

Lad  $\omega$  være en amerikaner. Definer

$F_i(\omega)$  = værdien af faktoren  $i$ 'te faktor for  $\omega$

$D(\omega)$  = dødsårsag for  $\omega$

$L(\omega)$  = dødsalder for  $\omega$

I det kommende undertrykkes afhængigheden af  $\omega$ .

## Faktorsandsynligheder

Vi har blandt andet brug for størrelserne

$P(F_i \in f)$ , for forskellige mængder  $f \in \mathcal{F}_i$

$P(F_{i_1} \in f_1, F_{i_2} \in f_2, \dots, F_{i_n} \in f_n)$  for  $f_i, i = 1, \dots, n$

hvor  $f_i$  er mængder der giver mening for deres tilhørende faktor. I filerne i mappen Factor\_frequencies ligger filer af typen.

$i_1$ 'te faktor	$\dots$	$i_n$ 'te faktor	freq
$f_1^0$	$\dots$	$f_n^0$	$P(F_{i_1} \in f_1^0, F_{i_2} \in f_2^0, \dots, F_{i_n} \in f_n^0)$
$f_1^1$	$\dots$	$f_n^1$	$P(F_{i_1} \in f_1^1, F_{i_2} \in f_2^1, \dots, F_{i_n} \in f_n^1)$
$\vdots$	$\dots$	$\vdots$	$\vdots$
$f_1^{n_k}$	$\dots$	$f_n^{n_k}$	$P(F_{i_1} \in f_1^{n_k}, F_{i_2} \in f_2^{n_k}, \dots, F_{i_n} \in f_n^{n_k})$
$\vdots$	$\dots$	$\vdots$	$\vdots$
$f_1^{n_1}$	$\dots$	$f_n^{n_1}$	$P(F_{i_1} \in f_1^{n_1}, F_{i_2} \in f_2^{n_1}, \dots, F_{i_n} \in f_n^{n_1})$

Alternativt kan man skrive  $F = (F_{i_1}, \dots, F_{i_n})$

## Incidents

Dernæst har vi sandsynlighederne for at dø af en dødsårsag i løbet af et år.

$$p_d(l) = P(D = d, L \leq l \mid L > l - 1), d \in \mathcal{D}, l \in \mathbb{R}_+$$

hvor  $\mathcal{D}$  er en mængde af alle dødsårsager i programmet. For beregning, kender vi  $p_d(l)$  som stykvis konstant funktion på mængderne

$$l_1, l_2, \dots, l_{22} = [0, 1), [1, 5), [5, 10), \dots, [95, 100), [100, \infty)$$

Vi er interesserede i vektoren

$$p_d(l_1) \quad p_d(l_2) \quad \dots \quad p_d(l_{22})$$

hvor jeg, med den lidt misbrugte notation  $p_d(l_i)$ , mener  $p_d(l)$  for et  $l \in l_i$ . Disse estimeres med

$$p_d(l_i) \leftarrow \frac{\text{antal amerikanere døde af } d \text{ i aldersgruppen } l_i \text{ i år } Y_1}{\text{antal amerikanere i aldersgruppen } l_i \text{ i år } Y_2} =: \frac{a_{di}}{a_i}$$

(Lige nu har vi  $Y_1 = 2014, Y_2 = 2013$ ). Filerne af formen *ICDcode.txt* indikerer et  $d$  med deres titel og indholdet er

$$a_{d1} \quad a_{d2} \quad \cdots \quad a_{d22}$$

Og filen *population.txt* indeholder

$$a_1 \quad a_2 \quad \cdots \quad a_{22}$$

## Risk ratios

Lad  $F_1, \dots, F_k$  være nogle faktorer. Risk ratios i dette program fortolkes som

$$\text{RR}_d(f) := \frac{P(D = d, L \leq l \mid L > l - 1, (F_1, \dots, F_k) \in f)}{P(D = d, L \leq l \mid L > l - 1, (F_1, \dots, F_k) \in f_0)}$$

