Eötvös Loránd Tudományegyetem

Budapesti Corvinus Egyetem

Bende Botond

Cenzorált Élettartamok Statisztikai Vizsgálata

MSc Diplomamunka

Témavezető:

Dr. Kováts Antal

Valószínűségelméleti és Statisztika tanszék

Budapest, 2016

Tartalomjegyzék

Köszönetnyilvánítás

Szeretném megköszönni témavezetőmnek, Kováts Antalnak, hogy elvállalta a konzulensi teendőket. Köszönöm, hogy mindig rendelkezésemre állt és szakmai tanácsaival hozzájárult a szakdolgozatom elkészüléséhez.

Valamint köszönöm családomnak és barátaimnak, hogy mindig mellettem álltak és segítették a munkámat.

Bevezetés

A statisztikai szakirodalom széles körben vizsgálta a cenzorált mintákra vonatkozó becsléselméleti és hipotézisvizsgálati kérdéseket. Ezek között kiemelkedő jelentősége van az exponenciális eloszlásra vonatkozó elemzéseknek. Jelen dolgozatban egy a biztosítási gyakorlatban felmerülő, káridőpontokra vonatkozó problémából kiindulva az exponenciális eloszlásnak a nullában elfajult eloszlással való keverékére végzek vizsgálatokat felülről cenzorált mintákat feltételezve. A feladat nehézsége abban rejlik, hogy a mintában nem tudjuk megkülönböztetni a 0 értékű és a cenzorálási időnél nagyobb értékű elemeket.

A dolgozat első fejezetében olyan módszereket, és eljárásokat mutatunk be, amelyeket a modell építése során felhasználunk. A második fejezetben a káridőpontokra írunk fel modellt. Fontos probléma, hogy egy kárügy biztosító általi lezárása után is jelentkezhetnek az adott káresetre vonatkozó újabb kárigények. Ilyenkor a szakzsargon szerint a kárügyet „újranyitják”. Az aktuáriusok számára ez sok esetben gondot jelent, mert a kár lezárásakor a tételes kártartalék „kinullázódott”, és ha az újranyitás már a következő értékelési időszakban történik, akkor ez lebonyolítási veszteséget eredményez. Fontos tehát, hogy a korábbi tapasztalatok alapján meghatározzuk az újranyitás arányát és időpontjának eloszlását. A harmadik fejezetben a újranyitás arányára és az intenzitásra végzünk statisztikai becsléseket exponenciális eloszlást feltételezve. A negyedik fejezetben hipotézisvizsgálatot végzünk különböző öregedő osztályok ellen, ahol a null-hipotézis az exponenciális eloszlás.

Jelölések

$X_1, X_2, ..., X_n$	–	n elemű minta
$X_1^*, X_2^*, ..., X_n^*$	–	n elemű rendezett minta
F	–	Az elméleti eloszlásfüggvény
$\overline{F}$	–	Az elméleti túlélésfüggvény
F_n	–	Az n elemű minta empirikus eloszlásfüggvénye
$r(t)$	–	Hazárd ráta
$R(t)$	–	Kumulatív hazárd függvény
$\mathbb{I}$	–	Indikátorfüggvény
$\Gamma(\cdot)$	–	Gamma-függvény
$R(\cdot)$	–	Statisztika teljesítménye
$ARE(\cdot, \cdot)$	–	Relatív hatásfok
Φ	–	Próbafüggvény

1. fejezet

Alapozó ismeretek

Ebben a részben azon ismereteket tekintjük át, amelyek szükségesek a továbbiak megértéséhez, illetve olyan állításokat és tulajdonságokat fogalmazunk meg, amelyeket a modell építése közben felhasználunk. Először az öregedő eloszlások néhány osztályát definiáljuk felhasználva az [1]-es a [2]-es könyveket. A második részben a cenzorálást mutatjuk be. A harmadik részben olyan statisztikákat ismertetünk, amelyeknek létezik határeloszlása. Végül bevezetjük a Pitman-féle relatív hatásfok fogalmát, amely két statisztika összehasonlítására, illetve a tesztek teljesítményének a mérésére szolgál.

1.1. Öregedő eloszlások

Öregedés alatt olyan folyamatot értünk, amikor egy alkatrész vagy rendszer hátralévő élettartama rövidebb, mint egy újnak. Az öregedő osztályt definiálhatjuk a hazárd-függvénnyel, a feltételes eloszlásfüggvénnyel, vagy az átlagos hátralévő élettartammal. Ez a három fogalom információt nyújt a rendszer vagy az alkatrész élettartamáról. A következőkben bevezetjük az **IFR** és az **NBU** öregedő osztályokat.

1.IFR osztály (Increasing Failure Rate)

Az IFR osztályba tartozik az az eloszlás, amelynek hazárd rátája monoton, nem csökkenő, azaz $0 < t_1 < t_2$ esetén

$$r(t_1) \leq r(t_2).$$

Ez azt jelenti, hogy az idő múlásával az elem megbízhatósága csak csökkenhet. Vezessük be a

$$R(t) = \int_0^t r(s)ds$$

jelölést. Ekkor a hazárd rátára tett feltételezések miatt észrevehető, hogy az R függvény alulról konvex.
Ha $\overline{F}$ függvény egy IFR eloszlás túlélésfüggvénye, akkor

$$\frac{\overline{F}(x+y_1)}{\overline{F}(y_1)} < \frac{\overline{F}(x+y_2)}{\overline{F}(y_2)}$$

bármilyen $x \geq 0, y_1 > y_2$.

Szemléletesen ez azt jelenti, hogy az alkatrész időtartama sztochasztikusan csökken az idővel.
Ha feltesszük, hogy az alkatrész az x időpillanatban működött, akkor annak a valószínűsége, hogy további y időt túléli, x-ben monoton fogyó függvény. A következőkben bemutatunk néhány IFR eloszlást.

1.Gamma-eloszlás

Egy valószínűségi változó λ rendű, α paraméterű **Gamma-eloszlást** követ, ha sűrűségfüggvénye

$$f(t) = \frac{\lambda(\lambda t)^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda t},$$

ahol λ és α pozitív számok, és $\Gamma(\alpha)$ a Gamma-függvény az α helyen.
Speciálisan, ha $\alpha = 1$, akkor λ paraméterű exponenciális eloszlást, illetve ha $\alpha = \frac{n}{2}$ és $\lambda = \frac{1}{2}$, akkor n szabadságfokú χ^2 eloszlást kapunk. Gamma-eloszlású független valószínűségi változók összege és exponenciális eloszlású független valószínűségi változók összege Gamma-eloszlás lesz.

A **Gamma-eloszlás** hazárd függvénye $\alpha \leq 1$-re monoton csökkenő függvény, ezért csak $\alpha \geq 1$ esetén lesz az eloszlás **IFR**-beli.

2.Weibull-eloszlás

Egy valószínűségi változó **Weibull-eloszlást** követ, ha sűrűségfüggvénye

$$f(t) = \alpha\lambda(\lambda t)^{\alpha-1}e^{-(\lambda t)^{\alpha}},$$

ahol $\alpha > 0$ az alak, és $\lambda > 0$ a skálaparaméter. Speciálisan ha $\alpha = 1$, akkor a valószínűségi változó λ paraméterű exponenciális eloszlású lesz.
A Weibull-eloszlás érdekes tulajdonsága, hogy független, azonos eloszlású valószínűségi változók minimumának a határeloszlása.
Ha a t jelzi a meghibásodásig eltelt időt, akkor a Weibull-eloszlás az idővel arányos meghibásodási gyakoriságot jelzi. A α alakparaméter értelmezése a következő:

- Ha az $\alpha < 1$, akkor a a meghibásodási gyakoriság idővel csökken.
- Ha az $\alpha = 1$, akkor a meghibásodási gyakoriság időben állandó.
- Ha az $\alpha > 1$, akkor a meghibásodási gyakoriság időben növekszik.

Ezek alapján elmondható, hogy a **Weibull-eloszlás** csak akkor lesz **IFR**, ha az $\alpha \geq 1$.

2. NBU osztály (New Is Better Than Used)
"Jobb az új a használtnál" vagyis, az új alkatrész életkilátásai jobbak, mint egy olyané, amelyik már élt valamennyit.
Tehát azt mondhatjuk, hogy egy eloszlás NBU, ha

$$P(T > t + s|T > t) \leq P(T > s).$$

Ezt megfogalmazhatjuk úgy is, hogy $F \in NBU$, ha

$$\overline{F}(t + s) \leq \overline{F}(t)\overline{F}(s)$$

ahol $t, s > 0$.

Szemléletesen ez azt jelenti, hogy az alkatrész időtartama sztochasztikusan nő az idővel. Tehát az új alkatrész időtartama sztochasztikusan nagyobb, mint egy olyan alkatrész további időtartama, amely már túlélte az t időtartamot.

Ha $R(t)$ jelöli t időpillanatig a kumulált meghibásodási rátát, akkor

$$R(t+s) > R(t) + R(s).$$

Egy **NBU** eloszlásnak mindig létezik véges momentuma, és az F relatív szórása nem nagyobb 1-nél.
A következőkben ismertetek egy eloszlást, amely az **NBU** osztályban van.
1.Geometriai-eloszlás diszkrét időben
Egy valószínűségi változó p-paraméterű Geometriai-eloszlást követ, ha túlélésfüggvénye

$$\overline{F}(t) = (1-p)^{[t]-1}.$$

Megmutatjuk, hogy ez az eloszlás NBU-beli. A túlélésfüggvény tulajdonságai miatt elmondható, hogy

$$\overline{F}(t+s) \leq \overline{F}([t]+[s]-1) \leq \overline{F}([t])\overline{F}([s]) = \overline{F}(t)\overline{F}(s).$$

Tehát teljesül az **NBU** tulajdonság, és észrevehető, hogy R szuperadditív lépcsős függvény. Mivel a R nem folytonos, ezért ez az eloszlás nem **IFR**. Ugyanakkor a két osztály között a következő reláció teljesül:

$$IFR \subset NBU.$$

1.2. Cenzorálás

A cenzorálás a hiányzó adatok problémája, amely egy általános jelenség a megbízhatóság-elméletben. Az élettartam-adatok esetén nem hagyhatjuk figyelmen kívül a cenzorálást, különben torzulna az eloszlásról alkotott kép. Cenzorálás fordul elő abban az esetben, ha egy alkatrész élettartamának a

vizsgálata során elveszítjük a követést, és csak azt tudjuk, hogy a t-edik időpillanatban még élt az alkatrész.
Kétféle cenzorálásról beszélünk:

- **felülről való cenzorálás:** Amikor a megfigyelt esemény egy adott időpontnál később következik be.
- **alulról való cenzorálás:** Amikor nem tudunk az eseményekről semmit egy időpont előtt.

A kétféle cenzorálás közül, mi most a felülről levágással foglalkozunk, mert az jön elő a modellünkben.

