Eötvös Loránd Tudományegyetem
Természettudományi Kar

Házastársak élettartamának vizsgálata

Szakdolgozat

Töttösi Nikolett
Matematika BSc
Alkalmazott matematikus szakirány

Témavezető:
Csiszár Villő
adjunktus
Valószínűségelméleti és Statisztika Tanszék

Budapest, 2012

Tartalomjegyzék

Köszönetnyilvánítás

Ezúton szeretném megköszönni témavezető tanáromnak, Csiszár Villőnek, hogy elvállalta a konzulensi teendőket. Köszönöm, hogy a félév során figyelemmel kísérte munkámat, ötleteivel és szakmai tanácsaival nagyban segítette szakdolgozatom elkészülését.

Külön köszönettel tartozom családomnak, akik az évek során mellettem álltak és támogatták tanulmányaimat.

Bevezető

A túlélés-analízis egyik népszerű ága a házastársak élettartamával foglalkozik. Számos statisztika bizonyítja ugyanis, hogy azoknak az embereknek az élettartama, akik kapcsolatban állnak egy másik személlyel lényegesen hosszabb, mint azoké, akik egyedülállók.

A kapcsolatban állók között is vannak különbségek. Megkülönböztetünk rövid- és hosszú idejű függést. Rövid idejű függés van két élet között amikor a túlélő házastárs halálának az esélye növekszik az együtt eltöltött idő függvényében. Hosszú idejű függésnél pedig pont fordítva, azaz minél később hal meg a párja, annál kisebb lesz az ő halálának valószínűsége, vagy egyáltalán nem is változik. Szakdolgozatomban mindkét esetre ismertetek egy-egy példát.

Az első fejezet egy kisebb elméleti bevezető. Bemutatom a túlélési függvényt, a hazárdfüggvényt és a cenzorálás fogalmát is, amik nélkül a házastársak túlélésvizsgálata lehetetlen lenne. Kitérek két külön esetre is. Az egyik, amikor külön-külön nézem a két félt és az élettartamukról szeretnék mondani valamit, míg a másik esetben együtt vizsgálom őket és meghatározom, hogy az adott pillanatban hány év után várható az első halála és hány év után a másodiké.

A második fejezetben a hosszú idejű függés egy modelljét, pontosabban a kopula modellt vizsgálom. A kopula modell napjaink egyik legnépszerűbb eszköze a házastársak élettartamának vizsgálatában. Ez a fogalom nem tekint vissza hosszú múltra, ám mára már számos területen felhasználták. Népszerűségre könnyen kezelhetősége és követhető lépései miatt tett szert. Ebben a szakaszban ismertetem a fogalmakat, összefüggéseket és kitérek néhány speciális eset bemutatására. Ilyen például, amikor két élet független egymástól, vagy rendelkeznek egy adott tulajdonsággal. Ebben a fejezetben az ötleteket főként a [3]. forrásból merítettem.

A harmadik és egyben az utolsó fejezet a rövid idejű függés egy modelljét taglalja. Definiálom a Markov-modellt, amit kiterjesztek úgy, hogy kellőképp hiteles képet adjon nekünk erről a függésről. A fejezetben megmutatom, hogyan becsüljük a paramétereket, valamint adok egy olyan módszert, amivel meg lehet vizsgálni, hogy a férj és a feleség élettartamára ugyanannyira van-e hatással a másik halála. Ebben a részben a [4] forrásra támaszkodtam.

1. fejezet

Túlélés-analízis

A házastársak túlélésének vizsgálatához szükségünk van pár definícióra és azok tulajdonságaira, amiket a későbbiekben alkalmazni tudunk.

1.1. Alapdefiníciók

1.1.1. Definíció (túlélési függvény). *Legyen adott egy T valószínűségi változó (az egyén hátralévő ideje) és egy $t \in \mathbb{R}$ szám (időtartam). Legyen $S(t) = P(T > t)$, azaz annak valószínűsége, hogy a halál később, mint t idő múlva következik be. Ekkor $S(t)$-t* ***túlélési függvény****nek, vagy más néven túlélhetőség függvénynek nevezzük.*

Általában feltesszük, hogy $S(0) = 1$, habár ez a valószínűség igazából kisebb 1-nél, ha az azonnali halált is számba vesszük. Ez a függvény a 0-hoz tart, ugyanis még nem találták meg az örök élet titkát.

1.1.2. Definíció (hazárdfüggvény). *A* ***hazárdfüggvény*** *(lényegében a pillanatnyi kockázat) t időben a halálozás esélye feltéve, hogy a túlélés t-ig, vagy tovább tart, azaz $\mu(t) = P(t \leq T < t+\varepsilon \mid T > t)$. Vagyis annak a feltételes valószínűsége, hogy a megfigyelt személy $\varepsilon <$ időn belül meghal, feltéve, hogy a t időpontban még élt.*

Ezt a függvényt más módszerrel is meg tudjuk határozni.

$$\mu(t) = P(t \leq T < t+\varepsilon \mid T > t) =$$
$$= \frac{P(t \leq T < t+\varepsilon, T > t)}{P(T > t)} = \frac{P(t \leq T < t+\varepsilon)}{P(T > t)} = \frac{f(t)}{S(t)} = -\frac{\frac{\partial}{\partial t}S(t)}{S(t)},$$

ahol $f(t)$ a túlélési függvény sűrűségfüggvénye.

Cenzorálás

Előfordulhat egy adott vizsgálat során, hogy van néhány adat, amely nagyban kilóg a megfigyelés adott időtartamából. Ilyenkor csak annyit tudunk feljegyezni magunknak, hogy a várt jelenség korábban, vagy később következett be, mint ahogy mi vártuk. Ezt nevezzük cenzorálásnak. Bizonyos esetekben előfordulhat, hogy a cenzorált adatok nem hagyhatóak figyelmen kívül, mert akkor torzulna a kép.

A cenzorálásnak a következő típusait különböztethetjük meg:

- **felülről való levágás**: amikor egy várt esemény egy adott időpontnál később következik be. Jelen esetben a megfigyelt személy a megfigyelés időtartama alatt élve marad. Ilyenkor ezeket az adatokat szokás kihagyni a vizsgálatból.
- **alulról való levágás**: amikor nem tudjuk mióta állnak kapcsolatban az egyének, vagy mi történt a vizsgálat előtt. A megfigyelés kezdetén csak a koruk számít nekünk.

1.2. A függés típusai

A továbbiakban tekintsünk két életet (férj és feleség). Legyen a férj x és a feleség y korú a vizsgálat kezdetén. Jelölje T_x és T_y a hátralévő élettartamukat. Feltesszük, hogy ezek a valószínűségi változók abszolút folytonos eloszlásúak $\omega_x - x$ és $\omega_y - y$ felső határokkal, ahol ω_x és ω_y a házastársak maximum életkorát jelölik. Legyen $t \in [0, \omega_x - x)$ és $s \in [0, \omega_y - y)$ esetén $\mu_1(x+t)$ és $\mu_2(y+s)$ a házastársak hazárdfüggvénye külön-külön.

Legyen $\mu_1(x+t \mid T_y = t_y)$ a férj feltételes hazárdfüggvényét t időre (azaz $x+t$ korig), ha a feleség t_y év után meghalt, ahol $t_y \in [0, s)$ (vagyis t_y időt töltöttek el közösen). Hasonlóan jelölje $\mu_2(y+s \mid T_x = t_x)$ a feleség feltételes hazárdfüggvényét s év után, ha a férj t_x év után hunyt el, ahol $t_x \in [0, t)$. Ezek után definiálhatjuk a függés két típusát:

1.2.1. Definíció. *T_x és T_y hátralévő élettartamok **rövid idejű függést** mutatnak, ha $\mu_1(x+t \mid T_y = t_y)$ egy monoton növekedő függvény t_y-ra nézve (másképpen, ha $\mu_2(y+s \mid T_x = t_x)$ egy monoton növekedő függvény t_x-re nézve).*
*T_x és T_y hátralévő élettartamok **hosszú idejű függést** mutatnak, ha $\mu_1(x+t \mid T_y = t_y)$ konstans (állandó), vagy monoton csökkenő függvénye t_y-nak (ekvivalensen, ha $\mu_2(y+s \mid T_x = t_x)$ konstans, vagy monoton csökkenő függvénye t_x-nek).*

Ez a definíció tehát a következőt mondja ki: rövid idejű függés van két élet között amikor a túlélő házastárs halálának az esélye növekszik az együtt eltöltött idő függvényében. Hosszú idejű függésnél pedig pont fordítva, azaz minél később hal meg a párja, annál kisebb lesz az ő halálának valószínűsége, vagy egyáltalán nem is változik.

A továbbiakban 3 darab túlélési függvényre lesz szükségünk:

- $S_1(t_x) = P(T_x > t_x)$ a férj túlélési függvénye;
- $S_2(t_y) = P(T_y > t_y)$ a feleség túlélési függvénye;
- $S(t_x, t_y) = P(T_x > t_x, T_y > t_y)$ a közös túlélési függvényük.

A házastársak túlélésének vizsgálatánál a következő eseteket vizsgálhatjuk meg:

1. A feleség elhunyt és a férj még életben van. Ilyenkor a férj hátralévő élettartamát vizsgáljuk.
2. A feleség még életben van, de a férj elhunyt, amikor is a feleség fennmaradó idejére koncentrálunk.
3. Mindketten életben vannak. Ekkor további három dolgora lehetünk kíváncsiak:
   - A házasság alatt várhatóan kit ér előbb a halál és kit utóbb.
   - A házasság alatt a férj várhatóan meddig él, ha tudjuk, hogy a feleség megél egy adott időtartamot.
   - A házasság alatt a feleség várhatóan meddig él feltéve, hogy a férj túlél egy adott időt.

1.3. Egyéni túlélések vizsgálata

Ebben az alfejezetben egyének túléléseit vizsgáljuk.

