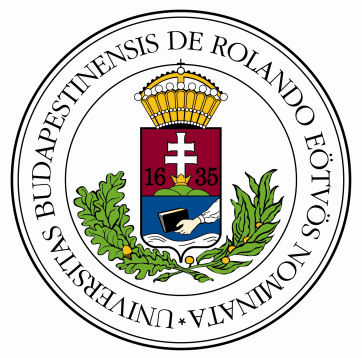
Eötvös Loránd Tudományegyetem
Matematika Intézet

Hosszú Klaudia

Biztosítási kárszámok modellezése

BSc szakdolgozat

Témavezető: Arató Miklós



Valószínűségelméleti és Statisztika Tanszék

2016. Budapest

Tartalomjegyzék

Bevezetés

A biztosítási matematika területén számos különböző módszert fejlesztettek már ki az éves baleseti kárszámok modellezésére. A biztosító társaságok által alkalmazott aktuáriusok több kísérletet is tesznek annak érdekében, hogy megtalálják azt a megfelelő valószínűségi eloszlást és a hozzáillő modellt, ami majd illeszkedni fog ezen kárszámokra. A legnevezetesebb kárszámeloszlások amit alkalmazni szoktak a binominális, negatív binominális, geometriai valamint a Poisson eloszlás. Ezek közül a Poisson alkalmazása a legelterjettebb.

A számítások általánosságban úgy működnek, hogy veszünk egy biztosított személyt. Ő N db esetet jelent be egy adott periódusban (ez általában 1 év), ami függhet egy vagy több tényezőtől.

Ilyen tényezők vagy más szóval változók lehetnek például a következők:

x_1: az ügyfél férfi vagy nő
x_2: városi zónában vagy külvárosban él
x_3: közepes veszélyességi zónában él
x_4: magas veszélyességi zónában él
x_5: a jogosítványa 4-14 éves
x_6: a jogosítványa 15 éves vagy annál idősebb
x_7: az ügyfél 3-5 éve kliens
x_8: az ügyfél 5 vagy annál több éve kliens
x_9: a sérült 30 éves vagy annál fiatalabb
x_{10}: ha nem tűzesetről van szó
x_{11}: ha anyagi kárról és/vagy tűzesetről van szó

Mint már korábban említettem, ezekre leggyakrabban Poisson eloszlást szoktak alkalmazni, de néha nagy az eltérés a Poisson által becsült, és a valós értékek között. Emiatt előfordulhat olyan, hogy fontos, egyénenkénti tulajdonságok figyelmen kívüli hagyása miatt adódik egy túlszóródás. (Ilyen tulajdonságok például a reflexek különböző működése, agresszivitás a kormány mögött, a drogok befolyása, stb..) Ennek kiküszöbölésére különböző modelleket szoktak alkalmazni, hogy megoldják ezt az illeszkedési problémát. Két ilyen modellcsoport például a zéró-inflált modellek, valamint a Hurdle modellek.

Szakdolgozatom célja az, hogy megismertessem ezt a két modellt, a köztük lévő különbségeket és hasonlóságokat példákon keresztül bemutatva, majd ezekre statisztikai becsléseket adjak.

A szakdolgozat felépítése a következőképpen fog kinézni: Az 1. és 2. fejezetekben bemutatásra kerül majd a zéró-inflált, valamint Hurdle modellek néhány alaptulajdonsága és lehetséges alkalmazásai Jean-Philippe Boucher, Michel Denuit & Montserrat Guillén ([2]) cikke alapján, majd ezekre néhány konkrét példa.

A 3. fejezet az összetett modellekről és a két modell közti különbségekről és hasonlóságokról fog szólni, melyet Mei-Chen Hu, Martina Pavlicova & Edward V. Nunes ([5]) cikke alapján fogok bemutatni.

A 4. fejezetben, miután már megismerkedtünk a 2 modell főbb tulajdonságaival, statisztikai becsléseket végzünk rájuk. Ezt pedig az 5. fejezetben egy összegzés fogja követni.

1. A Zéró-Inflált modellek

1.1. A Zéró-Inflált modellek jellemzése

Nézzük először a Zéró-Inflált modelleket! Ezt a modellt azért dolgozták ki, mert gyakran előfordult, hogy a nevezetes eloszlások - például a Poisson - alulbecsülték a kármentesség valószínűségét. Egy kevert modellt alkalmazunk, ami egy degenerált és egy standard elolszlás keveréke. Vagyis az eloszlásunk a következőképpen néz ki:

$$f_N(n) = \begin{cases} \phi + (1-\phi)g(0) & \text{ha } n = 0 \\ (1-\phi)g(n) & \text{ha } n = 1, 2, \dots \end{cases}$$

ahol g egy nemnegatív egész értékeken értelmezett eloszlás, és ϕ egy paraméter.

De az is előfordulhat, hogy a ϕ paraméter több tényezőtől is függ, és ilyenkor a modellünk így módosul:

$$f_{N_i}(n) = \begin{cases} \phi_i + (1-\phi_i)g_i(0) & \text{ha } n = 0 \\ (1-\phi_i)g_i(n) & \text{ha } n = 1, 2, \dots \end{cases}$$

ahol g_i-k eloszlások (persze ezek egyenlőek is lehetnek) a nemnegatív egész számokon, és $\phi_i = \frac{\exp(\underline{x_i}'\underline{\gamma})}{1+\exp(\underline{x_i}'\underline{\gamma})}$. Itt az $\underline{x_i}$ vektor az i-edik szerződés jellemzőit tartalmazza, a $\underline{\gamma}$ vektorban pedig a paramétereket találhatjuk.

Megvizsgáljuk az eloszlás néhány alaptulajdonságát, mint például a várható értékét, szórását, generátorfüggvényét, valamint ferdeségét és lapultságát. Ehhez definiálnunk kell a következő fogalmakat:

1. Definíció (Lapultság). *Az m várható értékű X valószínűségi változó lapultsága az $\frac{E[(X-m)^4]}{(E[(X-m)^2])^2} - 3$ kifejezés értékével egyenlő, ahol $E[.]$ a várható értéket jelöli. Úgy is megfogalmazhatjuk, hogy a lapultság a negyedik centrális momentum és a variancia négyzetének a hányadosánál pont hárommal kisebb szám.*

Az X valószínűségi változó lapultsága vagy lapultsági mutatója tulajdonképpen azt fogalmazza meg, hogy a valószínűségi változó sűrűségfüggvényének laposssága hogyan viszonyul a normális eloszláséhoz.

2. Definíció (Ferdeség). *Az m várható értékű X valószínűségi változó ferdesége az $\frac{E[(X-m)^3]}{(E[(X-m)^2])^3/2}$ kifejezés értékével egyenlő, ahol $E[.]$ a várható értéket jelöli. Úgy is mondhatjuk, hogy a ferdeség a harmadik centrális momentum és a szórás köbének a hányadosa.*

Az X valószínűségi változó ferdesége vagy ferdeségi együtthatója lényegében azt fogalmazza meg, hogy mennyire nem szimmetrikus a valószínűségi változó eloszlása.

3. Definíció (Generátorfüggvény). *Az X nemnegatív egész értékű valószínűségi változó generátorfüggvénye $G_X(z) = \sum_{k=0}^{+\infty} p_k z^k$. A generátorfüggvényt másképpen a várható értékből is tudunk számolni a következő módon, vagyis $G_X(z) = E(z^X) - el$.*

Mielőtt nekikezdenénk a számolásnak, vezessük be a következőt: Legyen η egy g eloszlású valószínűségi változó.
Ekkor az eloszlás várható értéke a következő lesz:

$$E(N) = \sum_{k=0}^{\infty} kP(N=k) = \sum_{k=0}^{\infty} k(1-\phi)P(\eta=k) = (1-\phi)\sum_{k=0}^{\infty} kP(\eta=k) = (1-\phi)E(\eta) \tag{1}$$

és szórása:

$$D^2(N) = (1-\phi)E(\eta^2) - ((1-\phi)E(\eta))^2 \tag{2}$$

ahol az $E(N^2) = \sum_{k=0}^{\infty} k^2 P(N=k) = (1-\phi)E(\eta^2)$.
Most nézzük az eloszlás lapultságát:

$$\begin{aligned} L(N) &= \frac{E(N^4)}{(D^2(N))^2} - 3 = \frac{(1-\phi)E(\eta^4)}{[(1-\phi)E(\eta^2) - ((1-\phi)E(\eta))^2]^2} - 3 = \\ &= \frac{(1-\phi)E(\eta^4)}{(1-\phi)^2E(\eta^2)^2 + (1-\phi)^4E(\eta)^4 - 2(1-\phi)^3E(\eta)^2E(\eta^2)} - 3 = \\ &= \frac{E(\eta^4)}{(1-\phi)E(\eta^2)^2 + (1-\phi)^3E(\eta)^4 - 2(1-\phi)^2E(\eta)^2E(\eta^2)} - 3 = \\ &= \frac{E(\eta^4)}{[(1-\phi)^{1/2}E(\eta^2) - (1-\phi)^{3/2}E(\eta)^2]^2} - 3. \end{aligned} \tag{3}$$

valamint ferdeségét:

$$F(N) = \frac{E(N^3)}{(D(N))^3} = \frac{(1-\phi)E(\eta^3)}{[(1-\phi)E(\eta^2) - ((1-\phi)E(\eta))^2]^{3/2}} \tag{4}$$

Mivel a generátorfüggvény definícióját már korábban kimondtuk, így most nézzük a számolást is:

$$G_N(z) = \sum_{k=0}^{n} P(N=k)z^k = P(N=0)z^0 + \sum_{k=1}^{n} P(N=k)z^k = \phi + (1-\phi)G_\eta(z) \tag{5}$$

Ennek a modellnek számos esete lehetséges. Ezek közül talán a legismertebbek a következők:

- **Zéró-Infált Poisson (ZIP)**:
  Ahol a g megegyezik a Poisson eloszlással, vagyis $g(n) = e^{-\lambda}\frac{\lambda^n}{n!}$. A Poisson eloszlásnál a várható érték és a szórás megegyezik, vagyis mindkettő λ-val egyenlő.