Egentlig skulle der et  $f_0$  på i notationen for  $\text{RR}_d(f)$ , for at indikere at det er riskratio med hensyn til baselinen  $f_0$ . Det er en antagelse, at riskratioen ikke afhænger af  $l$ . Risk ratioerne er kendt for en mængde af faktorinddelinger  $f \in \mathcal{F}$ . Vi kræver mere eller mindre at  $\mathcal{F}$  kan skrives på formen

$$\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \mathcal{F}_k$$

hvor

$$\mathcal{F}_i = \{f_1^0, f_1^1, \dots, f_1^{n_i}\}$$

og så er de kodet ved hjælp af

$i_1$ 'te faktor	$\cdots$	$i_k$ 'te faktor	RR
$f_1^0$	$\cdots$	$f_k^0$	$\text{RR}_d((f_1^0, \dots, f_k^0))$
$f_1^1$	$\cdots$	$f_k^1$	$\text{RR}_d((f_1^1, \dots, f_k^1))$
$\vdots$	$\cdots$	$\vdots$	$\vdots$
$f_1^{n_1}$	$\cdots$	$f_k^{n_k}$	$\text{RR}_d((f_1^{n_1}, \dots, f_k^{n_k}))$
$\vdots$	$\cdots$	$\vdots$	$\vdots$
$f_1^{n_1}$	$\cdots$	$f_k^{n_k}$	$\text{RR}_d((f_1^{n_1}, \dots, f_k^{n_k}))$

## 0.1 Udregning

I første omgang vil vi gerne udregne

$$P(D = d, L \leq l \mid L > l - 1, F = f), \text{ for } f \in \mathcal{F} \quad (1)$$

hvor  $F$  er en vektor af faktorer og  $\mathcal{F}$  er den endelige mængde af faktorsammensætninger. (1) kan senere (i javascript-delen) bruges til at udregne

$$P(D = d, L \leq l \mid L > l - 1, F = f_{\text{personal}}) \quad (2)$$

hvor  $f_{\text{personal}}$  ikke (nødvendigvis) ligger i  $\mathcal{F}$ . Det gøres ved

$$(1) = \frac{\text{RR}_d(f)}{\sum_{f \in \mathcal{F}} \text{RR}_d(f) \cdot P(F = f)} p_d(l) \quad (3)$$

Der er dog nogle forhindringer før vi bare kan stoppe tallene fra filerne ind i (3).

- Vi ikke har nok information til at kende  $P(F = f), f \in \mathcal{F}$  100%.
- Vi har flere riskratio filer for samme  $d$ .
- Hvordan man skal tage højde for alders specifikke riskratios og alders specifikke  $P(F = f)$ 'er

Lad os tage dem i voksende sværhedsgrad

### Flere risk ratio filer

Hvis vi der er to riskratiofiler baseret på to vektorer af faktorer  $F^1, F^2$  og to tilhørende krydsmængder,  $\mathcal{F}^1$  og  $\mathcal{F}^2$ , så ville den mest rigtige måde at kombinere dem på være

$$\begin{aligned} P(D = d, L \leq l \mid L > l - 1, (F^1, F^2) = (f^1, f^2)) \\ = \frac{g(\text{RR}_d(f^1), \text{RR}_d(f^2))}{\sum_{f^1, f^2 \in \mathcal{F}^1 \times \mathcal{F}^2} g(\text{RR}_d(f^1), \text{RR}_d(f^2)) \cdot P((F^1, F^2) = (f^1, f^2))} p_d(l) \end{aligned} \quad (4)$$

hvor  $g$  er en passende interaktionsfunktion. Hvis  $F^1$  og  $F^2$  er uafhængige og  $g(x, y) = x \cdot y$ , kan man dog skrive det som

$$\begin{aligned} P(D = d, L \leq l \mid L > l - 1, (F^1, F^2) = (f^1, f^2)) \\ = \frac{\text{RR}_d(f^1)}{\sum_{f^1 \in \mathcal{F}^1} \text{RR}_d(f^1) \cdot P(F^1 = f^1)} \frac{\text{RR}_d(f^2)}{\sum_{f^2 \in \mathcal{F}^2} \text{RR}_d(f^2) \cdot P(F^2 = f^2)} p_d(l) \end{aligned} \quad (5)$$