Nézzük először a legelső esetet, vagyis tegyük fel, hogy a feleség elhunyt t_y és $t_y + dt$ időtartam között (vagyis ez jelöli a halál pillanatát), ahol $t_y \in [0, t]$ (t a vizsgálat időtartama). Ekkor a férj túlélési függvényére vagyunk kíváncsiak, vagyis arra, hogy mennyi a valószínűsége annak, hogy még s időt élni fog.

$$\begin{aligned} S_{1;t}(s \mid T_y = t_y) &= P(T_x > s+t \mid T_x > t, T_y = t_y) = \frac{P(T_x > s+t, T_x > t, T_y = t_y)}{P(T_x > t, T_y = t_y)} \\ &= \frac{P(T_x > s+t, T_y = t_y)}{P(T_x > t, T_y = t_y)} = \frac{-\frac{\partial}{\partial t_y} P(T_x > t+s, T_y > t_y)}{-\frac{\partial}{\partial t_y} P(T_x > t+s, T_y > t_y)} = \frac{\frac{\partial}{\partial t_y} S(t+s, t_y)}{\frac{\partial}{\partial t_y} S(t, t_y)}. \end{aligned}$$

Következőleg vizsgáljuk meg a férj élettartamát akkor, amikor a feleség is még az élők sorában tartózkodik, és a vizsgálat tartson $t \geq 0$ ideig. Ekkor, ha tudjuk, hogy a feleség t időt él még, akkor a férj feltételes túlélési függvénye előáll az alábbi alakban:

$$\begin{aligned} S_{1;t}(s \mid T_y > t) &= P(T_x > t+s \mid T_x > t, T_y > t) \\ &= \frac{P(T_x > t+s, T_x > t, T_y > t)}{P(T_x > t, T_y > t)} = \frac{P(T_x > t+s, T_y > t)}{P(T_x > t, T_y > t)} = \frac{S(t+s, t)}{S(t,t)}. \end{aligned}$$

Hasonló meggondolással kaphatjuk az $S_{2;t}(s \mid T_x = t_x)$ feltételes túlélési függvényt, ami a feleség várható élettartamát jelöli, ha a férj elhunyt t_x közös időtartam után és $S_{2;t}(s \mid T_x > t)$ túlélési függvényt, ami a feleség várható élettartamát mutatja az esküvő pillanatától, ha tudjuk, hogy a férj a házasság alatt él még t időt.

Ezek után lehetőségünk van a közös túlélési függvény meghatározására. Egy adott t időtartamra tehát $S_t(s_1, s_2)$ felírható:

$$\begin{aligned} S_t(s_1, s_2) &= P(T_x > t+s_1, T_y > t+s_2 \mid T_x > t, T_y > t) \\ &= \frac{P(T_x > t+s_1, T_y > t+s_2, T_x > t, T_y > t)}{P(T_x > t, T_y > t)} \\ &= \frac{P(T_x > t+s_1, T_y > t+s_2)}{P(T_x > t, T_y > t)} = \frac{S(t+s_1, t+s_2)}{S(t,t)}. \end{aligned}$$

1.4. Együttes túlélés vizsgálata

Egy speciális problémája a túlélés-analízisnek a két különböző élet kapcsolatának tanulmányozása. A következő struktúra gyakran előfordul az egészségügyben (két vese élettartamának vizsgálata) és más demográfiai vizsgálatokban (ikerpárok élettartamának vizsgálata) is. A jelölések ebben a fejezetben is legyenek ugyanazok, vagyis T_x és T_y olyan valószínűségi változók, amik a férj és a feleség hátralévő élettartamát mutatják.

Tegyük fel, hogy a megfigyelés t ideig tart. Ennek a kezdetén legyen a férfi x, míg a nő y korú. Az egyének életének megfigyelése véget ér, ha a halál előbb következik be, mint a koruk és t összege, vagy ha lejár a t idő. Az utóbbi esetben csak azt tudjuk, hogy a halál mi után következett be, ami a felülről való levágást jelenti.

Legyen $v = (v_1, v_2, ..., v_n)$ olyan vektor, aminek az i-edik koordinátája jelöli az i-edik házaspár megfigyelését. Ezeket a koordinátákat egy vektor állítja elő, mégpedig:

$$v_i = (x_i, y_i, t_{ix}, t_{iy}, c_{ix}, c_{iy}),$$

ahol x_i, y_i a férj és a feleség kora a megfigyelés kezdetén, t_{ix}, t_{iy} a hátralévő életüket jelöli és c_{ij} $(j \in \{x, y\})$ a cenzoráló indikátor:

$$c_{ij} = \begin{cases} 0, & \text{ha } t_{ij} = t \text{ (cenzorálás)} \\ 1, & \text{ha } t_{ij} < t \text{ (nincs cenzorálás)} \end{cases}$$

Az egyik legfontosabb kérdés a sok lehetséges közül az, hogy vajon meddig fognak a házastársak élni. A következőkben becslést adok a jövőbeli élettartamok eloszlására a belépési korokat figyelembe véve.

Az együttes első élet túlélési függvény (*joint first-life survival function*):

$$p_{FL}(t; x, y) = P(\min\{X - x, Y - y\} > t \mid \min\{X - x, Y - y\} > 0)$$

Ez adja annak a valószínűségét, hogy együtt megélnek t időt, vagy máshogy fogalmazva a házasságban az első halál t idő után következik be.

Az együttes utolsó túlélő függvény (*joint last-survivor function*):

$$p_{LS}(t; x, y) = P(\max\{X - x, Y - y\} > t \mid \min\{X - x, Y - y\} > 0)$$

Annak a valószínűsége, hogy a tovább élő fél mikor hal meg.

Egy egzakt formulával is megadható ez a két függvény a kétváltozós közös túlélési függvénnyel, vagyis $S(t_x, t_y)$-nal:

$$p_{FL}(t; x, y) = \frac{S(x + t, y + t)}{S(x, y)}$$

és

$$p_{LS}(t; x, y) = \frac{S(x, y + t) + S(x + t, y) - S(x + t, y + t)}{S(x, y)}$$

2. fejezet

A kopula modell

A kopula modell napjaink egyik legnépszerűbb módszere a valószínűségi változók közötti összefüggések vizsgálatában, főként a biostatisztika, a biztosításmatematika és a pénzügy területén. A kopulák alapötlete, hogy többdimenziós eloszlás esetén szeretnénk az egydimenziós valószínűségi változók közötti függőséget modellezni úgy, hogy a peremeloszlások és az együttes eloszlás között keresünk olyan kapcsolatot, melyet egy többdimenziós függvénnyel kifejezhetünk.

Ebben a fejezetben ismertetem a kopula modellt, majd az előző fejezetben tárgyaltakkal összhangba hozom.

2.1. A kopula modell általában

A kopula függvény (*copula* - egyesülés, összekötés) összeköti az együttes eloszlásfüggvényt a peremeloszlás függvényekkel. A copula szó először Abe Sklar dolgozatában jelent meg 1956-ban.

2.1.1. Definíció (kopula függvény). *A $C : [0,1]^n \to [0,1]$ n változós függvény* ***kopula függvény****, ha igazak a következő tulajdonságok:*

1. $C(u_1, u_2, \ldots, u_{k-1}, 0, u_{k+1}, \ldots, u_n) = 0$, $\forall u_1, u_2, \ldots, u_n \in [0,1]$
2. $C(1, 1, \ldots, 1, u_k, 1, \ldots, 1) = u_k$, $\forall u_k \in [0,1]$, *ahol* $k = 1, \ldots, n$
3. *n-növő, azaz:*
   $\Delta_{\boldsymbol{a}}^{\boldsymbol{b}} C(\boldsymbol{u}) = \Delta_{a_1}^{b_1} \ldots \Delta_{a_n}^{b_n} C(\boldsymbol{u}) \geq 0$, $\forall \boldsymbol{a} < \boldsymbol{b} \in [0,1]^n$.

Itt $\Delta_{a_k}^{b_k} C(\boldsymbol{u}) = C(u_1, \ldots, u_{k-1}, b_k, u_{k+1}, \ldots, u_n) - C(u_1, \ldots, u_{k-1}, a_k, u_{k+1}, \ldots, u_n)$ *differencia.*

A következő állítás lényege, hogy egy többváltozós eloszlásfüggvény megadható a marginálisai függvényében egy kopula segítségével. Ezt az összefüggést Sklar vette észre (amit a [6] forrásban tanulmányoznak):

2.1.1. Tétel (Sklar tétele). *Legyen* $F \in \mathcal{F}(F_1, \ldots, F_n)$ *egy* n *dimenziós eloszlásfüggvény* $F_1, \ldots, F_n$ *peremeloszlásokkal. Ekkor* $\exists C$ *kopula függvény, amire igaz a következő:*

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$$

.

A bizonyításhoz szükségünk lesz egy másik tételre is, ami azt mondja ki, hogy ha adott egy valószínűségi változónk és azt a saját eloszlásfüggvényébe beírjuk, akkor egy $[0, 1]$-en egyenletes eloszlású változót kapunk.

2.1.2. Tétel (Eloszlás-transzformáció). *Legyen* U *egy eloszlás-transzformáltja az* X *valószínűségi változónak, azaz* $U := F(X, Y)$*, ahol* Y *egy* $(0, 1)$*-en egyenletes eloszlású valószínűségi változó és*

$$F(x, \lambda) := P(X < x) + \lambda P(X = x)$$

.

Ekkor U *egyenletes eloszlású valószínűségi változó a* $(0, 1)$ *intervallumon és* $X = F^{-1}(U)$ *m.m.*

2.1.1. Megjegyzés. *Speciálisan, ha* $\lambda = 1$*, akkor az* X *valószínűségi változó eloszlásfüggvényét kapjuk.*

Bizonyítás: Legyen $0 < \alpha < 1$ esetén $q_\alpha^-(X) := \sup\{x : P(X \le x) < \alpha\}$ egy α-kvantilis. Ekkor $F(X, V) \le \alpha$ akkor és csak akkor, ha

$$(X, V) \in \{(x, \lambda) : P(X < x) + \lambda P(X = x) \le \alpha\}.$$

Ha $\beta := P(X = q_\alpha^-(X)) > 0$ és $q := P(X < q_\alpha^-(X))$, akkor a fenti halmaz ekvivalens a következővel: $\{X < q_\alpha^-(X)\} \cup \{X = q_\alpha^-(X), q + V\beta \le \alpha\}$ és

$$P(U \le \alpha) = P(F(X,V) \le \alpha) = q + \beta P(V \le \frac{\alpha - q}{\beta}) = q + \beta \frac{\alpha - q}{\beta} = \alpha.$$

Ha $\beta = 0$, akkor

$$P(F(X,V) \le \alpha) = P(X < q_\alpha^-(X)) = P(X \le q_\alpha^-(X)) = \alpha.$$

A másik állítás bizonyításához vegyük észre, hogy U definíciója miatt fennáll, hogy $F(X-) \le U \le F(X)$.