  Ekkor $n = 0, 1, \ldots$-re az eloszlásunk:

  $$f_N(n) = \begin{cases} \phi + (1-\phi)e^{-\lambda} & \text{ha } n = 0 \\ (1-\phi)g(n) & \text{ha } n = 1, 2, \ldots \end{cases}$$

  lesz.

  Ennek a várható értéke $E(N) = (1-\phi)\lambda$, és a szórása

  $D^2(N) = \lambda(1-\phi)\left(1 + \lambda - \lambda(1-\phi)\right)$-vel lesz egyenlő.

- **Zéró-Inflált Negatív Binominális (ZI-NBk)**:

  Itt az eloszlásunk a következőképpen néz ki:

$$f_N(n) = \begin{cases} \phi + (1-\phi)p^k & \text{ha } n = 0 \\ (1-\phi)\binom{n+k-1}{n}p^k(1-p)^n & \text{ha } n = 1, 2, \dots \end{cases}$$

  ahol $p = \frac{k}{k+\lambda}$.

  Ennek a várható értéke $E(N) = (1-\phi)\lambda$, és a szórása

  $D^2(N) = (1-\phi)\lambda + (1-\phi)\frac{\lambda^2(1+k)}{k} - (1-\phi)^2(\lambda)^2$.

- **Általánosított Zéró-Inflált Poisson (ZIGP)**:

  Ebben az esetben az eloszlásunk a következő:

$$f_N(n) = \begin{cases} \phi + (1-\phi)e^{-\lambda} & \text{ha } n = 0 \\ (1-\phi)\frac{(1+\alpha n)^{n-1}}{n!}\frac{(\lambda e^{-\alpha\lambda})^n}{e^{\lambda}} & \text{ha } n = 1, 2, \dots \end{cases}$$

  Ezt az eloszlást szokták úgy is felírni, hogy

$$f_N(n) = \begin{cases} \phi + (1-\phi)e^{-\theta} & \text{ha } n = 0 \\ (1-\phi)\frac{(1+\alpha n)^{n-1}}{n!}\theta^n e^{-\theta(1+\alpha n)} & \text{ha } n = 1, 2, \dots \end{cases}$$

  Ekkor ha a θ helyére λ-t írunk, visszakapjuk az első eloszlást. Erről biztosan tudjuk, hogy ez eloszlás, Felix Famoye ([10]) cikkéből. Ekkor észrevehetjük, hogy az $n = 0$ esetre a ZIP és a $ZIGP$ modellek megegyeznek. A várható érték és variancia eltér, vagyis a várható érték

  $E(N) = (1-\phi)\frac{\lambda}{1-\alpha\lambda}$ és a szórás

  $D^2(N) = (1-\phi)\frac{\lambda}{1-\alpha\lambda}[\frac{1}{(1-\alpha\lambda)^2} + \frac{\lambda}{1-\alpha\lambda} - (1-\phi)\frac{\lambda}{1-\alpha\lambda}]$-vel egyenlő.

- **Zéró-Inflált Dupla Poisson**:

  Itt az eloszlásunk:

$$f_N(n) = \begin{cases} \phi + (1-\phi)(\theta^{1/2}e^{-\theta\lambda}) & \text{ha } n = 0 \\ (1-\phi)(\theta^{1/2}e^{-\theta\lambda})(\frac{e^{-n}n^n}{n!})(\frac{e^{\lambda}}{n})^{\theta n} & \text{ha } n = 1, 2, \dots \end{cases}$$

  Ekkor a várható értékünk $E(N) = (1-\phi)\frac{\lambda}{\theta}$ és szórásunk

  $D^2(N) = (1-\phi)\frac{\lambda}{\theta^2} + \lambda(1-\phi)\frac{\lambda}{\theta} - (1-\phi)^2\frac{\lambda^2}{\theta^2}$.

  Ez az eloszlás már nem annyira ismert, de ennek ellenére nagyon érdekes.

Most számoljuk ki az eloszlások loglikelihood függvényét. A maximum likelihood módszer a matematikai statisztika egyik leggyakrabban használt becslési eljárása mérési eredmények, minták kiértékelésére. A célja, hogy adott mérési értékekhez, az ismeretlen paramétereknek olyan becslését adja meg, amely mellett az adott érték a legnagyobb valószínűséggel következik be. Az eljárás a likelihood függvény maximalizálásával történik. A számítások egyszerűsítése céljából a gyakorlatban nem az eredeti likelihood-függvényt

használjuk, hanem annak a természetes alapú logaritmusát. Tehát a függvényeink a következők lesznek:

$$\begin{aligned}
\ln L_{ZIP(n_1,\ldots,n_n)} &= \sum_{i=1}^{n}\left[\chi_{\{n_i=0\}}\ln\left(\phi+(1-\phi)e^{-\lambda}\right)+\chi_{\{n_i>0\}}\ln\left((1-\phi)\frac{e^{-\lambda}\lambda^{n_i}}{n_i!}\right)\right]= \\
&= \sum_{i:n_i=0}\ln\left(\phi+(1-\phi)e^{-\lambda}\right)+\sum_{i:n_i>0}\left(\ln(1-\phi)+n_i\ln\lambda-\lambda-\ln n_i!\right)= \\
&= -n\ln(1+e^a)+\sum_{i:n_i=0}\ln(e^a+e^{-\lambda})+\sum_{i:n_i>0}\left(n_i\ln\lambda-\lambda-\ln n_i!\right)
\end{aligned}$$

ahol az a megegyezik $logit(\phi)$-vel. ($logit(\phi)=\ln(\frac{\phi}{1-\phi})$)

$$\begin{aligned}
\ln L_{ZINB(n_1,\ldots,n_n)} &= \sum_{i:n_i=0}\ln(\phi+(1-\phi)p^k)+ \\
&\quad+\sum_{i:n_i>0}\ln((1-\phi)\binom{n_i+k-1}{n_i}p^k(1-p)^{n_i})= \\
&= \sum_{i:n_i=0}\ln\left(\phi+(1-\phi)p^k\right)+ \\
&\quad+\sum_{i:n_i>0}\left(\ln(1-\phi)+\ln\frac{\Gamma(n_i+k)}{\Gamma(n_i+1)\Gamma(k)}+k\ln(p)+n_i\ln(1-p)\right)
\end{aligned}$$

$$\begin{aligned}
\ln L_{ZIGP(n_1,\ldots,n_n)} &= \sum_{i:n_i=0}\ln\left(\phi+(1-\phi)e^{-\lambda}\right)+\sum_{i:n_i>0}\ln\left((1-\phi)\frac{(1+\alpha n_i)^{n_i-1}}{n_i!}\frac{(\lambda e^{-\alpha\lambda})^{n_i}}{e^{\lambda}}\right)= \\
&= \sum_{i:n_i=0}\ln\left(\phi+(1-\phi)e^{-\lambda}\right)+ \\
&+ \sum_{i:n_i>0}\ln(1-\phi)+(n_i-1)\ln(1+\alpha n_i)-\ln(n_i!)+n_i\ln(\lambda)-\lambda(1+n_i\alpha)
\end{aligned}$$

és végül az

$$\begin{aligned}
\ln L_{ZIDP(n_1,\ldots,n_n)} &= \sum_{i:n_i=0}\ln\left(\phi+(1-\phi)(\theta^{1/2}e^{-\theta\lambda})\right)+ \\
&\quad+\sum_{i:n_i>0}\ln\left((1-\phi)(\theta^{1/2}e^{-\theta\lambda})(\frac{e^{-n_i}n_i^{n_i}}{n_i!})(\frac{e^{\lambda}}{n_i})^{\theta n_i}\right)
\end{aligned}$$