Legyenek $X_1, .., X_n$ a megfigyelt egyedek élettartamai, és c a cenzorálási időpont. Ekkor a cenzorált minta:

$$(X_1, \delta_1), (X_2, \delta_2), \ldots, (X_n, \delta_n),$$

ahol δ_i jelöli, hogy az i. esemény a cenzorálás időpontjáig bekövetkezett-e, tehát $\delta_i = 1$, ha az esemény a cenzorálásig bekövetkezett, illetve 0, ha nem.
A cenzorált minta empirikus eloszlásfüggvényére Kaplan-Meyer adott becslést. Bevezetjük az s_i és a t_i státusz és időváltozókat, amiket a következőképpen definiálunk:
s_i - Adott időpillanatban bekövetkezett-e a vizsgált esemény (0-nem, 1-igen);
t_i - A vizsgált élettartam cenzorálás idpontjáig ismert hossza (s_i = 0-ra $t_i = 0$);
Jelölje a $0 < \tau_1 < \tau_2 < ... < \tau_l < maxX_i$ az esemény bekövetkezésének időpontjait, és

$$\begin{aligned} d_k &= \#\{1 \leq i \leq l : t_i = \tau_k, s_i = 1\} \\ m_k &= \#\{1 \leq i \leq l : t_i \geq \tau_k\}. \end{aligned}$$

Ekkor:

$$F_n(t) = \prod_{j:\tau_j<t} \left(1 - \frac{d_j}{m_j}\right) = \prod_{j:\tau_j<t} \frac{m_j - d_j}{m_j}, 0 \leq t \leq maxX_i.$$

A Kaplan-Meyer becslés konzisztens, vagyis

$$\sup_t |F_n(t) - F(t)| \longrightarrow 0$$

sztochasztikusan. Cenzorálás esetén a hazárd ráta becslését lépcsős függvényekkel adjuk meg, egy-egy adott intervallumba eső megfigyelések, és meghibásodások számára adunk ML-becslést. Ennélfogva a túlélésfüggvény becslése

$$F(\tau_j) = \prod_{i=1}^{j} \left(1 - \frac{d_i}{m_{i-1} - \frac{q_i}{2}}\right)$$

ahol q_i jelöli a $[\tau_i, \tau_{i+1})$ intervallumba eső cenzorált megfigyelések számát. Ezt nevezzük aktuáriusi becslésnek.

Ha megfigyeljük az eloszlásfüggvényt, azt vehetjük észre, hogy olyan, mintha minden meghibásodás az intervallum közepén lenne, a cenzorálásoknak pedig fele előtte, fele utána.

1.3. Statisztikák határeloszlása

Ebben a részben bevezetjük az U-statisztika fogalmát, és olyan statisztikákat tárgyalunk, amelyeknek létezik határeloszlása a [3], [4] és [5] cikkek alapján.

Legyen $X_1, ..., X_n$ független, azonos eloszlású minta F eloszlásfüggvénnyel. Legyen a $\Phi(x_1, ..., x_m)$ egy szimmetrikus függvény az $x_1, ..., x_m$ változókban, és

$$\theta(F) = \int_{\mathbb{R}} ... \int_{\mathbb{R}} \Phi(x_1, ..., x_m) \prod_{i=1}^{m} F(dx_i)$$

feltéve, hogy a $\theta(\cdot)$ integrál létezik.

Ekkor a [4] cikk alapján elmondható, hogy a következőkben bevezetett statisztika torzítatlan becslés a $\theta(F)$-re:

$$U_n = \frac{1}{\binom{n}{m}} \sum \Phi(X_{i_1}, ..., X_{i_m})$$

ahol az összegben szereplő tagok száma $\binom{n}{m}$. Az ilyen alakú statisztikát U-statisztikának nevezzük.

A [5]-ös cikkben leírtak alapján vezessük be a következő jelöléseket:

$$\begin{aligned}
\Phi_c(x_1, ..., x_c) &= \mathbb{E}(x_1, ..., x_c, X_{c+1}, ..., X_m), c = 1, ..., m, \\
\Psi_c(t_1, ..., t_c) &= \Phi_c(t_1, ..., t_c) - \theta, c = 1, ..., m
\end{aligned}$$

feltéve, hogy a kifejezések léteznek.

Definiáljuk a következő függvényt:

$$\begin{aligned}\zeta_0 &= 0,\\ \zeta_c &= Var(\Psi_c(X_1,...,X_c)), c = 1,...,m.\end{aligned}$$

Legyen F_n az $X_1,...,X_n$ minta empirikus eloszlásfüggvénye, valamint

$$\theta(F_n) = \frac{1}{n^m}\sum_{i_1=1}^{n}...\sum_{i_m=1}^{n}\Phi(X_{i_1},...,X_{i_m})$$

egy statisztika. Ekkor

$$\begin{aligned}\theta(F_n) &= \tfrac{1}{n^m}\sum_{i_1=1}^{n}...\sum_{i_m=1}^{n}\Phi(X_{i_1},...,X_{i_m})\\ &= \tfrac{1}{n^m}\sum_{j=1}^{m}\sum_{1\le i_1\le...\le i_m} j!\Phi_{(j)}(x_{i_1},...,x_{i_j})\\ &= \tfrac{1}{n^m}\sum_{j=1}^{m} j!\tbinom{n}{j}U_n^{(j)},\end{aligned}$$

ahol a $\Phi_{(j)}(x_1,...x_j) = \frac{1}{j!}\sum\Phi(x_{i_1},...,x_{i_n})$, és $U_n^{(j)}$ a $\Phi_{(j)}$-hez tartozó U-statisztika. Megmutatható, hogy

$$\sigma^2(U_n) = \frac{1}{\binom{n}{m}}\sum_{c=1}^{m}\binom{m}{c}\binom{n-m}{m-c}\zeta_c.$$

A következőkben tegyük fel hogy n darab $U_n^{(1)}$, $U_n^{(2)},...,U_n^{(n)}$ statisztikánk van, és jelölje

$$\zeta_c^{(i,j)} = cov(\Phi_c^{(i)}(x_1,...,x_c),\Phi_c^{(j)}(x_1,...,x_c)).$$

Ekkor

$$cov(U_n^{(i)},U_n^{(j)}) = \frac{1}{\binom{n}{m(i)}}\sum_{c=1}^{m(i)}\binom{m(j)}{c}\binom{n-m(j)}{m(i)-c}\zeta_c^{(i,j)},$$

ahol $m(i)$ függvény a Φ_i függvény argumentumainak a számát jelöli. Ezek után elmondható, hogy ha van $X_1,...,X_n$ független, azonos eloszlású mintánk, $\Phi^{(j)}(x_1,...,x_{m(j)}), j = 1,...,k$ szimmetrikus függvények, és $U^{(j)}$ jelöli a megfelelő U-statisztikát, akkor

$$(\sqrt{n}(U^{(1)}-\theta^{(1)}),...,\sqrt{n}(U^{(k)}-\theta^{(k)}))\to N(0,\sum),$$

ahol

$$\sum = (m(i)m(j)\zeta^{(i,j)})_{i,j=1}^{k}.$$

A $\theta^{(j)}$ az $U_n^{(j)}$ statisztikához tartozó integrálokat jelöli.

1.4. Próbák összehasonlítása

Ebben a részben olyan fogalmakat ismertetünk, amelyek segítségével össze tudjuk hasonlítani a próbákat különböző ellenhipotézisek mellett. A próbák összehasonlítására úgynevezett erőfüggvényeket használunk. Vizsgáljuk a következő hipotézist:

$H_0 : \upsilon = \upsilon_0$

$H_1 : \upsilon > \upsilon_0$

A próbákat úgy tudjuk összehasonlítani, hogy veszünk egy υ_n sorozatot, ami υ_0-hoz tart, és megnézzük, hogy közel azonos erő eléréséhez azonos υ_n ellenhipotézis mellett a szükséges mintaelem hányadosa mekkora. Ez a szám annál nagyobb, minél jobb az első próba a másodikhoz képest. A határértéket aszimptotikus relatív hatásfoknak nevezzük. Dolgozatomban a Pitman-féle relatív hatásfokkal dolgoztam, ezért a következőkben azt ismertetem a [6]-os cikk alapján.

Legyen U_{1n}, U_{2n} két statisztika. Jelölje $\psi_{in}(\upsilon)$ és $\sigma_{in}^2(\upsilon)$ az i. statisztika várható értékét és szórását.

A következő hányadost

$$R_{in}(\upsilon) = \frac{\psi_{in}^{(m)}(\upsilon)}{\sigma_{in}(\upsilon)}$$

az i-edik statisztika teljesítményének nevezzük, ahol a számláló a statisztika várható értékének m. deriváltját jelöli. Jelölje $ARE(U_{1n}, U_{2n})$ a két statisztika közti relatív hatásfokot. Ekkor:

$$ARE(U_{1n}, U_{2n}) = \lim_{n\to\infty} \frac{R_{2n}^2(\upsilon_0)}{R_{1n}^2(\upsilon_0)}.$$

2. fejezet

Károk újranyitási időpontjának modellezése

2.1. Károk újranyitása

A biztosítási gyakorlatban felmerül az a probléma, hogy egy lezárt kárra az ügyfelek újabb kárigényt jelentenek be. Ekkor a törvény szerint a kárügyet újranyitják. Ez gondot jelent az aktuáriusok számára, ugyanis a kár lezárásakor a tételes kártartalék kinullázódott, és ha az újranyitás a következő értékelési időszakban történik, akkor ez veszteséget okoz a biztosítónak a lebonyolításban. Egy állomány esetén fontos, hogy a múltbeli kártapasztalatok alapján meghatározzuk a lezárt károk újranyitási arányát, és az újranyitási időpont eloszlását.

2.2. Exponenciális eloszlás keveréke nullában elfajult eloszlással

Feltételezzük, hogy vannak lezárt illetve nyitott károk. Ha egy lezárt kárra lezárás után újabb kárigények jelentkeznek, akkor a kárt újranyitják. Lesznek olyan károk, amiket sose fognak újranyitni, és lesznek olyanok, amelyeket egy bizonyos idő elteltével újranyitnak. Legyen $X_1, ..., X_n$ a megfigyelt mintánk, ahol a mintaelemek jelentése a következő:

- $X_i = 0$, ha a kár lezárás után nem volt újranyitva;
- $X_i = t$, ha t idő elteltével a kárt újranyitották.