Mivel bármilyen $u \in (F(x-), F(x)]$-re igaz, hogy $F^{-1}(u) = x$, a fenti egyenlőtlenségből következik, hogy $F^{-1}(U) = X$ m.m. □

Akkor most nézzük Sklar tételének bizonyítását.

Bizonyítás: Legyen $X = (X_1, ..., X_n)$ egy tetszőleges vektor az $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn. Legyen F az együttes eloszlásfüggvény, és V X-től független a $(0,1)$ intervallumon egyenletes eloszlású valószínűségi változó (azaz $V \sim U(0,1)$). Tekintsük az $U_i := F_i(X_i, V)$ eloszlástranszformációt. Ekkor az 2.1.2.Tétel miatt $U_i \sim U(0,1)$ és $X_i = F_i^{-1}(U_i)$ m.m., $\forall 1 \le i \le n$. Így az $U = (U_1, \ldots, U_n)$ eloszlásfüggvényeként definiált C segítségével írható fel F:

$F(x) = P(X \le x) = P(F_i^{-1}(U_i) \le x_i, 1 \le i \le n) = P(U_i \le F_i(x_i), 1 \le i \le n) = C(F_1(x_1), \ldots, F_n(x_n))$,

vagyis F-nek C kopula függvénye. □

Tekintsük Sklar tételét két változóban. Ekkor a tétel a következő:

2.1.3. Tétel. *Bármely kétváltozós $H(x,y)$ eloszlásfüggvény, melynek folytonosak az F_1, F_2 peremeloszlásai, egyértelműen reprezentál egy kopula függvényt, mégpedig a $C(u,v) = H(F_1^{-1}(u), F_2^{-1}(v))$ függvényt.*

Szükségünk lesz még a következőkre a kopulák alkalmazásához:

2.1.2. Definíció (Pozitívan kvadratikus összefüggés (PQD)). *Legyenek X és Y valószínűségi változók. Azt mondjuk, hogy X és Y pozitívan kvadratikusan összefüggnek, ha $\forall x, y$-ra fennáll, hogy $P(X \le x, Y \le y) \ge P(X \le x)P(Y \le y)$.*

.

2.1.1. Állítás (Fréchet-Hoeffing-határok). *Tetszőleges $C(\boldsymbol{u})$ n dimenziós kopulára fennáll az alábbi egyenlőtlenség:*

$$\max\left\{\sum_{i=1}^{n} u_i + 1 - d, 0\right\} \leq C(\boldsymbol{u}) \leq \min\{u_1, \ldots, u_n\}$$

Az egyenlőtlenség bal oldalát szokás Fréchet alsó határnak, a jobb oldalát, pedig Fréchet felső határnak nevezni.

2.2. Gauss- és t-kopula

Az egyik módja a kopula függvények definiálásának a kétváltozós eloszlásfüggvények invertálása. Az együttes viselkedések tanulmányozásánál rendkívül népszerű módszereket alakítottak ki. Ilyen például a 2-dimenziós Gauss kopula, ami a

$$\begin{aligned} C^G(u, v; \rho) &= \Phi_\rho^2(\Phi^{-1}(u), \Phi^{-1}(v)) \\ &= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2 - \rho x y + y^2}{2(1-\rho^2)}} dx dy \end{aligned}$$

függvény. Itt $\Phi(x)$ az egyváltozós standard normális eloszlásfüggvényt, és $\Phi_\rho^2(x, y)$ a kétváltozós normális eloszlás együttes eloszlásfüggvényét jelöli, melynek marginálisai standard normálisak, ρ korrelációval a peremeloszlások között.

A Student t-kopula hasonló alapokra épül, itt a

$$\begin{aligned} C^t(u, v; \rho, \nu) &= t_{\rho,\nu}^2(t_\nu^{-1}(u), t_\nu^{-1}(v)) \\ &= \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \frac{1}{(1 + \frac{x^2 - \rho x y + y^2}{\nu(1-\rho^2)})^{1+\frac{\nu}{2}}} dx dy \end{aligned}$$

kopula a reprezentáló függvény, ahol $t_\nu(x)$ a t-eloszlás ν szabadságfokkal, és $t_{\rho,\nu}^2(x, y)$ a kétváltozós t-eloszlás eloszlásfüggvénye ρ korrelációval.

2.3. Arkhimédészi kopulák

A kopula felépítésének módszere nem korlátozódik a Gauss és a t-eloszlásra, ugyanis ezek az eloszlások nem mindig írják le megfelelően a valóságot, így vannak sokkal alkalmasabbak is. Használhatjuk a peremeloszlás-függvények helyett a túlélési függvényeket is.

Tehát lehetnek a kopula függvény argumentumai egyváltozós túlélési függvények, vagyis $S_1(t_x) = P(X > t_x)$ és $S_2(t_y) = P(Y > t_y)$. Ekkor $S(t_x, t_y) = C(S_1(t_x), S_2(t_y))$ alakba írható, ahol a C kopula az $(S_1(t_x), S_2(t_y))$ együttes eloszlásfüggvénye.

Ezt az elgondolást használják fel az úgynevezett Arkhimédészi kopulák, amik már jól ismertek és széles körben elfogadottak a matematikai kezelhetőségük, valamint rugalmasságuk miatt. Ezek egy adott generáló függvény segítségével állíthatóak elő. Definiálunk egy $\Phi : [0,1] \to \mathbb{R}^+$ szigorúan monoton csökkenő függvényt, aminek létezik és folytonos az első, illetve a második deriváltja. Ezeket jelölje $\Phi'(\tau)$ és $\Phi''(\tau)$. A generáló függvényre a következőek legyenek jellemzőek:

- $\Phi(1) = 0$;
- $\Phi'(\tau) < 0$, ahol $0 \leq \tau \leq 1$;
- $\Phi''(\tau) > 0$, ahol $0 \leq \tau \leq 1$.

Ekkor az Arkhimédészi kopulák a következő módon állnak elő:

$$C(u, v; \alpha) = \Phi(\Phi^{[-1]}(u) + \Phi^{[-1]}(v)), u, v \in [0,1],$$

ahol α a kapcsolati paraméter ($\Phi(\cdot)$ képletében található) és $\Phi^{[-1]}(\cdot)$ pszeudoinverze $\Phi(\cdot)$-nek, vagyis

$$\Phi^{[-1]}(\tau) = \begin{cases} \Phi^{-1}(\tau), & \text{ha } 0 \leq \tau \leq \Phi(0) \\ 0, & \text{ha } \Phi(0) \leq \tau \leq \infty \end{cases}$$

Jelen esetben csak olyan függvényeket vizsgálunk, ahol a függvény pszeudonverze egybeesik az inverzével.

Az Arkimédészi kopulákra továbbá meghatározható a Kendall-féle korrelációs együttható (τ).

2.3.1. Definíció (Kendall-féle korrelációs együttható). *A **Kendall-féle** τ a következő:*

$$\tau = P((X_2 - X_1)(Y_2 - Y_1) > 0) - P((X_2 - X_1)(Y_2 - Y_1) < 0),$$

ahol (X_1, Y_1), (X_2, Y_2) két független példány az együttes eloszlásból.

Ennek van egy tapasztalati verziója is, amikor ez az együttható két szám különbségéből kapható. Az egyik a p_+, ami a konkordáns (egyirányú) párok aránya a populációban, a másik pedig a p_-, ami a diszkordáns (ellentétes irányú) párok aránya, így kapjuk, hogy $\tau = p_+ - p_-$. Minél nagyobb ennek a mennyiségnek az abszolút értéke, annál erősebb a kapcsolat a valószínűségi változók között. A Kendall-féle τ jellemzői:

- $-1 \leq \tau \leq 1$;
- ha X, Y független, akkor $\tau = 0$;
- $\tau = -1$: determinisztikusan fogyó kapcsolat;
- $\tau = 1$: determinisztikusan növő kapcsolat.

Ezek után tekintsünk néhágy jól ismert Arkhimédészi kopulát.

Clayton kopulája

Ezt a kopulát a

$$\Phi^{-1}(t) = t^{-\alpha} - 1$$

függvény generálja. Ekkor a kopula a következőképp néz ki:

$$C_C(u, v; \alpha) = (u^{-\alpha} + v^{-\alpha} - 1)^{-\frac{1}{\alpha}},$$

ahol α a kapcsolati paraméter. A kopula főbb jellemzői:

- *PQD*: ha $\alpha > 0$
- *függetlenség*: ha $\alpha \downarrow 0$
- *Kendall τ maximuma*: $1(\alpha \to \infty)$

Gumbel-Hougaard kopula

Itt a generáló függvény a következő alakú:

$$\Phi^{-1}(t) = (-\log t)^{\alpha}.$$

Ekkor kapjuk, hogy a kopula a

$$C_{GH}(u, v; \alpha) = e^{-[(-\log u)^{\alpha} + (-\log v)^{\alpha}]^{\frac{1}{\alpha}}}$$

függvény, ahol α szintén a kapcsolati paraméter. A kopula főbb jellemzői:

- *PQD*: ha $\alpha \geq 1$
- *függetlenség*: ha $\alpha = 1$
- *Kendall τ maximuma*: $1(\alpha \to \infty)$

Frank kopulája

Ezt a kopulát a

$$\Phi(t) = -log\frac{e^{\alpha t} - 1}{e^{\alpha} - 1}$$

függvény generálja. Ebből kapjuk a kopulát, ami a

$$C_F(u, v; \alpha) = -\frac{1}{\alpha}\log\left[1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1}\right]$$

alakban írható. A Frank kopula jellemzői:

- *PQD*: ha $\alpha < 0$
- *függetlenség*: ha $\alpha \uparrow 0$
- *Kendall τ maximuma*: $1(\alpha \to -\infty)$

Az Arkhimédészi kopuláknak még számos fajtája van, de a legnépszerűbb az imént tárgyalt 3, így többet nem sorolok fel.