1.2. A modellek alkalmazása

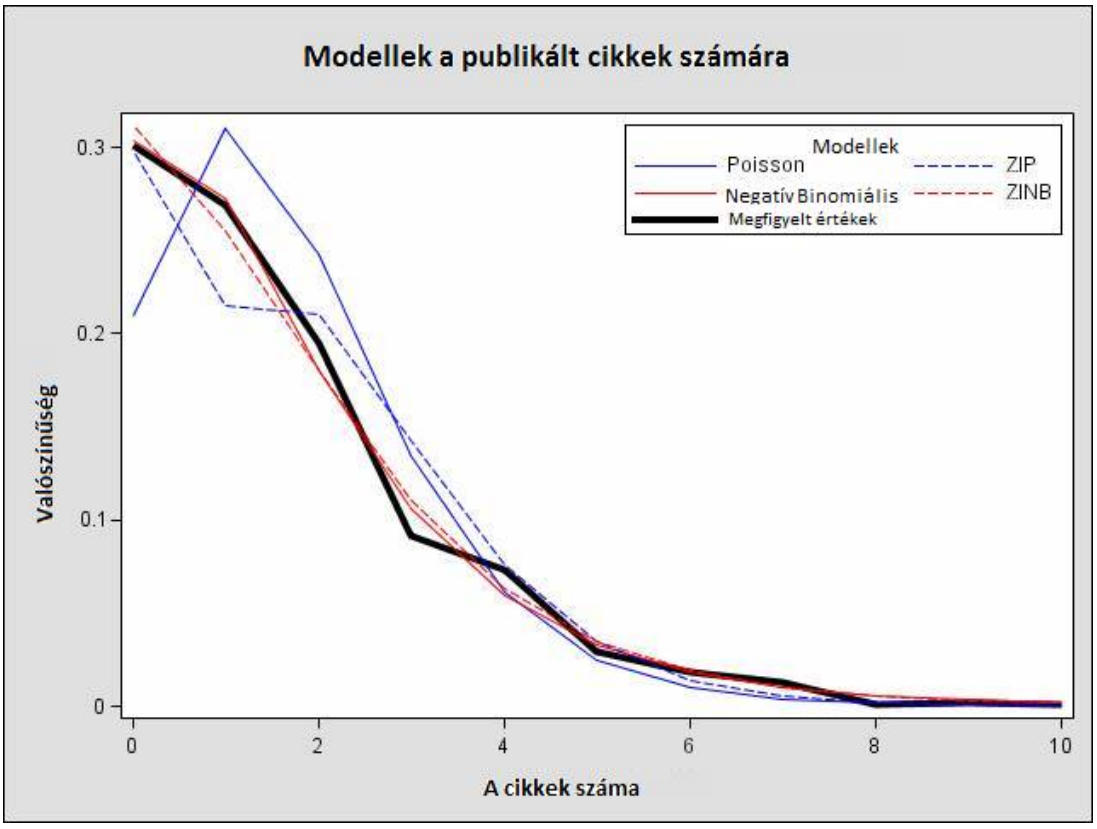
Ezt a modellt gyakran alkalmazzák szerződések kárszámainak modellezésére, különösen akkor, ha a kármentesség valószínűsége "túl nagy". A legtöbb biztosító társaságnak, főleg Európában, létezik egy tapasztalati osztályozó rendszere, mint például a bónusz-málusz rendszer. Ez a rendszer úgy működik, hogy a biztosítók igyekeznek a balesetmentes vezetést jutalmazni, vagyis megnézik egy meghatározott időszakban az adott személy

által bejelentett balesetek számát. Ez alapján határozzák majd meg, hogy a következő évben milyen kategóriába lesz besorolható a vezető. Így ha az adott személynek kevés balesete van, akkor a következő évben jutalmat kap, vagyis egy magasabb kategóriába sorolódik. Így ez egy úgynevezett "bónuszéhséget" okoz, ami azt jelenti, hogy ha a biztosítottak esetleg nem jelentenek be minden ügyet (vagy biztonságosabban vezetnek), akkor a jövőbeni jutalmuk magasabb lesz, mint a sérülésből járó hasznuk. Ezt a fajta viselkedést írja le jól a zéró-inflált modell.

További érdekes alkalmazási területek még a következők:

1986-ban Mullahyban arra használták ezeket a modelleket, hogy felmérjék az ottani italfogyasztási szokásokat.
1987-ben Lawlessben amiatt, hogy modellezni tudják a hajótörések gyakoriságát.
1992-ben Lambertben arra, hogy modellezzék a sérült alkatrészek előfordulását egy ipari előállítás során, míg 1995-ben van den Broek a HIV-vel fertőzött férfiak számának felbecsülésére.

Ezek és még sok más érdekes alkalmazási területe lehetséges ennek a modellnek, amiből néhány további példát találhatunk még Joseph Ngatchou-Wandji és Christophe Parisa ([6]) cikkében.



1. ábra. Egy példa a zéró-inflált eloszlásra

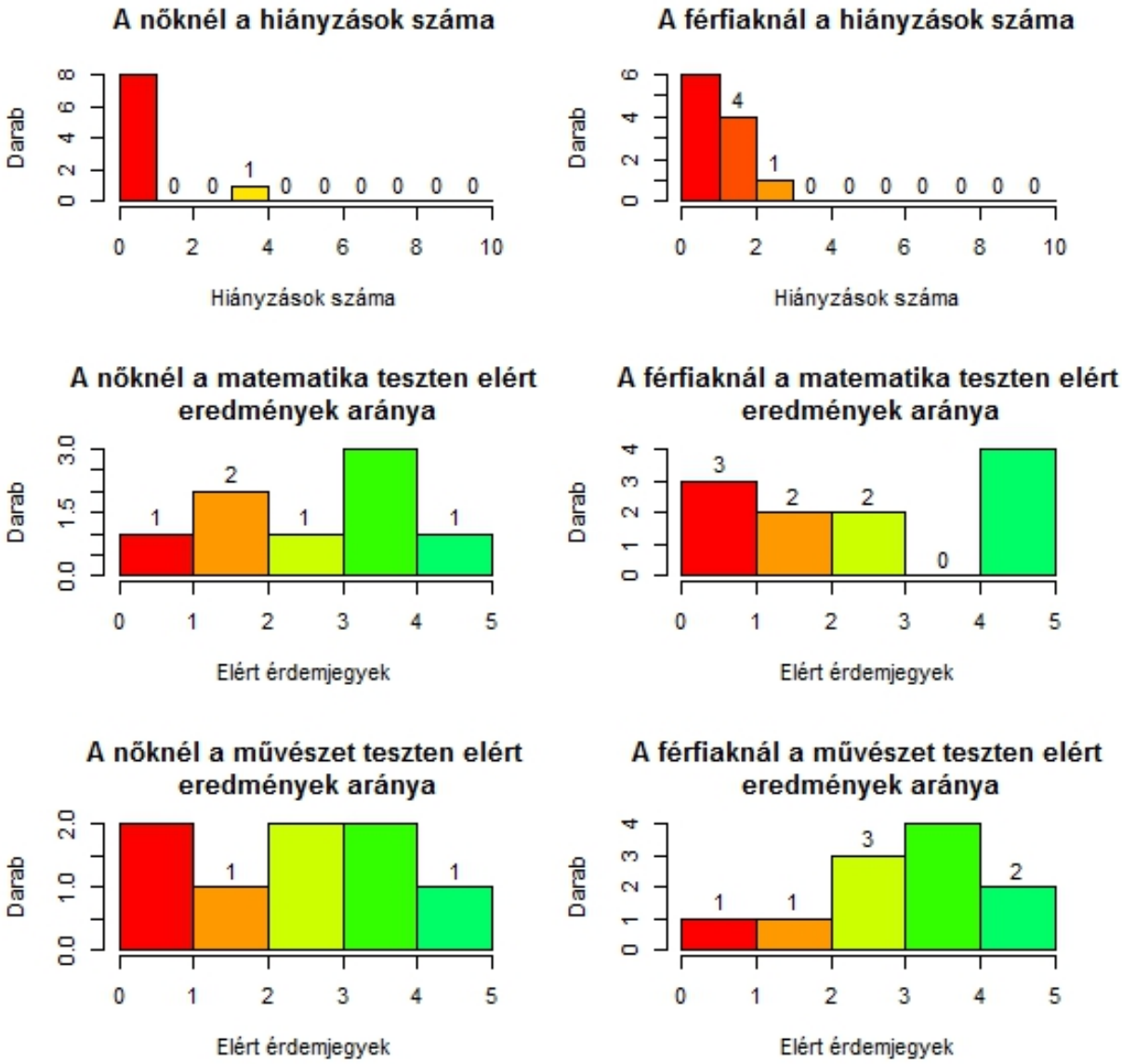
1.3. Néhány példa

1. Példa. *Iskolai adminisztrátorok 2 különböző középiskolában is azt tanulmányozták, hogy egy adott szemeter alatt, milyen a diákok órai látogatottsági viselkedése. A viselkedést az alapján mérték, hogy hány hiányzó nap volt (ez alatt azt értjük, hogy azon napok száma, amikor volt legalább egy hiányzó diák), valamint milyen volt az adott diák neme és 2 tárgyból elért eredménye (matematika és művészet). Azt vették észre, hogy a tanulók többségének nem volt hiányzása.*

Néhány konkrét esetre:

	Nem (F=1,N=0)	**Matematika jegy (1-5-ig)**	**Művészet jegy (1-5-ig)**	**Hiányzások száma(n)**
1,	0	3	4	0
2,	1	1	5	0
3,	1	2	3	1
4,	0	4	5	0
5,	0	1	1	3
6,	1	5	5	0
7,	1	2	4	0
8,	1	5	4	1
9,	0	2	1	0
10,	1	5	4	2
11,	1	3	4	0
12,	0	4	2	0
13,	0	4	4	0
14,	0	2	3	0
15,	1	1	1	1
16,	0	3	3	0
17,	0	5	4	0
18,	1	3	3	0
19,	1	1	2	1
20,	1	5	3	0