Fordelen ved (5) er, at man kan lægge et tallene

$$\frac{\text{RR}_d(f^1)}{\sum_{f^1 \in \mathcal{F}^1} \text{RR}_d(f^1) \cdot P(F^1 = f^1)}, f^1 \in \mathcal{F}^1$$

i en fil og tallene

$$\frac{\text{RR}_d(f^2)}{\sum_{f^2 \in \mathcal{F}^2} \text{RR}_d(f^2) \cdot P(F^2 = f^2)}, f^2 \in \mathcal{F}^2$$

i en anden fil og tallene

$$p_d(l_i), i = 1, \dots, 22$$

i en tredje fil. Selvom betingelserne for at bruge (5) ikke er helt opfyldt, kan det måske alligevel være en god ide at bruge den.

## Vi kender ikke $P(F = f), f \in \mathcal{F}$

Her tænkes  $\mathcal{F}$  som den mængde hvor vi kender  $\text{RR}_d(f)$  hvis og kun hvis  $f \in \mathcal{F}$ .

Der er flere slags udfordringer her.

1. Det hænder at vi kun kender

$$P(F = f), f \in \mathcal{F}'$$

hvor  $\mathcal{F}' \not\subseteq \mathcal{F}$ .

2. Vi har  $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$  og  $F = (F_1, F_2)$ , hvor  $F_1$  og  $F_2$  er hver deres faktor. Her hænder det at vi kun kender

$$P(F_i = f_i), f_i \in \mathcal{F}_i$$

for  $i = 1, 2$  og altså ingenting om den simultane fordeling af  $(F_1, F_2)$

3. Vi har  $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \mathcal{F}_3$  og  $F = (F_1, F_2, F_3)$ . Det hænder, at vi kun kender

$$\begin{aligned} P((F_1, F_2) = (f_1, f_2)), f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2 \\ P((F_1, F_3) = (f_1, f_3)), f_1 \in \mathcal{F}_1, f_3 \in \mathcal{F}_3 \end{aligned}$$

4. Vi har  $\mathcal{F} = \mathcal{F}_1$  og  $F = F_1$ . Det hænder, at vi kender

$$P((F_1, F_2) = (f_1, f_2)), f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2$$

men ikke (umiddelbart)

$$P(F_1 = f_1), f_1 \in \mathcal{F}_1$$

så vi så om sige har for meget

5. Vi har  $\mathcal{F} = \mathcal{F}_1$  og  $F = F_1$ . Det hænder, at vi slet ikke kender noget til

$$P(F_1 \in f)$$

Problemerne findes også i flere flere dimensioner og med kombinationer.

#### En løsning af problem 4

Vi laver en funktion som marginaliserer, dvs

$$P(F_1 = f_1) = \sum_{f_2 \in \mathcal{F}_2} P((F_1, F_2) = (f_1, f_2))$$

#### En løsning af problem 1

Hvis vi laver en funktion, som laver transformationen

$$\begin{aligned} w : \{P(F = f_1), P(F = f_2), \dots, P(F = f_n)\} \\ \mapsto \{P(F = f'_1), P(F = f'_2), \dots, P(F = f'_k)\} \end{aligned}$$

kan vi løse det første problem. Hvis vi antager

$$\bigcup_{i=1}^n f_i \subseteq \bigcup_{j=1}^k f'_j \tag{6}$$

kan man lave løsningen

$$\begin{aligned} P(F = f'_j) &= w(P(F = f_1), P(F = f_2), \dots, P(F = f_n)) \\ &= \sum_{i=1}^n \frac{P(F = f_i) \cdot |f_i \cap f'_j|}{|f_i|} \end{aligned}$$

hvor  $|\cdot|$  repræsenterer et mål. Det vil dog nok altid være muligt at bruge Lebesguemålet eller tælleområdet og nogle gange kan man måske være nødt til at bruge en mikstur af de to. Det ses eksempelvis ved rygning, hvor der er en kategori, der hedder 0 cigaretter. Betingelsen (6) er nødvendig for at  $P(F = f'_j), j = 1, \dots, k$  summer til 1 (det er nemlig antaget at  $P(F = f_i), i = 1, \dots, n$  summer til 1).

