Měření shody binárních pozorování

Václav Fidler, Med.Statistics, University of Groningen, NL

1. Úvod

Když se něco měří, tak nás dříve či později začne zajímat přesnost měření. Jaké odchylky lze očekávat když bude pozorovatel měření stejného objektu nezávisle opakovat; jak odlišná budou měření jiného pozorovatele? Odpověď na tyto otázky lze hledat na základě vhodně provedeného experimentu. Pokud lze pozorování považovat za realizace normálně rozdělené normální veličiny, tak můžeme použít standardních metod analýzy rozptylu a přesnost popsat variančními komponentami a koeficienty shody (neboli mezitřídními korelačními koeficienty). Třebaže binární pozorování jsou 'jednodušší' než pozorování normálně rozdělená, pro popsání jejich přesnosti není statistická dílna moc dobře vybavena. V tomto příspěvku načrtneme možný přístup a aplikujeme ho na konkrétní data.

2. Data a základní model

Představme si lékaře posuzující histologický preparát co se týče přítomnosti (1) či nepřítomnosti (0) zhoubných buněk. Při opětovném nezávislém posuzování stejného preparátu stejným lékařem, či jinými lékaři, nemusí být pozorování shodná. Za účelem vyhodnocení spolehlivosti takovýchto binárních pozorování byl realizován experiment, ve kterém dva lékaři nezávisle na sobě posuzovali každý dvakrát 45 preparátů. Data jsou v Tabulce 1. Pozorování značíme X_{ij}, X_{11} a X_{12} jsou pozorování lékaře 1, X_{21} a X_{22} lékaře 2.

Tabulka 1. Pozorování dvou lékařů (výběrové četnosti a četnosti předpovídané modelem z odstavce 6).

lékař 1		lékař 2, (X_{21},X_{22}): 00	01	10	11	+
(X_{11},X_{12}):	00	20	1	2	3	26
		17,9	2,7	2,7	2,4	
	01	2	1	0	1	4
		2,7	0,5	0,5	1,0	
	10	2	2	1	1	6
		2,7	0,5	0,5	1,0	
	11	2	2	0	5	9
		2,4	1,0	1,0	5,3	
	+	26	6	3	10	45

Jak vhodně popsat shodu lékaře se sebou samým, jak shodu mezi oběma lékaři? Přirozené se zdají pravděpodobnosti $P(X_{i2}=s|X_{i1}=s)$, $s=0,1$ (pro 'vnitřní' shodu lékařů) a $P(X_{1j}=s|X_{2k}=s)$ (pro shodu mezi lékaři). Kromě těchto pravděpodobností bychom též rádi disponovali nějakým jednoduchým indexem shody.

Preparáty budeme považovat za náhodně vybrané z jisté velké populace; pozorování jsou pak nezávislé, stejně rozdělené vektory $X=(X_{11},X_{12},X_{21},X_{22})$. X může nabýt 16 různých výsledků, rozdělení X má tedy 15 stupňů volnosti. Patnáct parametrů je neprakticky mnoho pro popsání intuitivně poměrně jednoduchého konceptu shody. Potřebujeme tedy flexibilní model pro $P(X=x)$.

3. Jednoduchý experiment

Nejdříve se však budeme zabývat jednodušším problémem, ve kterém buď oba lékaři pozorují jen jednou, nebo jen jeden lékař pozoruje dvakrát. Výsledek měření u náhodně zvoleného preparátu označíme $Y=(Y_1,Y_2)$. Rozdělení Y parametrizujeme notací Tabulky 2.

Tabulka 2. Notace pro rozdělení (Y_1,Y_2)

	$Y_2=0$	$Y_2=1$	+
$Y_1=0$	p_{00}	p_{01}	$p_{0.}$
$Y_1=1$	p_{10}	p_{11}	$p_{1.}$
+	$p_{.0}$	$p_{.1}$	1

Pro popis shody mezi Y_1 a Y_2 je často používán Cohenův koeficient shody κ,

$$\kappa = \frac{p_{00}+p_{11}-(p_{0.}p_{.0}+p_{1.}p_{.1})}{1-(p_{0.}p_{.0}+p_{1.}p_{.1})}.$$