1. táblázat. Néhány kitalált eset a tanulmányra



2. ábra. Néhány diagramm

2. Példa. *Az állami vadvilágot tanulmányozó biológusok azt akarták modellezni, hogy hány halat fogtak ki a horgászok egy állami parkból. Ezt úgy vizsgálták, hogy megkérdezték a látogatókat a következőkről: volt-e csalijuk, hányan voltak az adott csoportban, ebből hány gyermek volt, valamint, hogy így hány halat fogott a csoport összesen. A vizsgálatból az derült ki, hogy számos látogató aki horgászott, mégsem fogott egy halat sem, ezért túl sok nulla jelent meg az adatokban.*

Most nézzünk meg ezt konkrét számokkal:

	Csali	**Sátor**	**Emberek száma**	**Gyerekek száma**	**Fogott halak száma (n)**
1,	0	0	1	0	0
2,	1	1	1	0	0
3,	1	0	1	0	0
4,	1	1	2	1	0
5,	1	0	1	0	1
6,	1	1	4	2	0
7,	1	0	3	1	0
8,	1	0	4	3	0
9,	0	1	3	2	0
10,	1	1	1	0	1
11,	1	0	4	1	0
12,	1	1	3	2	0
13,	0	0	3	0	1
14,	1	0	3	0	2
15,	1	1	1	0	0
16,	1	1	1	0	1
17,	1	0	4	1	0
18,	1	1	3	2	0
19,	1	1	2	1	1
20,	1	0	3	1	0

2. táblázat. Különböző esetek a felmérésben

Ekkor legyenek az x változói a válaszokban megadott értékek, vagyis pl: $x_1 = (0, 0, 1, 0, 0)$ és ebből ϕ_i a következő képpen számolható ki: $\phi_i = \frac{\exp(x_i^{'}\gamma)}{1+\exp(x_i^{'}\gamma)}$

3. ábra. A fogott halak száma

2. Hurdle modellek

2.1. A Hurdle modellek jellemzése

Ez az eloszlás azért lesz érdekes számunkra, mert arra épül, hogy a biztosítottak nagyobb része (99,5%) évente legfeljebb csak 2 esetet jelent be a biztosító felé. Tehát itt 2 folyamat lesz érdekes a számunkra. Vagyis vegyük azt a legegyszerűbb hurdle modellt, amelyik beállítja a hurdle-t a nullába. Ekkor formálisan, adott 2 eloszlás: f_1 és f_2, továbbá a hurdle modell eloszlása:

$$f_N(n) = \begin{cases} q & \text{ha } n = 0 \\ \frac{1-q}{1-f_2(0)} f_2(n) = \Phi f_2(n) & \text{ha } n = 1, 2, \dots \end{cases}$$

ahol $\Phi = \frac{1-q}{1-f_2(0)}$ és $f_1(0) = q$. Ha $f_1 = f_2$, akkor csak f-el szokás jelölni a két eloszlást. Itt is vezessük be a következőt: Legyen $\eta \sim f_2$ eloszlású valószínűségi változó.
Ekkor a modell várható értéke a következőképpen alakul:

$$E(N) = \sum_k kP(N = k) = \sum_k k\Phi P(\eta = k) = \Phi \sum_k kP(\eta = k) = \Phi E(\eta), \tag{6}$$

szórása

$$D^2(N) = \Phi E(\eta^2) - (\Phi E(\eta))^2, \tag{7}$$

mivel $E(N^2) = \Phi E(\eta^2)$.

Tehát modellünk alul- vagy túlszóródó lesz, az f_1 és f_2 eloszlásoktól függően.

4. Definíció (Túlszóródás). *A túlszóródás vagy más néven overdispersion azt jelenti, hogy a megfigyelt variancia nagyobb, mint az adott modell várható értéke.*

Most számoljuk ki az eloszlás néhány tulajdonságát, mint például a lapultságát és ferdeségét. A számolás hasonlóan megy végbe, mint a Zéró-Inflált esetben (**5**. oldal), vagyis:

$$\begin{aligned} L(N) &= \frac{E(N^4)}{(D^2(N))^2} - 3 = \frac{(\Phi)E(\eta^4)}{[(\Phi)E(\eta^2) - ((\Phi)E(\eta))^2]^2} - 3 = \\ &= \frac{E(\eta^4)}{(\Phi)E(\eta^2)^2 + (\Phi)^3E(\eta)^4 - 2(\Phi)^2E(\eta)^2E(\eta^2)} - 3 = \\ &= \frac{E(\eta^4)}{[(\Phi)^{1/2}E(\eta^2) - (\Phi)^{3/2}E(\eta)^2]^2} - 3. \end{aligned} \tag{8}$$

és a ferdeség:

$$F(N) = \frac{E(N_i^3)}{(D(N_i))^3} = \frac{(\Phi)E(\eta^3)}{[(\Phi)E(\eta^2) - ((\Phi)E(\eta))^2]^{3/2}} \tag{9}$$

A generátorfüggvénynél is hasonlóan megy a számolás:

$$\begin{aligned} G_N(z) &= \sum_{k=0}^{n} P(N = k)z^k = P(N = 0)z^0 + \sum_{k=1}^{n} P(N = k)z^k = q + \sum_{k=1}^{n} \Phi\eta = \\ &= q + \Phi G_\eta(z) - \Phi \end{aligned} \tag{10}$$

Továbbá a log-likelihood függvénye a hurdle modellnek a következő:

$$l = \ln L = \sum_{i=1}^{n} I_{(n_i=0)} \ln(q) + \sum_{i=1}^{n} I_{(n_i>0)} \left[\ln(1-q) + \ln(f_2(n_i)) - \ln(1-f_2(0))\right] =$$

$$= \sum_{i=1}^{n} \left[I_{(n_i=0)} \ln(q) + I_{(n_i>0)} \ln(1-q)\right] + \sum_{i=1}^{n} I_{(n_i>0)} \left[\ln(f_2(n_i)) - \ln(1-f_2(0))\right]$$

2.2. A modellek alkalmazása

A hurdle modelleket nagyon gyakran az egészségügyi problémákkal kötik össze. Az ilyen modellek alkalmazása 2 dologtól függ: a változók jó megválasztásától, és attól, hogy melyik egészségügyi ellátót vesszük. Ez a modell azért illeszkedik olyan jól a zéró-inflált modellekre, mert itt is számos sérült vonakodik jelenteni a balesetét. Ezért könnyű azt is elhinni, hogy a sérültek viselkedése megváltozik addigra, mire jelentik az esetet. Ez ösztönözte azt, hogy két eljárást kell használni a teljes adatok elemzésére.

Egyszóval a Hurdle modellek azért különlegesek, mert a becslések 2 lépcsősek, vagyis van egy nulla és egy pozitív elemekre bontása az adatoknak. Ezért az egész modell ennek a két résznek a jó megválasztásától függ.

Ahogy láthatjuk a hurdle modell nagyobb részét a 0 rész teszi ki, és csak kis részét a pozitív rész, aminél a változókból olyanok derülhetnek ki, mint például a baleseti zóna területi elhelyezkedése, vagy a vezetési tapasztalatok. Emiatt a mostani eredmények azt mutatják, hogy az új biztosítottak rosszabb kárigény tapasztalatokat jelentenek, mint a régebbiek, amikor végre bejelentik a kárigényüket.

2.3. Néhány példa

3. Példa. *Egy kísérletben a mozik látogatottságát akarták modellezni. A kísérlet során az emberek először azt dönthették el, hogy egyáltalán akarnak-e moziba menni. Ezt főként az befolyásolta, hogy volt-e számukra érdekes film. Ha volt, akkor azt kellett megválaszolniuk, hogy havi szinten mennyit költenének mozira.*

A mérést az alapján végezték, hogy megnézték a heti ledolgozandó munkaórák számát, egy változóban tárolták azt, hogy hétvégén dolgozik-e az illető, és, hogy esetleg van-e újszülött a családban.

A felmérés azt mutatta ki, hogy azok az emberek akiknek újszülött gyerekük van, vagy hétvégente dolgoznak vagy esetleg csak magasabb óraszámban, azok kevésbé tudnak arról dönteni, hogy elmenjenek-e egy mozifilmet megnézni, vagy sem.

4. Példa. *Egy másik mérési lehetőség az előző esetre az, hogy azt nézik meg, hogy a tinédzserek mennyit költenek havi szinten a mozira, vagy, hogy éppen párkapcsolatban van-e az illető az adott időszakban, és van-e (és ha igen, mennyi) olyan 6-10 éves gyerek, akinek még szülői felügyelet kell.*

Itt a felmérésből az derült ki, hogy ha az ember éppen randevúra megy, vagy tinédzser korú, vagy éppen fiatal gyerekkel megy, hajlamosabb többet költeni egy filmért.