Azokat a károkat vesszük, amelyek c éven belül bekövetkeztek. Azt feltételezzük, hogy a károkat p valószínűséggel nem nyitják újra, és $1-p$ valószínűséggel pedig λ paraméterű exponenciális eloszlásnak megfelelően nyitják újra. Feltehető, hogy lesznek olyan károk, amelyek újranyitása c időpont után következik be, viszont a c-ben való cenzorálás miatt ezt figyelmen kívül hagyjuk, és nullának tekintjük. Tehát a mintában nem tudjuk megkülönböztetni a nulla értékű, és a cenzorálásnál nagyobb értékű elemeket, ezért a minta eloszlását egy exponenciális eloszlás és nullában elfajult eloszlás keverékével tudjuk leírni. Ha a c évig megfigyelt károkat vesszük, akkor $qe^{-\lambda c}$ lesz azok aránya, amiket nem látunk a c-ben való levágás miatt. A minta sűrűségfüggvényének a felírásához egy mértékelméleti lemmát használtunk fel, amit a következőkben ismertetünk:

2.2.1. Lemma. *Legyen egy (χ, B) teljesen szeparálható metrikus téren adva két egymásra szinguláris, véges mérték, ν_1 és ν_2. Tegyük fel, hogy $\mu_1 << \nu_1$ és $\mu_2 << \nu_2$. Ekkor a $\mu_1 + \mu_2 << \nu_1 + \nu_2$, és*

$$\frac{d(\mu_1 + \mu_2)}{d(\nu_1 + \nu_2)} = \begin{cases} \frac{d\mu_1}{d\nu_1} & a\ B_1\text{-en} \\ \frac{d\mu_2}{d\nu_2} & a\ B_2\text{-en} \end{cases}$$

ahol $B = B_1 \cup B_2$.

Azzal a feltételezéssel élünk, hogy a meghibásodások függetlenek a cenzorálási időtől, a mintaelemek eloszlása pozitív súlyt helyez a nullára, azon kívül pedig abszolút folytonos. Ezért mondhatjuk azt, hogy a domináló mérték a Lebesgue-mérték, plusz a 0-ra koncentrált elfajult eloszlás. A lemmát felhasználva a mintaelemek sűrűségfüggvénye a következőképpen írható fel:

$$f(t) = \begin{cases} q\lambda e^{-\lambda t} & \text{ha } t > 0 \\ p + qe^{-\lambda c} & \text{ha } t = 0 \end{cases}$$

A túlélésfüggvény:

$$P(X > t) = \begin{cases} q(e^{-\lambda t} - e^{-\lambda c}) & \text{ha } 0 < t < c \\ 0 & \text{ha } t > c \end{cases}$$

Az egyszerűség kedvéért vezessük be a következő jelölést:

$$\theta = p + qe^{-\lambda c}.$$

Megfigyelhető, hogy a θ paraméternek szemléletes jelentése van: A p annak a valószínűsége, hogy nem nyitják újra a kárt c-ig, a $qe^{-\lambda c}$ azok aránya, amiket nem látunk a c miatt.

Ennek következtében a mintaelemek sűrűségfüggvénye a következőképpen alakul:

- a 0 valószínűsége θ,
- $1 - \theta$ valószínűséggel c-ben levágott λ exponenciális eloszlás,

vagyis

$$f(t) = \begin{cases} (1-\theta)\frac{\lambda}{1-e^{-\lambda c}}e^{-\lambda t} & \text{ha } t > 0 \\ \theta & \text{ha } t = 0 \end{cases}$$

Látható, hogy a sűrűségfüggvény a λ és θ paraméterektől függ.

3. fejezet

Paraméterbecslés

A 2. fejezetben ismertetett modell paramétereire végzünk statisztikai vizsgálatokat. Az első részben meghatározzuk a paraméterek pontbecslését a likelihood függvény maximalizálásából, majd a 2. részben kiszámítjuk a paraméterek Fisher-féle információját, és harmadik részben a (θ, λ) paraméterpárra konfidencia ellipszoidot írunk fel.

3.1. Pontbecslés

Vezessük be a következő jelöléseket: Legyen

- $T = \sum_{i=1}^{n} I(X_i = 0)$ - azon károk száma, amelyeket c évig nem nyitottak újra;
- $S = \sum_{i=1}^{n} X_i$-az "összműködési idő" c évig;

Mivel a meghibásodások száma indikátor eloszlások összege, ezért nagy mintaelemszám esetén elmondhatjuk, hogy a T binomiális eloszlású n és θ paraméterekkel. Ezen jelöléseket használva a következőkben a (θ, λ) párosra adunk becsléseket. A likelihood-függvény a következőképpen alakul:

$$L(\theta,\lambda) = \theta^T(1-\theta)^{n-T}\left(\frac{\lambda}{1-e^{-\lambda c}}\right)^{n-T} e^{-\lambda S},$$

ahonnan a log-likelihood függvény:

$$l(\theta,\lambda) = Tln(\theta) + (n-T)ln(1-\theta) + (n-T)ln\left(\frac{\lambda}{1-e^{-\lambda c}}\right) - \lambda S. \qquad (3.1)$$

A log-likelihood függvényt θ majd λ paraméterek szerint deriválva a következő egyenletrendszerhez jutunk:

$$\begin{cases} \frac{\partial l(\theta,\lambda)}{\partial \theta} = \frac{T}{\theta} + \frac{n-T}{1-\theta} \\ \frac{\partial l(\theta,\lambda)}{\partial \lambda} = \frac{1}{\lambda} - \frac{c}{e^{\lambda c}-1} - \frac{S}{n-T} \end{cases}$$

Látható, hogy a (T,S) elégséges statisztika, és a maximum likelihood becslés alapján:

$$\hat{\theta} = \frac{T}{n}.$$

A log-likelihood függvény λ szerinti deriváltjából a következő egyenlethez jutunk:

$$\frac{1}{\lambda} - \frac{c}{e^{\lambda c}-1} = \frac{S}{n-T}.$$

Az egyenlet bal oldala egy nem csökkenő függvény a λ-ban, a jobboldal pedig konstans, tehát létezik egyértelmű megoldás, és ennek az egyenletnek a megoldása legyen $\hat{\lambda}$.

A $\hat{\lambda}$ és $\hat{\theta}$ becsült értékének felhasználásával meg tudjuk határozni az illesztéshez használandó eloszlásfüggvényt.

3.2. Fisher-féle Információ

Észrevehető, hogy a (3.1)-es függvényt a következőképpen lehet felírni:

$$l(\theta,\lambda) = l_1(\theta) + l_2(\lambda),$$

ahol

$$
\begin{aligned}
l_1(\theta) &= Tln(\theta) - (n-T)ln(1-\theta) \\
l_2(\lambda) &= (n-T)ln\left(\tfrac{\lambda}{1-e^{-\lambda c}}\right) - \lambda S.
\end{aligned}
$$

Tehát

$$\partial^2 l(\theta,\lambda) = \partial^2 l_1(\theta) + \partial^2 l_2(\lambda),$$

ami azt jelenti, hogy a Fisher-féle információk összeadódnak.

Felhasználva, hogy a T binomiális eloszlású n és θ paraméterekkel, a θ Fisher-féle információja:

$$
\begin{aligned}
I_1(\theta) &= -\mathbb{E}\left(\tfrac{\partial^2}{\partial\theta^2} l_1(\theta)\right) \\
&= -\mathbb{E}\left(-\tfrac{T}{\theta^2} - \tfrac{n-T}{(1-\theta)^2}\right) \\
&= \tfrac{n}{\theta(1-\theta)}.
\end{aligned}
$$

Hasonlóan kimutatható, hogy

$$I_2(\lambda) = n(1-\theta)\left(\tfrac{1}{\lambda} - \tfrac{c^2 e^{\lambda c}}{(e^{\lambda c}-1)^2}\right).$$

3.3. Konfidencia Ellipszoid

Ebben a részben konfidencia ellipszoidot írunk fel a (λ, θ) paraméterpárra. Ha a jelöli a fél nagytengely hosszát, illetve a b állandó a fél kistengelyét, akkor az ellipszis általános egyenlete a következőképpen írható fel:

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1.$$

Ha az ismeretleneket valószínűségi változóknak tekintjük, ahol a valószínűségi változók egy-egy paramétertől függnek, és ismerjük a változók négyzetösszegének az eloszlását, akkor egy α konfidenciaszint mellett konfidencia ellipszoidot tudunk felírni az ismeretlen paraméterekre.

Ismert, hogy

$$
\begin{cases}
\theta - \hat{\theta} \sim N(0, I_{1n}^{-1}(\hat{\theta})) \\
\lambda - \hat{\lambda} \sim N(0, I_{2n}^{-1}(\hat{\lambda}))
\end{cases}
$$

Vezessük be a következő jelöléseket:

$$X = \sqrt{I_{1n}(\hat{\theta})}(\theta - \hat{\theta}),$$

$$Y = \sqrt{I_{2n}(\hat{\lambda})}(\lambda - \hat{\lambda}).$$

Ekkor az X és Y valószínűségi változók független, sztenderd normális eloszlásúak. Ismert, hogy két sztenderd normális eloszlású valószínűségi változó négyzetösszege χ^2 eloszlást követ 2 szabadságfokkal. A konfidencia ellipszishez meg kell határozni azt az x-et, amelyre teljesül a következő egyenlőség:

$$\mathbb{P}(X^2 + Y^2 \leq x) = 0,05.$$

Mivel az összeg χ^2 eloszlást követ 2 szabadságfokkal, ezért a $0,05$-ös szignifikancia szint mellett az $x = 5,9914$.

Ezek alapján elmondható, hogy a (λ, θ) paraméterpár értékei 95 százalék valószínűséggel teljesítik az

$$\frac{(\theta - \hat{\theta})^2}{(nI_1(\hat{\theta}))^{-1}} + \frac{(\lambda - \hat{\lambda})^2}{(nI_2(\hat{\lambda}))^{-1}} \leq 5.9914$$

egyenlőtlenséget.

Ezek alapján a $0,05$ szinthez tartozó konfidencia ellipszoid:

$$K = \left\{ (\lambda, \theta) : \frac{(\theta - \hat{\theta})^2}{(nI_{1n}(\hat{\theta}))^{-1}} + \frac{(\lambda - \hat{\lambda})^2}{(nI_{2n}(\hat{\lambda}))^{-1}} \leq 5.9914 \right\}.$$

4. fejezet

Hipotézisvizsgálat

A továbbiakban hipotézisvizsgálatot fogunk végezni különböző függvényosztályok ellen. A $H0$ hipotézis minden részben azonos, ugyanis azt tesszük fel, hogy az újranyitás exponenciális eloszlás szerint történik. A különböző hipotézisellenőrzési feladatokra próbastatisztikákat ismertetünk, majd a statisztikákra kiszámoljuk a kritikus értékeket.
A hipotézis ellenőrzés során kétféle hibát követhetünk el: elvetjük a nullhipotézist, pedig az igaz, ennek a valószínűségét viszonylag könnyű számolni, ha ismerjük a hipotézishez tartozó statisztikánk határeloszlását. A másodfajú hiba az, mikor elfogadjuk a null-hipotézist, pedig nem igaz. Ennek a valószínűsége összetett $H1$ hipotézis esetén függ a paraméter értékétől. A hipotézis vizsgálata úgy működik, hogy keresünk egy olyan függvényt, amelynek eloszlása a null-hipotézis fennállása esetén ismert. A döntéskor a mintateret két részre osztjuk: elfogadási- és kritikus tartományra.
A hipotézisvizsgálatban a döntést próbának nevezzük, és a kritikus tartományt próbafüggvénnyel definiáljuk. Ha a próbafüggvény értéke p, akkor a nullhipotézist p valószínűséggel elfogadjuk. A hipotézisek ellenőrzése során Kolmogorov-Smirnov, Cramer-Misses illetve rang teszteket használtunk.