2.4. Egyéni túlélések vizsgálata kopulák segítségével

A kopulákat alkalmazzák az aktuáriusi vizsgálatok során, amikor életbiztosításokhoz számolnak várható élettartamokat. A közös túlélési függvény tekinthető egy kétváltozós kopula függvénynek, ahol a változók helyére az egyéni túlélési függvényeket írjuk:

$$S(s_1, s_2) = C[S_1(s_1), S_2(s_2)]$$

Ebben az alfejezetben a korábban definiált túlélés-analízis-beli függvényeket adjuk meg kopula függvények segítségével

Először is nézzük mennyi a valószínűsége annak, hogy a férj él még s időt, ha a feleség meghal t_y idő után. Ekkor a túlélési függvény a következő lesz:

$$\begin{aligned} S_{1;t}(s \mid T_y = t_y) &= P(T_x > t + s \mid T_x > t, T_y = t_y) \\ &= \frac{\frac{\partial}{\partial t_y} S(t+s, t_y)}{\frac{\partial}{\partial t_y} S(t, t_y)} = \frac{C_2[S_1(t+s), S_2(t_y)]}{C_2[S_1(t), S_2(t_y)]}, \end{aligned}$$

ahol $C_2[\cdot, \cdot]$ jelöli a második változó szerinti parciális deriváltját $C[\cdot, \cdot]$-nak. Természetesen ez a meghatározás csak akkor értelmes, ha a nevező nem nulla, vagyis $C_2[S_1(t), S_2(t_y)] \neq 0$.

Jobban szeretjük a halálozást a hazárdfüggvényekkel szemléltetni, ha lehetőség van rá. Tegyük fel, hogy $C_2[S_1(t+s), S_2(t_y)] \neq 0$. Ekkor a hazárdfüggvény:

$$\begin{aligned}\mu_1(x+t+s \mid T_y = t_y) &= -\frac{\partial}{\partial s} \log C_2[S_1(t+s), S_2(t_y)] \\ &= \mu_1(x+t+s) \frac{S_1(t+s) C_{21}[S_1(t+s), S_2(t_y)]}{C_2[S_1(t+s), S_2(t_y)]},\end{aligned}$$

ahol $\mu_1(x+t+s)$ jelöli a férj hazárdfüggvényét jelöli $x+t+s$ korra T_x eloszlásának megfelelően, továbbá $C_{21}[\cdot,\cdot]$ a kopula függvény második parciális deriváltját, ahol először a második, majd az első változó szerint deriválunk.

2.4.1. Állítás. *Tegyük fel, hogy a hazárdfüggvény állandó, azaz konstans. Ekkor a két házastárs élettartama független.*

Bizonyítás: Tegyük fel, hogy a férj hazárdfüggvénye konstans. Legyen $u = S_1(t+s)$ és $v = S_2(t_y)$. Ekkor $\mu_1(x+t+s \mid T_y = t_y)$ képletében levő kopula függvényekre igaz, hogy:

$$\frac{C_{21}[u,v]}{C_2[u,v]} = \frac{\partial \log C_2[u,v]}{\partial u}; u, v \in [0,1],$$

itt a logaritmusfüggvényt, mint összetett függvényt deriváljuk és kapjuk ezt az összefüggést. A kezdetiérték feltétel itt simán csak a kopula definíciójából kapható, vagyis, hogy $C(0,v) = C(u,0) = 0$ és $C(u,1) = u, C(1,v) = v$. A fenti kifejezés független v-től, ezért $\log C_2[u,v]$ felírható $K_1(u) + K_2$ alakban. Itt $K_1(u)$ jelöl egy olyan valós értékű differenciálható függvényt, ami csak u-tól és a függés paraméterétől függ, de v-től nem, valamint $K_2 \in \mathbb{R}$.

Ezt visszaintegrálva megkapjuk a kopula függvényt, vagyis ekkor:

$$\begin{aligned}C_2[u,v] &= e^{K_1(u)+K_2} \\ C[u,v] &= ve^{K_1(u)+K_2} + K_3,\end{aligned}$$

ahol $K_3 \in \mathbb{R}$.

Abból a feltételből, hogy $C[u,0] = 0$ (ahol $u \in [0,1]$) kapjuk, hogy $K_3 = 0$. A másodikból ($C[u,1] = u, u \in [0,1]$ pedig, hogy

$$\begin{aligned}u &= e^{K_1(u)+K_2} \\ \log u &= K_1(u) + K_2 \\ K_1(u) &= \log u - K_2.\end{aligned}$$

Így kapjuk végül, hogy a kopula függvényünk a következő:

$$C[u, v] = uv,$$

ami egy független kopula. Más szavakkal, ha két élet függ egymástól és egy kopula modellel felírható ez a függés, akkor az egyéni hazárdfüggvény mindig függ a másik ember halálának idejétől. □

A függésnek ezt a típusát vizsgáljuk a továbbiakban néhány kopulacsalád esetén.

2.4.1. Speciális esetek

Függetlenség

Tudjuk, hogy ekkor a kopula alakja $C[u, v] = uv$. Ebben az esetben a túlélési függvény és a feltételes hazárdfüggvényt könnyen kiszámíthatjuk:

$$S_{1;t}(s \mid T_y = t_y) = \frac{C_2[S_1(t+s), S_2(t_y)]}{C_2[S_1(t), S_2(t_y)]} = \frac{\frac{\partial}{\partial S_2(t_y)}[S_1(t+s)S_2(t_y)]}{\frac{\partial}{\partial S_2(t_y)}[S_1(t)S_2(t_y)]} = \frac{S_1(t+s)}{S_1(t)}.$$

$$\begin{aligned}
\mu_1(x+t+s \mid T_y = t_y) &= \mu_1(x+t+s)\frac{S_1(t+s)C_{21}[S_1(t+s), S_2(t_y)]}{C_2[S_1(t+s), S_2(t_y)]}\times \\
&= \mu_1(x+t+s)\frac{S_1(t+s)\frac{\partial^2}{\partial S_1(t+s)\partial S_2(t_y)}[S_1(t+s)S_2(t_y)]}{\frac{\partial}{\partial S_2(t_y)}[S_1(t+s), S_2(t_y)]} \\
&= \mu_1(x+t+s)\frac{S_1(t+s)}{S_1(t+s)} \\
&= \mu_1(x+t+s)
\end{aligned}$$

Erre számítottunk, vagyis arra, hogy a feltételes hazárdfüggvénye az egyénnek nem változik, ha a két élet független egymástól.

Hasonlóan ki tudjuk számolni a feleség túlélési függvényét és hazárdfüggvényét.

Fréchet felső határ

Ebben az esetben a kopula függvényünk $C[u, v] = \min\{u, v\}$ alakú, és így a második változó szerinti parciális deriváltja a $C_2[u, v] = I_{u>v}$. Tehát $S_{1;t}(s \mid T_y = t_y)$ akkor és csak akkor létezik, ha $S_2(t_y) < S_1(t)$. Most is meg tudjuk adni a férj és a feleség túlélési függvényét:

$$S_{1;t}(s|T_y = t_y) = \begin{cases} 1, & \text{ha } s < S_1^{[-1]}(S_2(t_y)) - t \\ 0, & \text{ha } s > S_1^{[-1]}(S_2(t_y)) - t \end{cases},$$

ahol

$$S_1^{[-1]}(\kappa) = \begin{cases} S_1^{-1}(\kappa), & \text{ha } 0 < \kappa \leq 1 \\ \omega_x, & \text{ha } \kappa = 0 \end{cases}.$$

Azaz, ha a feleség t_y idő után meghal, akkor a férj $x + S_1^{[-1]}(S_2(t_y))$ korban fog.

Fréchet alsó határ

Fréchet alsó határ esetén a kopula $C[u,v] = \max\{u+v-1,0\}$ alakban adható meg és ekkor a második változó szerinti parciális derivált $C_2[u,v] = I_{v>u-1}$, ami miatt $S_{1;t}(s|T_y = t_y)$ akkor és csak akkor létezik, ha $S_2(t_y) > 1 - S_1(t)$. Ekkor a feltételes túlélési függvény:

$$S_{1;t}(s|T_y = t_y) = \begin{cases} 1, & \text{ha } s < S_1^{[-1]}(1 - S_2(t_y)) - t \\ 0, & \text{ha } s > S_1^{[-1]}(1 - S_2(t_y)) - t \end{cases}.$$

Vagyis szavakkal megfogalmazva, ha a feleség t_y idő után hal meg, akkor a férj $x + S_1^{-1}(1 - S_2(t_y))$ korban fog.

2.4.2. Arkhimédészi kopulák alkalmazásával

Tudjuk, hogy az Arkhimédészi kopulák $C[u,v] = \Phi^{-1}(\Phi(u) + \Phi(v))$ alakban állnak elő. Most ezt helyettesítsük be a feltételes hazárdfüggvénybe:

$$\begin{aligned} \mu_1(x+t+s \mid T_y = t_y) = &\mu_1(x+t+s)S_1(t+s)(-\Phi'(S_1(t+s)))\times \\ &\left(-\frac{(\Phi^{-1})''(\Phi(S(t+s,t_y)))}{(\Phi^{-1})'(\Phi(S(t+s,t_y)))}\right). \end{aligned}$$

Ez függ t_y-tól. Ha $-\frac{(\Phi^{-1})''(\Phi(S(t+s,t_y)))}{(\Phi^{-1})'(\Phi(S(t+s,t_y)))}$ monoton csökkenő (növekedő) t_y-ban, akkor hosszú idejű (rövid idejű) függésről beszélünk.

Most tegyük fel, hogy hosszú idejű a függés a házastársak között, ugyanis a legtöbb kopula ezen a függésen alapszik. Egy alap példája a hosszú idejű függésnek a ***törékenység***, amit a következőekben tárgyalunk.

Ha a generátor inverze egy Laplace transzformált, akkor előáll a következő alakban:

$$\Phi^{-1}(v) = \int e^{-zv} dF(z),$$

ahol $F(z)$ a törékenység eloszlásfüggvénye. Ekkor a fenti függvény monoton növekedő (amit a [?]-ban bizonyítanak is). Ebből látható, hogy a törékenység tényleg a hosszú idejű függés egy speciális esete.