5. Példa. *A Zéró-Inflált esetben látott második példát (**11**. oldalon) is ki lehet számolni Negatív Binominális Hurdle eloszlással.*

3. Összetett modellek

3.1. Az Összetett modellek

Összetett modellnek nevezzük azt, amikor a $Z = \sum_{i=1}^{N} X_i$, ahol az X_i -k egymástól független azonos eloszlású nemnegatív egész értékű valószínűségi változók. Továbbá az N és az X_i-k is függetlenek egymástól. Klugman, Panjer & Willmot (2004) számos példát írt le az összetett modellekről. Az egyik ilyen:

6. Példa. *Legyen N Poisson eloszlású λ paraméterrel és az X_i-k pedig logaritmikusak θ paraméterrel. Ekkor Z Negatív Binominális eloszlású.*

A logaritmikus eloszlás a következőképpen néz ki: Legyen X logaritmikus eloszlású p paraméterrel, ahol $p \in (0,1)$, és az

$$f(n) = \frac{1}{-\ln(1-p)} \frac{p^n}{n}, n \in N_+.$$

Ekkor tudjuk a következőt:

$$E\left(X^{(k)}\right) = \frac{(k-1)!}{-ln(1-p)} \left(\frac{p}{1-p}\right)^k, k \in N_+.$$

Ugyanis:

$$\begin{aligned} E\left(X^{(k)}\right) &= \sum_{n=1}^{\infty} n^{(k)} \frac{1}{-\ln(1-p)} \frac{p^n}{n} = \frac{p^k}{-\ln(1-p)} \sum_{n=k}^{\infty} n^{(k)} \frac{p^{n-k}}{n} = \\ &= \frac{p^k}{-\ln(1-p)} \sum_{n=k}^{\infty} \frac{d^k}{dp^k} \frac{p^n}{n} = \frac{p^k}{-\ln(1-p)} \frac{d^k}{dp^k} \sum_{n=1}^{\infty} \frac{p^n}{n} = \\ &= \frac{p^k}{-\ln(1-p)} \frac{d^k}{dp^k}(-\ln(1-p)) = \frac{p^k}{-\ln(1-p)}(k-1)!(1-p)^{-k} \end{aligned} \quad (11)$$

Ekkor a logaritmikus eloszlás várható értéke és szórása úgy kapható, hogy:

$$E(X) = \frac{1}{-\ln(1-p)} \frac{p}{1-p}$$

és

$$D^2(X) = \frac{1}{-\ln(1-p)} \frac{p}{(1-p)^2} \left[1 - \frac{p}{-\ln(1-p)}\right].$$

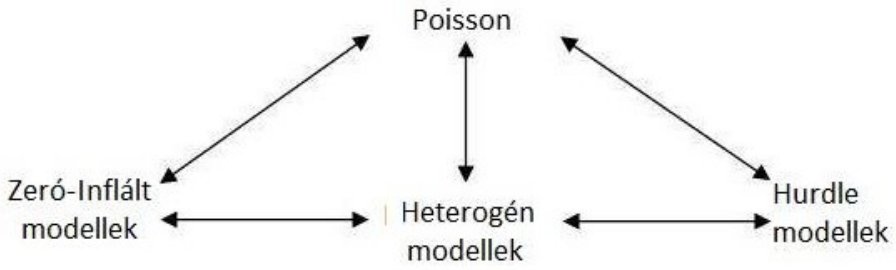
3.2. Az Összetett modellek alkalmazása

Santos Silva & Windmeijer (2001) arra használta a Negatív Binominális eloszlást, hogy modellezni tudják az orvost látogatók számát, ahol N a különböző betegségek számát jelöli, és X az adott betegség miatti látogatottság számát. Vagy az aktuáriusok arra, hogy modellezzék a sérült emberek számát egy adott balesetben.

Egy további alkalmazási lehetőség lehet az is, amikor viszonylag rövid idő alatt számos baleset történik egyszerre, és a biztosított csak egy esetet jelent be az összes kárra. Ennek az az oka, hogy igyekszik kerülni a büntetését, ami egyben a bónusz-málusz rendszer átverését jelenti. Ennek egy másik szemlélete lehet az is, hogy azért jelent be az adott illető kevés esetet, mert a vezető figyelmetlen volt és a saját hibájából adódóan kárt okozott. Viszont azt is tudja, hogy a biztosító cég akkor javíttatja meg gyorsabban a járművet, ha ez egy alkalommal történik meg.

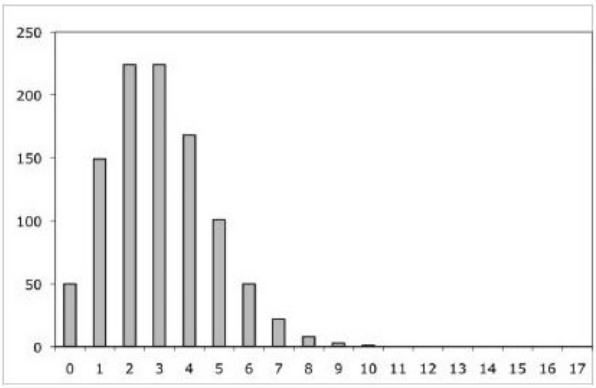
4. A két modell összehasonlítása és kapcsolata

Mind a Zéró-Inflátak, mind a Hurdle modellek azért jöttek létre, hogy kezelni tudják (elsősorban Negatív binominális, vagy Poisson eloszlással) a nagy mennyiségű 0 adatmennyiséget, ami (mint már korábban említettem) egyes esetekben túlszóródással is járhat. Habár mind a kettőnél nagy mennyiségű 0 jelenik meg az adatokban, egy fontos különbség mégis van a között, ahogyan elemzik és megoldják azt.



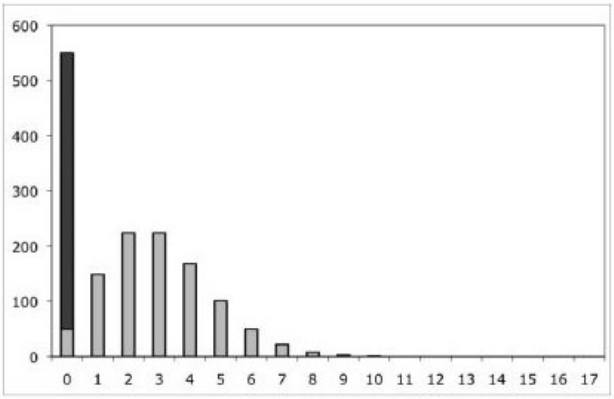
4. ábra. Kapcsolat a modellek között

Mivel általánosságban Poisson eloszlást szoktak alkalmazni az adatokra ezért induljunk ki ebből az ábrából:



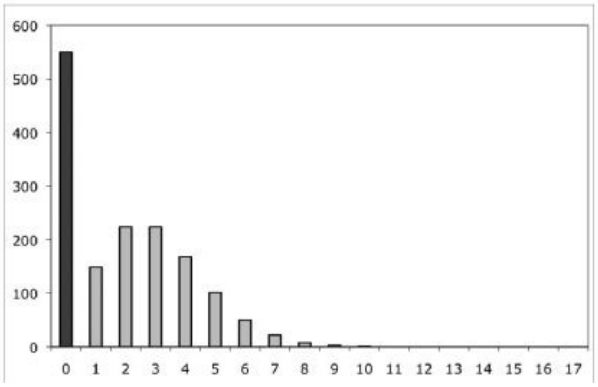
5. ábra. Egy Poisson modell

Ezzel ellentétben a Zéró-inflált esetben a 0 megfigyeléseknél 2 különböző származást veszünk: egy szerkezetit és egy mintavételest.



6. ábra. Zéró-Inflált modell az adatokra

Az előző ábra megmutatja, hogy hogyan is néz ki egy ilyen felosztás (itt $n = 1500$, és 500 szerkezeti, valamint 50 db mintavételes 0 szerepel rajta), ahol a sötétszürke színű jelöli a szerkezeti-, míg a világosszürke a mintavételes 0-ákat. A mintavételes 0-ák alapja általában a Poisson (vagy Negatív Binominális) eloszlás, ami azt feltételezi, hogy ezek véletlen megfigyelések. A Zéró-Inflált modellek azt feltételezik, hogy számos 0 megfigyelés speciális szerkezettel rendelkezik az adatokból. Vagyis vegyük példának azt a tanulmányt, ahol a magas rizikós szexuális viselkedést tanulmányozták. Ekkor néhány résztvevő 0 értéket ad, mivel nincs éppen szexuális partnerük. Ezek lesznek a szerkezeti 0-ák, mivel ők nem befolyásolják a védekezésmentes szexuális viselkedés mérését. Néhány résztvevő ugyan rendelkezik partnerrel, de mégis 0-át adnak, mivel valamivel elkerülik ezt a veszélyes viselkedést. Az ő rizikós viselkedésüket fogjuk tehát leírni Poisson vagy Negatív binominális eloszlással, amely tartalmazza a nulla esetet (mintavételes), és a nemnulla esetet is.



7. ábra. Poisson Hurdle modell az adatokra

Ezzel ellentétben a hurdle modell szerint minden 0 adat szerkezeti származású. Ellenben itt a pozitív adatoknak lesz mintavételes eredete és ezek eloszlása csonkított Poissont

vagy csonkított Negatív-binomiális eloszlást fog követni. A következő ábrán az n itt is egyenlő 1500-al, valamint 550 db szerkezeti és 0 db mintavételes 0 szerepel.