Jde o pravděpodobnost shody, $p_{00}+p_{11}$, škálovanou na interval $[-1,1]$ tak, aby nula odpovídala náhodné shodě nezávislých pozorování. Interpretace hodnot různých od 0, -1 a 1 je komplikována skutečností, že velikost κ závisí na marginálních rozděleních veličin Y_1 a Y_2 (k tomu se ještě vrátíme). Vezmemeli jako příklad data prvního lékaře, dostaneme tabulku

	0	1	+
0	26	4	30
1	6	9	15
+	32	13	45

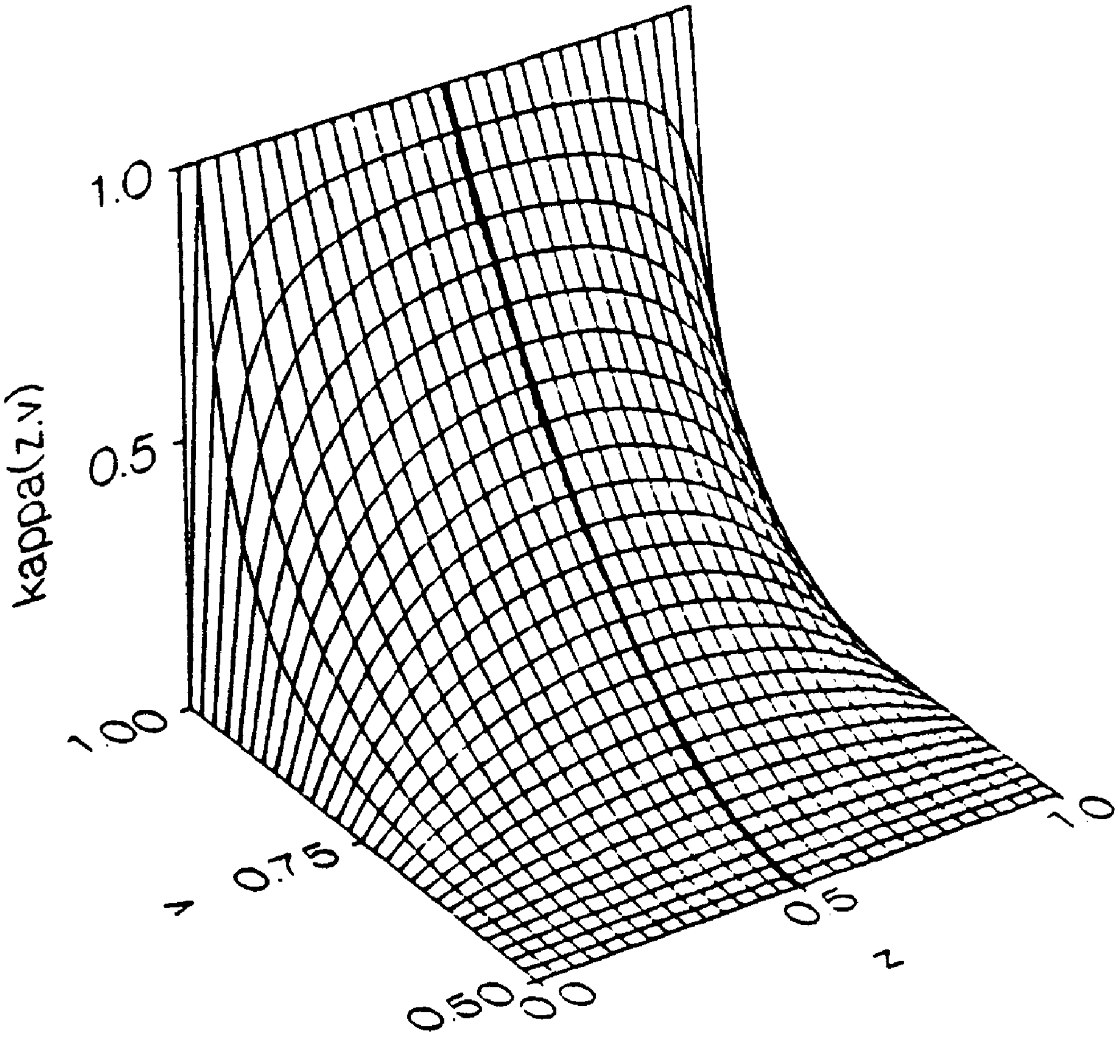
Nahradíme-li ve vzorci pro κ pravděpodobnosti četnostmi, dostaneme $\hat{\kappa}=0,48$.