Legyen $X_1, \ldots, X_n$ a megfigyelt mintánk, amelynek sűrűségfüggvénye a következő: $\hat{\theta}$ ha $t = 0$ és $1 - \hat{\theta}$ valószínűséggel $\hat{\lambda}$ intenzitású exponenciális eloszlás sűrűségfüggvénye.

A hipotézisek ellenőrzéséhez a következő módosításokat végezzük el: Ismert, hogy $\hat{\theta} = \hat{p} + \hat{q}e^{-\hat{\lambda}c}$, ahonnan

$$\hat{q} = \frac{n - T}{n(1 - e^{-\hat{\lambda}c})}.$$

Mivel $\hat{q}$ jelöli annak a valószínűségét, hogy egy kárt újranyitnak, ezért azok aránya, amiket nem látunk c miatt:

$$\hat{q}e^{-\hat{\lambda}c} = \frac{n - T}{n} \frac{1}{e^{\hat{\lambda}c} - 1},$$

tehát

$$n\hat{q}e^{-\hat{\lambda}c} = \frac{n - T}{e^{\hat{\lambda}c} - 1}.$$

Az exponencialitás teszthez előbb a pozitív megfigyeléseket kiegészítjük $[\frac{n-T}{e^{\hat{\lambda}c}-1}]$ darab $c + Y_i$ valószínűségi változóval, ahol Y_i exponenciális eloszlású $\hat{\lambda}$ intenzitással. Így a kiegészített mintát úgy tudjuk kezelni, mintha nem lenne nullában való elfajulás és c-ben való cenzorálás. Az új minta elemszámát jelöljük m-el.

4.1. Próbák IFR ellen

Ebben a részben hipotézisvizsgálattal foglalkozunk az IFR osztály ellen. A statisztikai próbákat a [11] cikkből vettük. Először meghatároztuk a statisztikák határeloszlását a $H0$ illetve a $H1$ hipotézisek alatt, majd alternatív eseteket vizsgálva, megnéztük, hogy milyen esetben fogadjuk el a hipotézist, ha az alternatív hipotézis a Weibull, vagy a Gamma.

Ebben a részben a következő hipotézisvizsgálati feladattal foglalkozunk:

$$\begin{aligned} H0 &: r(t) = \hat{\lambda}, t > 0 \\ H1 &: r(t_1) < r(t_2), t_1 < t_2 \end{aligned} \tag{4.1}$$

Ami pontosan azt jelenti, hogy:

$$\begin{aligned} H0 &: \text{ A mintánk } \hat{\lambda} \text{ intenzitású exponenciális eloszlást követ} \\ H1 &: \text{ A mintánk eloszlása IFR osztályból származik} \end{aligned}$$

Jelölje $X_1^*, X_2^*, ..., X_m^*$ a növekvő sorrendbe rendezett mintát. Bevezetjük a következő valószínűségi változókat:

$$\begin{aligned} D_1 &= mX_1^* \\ D_2 &= (m-1)(X_2^* - X_1^*) \\ &... \\ D_m &= X_m^* - X_{m-1}^*. \end{aligned} \tag{4.2}$$

Legyen

$$V_{ij} = \begin{cases} 1 & \text{ha } D_i \leq D_j \\ 0 & \text{különben} \end{cases}$$

Ekkor, a [11] cikk alapján ismert, hogy tesztstatisztika:

$$V_m = \sum_{1 \leq i < j \leq m} V_{ij} \tag{4.3}$$

A $H0$-t elvetjük az α szignifikancia szint mellett, ha $V_m > \hat{\theta}_{\alpha,m}$, ahol

$$P(V_m > \hat{\theta}_{\alpha,m} | H0) = \alpha.$$

Befogjuk látni, hogy a $H0$ alatt a $D_1, ..., D_m$ exponenciális eloszlást követ $\hat{\lambda}$ intenzitással, és $P(V_{ij} = 1) = \frac{1}{2}$, ha $i \neq j$. Azonban az alternatív hipotézis alatt a $P(V_{ij} = 1) > \frac{1}{2}$. Így a V_m nagy értékei mellett fogjuk elutasítani a null hipotézist. A következőkben meghatározzuk a V_m statisztika eloszlását a H_0 hipotézis alatt. Mivel a $D_1, ..., D_m$ minden sorrendje egyenlő valószínűséggel fordul elő, ezért a V_m eloszlása könnyen megadható. Jelölje $P_m(k)$ az m szám azon rendezéseinek számát, ahol k inverzió szerepel. Először belátjuk, hogy $(D_1, ..., D_m)$ együttesen $exp(\hat{\lambda})$ eloszlást követ, feltéve, hogy az $(X_1, ..., X_m)$ is az.

Legyenek $0 < x_1 < x_2 < ... < x_m$, és ϵ olyan kicsi, hogy az $(x_i, x_i + \epsilon)$ intervallumok diszjunktak. Ekkor

$$\begin{aligned} P(x_1 < X_1^* < x_1 + \epsilon, x_2 < X_2^* < x_2 + \epsilon, , ..., x_m < X_m^* < x_m + \epsilon) = \\ = m(m-1)...1 \prod_{i=1}^{m} \left(e^{-\hat{\lambda} x_i} - e^{-\hat{\lambda}(x_i + \epsilon))} \right) + O(\epsilon^m) \\ = m! \prod_{i=1}^{m} \left(e^{-\hat{\lambda} x_i} - e^{-\hat{\lambda}(x_i + \epsilon))} \right) + O(\epsilon^m) \\ = m! \hat{\lambda}^m e^{-\hat{\lambda}(x_1 + x_2 + ... + x_m)} \epsilon^m + O(\epsilon^m), \end{aligned}$$

ahonnan az $(X_1^*, ..., X_m^*)$ minta együttes sűrűségfüggvénye:

$$f(x_1, ..., x_m) = \hat{\lambda}^m e^{-\hat{\lambda}(x_1 + x_2 + ... + x_m)} \mathbb{I}(x_1 < x_2 < ... < x_m)$$

A (4.2)-es alapján a megfelelő Jacobi mátrix determinánsa $\frac{1}{m!}$, és $\sum T_i^* = \sum D_i$, ezért a $(D_1, ..., D_m)$ együttes sűrűségfüggvénye megegyezik az $(X_1^*, ..., X_m^*)$ sűrűségfüggvényével, tehát a $(D_1, ..., D_m)$ valóban exponenciális eloszlású.

Ekkor megmutatható, hogy $p_m(k) = P_{m-1}(k) + P_{m-1}(k-1) + ... + P_{m-1}(k-m+1)$ és $P_m(k) = 0, k < 0$, valamint

$$P(V_m = k) = \frac{p_m(k)}{m!}.$$

A V_m generátorfüggvénye:

$$G_m(z) = \sum_{i=0}^{+\infty} z^i P(V_m = i) = \sum_{i=0}^{+\infty} z^i \frac{p_m(i)}{m!}.$$

Legyen

$$\begin{aligned} (m+1)! G_{m+1}(z) &= (m+1)! \textstyle\sum_{i=0}^{+\infty} z^i P(V_{m+1} = i) \\ &= \textstyle\sum_{i=0}^{+\infty} z^i P_{m+1}(i) \\ &= \textstyle\sum_{i=0}^{+\infty} z^i (P_m(i) + P_m(i-1) + ... + P_m(i-m)) \\ &= (1 + z + ... + z^m) m! G_m(z), \end{aligned}$$

tehát észrevehető, hogy:

$$(m+1)!G_{m+1}(z) - (1+z+\ldots+z^m)m!G_m(z) \quad = \quad 0.$$

Ahonnan

$$G_m(z) = \frac{1}{m!}\Pi_{i=1}^{m}(1+z^2+\ldots+z^{i-1}). \tag{4.4}$$

A (4.4)-es kifejezés az $U_1+\ldots+U_m$ összeg generátorfüggvénye, ahol $U_i \sim U(0,i)$. Mivel az összeg generátorfüggvénye megegyezik a V_m generátorfüggvényével, ezért

$$V_m \sim \sum_{i=1}^{m} U_i,$$

tehát a V_m aszimptotikusan normális eloszlású, és

$$\begin{aligned} \mathbb{E}(V_m) &= \tfrac{m(m-1)}{4} \\ \mathbb{D}^2(V_m) &= \tfrac{m(m-1)(2m+5)}{72} \end{aligned}$$

Proschan és Pyke bebizonyították, hogy V_m-re alapozott próba konzisztens.

A következőkben nézzük meg, hogy mi lesz a V_m statisztika eloszlása, ha feltesszük, hogy a mintánk együttes eloszlásának a hazárd rátája nem-csökkenő és nem konstans függvény.

Legyen

$$G_m = \sum_{i=2}^{n-1}\sum_{j=i+1}^{n} g(D_i, D_j),$$

Ahol a g egy korlátos, nemnegatív függvény. Ha $x \geq y$, akkor $g(x,y) = 1$, különben nulla. Észrevehető, hogy G_m reprezentálja V_m-et.

Bemutatunk néhány konstrukciót, amely megtalálható a [11] cikkben. Legyen $\{Y_i, i > 0\}$ egy 1-intenzitású exponenciális eloszlásból származó valószínűségi változók sorozata H eloszlásfüggvénnyel, és legyen

$$Y_{i,m} = \sum_{j=1}^{i} \frac{Y_j}{m-j+1}.$$

Bevezetjük a következő valószínűségi változókat:

$$U_{m,i} = H(Y_{m,i}) = 1 - e^{Y_{m,i}}.$$

Ekkor az $U_{m,1}, \dots, U_{m,m}$ statisztikák egyenletes eloszlást követnek a $(0,1)$-en. Legyen $K = F^{-1} \circ H$. Ekkor $X_i^* = K(Y_{m,i})$. Ha h és k a megfelelő eloszlások sűrűségfüggvényei, akkor

$$k(\mu) = \frac{h(\nu)}{f(K(\nu))}, \forall \nu > 0.$$

A fenti jelöléseket használva, felírhatjuk:

$$D_i = (m - i + 1)[K(Y_{m,i}) - K(Y_{m,i-1})].$$

Ha $k(H^{-1}(u)) = w(u)$, akkor

$$w(u) = \frac{1-u}{f(F^{-1}(u))} = \frac{1}{r(F^{-1}(u)}.$$

A határeloszlás kiszámításához tegyük fel a következőket:

- F abszolút folytonos függvény;
- Az f folytonos;
- Létezik az w függvénynek deriváltja, és az folytonos.