Vegyük a Hurdlere példának azt a tanulmányt, ahol a dohányzási szokásokat vizsgálják. Itt csak a nem dohányzók fognak 0 értéket adni, vagyis ez azt jelenti, hogy ők nem szívnak el egy szál cigarettát sem. Emiatt itt a dohányzók adják majd a pozitív részt, vagyis, hogy hány szál cigarettát szívtak el az utolsó 1 hónapban.

Egyszóval a különbség itt jól látható a szerkezeti és mintavételes 0-ák miatt, mivel így a két modell lehet, hogy 2 különböző eredményt fog adni két különböző értékeléssel, ugyanis mások lehetnek a kezdeti feltételek.

5. Becslések a modellekre

A fejezetben a korábban bemutatott modellek esetében határozzuk meg a paraméterek maximum likelihood és momentum módszeres becslését. Feltételezzük, hogy megfigyeléseink független, azonos eloszlásúak. A szerződések jellemzőitől való függést itt nem vizsgáljuk.

1. **Maximum likelihood becslés**

   Zeró-Inflált Poissonra:

   Mint már korábban felírtuk (**7**. oldal) az eloszlás log-likelihood függvénye

$$l = \sum_{i:n_i=0} \ln\left(\phi + (1-\phi)e^{-\lambda}\right) + \sum_{i:n_i>0} \left(\ln(1-\phi) + n_i \ln\lambda - \lambda - \ln n_i!\right)$$

   Erre írjuk most fel a likelihood egyenleteket vagyis, hogy a $\partial_\lambda l$ és $\partial_\phi l$ mikor lesz 0:

$$0 = \partial_\phi l = \frac{1}{\phi + (1-\phi)e^{-\lambda}}(1-e^{-\lambda})|\{n_i = 0\}| + \frac{(-1)}{1-\phi}|\{n_i > 0\}|$$

$$0 = \partial_\lambda l = \frac{(-1)}{\phi + (1-\phi)e^{-\lambda}}(1-\phi)e^{-\lambda}|\{n_i = 0\}| + \frac{1}{\lambda}\sum_{i:n_i>0} n_i - |\{n_i > 0\}|$$

   Ekkor vezessük be a következő jelöléseket: k a megfigyelések száma, m legyen egyenlő $|\{n_i = 0\}|$-vel, és a $\sum_{\{i:n_i>0\}} n_i = d$-vel.

   Vagyis az első és második egyenletből azt kapjuk, hogy:

$$\frac{(1-e^{-\lambda})m}{\phi + (1-\phi)e^{-\lambda}} = \frac{k-m}{1-\phi} \tag{12}$$

$$\frac{-(1-\phi)e^{-\lambda}m}{\phi + (1-\phi)e^{-\lambda}} + \frac{1}{\lambda}d = (k-m). \tag{13}$$

   Ekkor ha a 2. egyenletből ki tudjuk fejezni ϕ-t, akkor azt az első egyenletbe beírva, kapunk λ-ra egy egyenletet.

$$\frac{-(1-\phi)e^{-\lambda}m}{\phi + (1-\phi)e^{-\lambda}} + \frac{1}{\lambda}d = (k-m)$$

$$(\phi-1)e^{-\lambda}m\lambda + (\phi + (1-\phi)e^{-\lambda})d = (\phi + (1-\phi)e^{-\lambda})\lambda(k-m)$$

$$\phi e^{-\lambda}m\lambda - e^{-\lambda}m\lambda + \phi d + e^{-\lambda}d - \phi e^{-\lambda}d = \phi\lambda(k-m) + e^{-\lambda}\lambda(k-m) - \phi e^{-\lambda}(k-m)$$

$$e^{-\lambda}\lambda k - e^{-\lambda}\lambda m - e^{-\lambda}d + e^{-\lambda}\lambda m =$$

$$= -\phi e^{-\lambda}m\lambda - \phi d + \phi e^{-\lambda}d + \phi\lambda k - \phi\lambda m - \phi e^{-\lambda}\lambda k + \phi e^{-\lambda}\lambda m$$

$$
\begin{aligned}
e^{-\lambda}\lambda k - e^{-\lambda}d &= -\phi d + \phi e^{-\lambda}d + \phi\lambda k - \phi\lambda m - \phi e^{-\lambda}\lambda k \\
e^{-\lambda}(\lambda k - d) &= \phi(-d + e^{-\lambda}d + \lambda k - \lambda m - e^{-\lambda}\lambda k) \\
e^{-\lambda}(\lambda k - d) &= \phi(-d + e^{-\lambda}(d - \lambda k) + \lambda(k - m))
\end{aligned}
$$

$$
\Rightarrow \phi = \frac{e^{-\lambda}(\lambda k - d)}{e^{-\lambda}(d - \lambda k) + \lambda(k - m) - d} \tag{14}
$$

és ebből a λ-ra az egyenlet (ahol ϕ egyenlő az előző sorban kiszámolt értékkel):

$$
\frac{(1 - e^{-\lambda})}{\frac{\phi}{1-\phi} + e^{-\lambda}} = \frac{(k - m)}{m} \tag{15}
$$

Hurdle Poissonra:

A korábbiaknak megfelelően a log-likelihood függvény:

$$
l = \sum_{i:n_i=0} \ln(q) + \sum_{i:n_i>0} \left[\ln(1 - q) - \ln(1 - e^{-\lambda}) - \lambda + n\ln\lambda - \ln(n!)\right] \tag{16}
$$

Ekkor az előzőhöz hasonlóan

$$
0 = \partial_\lambda l = \left(\frac{1}{1 - e^{-\lambda}} e^{-\lambda}(-1) - 1\right) |\{n_i > 0\}| + \frac{1}{\lambda} \sum_{\{i:n_i>0\}} n_i \tag{17}
$$

és

$$
0 = \partial_q l = \frac{1}{q}|\{n_i = 0\}| + \frac{(-1)}{1 - q}|\{n_i > 0\}| \tag{18}
$$

Vagyis azt kaptuk, hogy

$$
\frac{-e^{-\lambda}(k - m)}{1 - e^{-\lambda}} + \frac{1}{\lambda} d = (k - m). \tag{19}
$$

$$
\frac{1}{1 - q}(k - m) = \frac{1}{q} m \tag{20}
$$

Ekkor a második egyenletből kiszámolhatjuk q-t.

$$
q(k - m) = (1 - q)m \Rightarrow qk - qm = m - qm \Rightarrow qk = m
$$

Tehát $q = \frac{m}{k}$ és az első egyenletből a λ-ra azt kapjuk, hogy:

$$
\frac{e^{-\lambda}\lambda}{1 - e^{-\lambda}} + \lambda = \frac{d}{(k - m)} \tag{21}
$$

2. **Momentum módszer**

A momentum módszer a maximum likelihood módszer mellett a matematikai statisztikában az egyik leggyakrabban használt becslés a minta ismeretlen paramétereire.

A becslést úgy végezzük, hogy adott egy X valószínűségi változó és hozzá egy eloszlásfüggvény $F(X, \theta)$, valamint X-re a momentumok: $\mu_i = EX^i$, $i > 1$. Ha a θ paraméter megadható a μ_i, $1 \leq i \leq m$ momentumok $\theta = h(\mu_1, \ldots, \mu_m)$ függvényeként, akkor a θ paraméter becsléseként vehetjük a

$$\hat{\theta}_k = h(\hat{\mu}_1, \ldots, \hat{\mu}_m)$$

értéket, ahol

$$\hat{\mu}_i = \frac{1}{k}(X_1^i + \ldots + X_k^i), 1 \leq i \leq m$$

jelöli az i-edik tapasztalati momentumot. A következőkben nézzük meg a becsléseket:

Zeró-Inflált Poissonra:
Az 1. egyenlet:

$$E(N) = (1 - \phi)\lambda = \overline{X}, \tag{22}$$

ahol az $\overline{X} = \frac{1}{k}\sum_{i=1}^{k} X_i$. Vagyis

$$\hat{\lambda} = \frac{\overline{X}}{(1 - \phi)} \tag{23}$$

A 2. egyenlet:

$$E(N^2) = (1 - \phi)E(\eta^2) = (1 - \phi)(\lambda + \lambda^2) = \hat{\mu}_2, \tag{24}$$

ahol a $\hat{\mu}_2 = \frac{1}{k}\sum_{i=1}^{k} X_i^2$. Vagyis ekkor azt kaptuk, hogy:

$$(1 - \phi)(\lambda + \lambda^2) = \hat{\mu}_2 \Rightarrow (1 - \phi)\left(\frac{\overline{X}}{(1 - \phi)} + \frac{\overline{X}^2}{(1 - \phi)^2}\right) = \hat{\mu}_2 \Rightarrow \overline{X} + \frac{\overline{X}^2}{(1 - \phi)} = \hat{\mu}_2$$

$$\Rightarrow (1 - \phi)\overline{X} + \overline{X}^2 = \hat{\mu}_2(1 - \phi) \Rightarrow (1 - \phi)(\overline{X} - \hat{\mu}_2) = -\overline{X}^2 \Rightarrow (1 - \phi) = \frac{-\overline{X}^2}{(\overline{X} - \hat{\mu}_2)}$$