### En løsning af problem 2,3 og 5

For at løse problem 2,3 og 5 kan man bruge tilpasning af marginaler. Man har den ønskede fordeling

$$P(F = f), f \in \mathcal{F}$$

hvor  $F = (F_1, \dots, F_k)$  og  $\mathcal{F} = \mathcal{F}_1 \times \dots \times \mathcal{F}_k$ , og man kender

$$\begin{aligned} P(F^1 = f^1), f^1 \in \mathcal{F}^1 \\ \vdots \\ P(F^m = f^m), f^m \in \mathcal{F}^m \end{aligned}$$

hvor  $\mathcal{F}^l = \mathcal{F}_{i_1^l} \times \mathcal{F}_{i_2^l} \times \dots \times \mathcal{F}_{i_{r_l}^l}$ . Man starter med en standard uniformfordeling

$$p(f) = \frac{1}{\#\mathcal{F}}, f \in \mathcal{F}$$

som estimer for  $P(F = f)$ . Dernæst *tilpasser man med marginalen  $\mathcal{F}^1$*

$$p^{ny}(f) = p(f) \frac{P(F^1 = f^1(f))}{\sum_{f': f^1(f') = f^1(f)} p(f')}$$

og derefter med marginalen  $\mathcal{F}^2, \mathcal{F}^3$  og så videre (indtil man når til hvad?).

### Forskellige credibilities

Vi har mængder,  $f \in \mathcal{F}$ , for hvilke vi vil finde  $P(F = f)$ .  $F$  kan her skrives  $(F_i)_{i \in I}$ , hvor  $I$  er en mængde i  $\{1, \dots, n\}$ . Vi kender da 'binnede' fordelinger af  $(F_i)_{i \in I_j}$  for  $j = 1, \dots, k$ , hvor  $I_j$  også er mængder i  $\{1, \dots, n\}$ . Credibilityscoren er en funktion  $c : \{I_j\}_{j=1, \dots, k} \rightarrow \mathbb{R}_+$ . De binnede fordelinger, der indgår i konstruktionen  $P(F = f)$  er

$$\left\{ (F_i)_{i \in I_j} \mid I_j \cap I \neq \emptyset \wedge (\nexists k : I \cap I_j \subseteq I \cap I_k \wedge c(I_k) > c(I_j)) \right\}$$

•

### Aldersspecifikke $P(F = f)$ 'er eller riskratioer

I princippet burde alle ovenstående udregninger laves separat for alle aldersgrupper. Det gøres også, og der er nogle genveje. Definer  $A$  til at være

den stokastiske variabel der angiver alderen på personen. De nye faktorsandsynlighedsfiler har specificerende kolonner ved

$i_1$ 'te faktor	$\dots$	$i_n$ 'te faktor	aldersgr.
$f_1^0$	$\dots$	$f_n^0$	$A_1$
$f_1^0$	$\dots$	$f_n^1$	$A_1$
$\vdots$	$\dots$	$\vdots$	
$f_1^0$	$\dots$	$f_n^{n_k}$	$A_1$
$\vdots$	$\dots$	$\vdots$	
$f_1^{n_1}$	$\dots$	$f_n^{n_k}$	$A_h$

og freq kolonnen indeholder

$$\begin{aligned}
 &\text{freq} \\
 &P(F_{i_1} \in f_1^0, F_{i_2} \in f_2^0, \dots, F_{i_n} \in f_n^0 \mid A \in A_1) \\
 &P(F_{i_1} \in f_1^0, F_{i_2} \in f_2^0, \dots, F_{i_n} \in f_n^1 \mid A \in A_1) \\
 &\vdots \\
 &P(F_{i_1} \in f_1^0, F_{i_2} \in f_2^0, \dots, F_{i_n} \in f_n^{n_k} \mid A \in A_1) \\
 &\vdots \\
 &P(F_{i_1} \in f_1^{n_1}, F_{i_2} \in f_2^{n_2}, \dots, F_{i_n} \in f_n^{n_k} \mid A \in A_h)
 \end{aligned}$$