Legyen $L(u,v) = \mathbb{E}(Y_1 w(u), Y_2 w(v)), 0 < u < v < 1$. Ekkor a $\delta > 0$-ra definiáljuk a következő mennyiségeket:

$$\begin{aligned}
\overline{L}(u,v,\delta) &= \sup_{|u-x|<\delta, |v-y|<\delta} L(x,y), \\
\overline{L}(u,v,\delta) &= \inf_{|u-x|<\delta, |v-y|<\delta} L(x,y).
\end{aligned}$$

Megmutatható, hogy

$$K(u,v) = |L_1(u,v)| + |L_2(u,v)|,$$

$$\begin{aligned}
L_1(u,v) &= \tfrac{w'(u)}{r(u)} \mathbb{E}(g(Y_1 w(u), Y_2 w(v))(1 - Y_1)), \\
L_2(u,v) &= \tfrac{w'(v)}{r(v)} \mathbb{E}(g(Y_1 w(u), Y_2 w(v))(1 - Y_2)).
\end{aligned}$$

Minden $1 \leq i < j \leq m$-re legyenek:

$$
\begin{aligned}
G_{m,i,j} &= g(D_i^*, D_j^*) \\
S_{m,i,j} &= g\left(Y_i w\left(\tfrac{i}{m}\right), Y_j w\left(\tfrac{j}{m}\right)\right) \\
T_{m,i,j} &= G_{m,i,j} - S_{m,i,j} \\
R_{m,i,j} &= \left(U_{m,i} - \tfrac{i}{m}\right) L_1\left(\tfrac{i}{m}, \tfrac{j}{m}\right) + \left(U_{m,j} - \tfrac{j}{m}\right) L_2\left(\tfrac{i}{m}, \tfrac{j}{m}\right)
\end{aligned}
$$

4.1.1. Tétel. *A fenti feltételezések alatt igaz, hogy*

$$m^{(-3/2)}(G_m - S_m - R_m) \longrightarrow 0.$$

eloszlásban.

Mivel ismerjük a statisztika határeloszlásait, ezért meg tudjuk határozni α szignifikancia szint mellett a kritikus értéket, azonban a $P(V_m = k)$ valószínűség kiszámítása nagy m-re és nagy k-ra bonyolult, ezért nem ajánlott ezzel a teszttel dolgozni. A következőkben ismertetek egy ún. rang-tesztet.

A Φ-függvényt próbának nevezzük, ha értékeit a $[0, 1]$-ből veszi fel, és a $H0$ elvetésének a valószínűségét jelöli. A következőkben a próbát a $D_1, ..., D_m$ sztenderdizált növekményekre fogjuk vetíteni.

A Φ próbát monotonnak nevezzük, ha

$$\Phi(D_1^{'}, ..., D_m^{'}) \leq \Phi(D_1^*, ..., D_m^*),$$

$\forall (D_1^{'}, ..., D_m^{'}), (D_1^*, ..., D_m^*)$ sztenderdizált növekményekre, ha igaz, hogy $\forall i < j$-re $D_i^{'} \geq D_j^{'}$, és $D_i^* \geq D_j^*$

Az F_1 függvényt az F_2-re konvexnek nevezzük, ha

$$F_1^{-1} \circ F_2 konvex.$$

Ha F_1 konvex az F_2-re, akkor

$$\mathbb{E}(\Phi | F_1) \leq \mathbb{E}(\Phi | F_2).$$

Egy tesztet rang-tesztnek nevezünk, ha próbafüggvénye

$$\Phi = \Phi(R_1, ..., R_m),$$

ahol R_i a D_i rangja. Az R_i jelöli, hogy a D_i hányadik a $D_1^* \leq D_2^* \leq ... \leq D_m^*$ mintában. P.J.Bickel [14] bebizonyította, hogy minden monoton rang-teszt torzítatlan a (4.1)-es feladatra nézve.
A továbbiakban a következő eloszláscsaládot fogjuk vizsgálni:

$$\{f_{\nu,\hat{\lambda}}, \nu \geq 0, \hat{\lambda} > 0\}.$$

Továbbá feltesszük azt is, hogy $f_{0,\hat{\lambda}}(t) = \hat{\lambda}e^{-\hat{\lambda}t}$, és $\nu > 0$-ra az $f_{\nu,\hat{\lambda}}$ IFR-beli. Ekkor az eredeti feladatot a következő formában tudjuk felírni:

$$\begin{aligned} H0 &: \nu = 0 \\ H1 &: \nu > 0 \end{aligned}$$

Rang tesztek használata esetén a következő statisztikákat szokták használni:

$$\begin{aligned}
W_0 &= \textstyle\sum_{i=1}^{m} -(\frac{i}{m+1})(-\frac{R_i}{m+1}); \\
W_1 &= \textstyle\sum_{i=1}^{m} -(1-\frac{i}{m+1})(-log(\frac{R_i}{m+1})); \\
W_2 &= \textstyle\sum_{i=1}^{m} -(log(1-\frac{i}{m+1})(-log(\frac{R_i}{m+1})); \\
W_3 &= \textstyle\sum_{i=1}^{m} -log(-(log(1-\frac{i}{m+1}))(-log(\frac{R_i}{m+1})); \\
S_1 &= \textstyle\sum_{i=1}^{m} -\frac{i}{m+1}D_i; \\
S_2 &= \textstyle\sum_{i=1}^{m} log(-\frac{i}{m+1})D_i; \\
S_3 &= \textstyle\sum_{i=1}^{m} -log(-log(-\frac{i}{m+1}))D_i;
\end{aligned}$$

A próbák közti választásban az erőfüggvény segíthet. A [12] cikk alapján elmondható, hogy az $S_i^* = S_i/(\sum D_i)$ és a W_i teljesítménye ugyanaz, és a W_0 aszimptotikusan ekvivalens a V_m-el.

A következőkben meghatározzuk az ARE adatokat különböző eloszláscsaládok mellett.
Legyen az exponencialitást ellenőrző teszt esetén az alternatív hipotézis a Weibull-eloszlás. Ezek alapján a (4.1)-es feladat a következőképpen írható fel:

$$\begin{aligned} H0 &: F(x) = 1 - e^{-\hat{\lambda}x}, \hat{\lambda} > 0 \\ H1 &: F(x) = 1 - e^{-\hat{\lambda}x^{\alpha}}, \hat{\lambda} > 0, \alpha \geq 1 \end{aligned}$$

Likelihood-arány tesztet használva, elutasítjuk a $H0$ hipotézist, ha

$$\frac{\max_{\alpha\geq 1}\hat{\lambda}^m\alpha^m\left(\prod_{i=1}^m X_i^*\right)e^{-\hat{\lambda}\sum_{i=1}^m (X_i^*)^\alpha}}{\hat{\lambda}^m e^{-\hat{\lambda}\sum_{i=1}^m X_i^*}} > \hat{\theta}_\alpha.$$

Legyen V_m a sztenderdizált növekményekre épített minta, és

$$T_m = \sum_{i=1}^m (1-\hat{\lambda}X_i^*)ln(X_i^*).$$

Ha $\alpha = 1$, akkor az $F(x) = 1 - e^{-\hat{\lambda}x}$, ezért

$$ARE(V_m, T_m) = \frac{\frac{[\mu_T'(1)]^2}{\sigma_T^2(1)}}{\frac{[\mu_W'(1)]^2}{\sigma_W^2(1)}}.$$

A T_m statisztika várható értéke:

$$\mu_T(\alpha) = \int_0^{+\infty}(1-\hat{\lambda}x)ln(x)\alpha\hat{\lambda}x^{\alpha-1}e^{-\hat{\lambda}x^\alpha}dx,$$

ahonnan

$$\begin{aligned}\mu_T'(1) &= \int_0^{+\infty}(1-\hat{\lambda}x)lnx\hat{\lambda}e^{-\hat{\lambda}x}(1+lnx-\hat{\lambda}xlnx)dx\\ &= (ln\hat{\lambda}+\gamma-1)^2+\tfrac{\pi^2}{6}\end{aligned}$$

és

$$\sigma_T^2(1) = E(T_m^2) - \mu_T^2(1) = (ln\hat{\lambda}+\gamma-1)^2 + \frac{\pi^2}{6},$$

ahol $\gamma = 0,5772...$

A V_m statisztika várható értéke:

$$\mu_W(\alpha) = \int_0^{+\infty}\int_0^{+\infty}\frac{r(y)}{r(x)+r(y)}f_W(x)f_W(y)dxdy$$

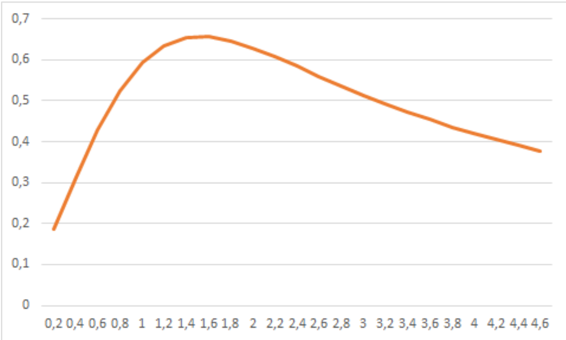
A kettős integrál kiszámításokból következik:

$$\begin{aligned}\mu_W'(1) &= \tfrac{1}{4}ln2,\\ \sigma_W^2(1) &= \tfrac{1}{36}.\end{aligned}$$

Tehát

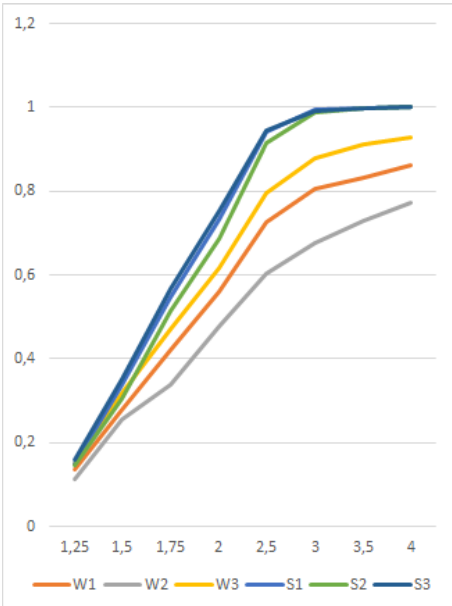
$$ARE(V_m, T_m) = \frac{1.0809}{(ln\hat{\lambda}-0.4228)^2+1.6449}.$$

A lenti ábrán az erőfüggvény grafikus képe látható. Észrevehető, hogy az erőfüggvény $ln\hat{\lambda} = 0,4228$-ben veszi fel a legnagyobb értékét, valamint ha $\hat{\lambda}\to 0$, akkor $ARE \to 0$, illetve $ARE(V_m, T_m) \leq 0,6571$.



4.1. ábra. Relatív hatásfok

A függelék 4.1-es táblázat tartalmazza néhány paraméter mellett a W_i statisztikák teljesítményeit. Ezen értékek Monte-Carlo szimulációval lettek elkészítve 2000 szcenárióra, ahol az $m = 10$. A táblázat a [12] cikkből van. A statisztikák teljesítmény-függvénye a következő ábrán látható:



4.2. ábra. Teljesítmény Weibull alternatívára

A teljesítmény becsléseiből az látszik, hogy a felsorolt statisztikák közül teljesítmény alapján az S_1 illetve az S_3 statisztika tűnik a legjobbnak. Ugyanakkor az is elmondható, hogy a rang tesztek rosszabbul teljesítenek, mint a D_i-re alapozott tesztek.

