177–183.

[3] I. Kopocińska, B. Kopociński and A. Okulewicz, A class of distributions applicable to the description of the number of nematodes parasitizing birds, *Mathematical Medicine and Biology*, **21** (2004), 35–48.

[4] H. Steinhaus, *Wspomnienia i zapiski*, Aneks, Londyn 1990.

Instytut Matematyczny
Uniwersytet Wrocławski
Plac Grunwaldzki 2/4, 50-384 Wrocław
E-mail: ibk@math.uni.wroc.pl

Hugo Steinhaus problem of estimation of the war casualities on the base of contemporary press obituaries

Abstract. In the memoirs of Professor Hugo Steinhaus the problem of estimation of the war casualties of German Army during the Second World War on the base of contemporary press obituaries is mentioned. Assuming the inflated geometric probability distribution function for the size of German family, introducing some transformations of this random variable and taking into account the probability distribution function of the contens of obituaries we propose in the note a solution of the problem.

Key words: Steinhaus of obituaries method, inflated geometric probability distribution, binomial dilution, parameter estimation.

(*wpłynęło 31 maja 2007 r.*)

MATEMATYKA STOSOWANA 8, 2007

Errata do artykułu I. Czochralskiej, Matematyka Stosowana 7, 2006,

Metoda rozwiązywania układu równań liniowych z symetryczną, nieokreśloną macierzą współczynników

Strona	Jest	Powinno być
45^5	źle uwarunkowanych, nieosobliwych	nieosobliwych
45^6	symetryczną	symetryczną, nieokreśloną
46^2	działań) i obciążenie pamięci komputera	działań i obciążenie pamięci komputera)
46_8	$\varepsilon \ll 1$	$\lvert\varepsilon\rvert \ll 1$
47_4	$\boldsymbol{M}^{k-1}$	$\boldsymbol{M}^{(k-1)}$
47_7	$\overline{0, n-1})$	$\overline{1, n-1})$
48^{17}	$2\lvert\bar{\bar{a}}_{pq}^{(k-1)}\rvert \geq \max\{a_{ij}^{(k-1)}$:	$2\lvert a_{pq}^{(k-1)}\rvert \geq \max\{\bar{\bar{a}}_{ij}^{(k-1)}$:
48_{10}	$\boldsymbol{L}_k^{-1} k$	$\boldsymbol{L}_k^{-1} -$
48_5	$\overline{\boldsymbol{M}}^{(k-1)}$	$\overline{\overline{\boldsymbol{M}}}^{(k-1)}$
49_8	$\boldsymbol{L} =$	$\boldsymbol{C} =$
50^1	$\lvert a_{kk}^{(k)}\rvert$	$\lvert a_{kk}^{(k-1)}\rvert$
$50_{2,3}$	$\mathbb{R}^{n\times n}$	$\mathbb{R}^{m\times n}$
51_8	solving indefinite	solving
51_7	symmetric	symmetric indefinite

Innych zmian stylistycznych nieakceptowanych przez Autorkę nie podajemy za jej zgodą. Za powyższe błędy, powstałe w trakcie wprowadzania korekty, wydawca([1]) przeprasza Autorkę i Czytelników. Oryginalny tekst artykułu jest dostępny w pdf na stronie Pisma: http://matstos.pjwstk.edu.pl/.../.

(*wpłynęło 2 października 2006 r.*)

([1]) Dział Wydawnictw IM PAN – drukujący nr 7.

MATEMATYKA STOSOWANA 8, 2007

Kongres ICIAM 2007 – wrażenia uczestnika

W dniach 16–20 lipca 2007 r. odbył się w Zurichu Szósty Międzynarodowy Kongres Matematyki Stosowanej i Przemysłowej (6th International Congress on Industrial and Applied Mathematics). Organizatorami Kongresu były Uniwersytet w Zurichu i ETH, a imiennie były Prezydent EMS-u a jednocześnie były Przewodniczący Rady Naukowej Centrum Banacha – Rolf Jeltsch. Była to impreza ze wszech miar ogromna. Mieliśmy 73 równoległe sesje, około 2900 wystąpień naukowych i ponad 3000 uczestników. Szczęśliwie kampusy Uniwersytetu i ETH przylegają do siebie co umożliwiło organizację Kongresu na małej przestrzeni. Nie zmienia to faktu, że wszyscy ciągle gdzieś biegali, gdyż wiele ciekawych sesji niemalże pokrywało się. Była to przede wszystkim impreza towarzyska i tak ją chciał widzieć w rozmowach ze mną prof. Rolf Jeltsch. Trzeba przyznać, że Zurich jest miastem bardzo wdzięcznym do organizacji tego rodzaju konferencji. Znakomite połączenia kolejowe z lotniska do centrum, wygodna i bardzo regularna komunikacja miejska, z której korzystali za darmo wszyscy uczestnicy Kongresu. Wreszcie ogromny ośrodek naukowy z dużą liczbą laureatów nagrody Nobla. Kiedyś nocowałem w hotelu w Zurichu, w którym każdy pokój był poświęcony innemu laureatowi nagrody Nobla związanemu z Zurichem. Mała Szwajcaria jest więc światowym potentatem naukowym.

ICIAM, czyli International Council for Industrial and Applied Mathematics, jest stowarzyszeniem bez indywidualnych członków. Członkami są towarzystwa w dwóch kategoriach: *full members* – towarzystwa, których głównym celem działalności jest matematyka stosowana i przemysłowa, oraz *associate members* – stowarzyszenia matematyków obejmujące swym zakresem działania matematykę stosowaną i przemysłową. Członkami stowarzyszonymi (associate members) są między innymi AMS, EMS, London Mathematical Society i również od dwóch lat PTM. Głównym zadaniem ICIAM-u jest organizacja Kongresów co 4 lata – poprzedni Kongres odbył w 2003 r. w Sydney, następny planowany jest w 2011 r. w Vancouver. Zwyczajowo organizator Kongresu zostaje prezydentem ICIAM-u. W tym roku ustępuje Ian Sloan z Sydney na rzecz Rolfa Jeltscha.