Tehát ekkor a

$$\phi = 1 + \frac{\overline{X}^2}{(\overline{X} - \hat{\mu}_2)}. \tag{25}$$

Amiből következik, hogy

$$\hat{\lambda} = 1 - \frac{\hat{\mu}_2}{\overline{X}}. \tag{26}$$

Hurdle Poissonra:
Az 1. egyenlet:

$$E(N) = \Phi\lambda = \overline{X} \tag{27}$$

ahol $\Phi = \frac{1-q}{1-e^{-\lambda}}$, és ebből a λ-ra a következő egyenletet kapjuk:

$$\frac{\overline{X}e^{-\lambda}}{1-q} + \lambda = \frac{\overline{X}}{1-q}, \tag{28}$$

A 2. egyenlet:

$$E(N^2) = \Phi(\lambda + \lambda^2) = \mu_2 = \hat{\mu}_2 \tag{29}$$

Ekkor az előzőhöz hasonlóan

$$\Phi(\lambda + \lambda^2) = \hat{\mu}_2 \Rightarrow \frac{1-q}{1-e^{-\lambda}}(\lambda + \lambda^2) = \hat{\mu}_2 \Rightarrow \ldots \Rightarrow (1-q) = \frac{\hat{\mu}_2(1-e^{-\lambda})}{(\lambda + \lambda^2)}$$

$$q = 1 - \frac{\hat{\mu}_2(1-e^{-\lambda})}{(\lambda + \lambda^2)} \tag{30}$$

3. **Fischer Információ**

Zeró-Inflált Poissonra:

Legyenek $X_1, \ldots X_k \sim$ Zéró-Inflált Poisson eloszlásúak ϕ és λ paraméterekkel.

$$f_{\lambda,\phi}(n) = \begin{cases} \phi + (1-\phi)e^{-\lambda} & \text{ha } n = 0 \\ (1-\phi)e^{-\lambda}\frac{\lambda^n}{n!} & \text{ha } n = 1, 2, \ldots \end{cases}$$

$$\log f_{\lambda,\phi}(\underline{n}) = \sum_{i:n_i=0} \ln\left(\phi + (1-\phi)e^{-\lambda}\right) + \sum_{i:n_i>0} \left(\ln(1-\phi) + n_i \ln\lambda - \lambda - \ln n_i!\right)$$

$$\partial_\lambda \log f_{\lambda,\phi}(\underline{n}) = \frac{(-1)}{\phi + (1-\phi)e^{-\lambda}}(1-\phi)e^{-\lambda}m + \frac{1}{\lambda}d - (k-m)$$

és

$$\partial_\phi \log f_{\lambda,\phi}(\underline{n}) = \frac{1}{\phi + (1-\phi)e^{-\lambda}}(1-e^{-\lambda})m + \frac{(-1)}{1-\phi}(k-m)$$

ahol m és d megegyezik a korábbi jelöléssekkel (**20**. oldalon).
Ekkor a Fisher információt úgy kapjuk, hogy

$$\mathbf{I_n}(\lambda, \phi) = \begin{pmatrix} \partial_\lambda^2 \log f_{\lambda,\phi}(\underline{n}) & \partial_\lambda\partial_\phi \log f_{\lambda,\phi}(\underline{n}) \\ \partial_\phi\partial_\lambda \log f_{\lambda,\phi}(\underline{n}) & \partial_\phi^2 \log f_{\lambda,\phi}(\underline{n}) \end{pmatrix}$$

ahol a

$$\begin{aligned} \partial_\lambda^2 \log f_{\lambda,\phi}(\underline{n}) &= \left[\frac{(-1)}{\phi + (1-\phi)e^{-\lambda}}(1-\phi)e^{-\lambda}m + \frac{1}{\lambda}d - (k-m)\right]^2 = \\ &= \frac{((1-\phi)e^{-\lambda}m)^2}{(\phi + (1-\phi)e^{-\lambda})^2} - 2\frac{d}{\lambda}\frac{(1-\phi)e^{-\lambda}m}{(\phi + (1-\phi)e^{-\lambda})} + (k-m)^2 + \\ &\quad +2(k-m)\frac{(1-\phi)e^{-\lambda}m}{(\phi + (1-\phi)e^{-\lambda})} + \frac{d^2}{\lambda^2} - 2\frac{d}{\lambda}(k-m) = \end{aligned}$$

$$= \ldots =$$

$$\begin{aligned}
&= \frac{d^2(\phi+(1-\phi)e^{-\lambda})^2 - 2d\lambda k\phi^2 - 2d\lambda k(1-\phi)^2(e^{-\lambda})^2 - 4d\lambda k\phi(1-\phi)e^{-\lambda}}{\lambda^2(\phi+(1-\phi)e^{-\lambda})^2} + \\
&+\frac{2d\lambda m\phi(1-\phi)e^{-\lambda} + \lambda^2k^2(1-\phi)^2(e^{-\lambda})^2 + 2\lambda^2k^2\phi(1-\phi)e^{-\lambda}}{\lambda^2(\phi+(1-\phi)e^{-\lambda})^2} + \\
&+\frac{\lambda^2k^2\phi^2 + \lambda^2m^2\phi^2 - 2\lambda^2km\phi^2 - 2\lambda^2km\phi(1-\phi)e^{-\lambda}}{\lambda^2(\phi+(1-\phi)e^{-\lambda})^2} = \\
&= \frac{(d^2 - 2d\lambda k + \lambda^2k^2)(\phi+(1-\phi)e^{-\lambda})^2 + \lambda^2m\phi^2(m-2k)}{\lambda^2(\phi+(1-\phi)e^{-\lambda})^2} + \\
&+\frac{2\lambda m\phi(1-\phi)e^{-\lambda}(d-\lambda k)}{\lambda^2(\phi+(1-\phi)e^{-\lambda})^2} \\
&= \frac{(d^2 - 2d\lambda k + \lambda^2k^2)}{\lambda^2} + \frac{m\phi^2(m-2k)}{(\phi+(1-\phi)e^{-\lambda})^2} + \frac{2m\phi(1-\phi)e^{-\lambda}(d-\lambda k)}{\lambda(\phi+(1-\phi)e^{-\lambda})^2} \qquad (31)
\end{aligned}$$

$$\partial_\lambda\partial_\phi \log f_{\lambda,\phi}(\underline{n}) =$$

$$\begin{aligned}
&= \left[\frac{(-1)(1-\phi)e^{-\lambda}m}{\phi+(1-\phi)e^{-\lambda}} + \frac{1}{\lambda}d - (k-m)\right]\left[\frac{(1-e^{-\lambda})m}{\phi+(1-\phi)e^{-\lambda}} + \frac{(-1)(k-m)}{1-\phi}\right] = \\
&= -\frac{(1-\phi)(1-e^{-\lambda})e^{-\lambda}m^2}{(\phi+(1-\phi)e^{-\lambda})^2} + \frac{(1-\phi)e^{-\lambda}m(k-m)}{(1-\phi)(\phi+(1-\phi)e^{-\lambda})} + \frac{(1-e^{-\lambda})md}{\lambda(\phi+(1-\phi)e^{-\lambda})} - \\
&-\frac{(k-m)d}{\lambda(1-\phi)} - \frac{(k-m)(1-e^{-\lambda})m}{(\phi+(1-\phi)e^{-\lambda})} + \frac{(k-m)^2}{1-\phi} =
\end{aligned}$$

$$= \ldots =$$

$$\begin{aligned}
&= \frac{(e^{-\lambda})^2[-kd(1-\phi)^2 + m^2\lambda(\phi^2-2\phi) + km\lambda\phi(1-\phi)]}{(\phi+(1-\phi)e^{-\lambda})^2\lambda(1-\phi)} + \frac{k^2}{(1-\phi)} + \\
&+\frac{e^{-\lambda}(md(1-\phi) - 2kd\phi(1-\phi) - m(k\lambda - \lambda\phi^2m + 2k\lambda\phi^2)}{(\phi+(1-\phi)e^{-\lambda})^2\lambda(1-\phi)} + \\
&+\frac{\phi(md - k(d\phi - m\lambda - m\lambda\phi) + 2m^2\lambda - m^2\lambda\phi)}{(\phi+(1-\phi)e^{-\lambda})^2\lambda(1-\phi)} \qquad (32)
\end{aligned}$$

$$\begin{aligned}
\partial_\phi^2 \log f_{\lambda,\phi}(\underline{n}) &= \left[\frac{(1-e^{-\lambda})m}{\phi+(1-\phi)e^{-\lambda}} + \frac{(-1)(k-m)}{1-\phi}\right]^2 = \\
&= \frac{(1-e^{-\lambda})^2m^2}{\phi+(1-\phi)e^{-\lambda}} - 2\frac{(1-e^{-\lambda})m(k-m)}{(\phi+(1-\phi)e^{-\lambda})(1-\phi)} + \frac{(k-m)^2}{(1-\phi)^2} =
\end{aligned}$$

$$= \ldots =$$

$$\begin{aligned}
&= \frac{\left(m(1-(e^{-\lambda})^2)+2ke^{-\lambda}(e^{-\lambda}-1)\right)m(1-\phi)^2+(m-2k)m\phi}{(1-\phi)^2(\phi+(1-\phi)e^{-\lambda})^2}+\\
&\quad+\frac{(1-\phi)2m\phi(-k+e^{-\lambda}+m)}{(1-\phi)^2(\phi+(1-\phi)e^{-\lambda})^2}+\frac{k^2(\phi+(1-\phi)e^{-\lambda})}{(1-\phi)^2(\phi+(1-\phi)e^{-\lambda})^2}=\\
&= \frac{\left(m(1-(e^{-\lambda})^2)+2ke^{-\lambda}(e^{-\lambda}-1)\right)m}{(\phi+(1-\phi)e^{-\lambda})^2}+\frac{(m-2k)m\phi}{(1-\phi)^2(\phi+(1-\phi)e^{-\lambda})^2}+\\
&\quad+\frac{2m\phi(-k+e^{-\lambda}+m)}{(1-\phi)(\phi+(1-\phi)e^{-\lambda})^2}+\frac{k^2}{(1-\phi)^2(\phi+(1-\phi)e^{-\lambda})} \qquad (33)
\end{aligned}$$