Végezetül azt a következtetést vonhatjuk le a teljesítményre vonatkozóan, hogy ha az alternatív hipotézis Weibull, akkor az S_1 illetve az S_3 próbák optimálisak.

Most vizsgáljuk azt az esetet, ha az alternatív hipotézis a Gamma-eloszlás. Ekkor a (4.1)-es feladatot a következőképpen lehet felírni:

$$\begin{aligned} H0 &: F(x) = 1 - e^{-\hat{\lambda}x}, \hat{\lambda} > 0 \\ H1 &: F(x) = \int_0^x \frac{\hat{\lambda}(\hat{\lambda}t)^{\alpha-1}}{\Gamma(\alpha)} e^{-\hat{\lambda}t} dt, \hat{\lambda} > 0, \alpha \geq 1 \end{aligned}$$

Legyen a V_m a D_i-re épített minta, és

$$T_m = \sum_{i=1}^{m} ln(X_i^*).$$

A T_m statisztika várható értéke az α függvényében:

$$\begin{aligned} \mu(\alpha) &= \int_0^\infty ln(t) f(t) dt \\ &= \int_0^\infty ln(t) \frac{\hat{\lambda}(\hat{\lambda}t)^{\alpha-1}}{\Gamma(\alpha)} e^{-\hat{\lambda}t} dt \end{aligned}$$

Ahonnan:

$$\mu^{'}(\alpha) = \int_0^{+\infty} ln(t) \frac{[\alpha(\hat{\lambda}t)^{\alpha-1} + (\alpha-1)\hat{\lambda}^{\alpha} t^{\alpha-2}]\Gamma(\alpha) + \Gamma^{'}(\alpha)\hat{\lambda}^{\alpha} t^{\alpha-1}}{\Gamma^2(\alpha)} dt$$

Numerikusan igazolható, hogy

$$\begin{aligned} \mu^{'}(1) &= (1 - \gamma\hat{\lambda}) \int_0^\infty e^{-\hat{\lambda}t} ln(t) dt \\ &= (1 - \hat{\lambda}\gamma) \frac{pi}{\sqrt{6}} \end{aligned}$$

A T_m statisztika szórása az α függvényében:

$$\sigma^2(1) = \int_0^\infty ln^2(t) f(t) dt|_{\alpha=1} - \mu^2(1).$$

Numerikusan igazolható, hogy

$$\sigma^2(1) = (1 - \hat{\lambda}\gamma)^2.$$

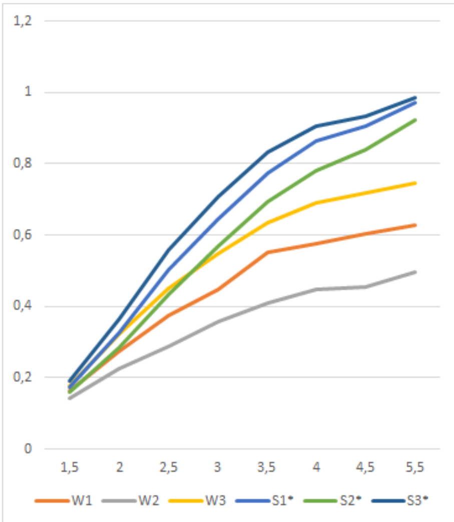
A V_m statisztikára:

$$\frac{\mu^{'}(1)}{\sigma(1)} = 9 \left(ln2 - \frac{1}{2} \right)^2,$$

ahonnan

$$ARE(V_m, T_m) = 0,2040.$$

A függelék 4.2-es táblázata tartalmazza néhány paraméter mellett a becsült erőfüggvények értékeit a W_i statisztikákra. Ezen értékek is szintén Monte-Carlo szimulációval lettek elkészítve 2000 szcenárióra, ahol az $m = 10$. A táblázat a [12] cikkből van. A tesztek teljesítménye a következő ábrán látható:



4.3. ábra. Teljesítmény Gamma alternatívára

Megfigyelhető, hogy az S_3 teszt a legerősebb, és a W_2 a leggyengébb, bármilyen paraméter mellett. Akárcsak a Weibull alternatívánál itt is levonhatjuk azt a következtetést, hogy a D_i-kre épített minta a teljesítmény szempontjából optimálisabbak, mint a rang-tesztek.

4.2. Próbák NBU ellen

Ebben a részben a következő hipotézisvizsgálattal foglalkozunk:

$$\begin{aligned} H0: &\quad \text{A minta exponenciális eloszlást követ} \\ H1: &\quad \text{A minta eloszlása NBU-beli} \end{aligned} \tag{4.5}$$

Először Cramer-Misses típusú próbákra szorítkozunk. A [8], [9] cikket alapján a próbastatisztika legyen:

$$
\begin{aligned}
\gamma(F) &= \int_0^{+\infty}\int_0^{+\infty}\left(\overline{F}(x)\overline{F}(y) - \overline{F}(x+y)\right)dF(x)dF(y) \\
&= \tfrac{1}{4} - \int_0^{+\infty}\int_0^{+\infty}\overline{F}(x+y)dF(x)dF(y) \qquad (4.6)\\
&= \tfrac{1}{4} - \Delta(F).
\end{aligned}
$$

Látható, hogy a $\gamma(F)$ mennyiség az exponenciálistól való eltérés mérésére szolgál, ugyanis ha a minta exponenciális eloszlást és az F nem lépcsős, akkor a $\gamma(F) = 0$, vagyis a $\Delta(F) = \frac{1}{4}$. Ha $F \in NBU$, akkor a $\gamma(F) > 0$, tehát $\Delta(F) < \frac{1}{4}$. Mivel az elméleti eloszlásfüggvényt lépcsős függvényekkel tudjuk közelíteni, ezért a a (4.5)-ös feladat helyett a következővel fogunk dolgozni:

$$
\begin{aligned}
H0:&\quad \Delta(F) = \tfrac{1}{4}\\
H1:&\quad \Delta(F) < \tfrac{1}{4}
\end{aligned} \qquad (4.7)
$$

Legyen F_m olyan, hogy $F_m \longrightarrow F$ eloszlásban. Tekintsük a következő mennyiséget:

$$
J_m = \frac{2}{m(m-1)(m-2)} \sum_{\alpha_1 \neq \alpha_2, \alpha_1 < \alpha_3, 1 \leq \alpha_1 \leq m} \psi(x_{\alpha_1}, x_{\alpha_2+\alpha_3}).
$$

Ekkor a J_m statisztika aszimptotikusan egyenlő $\Delta(F_m)$-el, ahol

$$
\psi(a,b) = \begin{cases} 1 & \text{ha } a > b \\ 0 & \text{ha } a < b \end{cases}
$$

Legyen $X = (X_1, ..., X_m)$ és $Y = (Y_1, ..., Y_m)$ két minta, amelyek F és G eloszlásfüggvénnyel rendelkeznek, és F szuperadditív a G-re nézve. Ekkor

$$
J_m(X) \leq J_m(Y)
$$

és a J_m teszt torzítatlan.

Észrevehető, hogy a J_m statisztika egy U-statisztika. Az aszimptotikus normalitáshoz a 2. fejezetben leírt módszereket használjuk fel.

Legyen $\Phi(x_1,x_2,x_3) = \frac{1}{3}\{\psi(x_1,x_2+x_3)+\psi(x_2,x_1+x_3)+\psi(x_3,x_1+x_2)\}$ és $\Phi_1(x_1) = \mathbb{E}\Phi(x_1,X_2,X_3)$, $\Phi_2(x_1,x_2) = \mathbb{E}\Phi(x_1,x_2,X_3)$, $\Phi_3(x_1,x_2,x_3) = \mathbb{E}\Phi(x_1,x_2,x_3)$, valamint $\xi_k = \mathbb{E}\Phi_k(X_1,...,X_k) - \Delta^2, k = 1,2,3$, ahol a $\Delta(F)$ a (4.6)-ben definiált mennyiséget jelöli.

Ekkor

$$Var(J_m) = \frac{1}{\binom{m}{3}}\sum_{k=1}^{3}\binom{3}{k}\binom{m-3}{3-k}\xi_k$$

és

$$\lim_{m\to\infty} mVar(J_m) = 9\xi_1.$$

Ha $\xi_1(F) > 0$, akkor $\sqrt{m}(J_m - \Delta(F_m)) \to N(0,9\xi_1)$. Tehát ha $H0$-on vagyunk, akkor $\sqrt{m}(J_m - \frac{1}{4}) \to N(0,\frac{5}{432})$. Az egyszerűség kedvéért a J_m teszt helyett a T_m teszttel szokás számolni, ahol

$$T_m = \frac{m(m-1)(m-2)J_m}{2}\sum \psi_m(X_{\alpha_1},X_{\alpha_2}+X_{\alpha_3}).$$

Legyen $X_1^*,...,X_m^*$ a rendezett minta. Minden $i \leq max(i,j)$-re $\psi(X_i^*,X_j^*+X_k)^* = 0$, ezért a T_m statisztikát a következőképpen tudjuk felírni:

$$T_m = \sum_{i>j>k}\psi(X_i^*,X_j^*+X_k^*)$$

A következőkben meghatározzuk a T_m statisztika eloszlását. Ehhez kiszámoljuk azt, hogy a statisztika milyen valószínűséggel veszi fel a $0,1,...,\frac{m(m-1)(m-2)}{6}$ értékeket a $H0$ hipotézis alatt. Tekintsük a $D_i = (m-i+1)(X_i^* - X_{i-1}^*)$ sztenderdizált növekményeket. Az előző részben beláttuk, hogy ezek a valószínűségi változók függetlenek, és a D_i exponenciális eloszlást követ $m-i+1$ intenzitással.

$$
\begin{aligned}
\mathbb{P}(T_m=0) &= \mathbb{P}\left(\sum_{i>j>k}\psi(X_i^*,X_j^*+X_k^*)=0\right)\\
&= \mathbb{P}\left(\bigcap_{i>j>k}\{X_i^*<X_j^*+X_k^*\}\right)\\
&= \mathbb{P}\left(X_m^*<X_1^*+X_2^*\right)\\
&= \mathbb{P}\left(\sum_{i=3}^{m}D_i<D_1\right)\\
&= m!\int_0^{+\infty}\int_0^{+\infty}\dots\int_{a_3+\dots a_m}^{+\infty}\prod_{i=1}^{m}e^{-(m-i+1)a_i}da_1\dots da_m\\
&= \binom{2m-2}{m}^{-1}.
\end{aligned}
$$

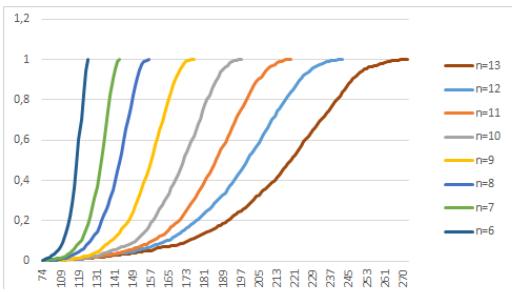
Ugyanezzel a gondolatmenettel megmutatható, hogy

$$
\mathbb{P}(T_m\leq 1)=\binom{2m-3}{m-3}^{-1}\frac{(3m-1)(m-2)}{(2m-2)(2m-1)}.
$$