Det vil sige, at freq-kolonnen skal summe til  $h$ . Når man konstruerer  $P(F = f)$ , bør man konstruere

## 0.2 Javascript delen

Til hver cause hører noget data på formen

$$d, (p_d(l_i))_{i=1, \dots, 22} \text{ dataframe, risk ratio data}_d$$

hvor  $d$  bare er en string,  $(p_d(l_i))_{i=1, \dots, 22}$  er en data frame udskrevet som liste og risk ratio data'et har formen

$$(\text{Risk ratio datagruppe})_1^d, (\text{Risk ratio datagruppe})_2^d, \dots, (\text{Risk ratio datagruppe})_{n_d}^d$$

En risk ratio datagruppe - indekseret ved  $(j, d)$  - består af

$$(\text{norm}^{j,d}(l_i))_{i=1, \dots, 22}, (\text{risk ratio dataframe}_i^{j,d})_{i=1, \dots, k_{j,d}}, g_{j,d}$$

hvor  $(\text{norm}^j(l_i))_{i=1,\dots,22}$  er en liste af 22 tal som normaliserer for hver af de 22 aldersgrupper, list of risk ratio dataframes er en liste af risk ratio dataframes udskrevet som lister og  $g_{j,d}$  er en string, som siger hvilken interaktion der mellem  $(j, d)$  dataframesne. En risk ratio dataframe for en faktor  $F^{i,j,d}$  har formen

$$(r^{i,j,d}(f))_{f \in \mathcal{F}^{i,j,d}}$$

hvor  $\mathcal{F}^{i,j,d}$  er en endelig mængde af faktor levels. Dette  $r$  er blot en diskretisering af den underliggende riskfaktor funktion

$$R^{i,j,d}(\psi), \psi \in \Psi^{i,j,d}$$

hvor  $\Psi^{i,j,d}$  er en mængde af alle tænkelige værdier af faktoren  $F^{i,j,d}$ , og derfor kan den være uendelig. Vi vil lave en *polating* funktion,  $\text{pol}$ , til at evaluere funktionen

$$R^{i,j,d}(\psi) = \text{pol}((r^{i,j,d}(f))_{f \in \mathcal{F}^{i,j,d}}, \psi)$$

Lad nu  $\psi$  være alle en persons faktorværdier, og lad  $\phi^{i,j,d}$  være vektoren af faktorer der er relevante for den  $(i, j, d)$ 'te risk ratio fil, dvs. en delvektor af  $\psi$ . Lad  $a$  være en vilkårlig alder og lad  $l(a)$  være den af de 22 alderskategorier, som  $a$  falder i. Så defineres

$$P_d(a, \psi) = P_d(a) \cdot \prod_{j=1}^{n_d} \frac{g_{j,d}(R^{1,j,d}(\psi^{1,j,d}), \dots, R^{k_{j,d},j,d}(\psi^{k_{j,d},j,d}))}{\text{norm}^{j,d}(l(a))} \quad (7)$$

hvor  $P_d(a)$  er 'afdiskretiseringen' af  $p_d(l)$ . Det kunne måske defineres som

$$P_d(a) = \max(1, \text{pol}((p_d(l_i))_{i=1,\dots,22}, a))$$

Størrelsen  $P_d(a, \psi)$  tænkes at være det samme som (2), dvs.

$$P_d(a, \psi) = P(D = d, L \leq a \mid L > a - 1, F = \psi) \quad (8)$$

## Kombinationer af $P_d(a)$ 'er

Vi definerer nu

$$p(a, \psi) = \sum_{d \in \mathcal{D}} p_d(a, \psi)$$

hvor  $\mathcal{D}$  er mængden af alle dødsårsager. Hvis (8) faktisk gælder, er

$$p(a, \psi) = P(L \leq a \mid L > a - 1, F = \psi)$$



### Liste over mest interessante kombinationer

Lad nu  $\tilde{\mathcal{D}} \subseteq \mathcal{D}$  være en vilkårlig delmængde af dødsårsagerne. Alle de følgende størrelser kan varieres ved at tilføje  $d \in \tilde{\mathcal{D}}$  og/eller  $L \in [a, b]$  til betingningen