2.4.1. Példa (Clayton kopula esetén). *A generátor inverze,* $\Phi^{-1}(\tau) = (\tau+1)^{-\frac{1}{\alpha}}$ *a Laplace transzformáltja a Gamma(a, b) eloszlásnak, ahol* $a = \alpha^{-1}$ *és* $b = 1$*. Általában a Gamma(a, b) eloszlás Laplace transzformáltja* $\Phi^{-1}(\tau) = (\frac{\tau}{b}+1)^{-\frac{1}{\alpha}}$*, ahonnan a generátor* $\Phi(\tau) = b(t^{-\alpha}-1)$*. Jelen esetben* $b = 1$*, így ez nem befolyásolja az eredményt, vagyis a hazárdfüggvénye a férjnek:*

$$\mu_1(x+t+s \mid T_y = t_y) = \mu_1(x+t+s)(\alpha+1)\frac{(S_1(t+s))^{-\alpha}}{(S_1(t+s))^{-\alpha}+(S_2(t_y))^{-\alpha}-1},$$

ami csökkenő t_y*-ban. Ha* $t_y = 0$ *(vagyis a házasságkötés után azonnal meghal), abban az esetben eltűnik a törtes kifejezés és ekkor* $\mu_1(x+t+s \mid T_y = 0) = \mu_1(x+t+s)(\alpha+1)$*. Vagyis szavakkal, ha a feleség a házasságot követően azonnal meghal, akkor a férj hazárdfüggvénye a perem hazárdfüggvénye* $(\alpha+1)$*-szer.*

2.4.2. Példa (Gumbel-Hougaard kopula esetén). *A generátor inverze,* $\Phi^{-1}(\tau) = e^{-\tau^{\frac{1}{\alpha}}}$ *a Laplace transzformáltja a pozitív stabilis eloszlásnak. Ekkor a feltételes hazárdfüggvényre a következőt kapjuk:*

$$\begin{aligned}\mu_1(x+t+s \mid T_y = t_y) =& \mu_1(x+t+s)\left(\frac{(-\log S_1(t+s))^{\alpha}}{(-\log S_1(t+s))^{\alpha}+(-\log S_2(t_y))^{\alpha}}\right)^{1-\frac{1}{\alpha}} \times \\ & \left(1+(\alpha-1)((-\log S_1(t+s)^{\alpha}+(-\log S_2(t_y))^{\alpha})^{-\frac{1}{\alpha}}\right),\end{aligned}$$

ami szintén monoton csökkenő, mint t_y *függvénye.*

2.4.3. Példa (Frank kopulája esetén). *A generátor inverze,* $\Phi^{-1}(\tau) = \frac{\log\{1+(e^{\alpha}-1)e^{-\tau}}{}\}\alpha$ *Laplace transzformáltja a lognormális eloszlásnak. Ekkor a feltételes hazárdfüggvény:*

$$\begin{aligned}\mu_1(x+t+s \mid T_y = t_y) =& \mu_1(x+t+s)S_1(t+s)(-\alpha)\frac{e^{\alpha S_1(t+s)}}{1-e^{\alpha S_1(t+s)}} \times \\ & \left(\frac{1-e^{\alpha}}{1-e^{\alpha}-(1-e^{\alpha S_1(t+s)})(1-e^{\alpha S_2(t_y)})}\right),\end{aligned}$$

ami szintén monoton csökkenő a t_y*-ban.*

2.5. Közös túlélés vizsgálata kopulák segítségével

2.5.1. Speciális esetek

Függetlenség

Ekkor a kopula függvényünk $C[u, v] = uv$ alakú, amiből a hazárdfüggvény:

$$\mu_1(x+t+s \mid T_y > t) = \mu_1(x+t+s)\frac{S_1(t+s)C_1[S_1(t+s), S_2(t)]}{C[S_1(t+s), S_2(t)]} = \mu_1(x+t+s),$$

amire számítottunk is, ugyanis függetlenség esetében a hazárdfüggvény nem függ a másik élet történetétől. Ekkor kapjuk, hogy a közös túlélési függvény

$$S_t(s_1, s_2) = \frac{S_1(t+s_1)S_2(t+s_2)}{S(t,t)} = S_{1;t}(s_1 \mid T_y > t)S_{2;t}(s_2 \mid T_x > t).$$

Fréchet felső határ

Tegyük fel, hogy a kopula függvényünk $C[u,v] = \min\{u,v\}$ alakú. Ekkor igaz a következő:

$$S_{1;t}(s \mid T_y > t) = \frac{C[S_1(t+s), S_2(t)]}{C[S_1(t), S_2(t)]} = \frac{\min\{S_1(t+s), S_2(t)\}}{\min\{S_1(t), S_2(t)\}}.$$

Hasonlóan kapható $S_{2;t}(s \mid T_y > t)$-ra. Ebben az esetben a közös túlélés kopula függvénye:

$$\begin{aligned} C_t[S_{1;t}(s_1 \mid T_y > t), S_{2;t}(s_2 \mid T_x > t)] &= \frac{C[S_1(t+s_1), S_2(t+s_2)]}{S(t,t)} = \frac{\min(S_1(t+s_1), S_2(t+s_2))}{S(t,t)} \\ &= \min(S_{1;t}(s_1 \mid T_y > t), S_{2;t}(s_2 \mid T_x > t)). \end{aligned}$$

Vagyis ha a közös túlélési függvény kezdetben eléri a Fréchet felső határt, mint kopula, akkor a jövőben is el fogja érni a Fréchet felső határt.

Fréchet alsó határ

Tegyük fel, hogy a kopula rendelkezik Fréchet alsó határ tulajdonságával, vagyis $C[u,v] = \max\{u+v-1, 0\}$ alakú. Ekkor

$$S_{1;t}(s \mid T_y > t) = \frac{\max\{S_1(t+s) + S_2(t) - 1, 0\}}{S(t,t)}$$

és $S_{2;t}(s \mid T_x > t)$ hasonlóan néz ki. Ebben az esetben a közös túlélési függvény, mint kopula függvény a következő alakban írható fel:

$$\begin{aligned} C_t[S_{1;t}(s_1 \mid T_y > t), S_{2;t}(s_2 \mid T_y > t)] &= \frac{\max\{S_1(t+s) + S_2(t+s) - 1, 0\}}{S(t,t)} \\ &= \max(S_{1;t}(s_1 \mid T_y > t) + S_{2;t}(s_2 \mid T_x > t) - 1, 0), \end{aligned}$$

azaz ha a közös túlélési függvény eléri a Fréchet alsó határt a kezdetekkor, akkor a továbbiakban is el fogja érni.

2.5.2. Arkhimédészi kopulák alkalmazásával

Alkalmazzuk az együttes eloszlás meghatározásához az Arkhimédészi kopulákat. Nézzük először, hogy számítható ki a férj hátralévő idejének feltételes túlélési függvénye ezen kopulák segítségével, ami megadja, hogy mennyi a valószínűsége annak, hogy a férj él még $s+t$ ideig, ha tudjuk, hogy a feleség él t időt:

$$S_{1;t}(s \mid T_y > t) = \frac{\Phi^{-1}(\Phi(S_1(t+s)) + \Phi(S_2(t)))}{S(t,t)}.$$

Hasonló egyenlőséget kapunk a feleség feltételes túlélési függvényére, vagyis $S_{2;t}(s \mid T_x > t)$-re.

A hazárdfüggvény annyiban fog változni az korábbi esethez képest, hogy most nem azt nézzük, hogy a feleség megélt t_y időt, hanem, hogy tovább él, mint t idő, vagyis:

$$\mu_1(x+t+s \mid T_y > t) = \mu_1(x+t+s)S_1(t+s)(-\Phi'(S_1(t+s)))\left(-\frac{(\Phi^{-1})'(\Phi(S(t+s,t))}{\Phi^{-1}(\Phi(S(t+s,t)))}\right).$$

2.5.1. Tétel. *Tegyük fel, hogy az együttes túlélési függvény kopulája Arkhimédészi kopula a kezdetekkor ($t=0$). Ekkor a feltételes együttes túlélési függvény kopulája (azzal a feltétellel, hogy T_x és T_y is nagyobb, mint t) szintén Arkhimédészi.*
Precízebben: Legyen $\Phi(\cdot)$ az Arkhimédészi kopula generátorfüggvénye $t=0$ időpontban. Ekkor $\Phi_t(\cdot)$ (generálófüggvény t időpontban) a következő alakban áll elő:

$$\Phi_t(\tau) = \Phi(\tau S(t,t)) - \Phi(S(t,t)),$$

ahol $\tau \in [0,1]$.

Bizonyítás: Megmutatjuk, hogy az a kopula, amit $\Phi_t(\cdot)$ generál mindkét élet túlélését adja t idő után.

Először vizsgáljuk meg, hogy $\Phi_t(\cdot)$ rendelkezik-e az Arkhimédészi kopulák generátorfüggvényének tulajdonságaival ($\Phi_t(1)=0$, $\Phi_t'(\tau)<0$ és $\Phi_t''(\tau)>0$):

$$\Phi_t(1) = \Phi(1S(t,t)) - \Phi(S(t,t)) = 0$$

$$\Phi_t'(\tau) = \Phi'(\tau S(t,t))S(t,t) < 0,$$

mert $\Phi'(\tau S(t,t)) < 0$, hiszen $\Phi(\cdot)$ generáló függvény és $S(t,t) \in [0,1]$, mivel valószínűség.

$$\Phi_t''(\tau) = \Phi''(\tau S(t,t))(S(t,t))^2 > 0,$$

mert $\Phi''(\tau S(t,t)) > 0$, hiszen $\Phi(\cdot)$ generáló függvény és $(S(t,t))^2 \geq 0$, mivel négyzetszám nem lehet negatív. A következő lépésben meghatározzuk a kopulát, amit $\Phi_t(\cdot)$ generál. Legyen $\Phi_t(\cdot)$ inverze $\Phi_t^{-1}(\cdot)$. Ekkor:

$$\Phi_t^{-1}(\tau) = \frac{\Phi^{-1}(\tau + \Phi(S(t,t)))}{S(t,t)}.$$

Innen már meg tudjuk adni az új kopulát, amit értelemszerűen $C_t[\cdot,\cdot]$-tal jelölünk:

$$C_t[u,v] = \Phi_t^{-1}(\Phi_t(u) + \Phi_t(v)) = \frac{\Phi^{-1}(\Phi(uS(t,t)) + \Phi(vS(t,t)) - \Phi(S(t,t)))}{S(t,t)}.$$

Most helyettesítsünk u és v helyére $S_{1;t}(s \mid T_y > t)$ és $S_{2;t}(s \mid T_x > t)$ feltételes túlélési függvényeket. Ekkor kapjuk, hogy

$$\begin{aligned} C_t[S_{1;t}(s \mid T_y > t), S_{2;t}(s|T_x > t)] &= \frac{\Phi^{-1}(\Phi(S_1(t+s)) + \Phi(S_2(t+s)))}{S(t,t)} \\ &= P(T_x > t+s, T_y > t+s \mid T_x > t, T_y > t). \end{aligned}$$

Azaz $\Phi_t[\cdot]$ tényleg mindkét élet túlélését adja t idő után. □

A továbbiakban a kapcsolat többféle időtől függő mértékéről tárgyalunk, amik közül az egyik a Kendall-féle korrelációs együttható, másik pedig a kereszt-arány (*cross-ratio*) függvény. Először tekintsük a Kendall-féle τ-t kopulák esetén, ami népszerű azon tulajdonsága miatt, hogy független a marginálisok eloszlásától.