Hurdle Poissonra:

Legyenek itt az $X_1, \ldots X_k \sim$ változók Hurdle Poisson eloszlásúak λ és q paraméterekkel.

$$f_{\lambda,q}(\underline{n}) = \begin{cases} q & \text{ha } n=0 \\ \frac{1-q}{1-e^{-\lambda}}e^{-\lambda}\frac{\lambda^n}{n!} & \text{ha } n=1,2,\ldots \end{cases}$$

$$\log f_{\lambda,q}(\underline{n}) = \sum_{i=1}^{k}\left[I_{(n_i=0)}\ln(q)+I_{(n_i>0)}\ln(1-q)\right]+$$

$$+\sum_{i=1}^{k}I_{(n_i>0)}\left[n_i\ln(\lambda)-\lambda-\ln(n_i!)-\ln(1-e^{-\lambda})\right]$$

$$\begin{aligned}
\partial_\lambda \log f_{\lambda,q}(\underline{n}) &= \left(-\frac{e^{-\lambda}}{1-e^{-\lambda}}-1\right)(k-m)+\frac{d}{\lambda}\\
\partial_q \log f_{\lambda,q}(\underline{n}) &= \frac{m}{q}-\frac{(k-m)}{1-q}
\end{aligned}$$

Ekkor az előzőhöz hasonlóan a

$$\partial_\lambda^2 \log f_{\lambda,q}(\underline{n}) = \left[\left(-\frac{e^{-\lambda}}{1-e^{-\lambda}}-1\right)(k-m)+\frac{d}{\lambda}\right]\left[\left(-\frac{e^{-\lambda}}{1-e^{-\lambda}}-1\right)(k-m)+\frac{d}{\lambda}\right]=$$

$$= \ldots =$$

$$\begin{aligned}
&= \frac{(e^{-\lambda})^2}{(1-e^{-\lambda})^2}(k-m)^2+2\frac{e^{-\lambda}}{1-e^{-\lambda}}(k-m)^2-2\frac{e^{-\lambda}}{1-e^{-\lambda}}(k-m)\frac{d}{\lambda}+\\
&\quad+\frac{(1-e^{-\lambda})}{1-e^{-\lambda}}(k-m)^2-2\frac{(1-e^{-\lambda})}{1-e^{-\lambda}}(k-m)\frac{d}{\lambda}+\left(\frac{d}{\lambda}\right)^2=\\
&= \frac{(e^{-\lambda})^2}{(1-e^{-\lambda})^2}(k-m)^2+\frac{(1+e^{-\lambda})}{1-e^{-\lambda}}(k-m)^2-\frac{2(k-m)d}{(1-e^{-\lambda})\lambda}+\left(\frac{d}{\lambda}\right)^2 \qquad (34)
\end{aligned}$$

$$\partial_\lambda \partial_q \log f_{\lambda,q}(\underline{n}) = \left[\left(-\frac{e^{-\lambda}}{1-e^{-\lambda}} - 1\right)(k-m) + \frac{d}{\lambda}\right]\left[\frac{m}{q} - \frac{(k-m)}{1-q}\right] =$$

$$= \frac{md}{q\lambda} - \frac{(mk-m^2)e^{-\lambda}}{q(1-e^{-\lambda})} - \frac{(mk-m^2)}{q} - \frac{(k-m)d}{(1-q)\lambda} - \frac{(k-m)^2}{1-q} + \frac{(k-m)^2 e^{-\lambda}}{(1-q)(1-e^{-\lambda})} =$$

$$= \ldots =$$

$$\begin{aligned} &= \frac{md - mde^{-\lambda} - mk\lambda + 3mkq\lambda + m^2\lambda - 2m^2q\lambda - kdq + kdqe^{-\lambda}}{q(1-q)\lambda(1-e^{-\lambda})} + \\ &+ \frac{2m^2q\lambda e^{-\lambda} - 4kmq\lambda e^{-\lambda} - k^2q\lambda + k^2q\lambda e^{-\lambda}}{q(1-q)\lambda(1-e^{-\lambda})} = \end{aligned}$$

$$= \frac{md}{q(1-q)\lambda} - \frac{kd}{q(1-q)} + \frac{m(m-k)}{q(1-q)(1-e^{-\lambda})} - \frac{(2m^2+k^2)}{(1-q)} + \frac{mk(3-4e^{-\lambda})}{(1-q)(1-e^{-\lambda})} \quad (35)$$

és végül a

$$\begin{aligned} \partial_q^2 \log f_{\lambda,q}(\underline{n}) &= \left[\frac{m}{q} - \frac{(k-m)}{1-q}\right]\left[\frac{m}{q} - \frac{(k-m)}{1-q}\right] = \frac{m^2}{q^2} - 2\frac{(k-m)m}{q(1-q)} + \frac{(k-m)^2}{(1-q)^2} = \\ &= \frac{(1-q)^2m^2}{(1-q)^2q^2} - 2\frac{(k-m)mq(1-q)}{q(1-q)} + \frac{(k-m)^2q^2}{(1-q)^2q^2} = \\ &= \frac{m^2 + 2mq^2 - 4kmq^2 - 4m^2q + 2kmq + 2mq^2 + k^2q^2}{q^2(1-q)^2} \end{aligned} \quad (36)$$

6. Összefoglalás

A szakdolgozat során bemutatásra került a 2 modell főbb tulajdonságai, a köztük lévő különbségek és hasonlóságok, valamint egy példán keresztül a Poisson eloszlástól való eltérésük is. A bemutatott modelleket konkrét példákra alkalmaztuk, majd a modellekre statisztikai méréseket folytattunk. Ezekből látható, hogy habár a két modell nagyon hasonló, mert mind a kettő a nagymennyiségű 0 adatszerkezettel foglalkozik, sokszor mégis más eredményeket adhatnak, ezért mindkét modell ismerete fontos lehet számunkra.

Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani mindazoknak, akik segítettek abban, hogy ez a dolgozat létrejöhessen.

Elsősorban hálás köszönettel tartozom konzulensemnek, Arató Miklósnak, aki felhívta figyelmem erre a rendkívűl érdekes témára, és aki sokat segített mind útmutatásaival, hasznos ötleteivel valamint észrevételeivel és fáradhatalan biztatásával.

Szeretném megköszönni családomnak a folyamatos támogatásukat és azt, hogy mindig számíthattam rájuk.

Végül, de nem utolsó sorban szeretném megköszönni tanáraimnak azt, hogy megismertettek engem a valószínűségszámítás és matematika szépségeivel és, hogy segítettek abban, hogy ideig eljuthassak.

Ábrák jegyzéke

Táblázatok jegyzéke

Irodalomjegyzék

Hivatkozások

[1] Arató Miklós. Nem élet biztosítási matematika. *ELTE Eötvös Kiadó*, 2003

[2] Jean-Philippe Boucher, Michel Denuit & Montserrat Guillén (2007). Risk Classification for Claim Counts: A Comparative Analysis of Various Zero-Inflated Mixed Poisson and Hurdle Models. North American Actuarial Journal **11**, Issue 4, 1-24.

[3] Michel Denuit, Xavier Marechal, Sandra Pitrebois and Jean-Francois Walhin. Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems. *Wiley*, 2007.

[4] Karen C.H. Yip and Kelvin K.W. Yau (2005). On modelling claim frequency data in general insurancewith extra zeros. Insurance: Mathematics and Economics **36**, Issue 2, 153-163 .

[5] Mei-Chen Hu, Martina Pavlicova & Edward V. Nunes (2011): Zero-Inflated and Hurdle Models of Count Data with Extra Zeros: Examples from an HIV-Risk Reduction Intervention Trial. The American Journal of Drug and Alcohol Abuse **37**, Issue 5, 367-375.

[6] Joseph Ngatchou-Wandji and Christophe Paris (2011). On the Zero-Inflated Count Models with Application to Modelling Annual Trends in Incidences of Some Occupational Allergic Diseases in France. Journal of Data Science **9**, 639-659.

[7] Noriszura Ismail and Abdul Aziz Jemain (2007). Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models. Casualty Actuarial Society Forum, Winter 2007, 103-158.

[8] Fazekas I. (szerk.): Bevezetés a matematikai statisztikába, *Kossuth Egyetemi Kiadó*, 2000.

[9] Tómács Tibor. Matematikai statisztika. Eszterházy Károly Főiskola Matematikai és Informatikai Intézet, 2012.

[10] Felix Famoye (1993). Restricted generalized Poisson regression model. Communication in Statistics- Theory and Methods **22**, Issue 5, 1335-1354.