Végül

$$
\begin{aligned}
\mathbb{P}(T_m=\tfrac{m(m-1)(m-2)}{6}) &= \mathbb{P}(X_3^*>X_1^*+X_2^*,\dots,X_m^*>Xm-1^*+X_{m-2}^*)\\
&= m!\int_0^{+\infty}\int_{a_1}^{+\infty}\int_{a_1+a_2}^{+\infty}\dots\int_{a_{m-1}+a_{m-2}}^{+\infty}\prod_{i=1}^{m}e^{-a_{m-i+1}}da_{m-i+1}.
\end{aligned}
$$

Ezek a képletek nagy m-re igen nehezen számolhatók, ezért a kritikus értékeket és az eloszlásfüggvényeket Monte-Carlo szimulációval adtuk meg, összesen 2000 kimenetelt vizsgálva. A kritikus értékek számítását Visual Basicben végeztük, a programkód megtalálható a függelékben. A 4.3-as táblázat a T_m-statisztika eloszlásfüggvényeinek az értékeit tartalmazza néhány m-re. Ezek alapján a statisztika eloszlásfüggvényéi:



4.4. ábra. Rao-Cramer statisztika eloszlásfüggvényei

A (4.4)-es ábrán megfigyelhető, hogy az elemszám növekedésével egyre valószínűbb, hogy a statisztika nagy értékeket fog felvenni.
A statisztikára alapozott próba kritikus értékei megtalálhatóak a függelék 4.4-es táblázatában. Ha a T kisebb, mint a kritikus érték, akkor a null-hipotézist elvetjük.

A [10]. cikket felhasználva Kolmogorov-Smirnov típusú próbát fogunk felírni a mintánkra. Definiáljuk a

$$D(F) = \inf_{x,y \geq 0} (\overline{F}(x+y) - \overline{F}(x)\overline{F}(y))$$

mennyiséget. Ha $H0$-on vagyunk, akkor a $D(F) = 0$, illetve $H1$-en a $D(F) < 0$. Jelölje F_m a minta tapasztalati eloszlásfüggvényét. Ekkor

$$D(F_m) = \inf_{x,y \geq 0} (\overline{F}_m(x+y) - \overline{F}_m(x)\overline{F}_m(y)).$$

A hipotézis tesztelése során a $D(F_m)$ statisztikával fogunk dolgozni. A $D(F_m)$ statisztika a következőképpen írható fel:

$$D(F_m) = \min[\overline{F}_m(X_i^* + X_j^*) - \overline{F}_m(X_i^*)\overline{F}_m(X_j^*)].$$

Bevezetjük a következő jelöléseket:

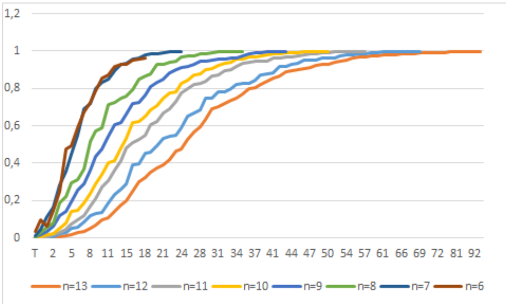
$$\begin{aligned} s_{ij} &= \sum_{k=1}^{m} \mathbb{I}(X_k^* > X_i^* + X_j^*), \\ t_{ij} &= [ms_{ij} - (m-i)(m-j)]. \end{aligned}$$

Ekkor

$$\begin{aligned} m^2 D(F_m) &= \min_{1 \leq i < j \leq m} [ms_{ij} - (m-i)(m-j)] \\ &= \min_{1 \leq i < j \leq m} t_{i,j} \\ &= T_m. \end{aligned}$$

Ezek alapján elmondhatjuk, hogy a $H0$ hipotézist elutasítjuk a $H1$-el szemben, ha a T_m statisztika értéke kicsi. Észrevehető, hogy a T_m könnyebben számolható, mint a $D(F_m)$, ezért a következőkben a tesztjeinket a T_m statisztikára alapozzuk.

A T_m eloszlása expliciten nagyon nehezen számolható, ezért a kritikus értékeket Monte-Carlo szimulációval határoztuk meg. Először a T_m statisztika empirikus eloszlásfüggvényét írtuk fel, ha m 6-tól 13-ig vesz fel értékeket. Az eloszlásfüggvények alakjából ((4.5)-ös ábra) azt a következtetést vonhatjuk le, hogy az elemszám növekedésével egyre valószínűbb, hogy a statisztika nagy értékeket fog felvenni.



4.5. ábra. Kolmogorov-Smirnov statisztika eloszlásfüggvényei

A kritikus értékeke meghatározására 50 m-re készítettünk 2000 szcenáriót, majd meghatároztuk a kimenetelek eloszlásának az eloszlásfüggvényét, az eloszlásfüggvényből pedig kiszámoltuk az értékeket $0.01, 0.05$ és 0.1-es szignifikancia szinteken. Ezen értékeket az 5. táblázat tartalmazza. A próba elveti $H0$-t, ha a T_m kisebb, mint a táblázatbeli érték (-1)-szerese.

Összefoglalás

Egy káresemény lezárása után is jelenthetnek be a kárra kárigényt. Ekkor a törvény szerint a kárt újranyitják, amely a biztosítónak veszteséget okoz a lebonyolításban. Ezen veszteség kikerülésére tartalékot kell képezni, amihez ismerni kell az újranyitás valószínűségét, illetve időpontjának eloszlását.

A dolgozatban e két probléma került bemutatásra. Azzal a feltételezéssel éltem, hogy az időpontok eloszlása exponenciális eloszlást követ. A feladat nehézsége abban rejlik, hogy a mintában nem tudjuk megkülönböztetni a nulla értékű és a cenzorálási időnél nagyobb elemeket. Ezért a minta együttes sűrűségfüggvénye nullában elfajult eloszlás keveréke exponenciális eloszlással. Ezen sűrűségfüggvényt átparamétereztük, és bevezettük a $\hat{\theta}$ paramétert. A maximum-likelihood becsléssel kapott értékekből látszik, hogy a $\hat{\theta}$ paraméternek szemléletes jelentése van: a nullák száma a mintában, és a mintaelemszám aránya. A $\hat{\lambda}$ paraméter értékét nem tudtuk expliciten meghatározni, de a kapott egyenletnek létezik egyértelmű megoldása. Ezután a kiszámoltam a paraméterek Fisher-féle információját, és a paraméterpárra konfidencia ellipszoidot illesztettem. A negyedik részben az időpontok eloszlására végeztem hipotézis ellenőrzéseket. Először a mintához hozzávettem bizonyos számú eltolt exponenciális eloszlású valószínűségi változót, amellyel a c-ben való cenzorálást, illetve a nullában való elfajulást tudjuk kikerülni, és a hipotézis ellenőrzéseket az új mintára csináltuk. Elmondható, hogy ha a null-hipotézis elutasításra kerül az új mintára, akkor az eredeti feladatban sem fogadható el. Azokat az eseteket vizsgáltam, amikor az alternatív hipotézis egy NBU vagy IFR osztálybeli függvény. Az IFR osztályra ismertettem néhány statisztikát, valamint Weibull és a Gamma alternatívákra kiszámoltam a sztenderdizált növekményekre épített statisztika, és egy Cramer-Misses típusú statisztika erejét. Mindkét alternatíva esetén elmondható, hogy a sztenderdizált növekményekre épített minták jobban teljesítenek, mint a rang-tesztek. Mindkét osztály esetén néhány statisztikára Monte-Carlo szimulációval összesen 2000 szcenárióra, és 50 m-re meghatároztam a $0,01; 0,05$ és $0,1$-es szignifikancia szintekhez tartozó kritikus értékeket.

Irodalomjegyzék

[1] Gnyegyenko, B. V. Beljajev, Szolovej, *A megbizhatóságelmélet matematikai módszerei*, Műszai könyvkiadó, Budapest, 1970

[2] Móri Tamás, *Élettartam-adatok elemzése.* Typotex, Budapest, 2011.

[3] Richard E. Barlow and Kjell A. Doksum, *Isotonic Tests For Convex Orderings*, University of California, Berkeley,pg. 293-323

[4] Alan J. Lee, *On the Asymptotic Distribution Of U-statistics*, University Of Auckland and University of North California at Chapel Hill

[5] Wassily Hoeffding, *A class Of Statistics with asymptotically normal distribution*, University of North Carolina, Institute of Statistics, 1948, (293-325)

[6] Gottfried E. Noether, *On a theorem of Pitman*, Boston University, 1955 (64-68)

[7] Tómács Tibor, *Matematikai Statisztika*, Eszterházy Károly Főiskola, Matematikai és Informatikai Intézet, Eger, 2012

[8] Myles Hollander and Frank Proschan *Testing whether new is better than used*, Florida State University, 1972 (1136-1146)

[9] Yuan Yan Chen, Myles Hollander and Naftali A. Langberg *Testing Whether New Is Better Than Used With Randomly Censored Data*, Syracuse University, Florida State University, and University of Haifa, The Annals of Statistics 1983 (267-276)

[10] Hira L. Koul: *A Test For New Better Than Used*, Michigan State University, Commun. Statist.-Theor. Meth, 563-573

[11] Frank Proschan and Ronald Pyke: *Test For Monoton Failure Rate Boeing Scientific Research Laboratories*

[12] Peter J. Bickel And Kjell A. Doksum: *Test For Monotone Failure Rate Based On Normalized Spacings*, University of California, Berkeley, The Annals of Mathematical Statistics 1969, Vol. 40, No. 4, 1216-1235

[13] R. E. Barlow and F- Proschan: *A Note on Test For Monotone Failure Rate Based On Incomplete Data*, The Annals of Mathematical Statistics 1969, Vol. 40, No. 2, 595-600

[14] P.J.Bickel: *Test For monotone Failure Rate II*, University Of California, Berkeley, The Annals Of Mathematical Statistics, 1969, Vol. 4, 1250-1260

Függelék

t	1,25	1,5	1,75	2	2,5	3	3,5	4
W1	0,136	0,28	0,42	0,562	0,727	0,805	0,832	0,863
W2	0,112	0,255	0,34	0,477	0,605	0,675	0,731	0,774
W3	0,15	0,318	0,47	0,617	0,797	0,879	0,912	0,927
S1	0,149	0,337	0,546	0,734	0,941	0,994	0,999	1
S2	0,146	0,306	0,513	0,686	0,916	0,989	0,998	1
S3	0,161	0,352	0,566	0,752	0,945	0,992	0,999	1