2.5.1. Definíció. *Két valószínűségi változó (U,V) Kendall-féle korrelációs együtthatója kopulák segítségével (azaz $H(u,v) = C(F(u), G(v))$) az alábbi alakban adható meg:*

$$\tau(X_1, X_2) = 4\int_0^1\int_0^1 C[u,v]dC[u,v] - 1$$

Ez az együttható Arkhimédészi kopulák esetén megadható a generátorfüggvény segítségével is:

$$\widetilde{\tau}_t(U,V) = 4\int_0^1 \frac{\Phi_t(v)}{\Phi_t'(v)}dv + 1$$

Alkalmazzuk ezt most a házastársak hátralévő idejét tartalmazó valószínűségi változókra, azaz határozzuk meg $\tau(T_x, T_y)$-t:

$$\widetilde{\tau}_t(T_x, T_y) = 4\int_0^1 \frac{\Phi_t(v)}{\Phi_t'(v)}dv + 1 \ = \frac{4}{S(t,t)}\int_{u=0}^1 \frac{\Phi(uS(t,t)) - \Phi(S(t,t))}{\Phi_t'(uS(t,t))}dv + 1.$$

Ezt az összefüggést nevezik **csonka** τ-nak.

Most nézzük a másik időtől függő kapcsolati mértéket, azaz a kereszt-arány függvényt ($CR(\cdot)$).

2.5.2. Definíció. *A közös túlélésfüggvény ($S(t,t)$) kereszt-arány függvénye ($CR(S(t,t))$) megadható a következő alakban:*

$$CR(S(t_x,t_y)) = \frac{S(t_x,t_y)\frac{\partial^2}{\partial t_x \partial t_y}S(t_x,t_y)}{\frac{\partial}{\partial t_x}S(t_x,t_y)\frac{\partial}{\partial t_y}S(t_x,t_y)}.$$

Ezek után látható, hogy egy egyszerű összefüggést fel tudunk írni a csonka tau és a kereszt-arány függvény között:

$$\widetilde{\tau}_t(T_x,T_y) = \frac{CR(S(t_x,t_y))-1}{CR(S(t_x,t_y))+1}.$$

Két okból szeretjük jobban használni a kereszt-arány függvényt a csonka tau-nál. Egyik szempont, hogy meghatározza az egyének halálozási erejének relatív növekedési ütemét a partner haláláig, vagyis:

$$CR(S(t_x,t_y)) = \frac{\mu_1(x+t_x \mid T_y = t_y)}{\mu_1(x+t_x \mid T_y > t_y)}.$$

A másik szempont pedig az, hogy ezt a függvényt könnyebb becsülni, mint a csonka tau-t, mivel nincs benne integrálás. Tegyük fel ugyanis, hogy $t_x = t_y = t$, ekkor adódik a kereszt-arány függvényre, hogy

$$CR(S(t,t)) = \frac{\Phi^{-1}(\Phi(S(t,t)))(\Phi^{-1})''(\Phi(S(t,t)))}{((\Phi^{-1})'(\Phi(S(t,t))))^2}.$$

Azaz ez a függvény csak a generáló függvénytől függ.

A következőekben meghatározzuk pár kopula családnál ezt a kereszt-arány függvényt.

2.5.1. Példa (Clayton kopulája esetén). *Legyen a generáló függvény a következő: $\Phi(\tau) = \delta(\tau^{-\alpha}-1)$, ahol $\alpha > 0$ és $\delta > 0$. Ekkor a feltételes hazárdfüggvény*

$$\mu_1(x+t+s \mid T_y > t) = \mu_1(x+t+s)\frac{(S_1(t+s))^{-\alpha}}{S_1(t+s))^{-\alpha}+S_2(t))^{-\alpha}-1}$$

alakú. Így ha t időre nézzük a túlélést $\Phi_t(\cdot)$ a következő alakú:

$$\Phi_t(\tau) = \delta(S(t,t))^{-\alpha}(\tau^{-\alpha}-1) = \delta(S(t,t))^{-\alpha}\Phi(\tau),$$

vagyis a kopula lényegében nem változik az idő múlásával. Ilyenkor a kereszt-arány függvény konstans az adott időben és egyenlő $\alpha+1$-gyel.

2.5.2. Példa (Gumbel-Hougaard kopula esetén). *Ebben az esetben a generáló függvény* $\Phi(\tau) = (-\log\tau)^\alpha$ *alakú, ahol* $\alpha \geq 1$. *Ilyenkor a feltételes hazárdfüggvényre igaz, hogy*

$$\mu_1(x+t+s \mid T_y > t) = \mu_1(x+t+s)\left(\frac{(-\log S_1(t+s))^\alpha}{(-\log S_1(t+s))^\alpha + (-\log S_2(t))^\alpha}\right)^{1-\frac{1}{\alpha}}.$$

Ebből pedig következik, hogy $\Phi_t(\tau) = (-\log\tau S(t,t))^\alpha - (-\log S(t,t))^\alpha$, *vagyis a két élet közötti kapcsolat 0-hoz tart. Más szavakkal a két élet egyre kevésbé függ egymástól az évek múlásával. Így a kereszt-arány függvény a következő lesz:*

$$CR(S(t,t)) = 1 + \frac{\alpha - 1}{-\log S(t,t)}.$$

2.5.3. Példa (Frank kopulája esetén). *Frank kopulájánál a generáló függvény:* $\Phi(\tau) = -\log\frac{(e^{\alpha\tau}-1)}{e^\alpha - 1}$, *ahonnan*

$$\mu_1(x+t+s \mid T_y > t) = \mu_1(x+t+s) \times \frac{(1-e^{\alpha S_2(t)})e^{\alpha S_1(t+s)}(-\alpha)S_1(t+s)}{((e^\alpha - 1) + (e^{\alpha S_1(t+s)} - 1)(e^{\alpha S_2(t)} - 1))\log 1 + \frac{(e^{\alpha S_1(t+s)}-1)(e^{\alpha S_2(t)}-1)}{e^\alpha - 1}}.$$

Innen kapjuk a generáló függvényre t időben, hogy

$$\Phi_t(\tau) = -\log\frac{e^{\alpha S(t,t)\tau} - 1}{e^{\alpha S(t,t)} - 1}$$

. *Ahogy az idő változik, az* α *paraméter* 0*-hoz tart, utalva a függetlenségre. Így a kereszt-arány függvény*

$$CR(S(t,t)) = -\frac{\alpha S(t,t)}{1 - e^{\alpha S(t,t)}}.$$

2.6. A paraméterek becslése

A közös túlélési függvény becslése két lépésben történik. Először becsüljük a perem túlélési függvényeket ($S_1(t_x)$, $S_2(t_y)$), majd a becsült túlélési függvényeket beillesztjük az adott kopulákba, valamint becsüljük a kapcsolati paramétert.

A túlélési függvényt becsülhetjük nem paraméteres és paraméteres módon is. Egy nem paraméteres becslése a túlélési függvénynek a Kaplan-Meier becslés (aminek módszerét [5] forrásból merítettem). A vizsgálat során legyen n darab egyedünk, melyeket sorbarendezve $t_1^*, t_2^*, ..., t_n^*$ időpontokban ért a halál. Jelölje d_i a halálozások számát t_i^*-kor (vagyis

$\sum d_i = n$) és legyen r_j az élők száma kicsivel t_j^* előtt (azaz $r_{j+1} = r_j - d_j$). Ekkor a becsült eloszlásfüggvényünk

$$\widehat{F}(t) = \frac{1}{n} \sum_{j=1}^{s} d_j,$$

ahol $t_s^* \leq t \leq t_{s+1}^*$. Így kapjuk, hogy a becsült túlélési függvény

$$\widehat{S}(t) = 1 - \widehat{F}(t) = \frac{n - \sum_{j=1}^{s} d_j}{n},$$

ahol szintén fennáll, hogy $t_s^* \leq t \leq t_{s+1}^*$. A kapott túlélési függvény "szebb" alakra hozható, vagyis a következő alakban adható meg:

$$\widehat{S}(t) = \prod_{j=1}^{s} \left(1 - \frac{d_j}{r_j}\right).$$

Adatok cenzorálása esetén változik az élők száma, mégpedig: $r_1 = n - l_1$, $r_{j+1} = r_j - d_j - l_{j+1}$, ahol l_1 a t_1^*-ig cenzorált adatok száma, l_i pedig a t_{i-1}^* és t_i^* közöttiek száma.

Egy másik becslése a túlélési függvénynek egy paraméteres modell segítségével történik. Ezek lehetnek a Gompertz- vagy a Weibull-eloszlások.

2.6.1. Definíció (Gompertz-eloszlás). *Az X valószínűségi változó Gompertz-eloszlású, ha eloszlásfüggvénye*

$$F(x; \beta, \gamma) = \begin{cases} 1 - e^{-\gamma\left(e^{\beta x} - 1\right)}, & ha\ x \geq 0 \\ 0, & ha\ x < 0 \end{cases},$$

ahol $\gamma > 0$*az alakparaméter és* $\beta > 0$ *a skálaparaméter.*

2.6.2. Definíció (Weibull-eloszlás). *Az X valószínűségi változó Weibull-eloszlású, ha eloszlásfüggvénye*

$$F(x; \beta, \gamma) = \begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^{\gamma}}, & ha\ x \geq 0 \\ 0, & ha\ x < 0 \end{cases},$$

ahol $\gamma > 0$*az alakparaméter és* $\beta > 0$ *a skálaparaméter.*

Ezek után tekintsük a következő kétváltozós túlélési függvényt, mint kopulát:

$$S(t_x, t_y; \theta_x, \theta_y, \alpha) = C(S_1(t_x; \theta_x), S_2(t_y; \theta_y); \alpha),$$

ahol θ_x és θ_y a marginálisok paraméterei és α a kapcsolati paraméter.