4. Model pro jednoduchý experiment

Rozdělení náhodných veličin Y budeme modelovat za pomoci latentní veličiny Z, o které si můžeme představovat, že udává opravdový, nám neznámý, binární stav preparátu. Dále budeme předpokládat, že Y_1 a Y_2 jsou podmíněné nezávislé při dané hodnotě Z. Pro rozdělení těchto veličin zavedeme následující notaci (mimo to též uvádíme názvy běžně užívané v kontextu evaluace diagnostických testů):

$P(Z=1)=z$	prevalence (nemoci)
$P(Y_i=1\|Z=1)=v_i$	sensitivita (testu), $i=1,2$
$P(Y_i=0\|Z=0)=w_i$	specificita (testu), $i=1,2$

Rozdělení $P(y)=P(Y=y)$ je tímto parametrizováno pěti parametry: z, v_1, v_2, w_1, w_2. Platí například $P(00)=z(1-v_1)(1-v_2)+(1-z)w_1w_2$. Počet parametrů je pro praktické použití nutno omezit. Výhodné zjednodušení nastává například když $v_1=v_2=w_1=w_2=v$. Obrázek 1 ukazuje, jak v tomto případě vypadá $\kappa=\kappa(z,v)$.



Obrázek 1. Závislost koeficientu shody κ na prevalenci z a sensitivitě/specificitě v.

Obrázek ilustruje poznámku o závislosti κ na marginálních rozděleních, tedy na z. Jeden způsob jak tuto závislost zneškodnit by byl vyhodnocovat shodu při konstantní prevalenci z. Experimentálně je tento návrh obtížně uskutečnitelný, ale lze se opřít se o model a zavést

$$\kappa_{adjusted} = \kappa \text{ při prevalenci } z=0{,}5.$$

V našem dvouparametrovém případě dostáváme: $\kappa_{adjusted}(v)=(1-2v)^2$.

Maximálně věrohodnostní odhady jsou $\hat{z}=0{,}25$, $\hat{v}=0{,}87$, $\hat{\kappa}_{adjusted}=0{,}55$ (test dobré shody nenachází významné odchylky mezi výběrovými a modelem předpovídanými četnostmi).

Poznamenáváme, že interpretovat Z jako opravdový stav preparátu je lákavé, nicméně hypotetické: na modely používající latentní veličiny je třeba pohlížet pragmaticky.

5. Konstrukce modelu pro P(x)

Nyní se vrátíme k našemu původnímu problému: konstrukce modelu pro P(x). Představme si, že lékař *i* má pro jistý preparát latentní výsledek Y_i, že (Y_1,Y_2) má stejné rozdělení jako v předchozím případě, a že Y_i hraje podobnou roli pro (X_{i1},X_{i2}) jako Z pro (Y_1,Y_2). Výhodný způsob, jak tento model formálně definovat je následující:

$$Y_1=V_1Z+(1-W_1)(1-Z)$$
$$Y_2=V_2Z+(1-W_2)(1-Z)$$

Z, V_1, V_2, W_1, W_2 jsou nezávislé binární veličiny, pravděpodobnost jedničky označíme malým písmenkem, např. $P(W=1)=w$. Opětným použitím "Bernouilliovské" transformace dostáváme

$$X_{11}=A_{11}Y_1+(1-B_{11})(1-Y_1)$$
$$X_{12}=A_{12}Y_1+(1-B_{12})(1-Y_1)$$
$$X_{21}=A_{21}Y_2+(1-B_{21})(1-Y_2)$$
$$X_{22}=A_{22}Y_2+(1-B_{22})(1-Y_2)$$

kde $\{A_{ij}\}$, $\{B_{ij}\}$, $\{V_i\}$, $\{W_i\}$ a Z jsou nezávislé Bernoulliovské náhodné veličiny; $\{a_{ij}\}$, $\{b_{ij}\}$, $\{v_i\}$, $\{w_i\}$ a z jsou korespondující pravděpodobnosti jedniček. Při výpočtu P(x), například P(1111), lze postupovat následovně:

$$P(1111)=E(X_{11}X_{12}X_{21}X_{22})=Eh(\{A_{ij}\},\{B_{ij}\},\{V_i\},\{W_i\},Z),$$

nejdříve zjednodušíme výraz h() použitím vlastnosti "$U^2=U$, $U(1-U)=0$ pro binární U"; střední hodnotu pak dostaneme nahrazením velkých písmenek malými.