- Forventet levealder

$$E[L \mid \psi] = \sum_{a=0}^{\infty} a \cdot p(a, \psi) \prod_{b=0}^{a-1} (1 - p(b, \psi))$$

- Forventet antal år mistet til sygdom  $d$

$$E[L \mid \psi, D \in \mathcal{D} \setminus \{d\}] - E[L \mid \psi]$$

- Sandsynligheden for at dø af  $d$

$$P(D = d \mid \psi) = \sum_{a=0}^{\infty} p_d(a, \psi) \prod_{b=0}^{a-1} (1 - p(b, \psi))$$

- Ens overlevelseskurve

$$(P(L \geq a \mid \psi))_{a \in \mathbb{N}_0} = \left( \prod_{b=0}^{a-1} (1 - p(b, \psi)) \right)_{a \in \mathbb{N}_0}$$

### Forklare døden ved hjælp af ens faktorer

Vi vil nu også lave en optimal  $\psi$ -værdi som man kan måle brugerens  $\psi$  op imod. Dette er dog en udfordring, fordi en faktor som forårsager en sygdom kan hæmme fremkomsten af en anden sygdom. Det gælder foreksempel rygning, lungekræft og Parkinson's. Vi definerer derfor  $\psi_0$  så  $\psi_0^{i,j,d}$  minimerer  $R^{i,j,d}(\cdot)$ . Det vil sige at i en lungekræftssammenhæng har  $\psi_0$  rygning på 0, mens i en Parkinson's sammenhæng har  $\psi_0$  rygning på 2 (cig/day). Lad nu  $\psi_{F,0}$  opfylde at  $\psi_{F,0}^{i,j,d}$  minimerer  $R^{i,j,d}(\cdot)$  for faktorerne i vektoren af faktorer,  $F$ . Så hvis man vælger  $F$  stor nok får man  $\psi_{F,0} = \psi_0$ . Betragt nu følgende dekomposition

$$\begin{aligned} P(D = d \mid F = \psi) &= \sum_{a=0}^{\infty} p_d(a, \psi) \prod_{b=0}^{a-1} (1 - p(b, \psi)) \\ &= \sum_{a=0}^{\infty} (p_d(a, \psi) - p_d(a, \psi_{0,F})) \prod_{b=0}^{a-1} (1 - p(b, \psi)) \\ &\quad + \sum_{a=0}^{\infty} p_d(a, \psi_{0,F}) \prod_{b=0}^{a-1} (1 - p(b, \psi)) \end{aligned}$$

Definitionen af  $\psi_{0,F}$  gør begge de to led positive. Vi fortolker det første led som sandsynligheden for død på grund af ens faktorværdier, mens det andet led er *death by chance/age/destiny*. Bemærk at andet led ikke er identisk med  $P(D = d \mid F = \psi_{0,F})$  fordi parameteren  $\psi$  stadig indgår i leddet. Vi er nu interesserede i at dekomponere

$$p_d(a, \psi) - p_d(a, \psi_{0,F})$$

for da kan vi dekomponere  $P(D = d \mid F = \psi)$  i flere led.

Hvis  $F$  bare er en enkelt faktor dvs.,  $F = (F_1)$ , er den fuldt dekomponeret. Men hvis  $F$  er en vektor  $F = (F_1, F_2, \dots, F_n)$  ville vi ideelt have positive tal

$$s_d(a, \psi_{0,F_J}), J \subseteq \{1, \dots, n\}$$

sådan at

$$p_d(a, \psi) - p_d(a, \psi_{0,F_I}) = \sum_{J \subseteq I, J \neq \emptyset} s_d(a, \psi_{0,F_J}) \quad (9)$$