Jobbra cenzorált esetben a likelihood függvény $\theta = (\theta_x, \theta_y, \alpha)$ esetén a következő lesz:

$$l(\theta) = \prod_{i=1}^{n} f(x_i, y_i; \theta_x, \theta_y)^{c_{x_i} c_{y_i}} f_1(x_i, y_i; \theta_x, \theta_y)^{c_{x_i}(1-c_{y_i})}$$
$$\times f_2(x_i, y_i; \theta_x, \theta_y)^{(1-c_{x_i})c_{y_i}} S(x_i, y_i; \theta_x, \theta_y)^{(1-c_{x_i})(1-c_{y_i})},$$

ahol

$$f(x, y; \theta_x, \theta_y) = \frac{\partial^2}{\partial x \partial y} S(x, y; \theta_x, \theta_y);$$
$$f_1(x, y; \theta_x, \theta_y) = \frac{\partial}{\partial x} S(x, y; \theta_x, \theta_y);$$
$$f_2(x, y; \theta_x, \theta_y) = \frac{\partial}{\partial y} S(x, y; \theta_x, \theta_y);$$

és x_i =belépési kor $+t_{x_i}$, y_i =belépési kor $+t_{y_i}$.

Tegyük fel, hogy a túlélési függvényt a Weibull-eloszlással határoztuk meg, vagyis kapjuk, hogy $S(t_j) = P(T_j > t_j) = e^{-\left(\frac{t}{\beta_j}\right)^{\gamma_j}}$, ha $t \geq 0$, ahol β_j a skála- és γ_j az alakparaméter $j = x, y$-ra. Ezután helyettesítsük be a perem túlélési függvényeket egy adott kopulába. Nézzük mondjuk a Gumbel-Hougaard kopulát. Kapjuk, hogy:

$$S(t_x, t_y) = C_{GH}(S(t_x; \beta_x, \gamma_x), S(t_y; \beta_y, \gamma_y); \alpha) \quad = e^{-\left[\left(-\log e^{-\left(\frac{t_x}{\beta_x}\right)^{\gamma_x}}\right)^{\alpha} + \left(-\log e^{-\left(\frac{t_y}{\beta_y}\right)^{\gamma_y}}\right)^{\alpha}\right]^{\frac{1}{\alpha}}}$$
$$= e^{-\left[\left(\frac{t_x}{\beta_x}\right)^{\gamma_x \alpha} + \left(\frac{t_y}{\beta_y}\right)^{\gamma_y \alpha}\right]^{\frac{1}{\alpha}}}.$$

Ennek a függvénynek a maximum likelihood becslése elég bonyolult, így nem vezetem le, de megtalálható [**?**]-ben.

Ez az előállítás rendkívül elegáns és hatásos, azonban néha kevésnek bizonyul a kapcsolatok típusától függően. Ez azért lehet, mert egy kétváltozós túlélési függvényt használunk ahhoz, hogy egy 3-változós függvényt vizsgáljunk (*first-life, last-survivor*). Tekintsük a következő sűrűségfüggvényt:

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} S(x, y),$$

ami mutatja a változás mértékét $S(x, y)$-ban, amikor a halálok x és y korban bekövetkeznek. Bontsuk fel $S(x, y)$-t a következőképp: $S(x, y) = S(x_0 + t_x, y_0 + t_y)$, ahol x_0 és y_0 a házasságba belépéskor az életkor.

1.eset: $t_x = t_y$

$f(x, y)$ mutatja a változás mértékét $S(x, y)$-ban, amikor a halálok x és y korban bekövetkeznek. Ekkor a halál egyszerre, vagy szinte egymás után azonnal történik.

2.eset: $t_x \neq t_y$

Legyen ekkor például $t_y = t_x + t$, ahol $t \neq 0$. Itt szintén $f(x, y)$ mutatja a változás mértékét $S(x, y)$-ban, amikor a halálok x és y korban bekövetkeznek, de most a két halál között t idő telik el.

Egyik esetben sincs probléma akkor, ha a házasságba belépéskor mindkét fél egyforma idős, azaz $x_0 = y_0$ (ez áll fent mondjuk akkor, ha páros szerveink élettartamát vizsgáljuk). Akkor sincs baj, ha a két élet közötti kapcsolat csak a kortól függ és nincs külön befolyásoló tényező. Azonban házastársak életének vizsgálatánál egyik sem áll fent feltétlenül, ugyanis nem csak azonos korú emberek kötk össze életüket és házasságukra több külső hatás is befolyással lehet (például egy végzetes karambol, ahol mindkét fél elhunyt). Befolyásoló tényező lehet még az összetört szív szindróma, vagyis amikor az egyik fél meghal a másik nem sokkal utána követi őt a halálba. Ebből kifolyólag úgy tűnik, hogy nincs tökéletes becslés kopulákkal az együttes túlélési függvényre, valamint az első-élet és utolsó-túlélő függvények becslésére. (Ezt a problémát bővebben [?] tárgyalja és próbálja megoldani)

3. fejezet

Rövid idejű függőség vizsgálata

Két élettartam között rövid idejű függőségről beszélünk, ha a tovább élő ember hazárdfüggvénye magasabb lesz miután a partnere meghal. Ezt a függést egy összetett modellel lehet vizsgálni, amit ebben a fejezetben fogok bemutatni.

3.1. Markov modell

A rövid idejű függőség ezen modelljét *Norberg* és *Wolthuis* alakították ki, ami a Markov lánc elméletén alapszik.

3.1.1. Definíció (Markov-lánc). *Legyenek $X_1, X_2, ...$ valószínűségi változók, I állapottér, ami véges vagy megszámlálhatóan végtelen. Tegyük fel, hogy X_n $(n = 0, 1, 2, ...)$ értékei I-be esnek. Azt mondjuk, hogy ezen valószínűségi változók* ***Markov-láncot*** *alkotnak, ha rendelkeznek a* ***Markov-tulajdonsággal****, azaz $P(X_{n+1} = x_{n+1}|X_0 = x_0, ..., X_n = x_n) = P(X_{n+1} = x_{n+1}|X_n = x_n)$ (vagyis a jövő nem függ a múlttól, csak a jelentől).*

Jelen esetben a Markov-lánc 4 lépcsőből áll (amit a 3.1. ábra mutat). Kezdetben mind a ketten életben vannak, amit 0. állapotnak nevezünk. A második és a harmadik lépcsőre akkor kerülünk, ha valamelyik fél meghalt, vagyis ha a férj hal meg az első, míg ha a feleség, akkor a második állapotba jutunk. Ezután már csak egy eset lehetséges, ha az életben maradt fél is távozik az élők sorából. Ekkor érünk a lánc aljára. Legyen T_x a férj, T_y a feleség hátralevő élettartama a házasságkötéstől, valamint x és y a házasságba lépési koruk. Ekkor tehát az átmenetvalószínűségek:

$\mu_{01}(t) = \mu_1(x + t|T_y > t)$, vagyis a férfi hazárdfüggvénye $x + t$ időre, ha tudjuk, hogy a nő tovább él, mint t év.

$\mu_{02}(t) = \mu_2(y + t|T_x > t)$, vagyis a nő hazárdfüggvénye $y + t$ időre, ha tudjuk, hogy a férfi tovább él, mint t év.

$\mu_{13}(t) = \mu_2(y + t|T_x \leq t)$, vagyis a nő hazárdfüggvénye $y + t$ időre, ha tudjuk, hogy a férfi t éven belül elhunyt.

$\mu_{23}(t) = \mu_1(x + t|T_y \leq t)$, vagyis a férfi hazárdfüggvénye $x + t$ időre, ha tudjuk, hogy a nő t éven belül elhunyt.

Tegyük fel, hogy ezek a hazárdfüggvények előállnak a férfi és a nő hazárdfüggvényeiből a következő alakban:

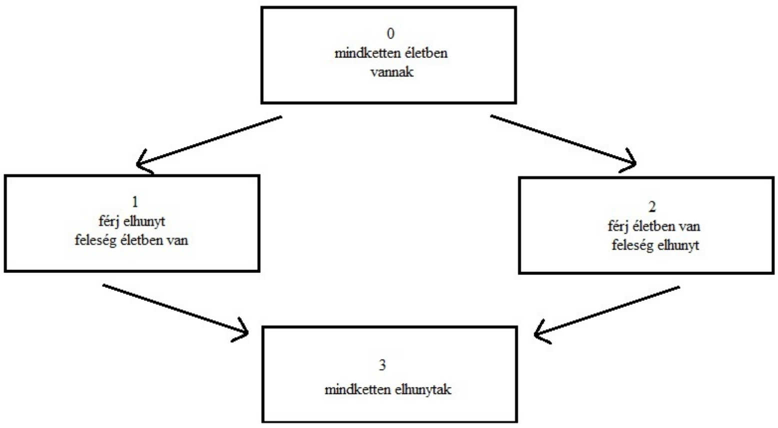
$$\mu_{01}(t) = (1 - \alpha^*_{01})\mu_1(x + t);$$

$$\mu_{02}(t) = (1 - \alpha^*_{02})\mu_2(y + t);$$

$$\mu_{13}(t) = (1 + \alpha^*_{13})\mu_2(y + t);$$

$$\mu_{23}(t) = (1 + \alpha^*_{23})\mu_1(x + t);$$

ahol $\alpha^*_{01}, \alpha^*_{02}, \alpha^*_{13}, \alpha^*_{23} \geq 0$ paraméterek.



3.1. ábra.

Így kaptunk egy 4 lépéses modellt. Azonban ez a modell még nem adja vissza kellőképpen a valóságot, szükség van némi változtatásra, amit a következő fejezetben mutatunk be.

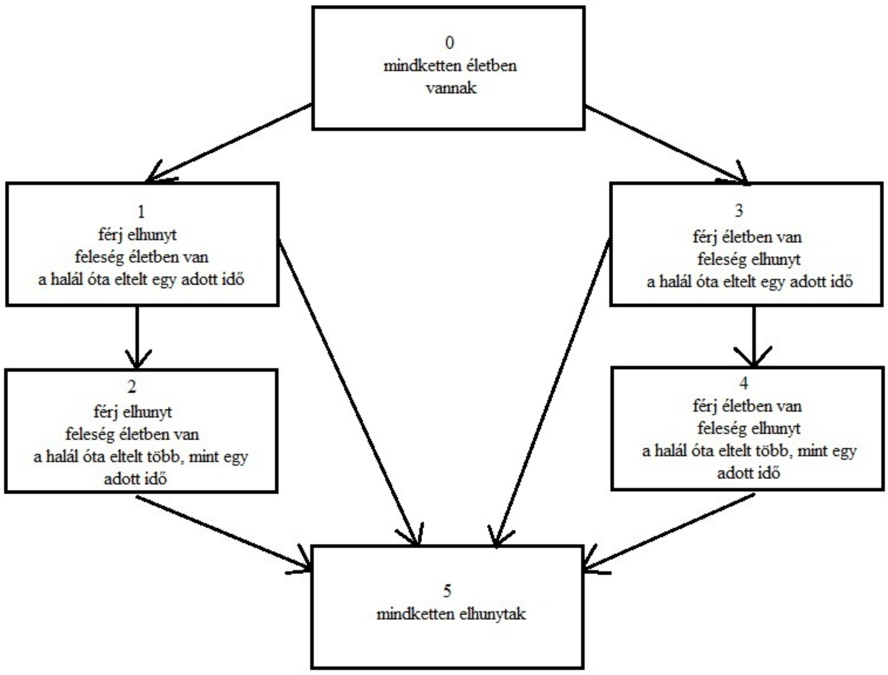
3.2. Kiterjesztett Markov modell

A kiterjesztett Markov modell azon alapszik, hogy a fennmaradó élet hazárdfüggvénye függ a házastárs halála óta eltelt időtől. Így további 2 állomást teszünk be az alapmodellbe és így kapunk egy 6 lépcsős kiterjesztett Markov modellt. Ezt a további két állomást úgy kapjuk, hogy felbontjuk azt az időintervallumot, amikor már az egyik fél elhunyt és a másik életben van.

