Eötvös Loránd Tudományegyetem

Természettudományi Kar

Romvári Petra

Generátorfüggvények a biztosításmatematikában

BSc szakdolgozat

Témavezető:

Arató Miklós

Valószínűségelméleti és Statisztika Tanszék

Budapest, 2015

Tartalomjegyzék

1. Előszó

Szakdolgozatom célja annak bemutatása, hogy miképpen lehetnek segítségünkre a generátorfüggvények egyes valószínűségi változók konvolúciójának meghatározásakor. Minthogy sokszor elengedhetetlenül fontos a konvolúció eloszlásának pontos ismerete (többek között biztosításmatematikai alkalmazások esetében), ezért ezen tárgykör tanulmányozása nem pusztán érdekes, hanem rendkívül hasznos is.

Először egy bevezető példán keresztül fogok rávilágítani arra, hogy egy lehetséges, a konvolúció eloszlásának megadására irányuló megközelítés –a normális közelítés alkalmazása– milyen pontatlan eredményeket tud adni.

A következő fejezetben egy 1985-ös Nelson De Pril-cikk ([2]) feldolgozásával áttekintek egy módszert, amely generátorfüggvények segítségével rekurziós eljárást ad független, azonos eloszlású, nemnegatív egész értékű valószínűségi változók konvolúciójának meghatározására, továbbá bemutatásra kerül ezen rekurzió Panjer-rekurzióval való kapcsolata. Ezt követően az eljárást két konkrét biztosítási példa megoldásán keresztül alkalmazom, majd részletesen elemzem a kapott eredményeket.

Végül egy, az előzőnél általánosabb esetre vonatkozó rekurzió kerül tárgyalásra. Nelson De Pril 1986-os cikke alapján ([5]) megnézzük, hogyan határozható meg független, nemnegatív egész értékű valószínűségi változók konvolúciójának eloszlása, amennyiben azok nem azonos eloszlásúak. Egy speciális eset vizsgálata után akárcsak az előző fejezetben, most is biztosításmatematikai példák megoldásán keresztül alkalmazom az elméleti eredményeket.

2. Bevezető példa

Tekintsük az alábbi egyszerűsített példát: egy biztosítóval két közösség köt biztosítást, közösségenként 100-100 fővel. Az első közösség egy tagjának esetében a káreset bekövetkezési valószínűsége 3%, ekkor a biztosító kifizetése 1 pénzegység. A második közösség egy tagjának esetében a káreset bekövetkezési valószínűsége 1%, viszont ekkor a biztosító kifizetése 100 pénzegység.

A felmerülő kérdés: mekkora pénzösszeggel kell rendelkeznie a biztosítónak ahhoz, hogy az legalább 99%, illetve 99,5% valószínűséggel elég legyen a kifizetésekre?

Modellezzük a feladatot!

Az első közösséget X_i, a második közösséget Y_j valószínűségi változók segítségével fogjuk reprezentálni, ezek fogják jelölni a közösségek tagjainak kifizetéseit. A feltételek szerint:

$$\begin{aligned} P(X_i = 0) &= 97\%, \qquad & P(X_i = 1) &= 3\%, \\ P(Y_j = 0) &= 99\%, \qquad & P(Y_j = 100) &= 1\%, \end{aligned}$$

ahol $i, j = 1, 2, \dots 100$.

Jelölje az összkárt S! Ekkor $S = \sum\limits_{i=1}^{100} X_i + \sum\limits_{j=1}^{100} Y_j$.

A feladat tehát azon legkisebb z_1 és z_2 pozitív egészek meghatározása, amelyekre $P(S \leq z_1) \geq 0,99$, illetve $P(S \leq z_2) \geq 0,995$.

Normális közelítés

A centrális határeloszlás-tétel nem azonos eloszlású valószínűségi változókra vonatkozó változata alapján normális eloszlással közelíthetjük S-t.

Legyenek $m_x = E(X_i)$, $m_y = E(Y_j)$, továbbá $\sigma_x^2 = D^2(X_i)$, $\sigma_y^2 = D^2(Y_j)$. A tétel alapján az alábbinak kellene teljesülnie a standardizált összegre, amennyiben a 200-as tagszám valóban elég nagynak bizonyul:

$$F(z) = P\left(\frac{S - 100\left(m_x + m_y\right)}{\sqrt{100\left(\sigma_x^2 + \sigma_y^2\right)}} < z \right) \approx \phi(z) \tag{1}$$

Ezt átrendezve, majd a megfelelő értékeket behelyettesítve kapjuk az alábbi közelítéseket:

$$\hat{z}_1 = 334$$
$$\hat{z}_2 = 360$$

Hibabecslés

A normális közelítés hibáját az Esséen-tétel segítségével becsülhetjük.