for alle  $I \subseteq \{1, \dots, n\}$ . Antag  $p_d$  splitter pænt multiplikativt op

$$p_d(a, \psi) = P_d(a) \prod_{j=1}^n \frac{R^{1,j,d}(a, \psi^{F_j})}{\text{norm}^{j,d}(l(a))} \quad (10)$$

hvor  $\psi^{F_j}$  er  $\psi$ -vektorens værdi hørende til  $F_j$ -faktoren. En løsning til (9) er da

$$\begin{aligned} s_d(a, \psi_{0,F_J}) &= P_d(a) \frac{1}{\prod_{j=1}^n \text{norm}^{j,d}(l(a))} \\ &\cdot \prod_{j \in J} [R^{1,j,d}(a, \psi^{F_j}) - R^{1,j,d}(a, \psi_{F_j,0}^{F_j})] \prod_{j \notin J} R^{1,j,d}(a, \psi_{F_j,0}^{F_j}) \end{aligned}$$

Jeg kan ikke lige bevise det på stående fod, men skal lige se hvad min ‘Hierarkiske og grafiske kontingenstabeller’- bog siger om et meget lignende resultat.

Konstanten  $P_d(a) \frac{1}{\prod_{j=1}^n \text{norm}^{j,d}(l(a))}$  skal altid ganges på så definer nu  $\tilde{s}_d$  ved

$$s_d(a, \psi_{0,F_J}) = P_d(a) \frac{1}{\prod_{j=1}^n \text{norm}^{j,d}(l(a))} \tilde{s}_d(a, \psi_{0,F_J})$$

I tilfælde hvor vi ikke har den pæne struktur i (10), kan man udnytte den pæne struktur risk ratio grupperne imellem. Indenfor riskratiogrupperne må man tilpasse dekompositionen interaktionsfunktionen - hvis interaktionsfunktionen er *multiplicative* er det hurtigt klaret. Inden for hver risk ratio fil, må vi finde en heuristik - helst sådan at der gælder

$$p_d(a, \psi) - p_d(a, \psi_{0,F_i}) = s_d(a, \psi_{0,F_i}) \quad (11)$$

for alle  $i$ . Det vigtigste krav er dog

$$p_d(a, \psi) - p_d(a, \psi_{0, F_{\{1, \dots, n\}}}) = \sum_{J \subseteq \{1, \dots, n\}, J \neq \emptyset} s_d(a, \psi_{0, F_J}) \quad (12)$$

fordi ellers summer alle komponenterne ikke til  $p_d(a, \psi)$ .

### Overvejelse

Betragt de simple riskratiointeraktioner

	$f_1^1$	$f_1^2$
$f_2^1$	1.0	1.1
$f_2^2$	1.1	2.1

Tabel 0.1: Her er der en positiv interaktion mellem  $F_1$  og  $F_2$

	$f_1^1$	$f_1^2$
$f_2^1$	1.0	2.0
$f_2^2$	2.0	2.1

Tabel 0.2: Her er der en negativ interaktion mellem  $F_1$  og  $F_2$

Ved at bruge (11) på den Tabel 0.1 får man

$$\begin{aligned} \tilde{s}_d(a, \psi_{F_1, 0}) &= 0.1 \\ \tilde{s}_d(a, \psi_{F_2, 0}) &= 0.1 \\ \stackrel{(12)}{\Rightarrow} \tilde{s}_d(a, \psi_{(F_1, F_2), 0}) &= 0.9 \end{aligned}$$

Reglen kan ikke bruges på Tabel 0.2 for da er

$$\begin{aligned} \tilde{s}_d(a, \psi_{F_1, 0}) &= 1.0 \\ \tilde{s}_d(a, \psi_{F_2, 0}) &= 1.0 \\ \stackrel{(12)}{\Rightarrow} \tilde{s}_d(a, \psi_{(F_1, F_2), 0}) &= -0.9 \end{aligned}$$

Og da der ikke må være negative  $\tilde{s}_d$  er den ikke gyldig.

### Løsninger

?

## **den polerende funktion**

Valget her er ikke taget, og der er sikkert mange gode valg