Először nézzük azt, amikor elsőként a férj távozik az élők sorából. Ekkor a feleség belép az 1 jelzésű állapotba és ott marad mindaddig, amíg a haláltól eltelt idő 0-tól nagyobb, de kisebb, mint egy adott t_y időtartam. Miután letelt a t_y idő és még mindig életben van a nő, akkor átkerül a 2-es jelzésű állapotba, ahol egészen addig marad, amíg meg nem hal.

Másodszor tekintsük azt az esetet, amikor a feleséget éri utol előbb a halál. Ilyenkor a férj kerül a 3-as jelzésű állapotba és egészen addig ott marad, amíg a feleség halála óta eltelt idő 0 és t_x között van. Amikor az idő átlépi t_x-et a férfi átkerül a 4 jelzésű állapotba és mindaddig ottmarad, amíg véget nem ér az élete.

Értelemszerűen az 1 és 3 jelzésű állapotból kerülhetünk a végállapotba, amikor a férj és a feleség sem él tovább, mint a halál után eltelt adott időtartam. Valamint nem tehető fel, hogy $t_x = t_y$, ugyanis ez a törött szív két külön tulajdonsága a férfiaknál és a nőknél.



3.2. ábra.

A kiterjesztett modell így a következő hazárdfüggvényekkel rendelkezik:

$$\mu_{01}(t) = \mu_1(x+t|T_y > t) = (1-\alpha_{01})\mu_1(x+t);$$
$$\mu_{03}(t) = \mu_2(y+t|T_x > t) = (1-\alpha_{03})\mu_2(y+t);$$
$$\mu_{15}(t) = \mu_2(y+t|0 \leq t - T_x < t_y) = (1+\alpha_{15})\mu_2(y+t);$$
$$\mu_{25}(t) = \mu_2(y+t|t - T_x > t_y) = (1+\alpha_{25})\mu_2(y+t);$$
$$\mu_{35}(t) = \mu_1(x+t|0 \leq t - T_y < t_x) = (1+\alpha_{35})\mu_1(x+t);$$
$$\mu_{45}(t) = \mu_1(x+t|t - T_y > t_x) = (1+\alpha_{45})\mu_1(x+t);$$

ahol $\alpha_{01}, \alpha_{03}, \alpha_{15}, \alpha_{25}, \alpha_{35}, \alpha_{45} \geq 0$ adott paraméterek.

Ennek a kiterjesztett modellnek az eredeti speciális esete, ha $\alpha_{15} = \alpha_{25} = \alpha_{13}^*$ és $\alpha_{35} = \alpha_{45} = \alpha_{23}^*$.

3.3. A paraméterek becslése

Tegyük fel, hogy a perem hazárdfüggvényeket a Gompertz-eloszlásból kaptuk, amit már korábban definiáltunk. Ekkor ezek a következő alakúak:

$$\mu_x = \frac{1}{\sigma_f} e^{\frac{x-m_f}{\sigma_f}}; \mu_y = \frac{1}{\sigma_n} e^{\frac{y-m_n}{\sigma_n}},$$

ahol m_f és m_n a férj és a feleség életkorának várható értékét,valamint σ_f és σ_n a férj és a feleség élettartamának szórását jelöli. Ezeket a paramétereket könnyen megadhatjuk a maximum likelihood becsléssel.

Az α paramlétereket szintén maximum likelihood módszerrel becsülhetjük. Legyen a vizsgálatban résztvevő nők száma n, míg a férfiaké m. Jelölje az egyes állapotokba belépés korát b, míg az onnan kilépését k. Ez alapján kapjuk, hogy ha a 0. állapotban vagyunk, akkor az 1.-be és a 3.-ba jutáshoz az α-k a következőek:

$$\widehat{\alpha}_{01} = 1 - \frac{d_{01}}{\sum_{i=1}^{m} \int_{b(i)}^{k(i)} \mu_{x_i} dx_i}; \widehat{\alpha}_{03} = 1 - \frac{d_{03}}{\sum_{i=1}^{n} \int_{b(i)}^{k(i)} \mu_{y_i} dy_i},$$

ahol d_{01} (d_{03}) jelöli a 0. lépcsőn a megfigyelt férfiak (nők) halálának számát. A megfigyelés azonnal véget ér, ha bekövetkezik az egyik fél halála, vagy cenzorálják az adatokat. A keletkezett hibát a következő formulával kaphatjuk meg: $o(\widehat{\alpha}_{0j}) = \frac{1-\widehat{\alpha}_{0j}}{\sqrt{d_{0j}}}$, $j = 1, 3$.

Nézzük azokat az eseteket, amikor az $1, 2, 3, 4$ állapotban vagyunk. Itt jelölje a j. állapotban éppen tartozkodó nők számát n_j, a férfiakét pedig m_j, ahol $j = 1, 2, 3, 4$. Ekkor a paraméterekre a következő becsléseket mondhatjuk:

$$\widehat{\alpha}_{l5} = \frac{d_{l5}}{\sum\limits_{i=1}^{n_j} \int\limits_{b(i)}^{k(i)} \mu_{y_i} dy_i} - 1; \widehat{\alpha}_{k5} = \frac{d_{k5}}{\sum\limits_{i=1}^{m_j} \int\limits_{b(i)}^{k(i)} \mu_{x_i} dx_i} - 1,$$

ahol $l = 1, 2$ és $k = 3, 4$, valamnt d_{l5} (d_{k5}) jelöli az l. (k.) lépcsőben a megfigyelt nők (férfiak) halálának száma. A keletkezett hibát itt is könnyedén meg tudjuk adni: $o(\widehat{\alpha}_{j5}) = \frac{1-\widehat{\alpha}_{j5}}{\sqrt{d_{j5}}}$, $j = 1, 2, 3, 4$.

A becslésnél látszik, hogy nagyon érzékeny a kor megválasztásától, hiszen az integrálás a belépési kortól a kilépési korig tart. A becslésben általában a belépési korokat a következő intervallumokból vesszük: férfiak esetén $[65, 85]$, míg nőknél $[60, 80]$, ugyanis ezek a leggyakoribb korok, amikor valaki megözvegyül. A meghatározásunk még alapszik a korok közötti függésről, vagyis az eredmények nagyban függnek a kapcsolattól.

3.4. A modell tesztelése

A következőkben teszteljük, hogy a túlélő házastárs hazárdfüggvénye szignifikánsan függ-e a párja halála óta eltelt időtől. Ezáltal megfogalmazhatunk egy nullhipotézist és egy alternatív (ellen) hipotézist:

$$H_0 : \alpha_{15} = \alpha_{25}$$

$$H_1 : \alpha_{15} \neq \alpha_{25}$$

Ha elfogadjuk H_0-t, akkor a hazárdfüggvény az 1. és a 2. szinten megegyezik, vagyis a túlélési függvényük ugyanazok, vagy kevés az adatunk ezt cáfolni. Amennyiben elvetjük H_0-t, akkor arra kapunk statisztikai bizonyítékot, hogy a hazárdfüggvény az 1. és a 2. szinten különbözik, vagyis a túlélési függvényük is eltér. Ezáltal át tudjuk fogalmazni a null- és az ellen hipotézisünket a következőképpen:

$$H_0 : S_{(y)15}(t) = S_{(y)25}(t)$$

$$H_1 : S_{(y)15}(t) \neq S_{(y)25}(t),$$

ahol $S_{(y)i5}(t)$ a túlélő házastárs túlélési függvénye y korban az i. állapotban ($i \in \{1, 2\}$).

Ugyanígy fogalmazható meg, a másik oldalra, vagyis amikor a feleség hunyt el korábban.

Ezek után alkalmazzuk a kétoldali, kétmintás Kolmogorov-Smirnov próbát, ami a következő alakú:

$$D_{mn} = \sup_t |S_{(y)m}(t) - S_{(y)n}(t)|,$$

ahol $S_{(y)m}(t)$ a megfelelő Kaplan-Meier függvénye $S_{(y)15}(t)$-nek és m a halálok száma, hasonlóan $S_{(y)n}(t)$. Így látszik, hogy a Kolmogorov-Smirnov teszt csak folytonos valószínűségi változóknál használható és az eloszlások közötti eltérésre kérdez rá.

A nullhipotézist elvetjük, ha

$$\left(\frac{mn}{m+n}\right)^{\frac{1}{2}} D_{mn} \geq c$$

, ahol c a kritikus értéket a Kolmogorov eloszlásból nyertük az előre meghatározott terjedelem mellett.

Irodalomjegyzék

[1] Móri Tamás: *Élettartamadatok elemzése*, Typotex Kft, Budapest (2011)

[2] Shemyakin A. and Youn, H.: *Copula Models of joint survivor analysis*, Applied Stohastic Models in Business and Industry, (2006)

[3] Spreeuw, J.: *Types of dependence and time-dependent association between two lifetimes in single parameter copula models*, Scandinavian Actuarial Journal, 286-309. (2006)

[4] Jaap Spreeuw and Xu Wang: *Modelling the short-term dependence between two remaining lifetimes* (March 27, 2008)

[5] Rob Allis, Amgen Ltd., Uxbridge, UK: *Comparing Kaplan-Meier curves - what are the (SAS) options?* PhUSE 2009, Paper SP02

[6] Ludger Rüschendorf: *On the distributional transform, Sklar's Theorem, and the empirical copula process*, Journal of Statistical Planning and Inference, 39213927. (November 1, 2009)