4.1.táblázat

T α+1	1,5	2	2,5	3	3,5	4 4,5	5,5
W1	0,162	0,275	0,375	0,447	0,553	0,576 0,603	0,628
W2	0,144	0,225	0,287	0,356	0,41	0,449 0,454	0,498
W3	0,178	0,323	0,452	0,548	0,635	0,691 0,72	0,746
S1*	0,175	0,328	0,502	0,647	0,775	0,864 0,905	0,971
S2*	0,161	0,283	0,434	0,57	0,694	0,782 0,841	0,922
S3*	0,19	0,365	0,559	0,707	0,834	0,906 0,934	0,987

4.2.táblázat

n ⌢	10	20	35	55	75	120	160	200	250
6	0,139	0,9995	1	1	1	1	1	1	1
7	0,0205	0,1535	0,9995	1	1	1	1	1	1
8	0,003	0,0195	0,203	0,9975	1	1	1	1	1
9	0,0015	0,019	0,024	0,224	0,9025	1	1	1	1
10	0	0	0,00021	0,0355	0,156	1	1	1	1

n/t	0,01	0,05	0,1	n/t	0,01	0,05	0,1
2	0	0	0	26	1426	1618	1712
3	0	0	0	27	1612	1862	1951
4	0	0	1	28	1858	2042	2176
5	0	3	4	29	2040	2301	2420
6	2	7	9	30	2283	2573	2691
7	7	14	18	31	2516	2907	3011
8	12	26	30	32	2798	3163	3325
9	21	40	48	33	3130	3453	3642
10	43	59	70	34	3516	3839	4019
11	58	86	98	35	3819	4215	4407
12	79	118	128	36	4222	4649	4791
13	128	156	172	37	4552	5066	5238
14	158	200	224	38	5001	5520	5735
15	215	259	279	39	5362	5956	6240
16	272	323	354	40	5972	6509	6701
17	336	388	426	41	6428	6952	7244
18	388	482	519	42	6895	7479	7795
19	476	576	617	43	7418	8101	8436
20	579	676	733	44	8152	8801	9042
21	678	790	862	45	8863	9389	9696
22	837	936	990	46	9074	9919	10372
23	959	1096	1149	47	9925	10694	11032
24	1029	1258	1325	48	10669	11389	11793
25	1174	1427	1521	49	11341	12251	12593

4.4.táblázat

n/t	0,01	0,05	0,1	n/t	0,01	0,05	0,1
1	0	0	0	26	218	160	142
2	1	1	1	27	225	174	157
3	4	4	2	28	232	185	164
4	9	6	5	29	228	194	168
5	12	11	8	30	265	210	184
6	19	14	12	31	287	220	191
7	25	18	16	32	288	228	201
8	28	24	20	33	301	242	211
9	38	29	25	34	325	253	224
10	44	34	30	35	331	266	235
11	53	40	34	36	342	276	244
12	64	48	40	37	366	294	257
13	69	55	45	38	362	298	264
14	79	58	51	39	393	321	283
15	87	68	60	40	416	336	296
16	96	73	66	41	414	339	298
17	110	84	72	42	424	352	310
18	116	90	79	43	462	376	329
19	130	99	86	44	476	385	340
20	132	104	90	45	486	404	362
21	153	114	103	46	501	415	363
22	166	126	110	47	508	419	366
23	170	134	119	48	518	420	378
24	184	145	124	49	570	463	406
25	200	156	135	50	569	447	404

4.5.*táblázat*

Programkódok

```
Sub nbu_kolmogorov()
Dim w()
Dim t()
Dim s()
Dim st(2000)
Dim n, l, i, also, felso, j, csere
For n = 1 To 50
ReDim w(n)
ReDim t(n, n)
For l = 1 To 2000
ReDim s(n, n)
For i = 1 To n
    w(i) = -Application.WorksheetFunction.Ln(1 - Rnd)
Next i

also = LBound(w, 1) + 1 'A nullas indexet nem hasznaljuk
felso = UBound(w, 1)

For j = 1 To felso - also + 1
    For i = also To felso - 1
        If w(i) > w(i + 1) Then
            csere = w(i)
            w(i) = w(i + 1)
            w(i + 1) = csere
        End If
    Next i
Next j
```

```
For i = 1 To n
    For j = 1 To n
        For k = 1 To n
            If (w(k) > (w(i) + w(j))) Then
                s(i, j) = s(i, j) + 1
            End If
        Next k
    Next j
Next i

For i = 1 To n
    For j = 1 To n
        t(i, j) = n * s(i, j) - (n - i) * (n - j)
    Next j
Next i

Min = t(1, 1)
For i = 1 To n
    For j = 1 To n
        If (t(i, j) < Min) Then
            Min = t(i, j)
        End If
    Next j
Next i

st(l) = Min
Next l
```

```
also = LBound(st, 1) + 1 'A nullas indexet nem hasznaljuk
felso = UBound(st, 1)

For j = 1 To felso - also + 1
    For i = also To felso - 1
        If st(i) > st(i + 1) Then
            csere = st(i)
            st(i) = st(i + 1)
            st(i + 1) = csere
        End If
    Next i
Next j

Cells(n, 1) = n
Cells(n, 2) = st(20)
Cells(n, 3) = st(100)
Cells(n, 4) = st(200)

Next n
End Sub

Sub nbu_cramer()

Dim w()
Dim st(1000)

For n = 2 To 50
ReDim w(n)
For l = 1 To 1000
```

```
For i = 1 To n
    w(i) = (-Application.WorksheetFunction.Ln(1 - Rnd))
Next i

also = LBound(w, 1)   'A nullas indexet nem hasznaljuk
felso = UBound(w, 1)

For j = 1 To felso - also + 1
    For i = also To felso - 1
        If w(i) > w(i + 1) Then
            csere = w(i)
            w(i) = w(i + 1)
            w(i + 1) = csere
        End If
    Next i
Next j

T = 0
For i = 1 To n
    For j = 1 To n
        For k = 1 To n
            If ((i > j) And (j > k)) Then
                If (w(i) > w(j) + w(k)) Then
                    T = T + 1
                End If
            End If
        Next k
    Next j
Next i
```

```
st(l) = T
Next l

also = LBound(st, 1)   'A nullas indexet nem hasznaljuk
felso = UBound(st, 1)

For j = 1 To felso - also + 1
    For i = also To felso - 1
        If st(i) > st(i + 1) Then
            csere = st(i)
            st(i) = st(i + 1)
            st(i + 1) = csere
        End If
    Next i
Next j
Cells(n, 1) = n
Cells(n, 2) = st(10)
Cells(n, 3) = st(50)
Cells(n, 4) = st(100)
Next n
End Sub

Sub nbu_cramer_10()

Dim w()
Dim st(1000)

n = 10
ReDim w(n)
```

```
For l = 1 To 1000

For i = 1 To n
    w(i) = (-Application.WorksheetFunction.Ln(1 - Rnd))
Next i

also = LBound(w, 1)   'A nullas indexet nem hasznaljuk
felso = UBound(w, 1)

For j = 1 To felso - also + 1
    For i = also To felso - 1
        If w(i) > w(i + 1) Then
            csere = w(i)
            w(i) = w(i + 1)
            w(i + 1) = csere
        End If
    Next i
Next j

T = 0
For i = 1 To n
    For j = 1 To n
        For k = 1 To n
            If ((i > j) And (j > k)) Then
                If (w(i) > w(j) + w(k)) Then
                    T = T + 1
                End If
            End If
        Next k
```

```
    Next j
Next i
Cells(l, 1) = T
Next l
End Sub

Sub nbu_cramer_szcenario()

Dim w()
Dim st(1000)

For n = 6 To 14
ReDim w(n)
For l = 1 To 2000

For i = 1 To n
    w(i) = (-Application.WorksheetFunction.Ln(1 - Rnd))
Next i

also = LBound(w, 1)   'A nullas indexet nem hasznaljuk
felso = UBound(w, 1)

For j = 1 To felso - also + 1
    For i = also To felso - 1
        If w(i) > w(i + 1) Then
            csere = w(i)
            w(i) = w(i + 1)
            w(i + 1) = csere
        End If
```

```
    Next i
Next j

T = 0
For i = 1 To n
    For j = 1 To n
        For k = 1 To n
            If ((i > j) And (j > k)) Then
                If (w(i) > w(j) + w(k)) Then
                    T = T + 1
                End If
            End If
        Next k
    Next j
Next i
Cells(1, n) = n
Cells(l, n) = T
Next l

Next n
End Sub

Sub illesztes()

Dim w()
Dim st(1000)

 n = 13
ReDim w(n)
```

```
For l = 1 To 2000

For i = 1 To n
    w(i) = (-Application.WorksheetFunction.Ln(1 - Rnd))
Next i

also = LBound(w, 1)   'A nullas indexet nem hasznaljuk
felso = UBound(w, 1)

For j = 1 To felso - also + 1
    For i = also To felso - 1
        If w(i) > w(i + 1) Then
            csere = w(i)
            w(i) = w(i + 1)
            w(i + 1) = csere
        End If
    Next i
Next j

T = 0
For i = 1 To n
    For j = 1 To n
        For k = 1 To n
            If ((i > j) And (j > k)) Then
                If (w(i) > w(j) + w(k)) Then
                    T = T + 1
                End If
            End If
        Next k
```

```
    Next j
Next i
Cells(1, n) = n
Cells(l, n) = T
Next l

End Sub

Sub nbu_kolmogorov()
Dim w()
Dim t()
Dim s()
Dim st(2000)
For n = 6 To 50
ReDim w(n)
ReDim t(n, n)
For l = 1 To 2000
ReDim s(n, n)
For i = 1 To n
    w(i) = -Application.WorksheetFunction.Ln(1 - Rnd)
Next i

also = LBound(w, 1) + 1 'A nullas indexet nem hasznaljuk
felso = UBound(w, 1)

For j = 1 To felso - also + 1
    For i = also To felso - 1
        If w(i) > w(i + 1) Then
            csere = w(i)
            w(i) = w(i + 1)
```

```
            w(i + 1) = csere
        End If
    Next i
Next j


For i = 1 To n
    For j = 1 To n
        For k = 1 To n
            If (w(k) > (w(i) + w(j))) Then
                s(i, j) = s(i, j) + 1
            End If
        Next k
    Next j
Next i

For i = 1 To n
    For j = 1 To n
        t(i, j) = n * s(i, j) - (n - i) * (n - j)
    Next j
Next i

Min = t(1, 1)
For i = 1 To n
    For j = 1 To n
        If (t(i, j) < Min) Then
            Min = t(i, j)
        End If
    Next j
```

```
Next i

st(l) = Min
Next l
also = LBound(st, 1) + 1 'A nullas indexet nem hasznaljuk
felso = UBound(st, 1)

For j = 1 To felso - also + 1
    For i = also To felso - 1
        If st(i) > st(i + 1) Then
            csere = st(i)
            st(i) = st(i + 1)
            st(i + 1) = csere
        End If
    Next i
Next j

Cells(n, 1) = n
Cells(n, 2) = st(20)
Cells(n, 3) = st(100)
Cells(n, 4) = st(200)

Next n
End Sub
```