W ramach programu naukowego Kongresu wygłoszono 27 zaproszonych wykładów. Główne wykłady i ceremonie były równocześnie transmitowane do trzech audytoriów. Wprowadzono również tzw. *Dni Przemysłu* poświęcone różnym przemysłowym zastosowaniom matematyki: obliczeniowej elektromagnetyce, symulowaniu przepływów w produkcji żywności, modelowaniu i symulacji systemów transportowych, optymalizacji infrastruktury telekomunikacyjnej, numerycznej optymalizacji dla przemysłowego modelowania samolotów, przewidywaniu sukcesu w farmaceutycznym modelowaniu, zarządzaniu ryzykiem na rynkach energii i rynkach finansowych.

Odbyły się różne spotkania panelowe i fora dyskusyjne. Między innymi obradowała komisja European Mathematical Society (EMS) do spraw zastosowań matematyki. Jej przewodniczący Mario Primicerio z Florencji poinformował o rozstrzygnięciu konkursu na letnie szkoły EMS-u z zastosowań. Dwustopniowa procedura wyłoniła dwóch laureatów: szkołę "**Risk Theory and Related Topics**" organizowaną w dniach 29.09–8.10.2008 r. w Będlewie przez A. Palczewskiego (UW) i Ł. Stettnera (IM PAN) i szkołę "**Mathematical modelling the manufacturing of glass, polymers and textiles**" organizowaną w dniach 8.09–18.09.2008 r. w Montecatini Terme w ramach CIME przez A. Fasona (Florencja) i J. Ockendona (Oxford).

W czasie Kongresu dominowała różnego rodzaju analiza numeryczna i równania różniczkowe cząstkowe. Istotną grupę tematyczną tworzyła biologia matematyczna. Duże wrażenie wywarł ładnie przygotowany wykład emerytowanego profesora z Purdue University Waltera Gautschi dotyczący Leonarda Eulera: jego życia i pracy naukowej. Wśród wykładów plenarnych wyróżniłbym „Numerical methods for nonlinear elliptic problems" R. Glowinskiego (Houston) i „Nonlinear problems involving integral operators" Luisa Caffarelliego (Austin). Oczywiście jest to wybór bardzo subiektywny. Idąc na te wykłady musiałem zrezygnować z czterech innych odbywających się równocześnie wykładów.

W czasie ceremonii otwarcia wręczono 5 nagród ICIAM-u. Otrzymali je: ICIAM Pioneer Prize – Ingrid Daubechies (Princeton) i Heinz Engl (Linz), ICIAM Collatz Prize – Felix Otto (Bonn), ICIAM Lagrange Prize – Joseph Keller (Stanford), ICIAM Maxwell Prize – Peter Deuflhard (Berlin), ICIAM Su Buchin Prize – Gilbert Strang (MIT).

Wzorem Kongresów Międzynarodowej Unii Matematycznej (IMU) codziennie pojawiał się kolorowy newsletter.

Stosunkowo liczna była polska reprezentacja (ponad 60 osób). Jest to dużo zważywszy na to, że w Polsce nie ma nadal dobrych warunków do rozwijania matematyki stosowanej i przemysłowej, choć posiadamy dobrą kadrę naukową. Nasze przedsiębiorstwa niestety nie palą się do współpracy, często opierając się na zagranicznych technologiach. Nie ulega jednak wątpliwości, że to wkrótce się zmieni, gdyż taka jest światowa tendencja. Liczę też, że istotną rolę w rozwoju zastosowań w Polsce odegra działające przy IM PAN Centrum Zastosowań Matematyki, którym mam przyjemność kierować.

Łukasz Stettner (IM PAN)

Wskazówki dla autorów

Maszynopisy w dwóch egzemplarzach w języku polskim należy przesyłać pod adresem:

Matematyka Stosowana
Instytut Matematyczny PAN
ul. Śniadeckich 8
00-956 Warszawa

Wszystkie artykuły w dziale naukowym podlegają recenzji. Dla jej przyspieszenia wersję PDF proszę przesłać e-mailem.

Materiały przeznaczone do druku powinny być przygotowane w systemie TEX (w dowolnym dialekcie), z użyciem czcionki 12 pt i szerokości tekstu 13,5 cm. Wydruki powinny być jednostronne.

Rysunki należy przygotować w wersji elektronicznej. Ich szerokość nie może przekraczać 13,5 cm.

Po przyjęciu artykułu do druku należy przesłać e-mailem plik źródłowy pod adresem matsto@impan.gov.pl.

Sprzedaż i prenumerata

Zgłoszenia w sprawie prenumeraty przyjmuje

Dom Handlowy Nauki Sp. z o.o.
Zakład ORPAN – Księgarnia
ul. Twarda 51/55
00-818 Warszawa
tel./fax (22)697-8914

Sprzedaż numerów aktualnych i archiwalnych prowadzi

Biuro Zarządu Głównego Polskiego Towarzystwa Matematycznego
ul. Śniadeckich 8
00-956 Warszawa
tel. (22)629-9592

Redakcja (matsto@impan.gov.pl) udostępnia prenumeratorom wersje elektroniczne (pliki PDF) opublikowanych artykułów.

MATEMATYKA STOSOWANA

Matematyka dla społeczeństwa

Spis treści