Takto získaný model má 13 parametrů; rozdělení P(x) má 15 stupňů volnosti. Pochopitelně máme zájem na dalším zjednodušení modelu. Výhodný je např. model se třemi parametry, z, a, v:

z	(prevalence)
$a_{ij}=b_{ij}=a$	(sensitivita/specificita 'uvnitř' lékaře)
$v_i=w_i=v$	(sensitivita/specificita 'mezi' lékaři)

Za tohoto modelu jsou κ 'uvnitř' (tj. na základě marginálního rozdělení (X_{11},X_{12})) a κ 'mezi' (na základě marginálního rozdělení (X_{21},X_{22})) překvapivě jednoduchými funkcemi z, a a v. Označme výrazem 'κ(čistě mezi)' koeficient κ 'mezi' při a=1, tj. při absolutní přesnosti 'uvnitř'. Mezi jmenovanými koeficienty platí následující jednoduché vztahy:

$$\kappa_{adjusted}(\text{uvnitř})=(1-2a)^2$$
$$\kappa_{adjusted}(\text{čistě mezi})=(1-2a)^2$$
$$\kappa_{adjusted}(\text{mezi})=(1-2a)^2(1-2v)^2$$
$$\kappa(\text{mezi})=\kappa(\text{čistě mezi}).\kappa(\text{uvnitř})$$

Pro statistické vyhodnocování popsaných modelů lze použít metody největší věrohodnosti. Optimalizace věrohodnostní funkce vyžaduje dost počítání.

Použitý způsob konstrukce modelu lze použít i ve složitějších experimentálních situacích: jde o pokus o 'binární' obdobu klasických modelů varianční analýzy s náhodnými efekty.

6. Aplikace

Data Tabulky 1 jsme analyzovali za pomoci výše popsaných modelů (pro výpočty byl napsán speciální program). Tabulka 1 ukazuje mimo vlastní pozorování též hodnoty očekávané za výše popsaného tříparametrového modelu: shoda je poměrně dobrá, což potvrzuje i test dobré shody na základě věrohodnostního poměru ($G^2=9,6$ při 15-3=12 stupních volnosti). V tomto případě popisuje tedy jednoduchý tříparametrový model data adekvátně. (Pochopitelně jsme vyhodnocovali i modely s více i s méně parametry).

Odhady pro z, v a a jsou 0,22, 0,92, 0,88; pro $\kappa_{adjusted}$(uvnitř) 0,58, pro $\kappa_{adjusted}$(čistě mezi)=0,71, pro $\kappa_{adjusted}$(mezi) 0,41 (=0,58.0,71). Pro $P_{uvnitř}(1|1)$ dostáváme 0,68, pro $P_{mezi}(1|1)=0,54$.

7. Závěrečné poznámky

Baker et al (1991) se zabývají stejným problémem a používají v podstatě stejný model. Náš (nezávisle vzniklý) přístup obsahuje následující nové prvky: a) konstrukce modelů s náhodnými efekty za pomoci "Bernouilliovské aritmetiky", b) odvození relace mezi κ(uvnitř) a κ(mezi) v tříparametrovém modelu, c) navrh na používání $\kappa_{adjusted}$.

Jiné použití metod pro latentní veličiny při vyhodnocování shody popisuje Uebersax a Grove (1990), tento článek obsahuje též obšírný přehled literatury. Relevantní je též Agresti (1992).

Literatura

Agresti, A. (1992). Modelling patterns of agreement and disagreement. Statistical Methods in Medical Research 1, 201-218.

Baker, S.G., Freedman, L.S. and Parmar, M.K.B. (1991). Using replicate observations in observer agreement studies with binary assessments. Biometrics 47, 1327-1338.

Uebersax, J.S. and Grove, W.M. (1990). Latent class analysis of diagnostic agreement. Statistics in Medicine, 9, 559-572.