A tételben szereplő konstansra a jelenlegi legjobb becslés Shevtsova 2010-es eredménye: 0,5600. ([1]) Ez alapján:

$$\sup_z |F(z) - \phi(z)| \leq 0{,}5600 \cdot \frac{\sum_{i=1}^{100} E|X_i - m_x|^3 + \sum_{j=1}^{100} E|Y_j - m_y|^3}{\left[100(\sigma_x^2 + \sigma_y^2)\right]^{\frac{3}{2}}} =$$

$$= 0{,}5600 \cdot \frac{\left[100 \cdot (0{,}03 \cdot 0{,}97^3 + 0{,}97 \cdot 0{,}03^3) + 100 \cdot (0{,}01 \cdot 99^3 + 0{,}99 \cdot 1^3)\right]}{\left[100 \cdot (0{,}03 \cdot 0{,}97 + 99 \cdot 1)\right]^{\frac{3}{2}}} \approx 0{,}55$$

A 0,55-os lehetséges eloszlásfüggvénybeli eltérés meglehetősen magas, ami óvatosságra int minket a normális közelítés eredményeinek használatával kapcsolatban. Mindenképpen érdemes pontosan is kiszámolni a keresett értékeket.

Mi tehát a valóság?

Keressük azon legkisebb z_1 és z_2 pozitív egészeket, amelyekre
$P(S \leq z_1) \geq 0{,}99$, illetve $P(S \leq z_2) \geq 0{,}995$.

Legyenek $r, t \in \mathbb{N}$, $0 \leq r \leq 100$, $0 \leq t \leq 99$.

Mikor nem fogja meghaladni a kifizetés a $100r + t$ értéket? Akkor, ha Y_j események közül legfeljebb $r - 1$ következik be, illetve akkor, ha pontosan r következik be, továbbá X_i események közül legfeljebb t. Azaz:

$$P(S \leq 100r + t) = P(\sum_{j=1}^{100} Y_j \leq 100(r-1)) + P(\sum_{j=1}^{100} Y_j = 100r) \cdot P(\sum_{i=1}^{100} X_i \leq t) =$$

$$= \sum_{v=0}^{r-1} \binom{100}{v} 0{,}01^v 0{,}99^{100-v} + \binom{100}{r} 0{,}01^r 0{,}99^{100-r} \cdot \sum_{w=0}^{t} \binom{100}{w} 0{,}03^w 0{,}97^{100-w}$$

A képletet felhasználván Matlab-program írható a keresett értékek meghatározására, ami alapján:

$$z_1 = 403$$
$$z_2 = 405$$

Megjegyzés

Első ránézésre érdekesnek tűnhet, hogy z_1 és z_2 értékek meglehetősen közel találhatóak egymáshoz, viszont ennek igen egyszerű a magyarázata.

Rögzített r esetén (most: $r = 4$) a viszonylag nagy valószínűségek kis t értékekre koncentrálódnak, hiszen t értéke azt mutatja meg, hogy X_i-k közül pontosan hány következett be. X_i valószínűségi változók konvolúciója binomiális eloszlású $\big(100,\ 0{,}03\big)$ paraméterekkel, 3 várható értékkel. Számítással ellenőrizhető, hogy kis, 3-hoz közeli t értékekre lesznek a legnagyobbak a valószínűségek, nagy értékekre pedig a legkisebbek.

A pontos értékek összevetése a normális közelítésből kapottakkal

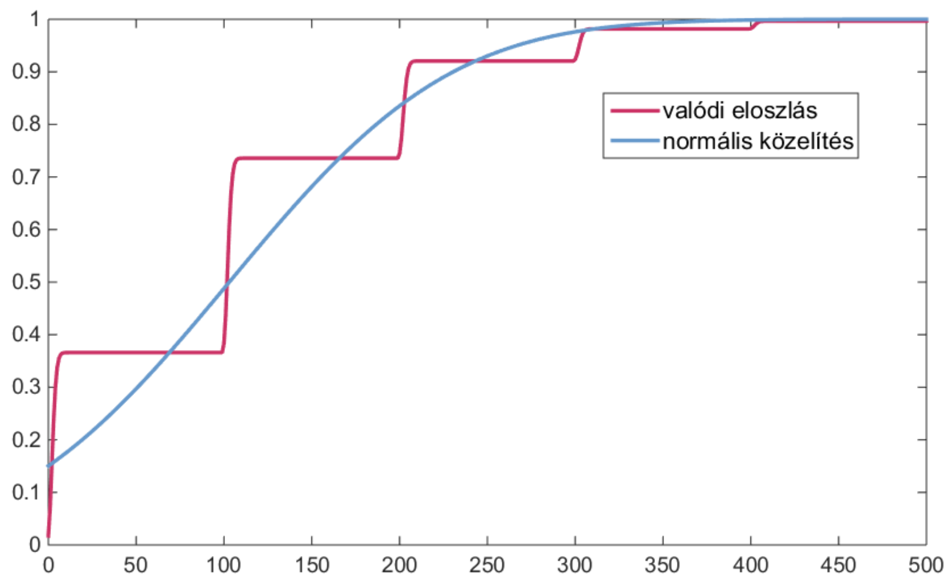
A keresett értékek pontos ismeretében mindenképpen érdemes összehasonlítani azokat a normális közelítésből kapott becslésükkel. A közelítés hibabecslésére vonatkozó eredmény alapján számíthatunk arra, hogy akár igen nagy eltérések is lehetnek az értékek között. Nézzük meg, hogy a vizsgált két konkrét esetben mekkora különbségek adódtak:

	Pontos érték	Normális közelítésből kapott érték	A kettő eltérése
z_1	403	334	69
z_2	405	360	45

1. táblázat. Bevezető példa: a pontos értékek összevetése a normális közelítésből kapottakkal

A táblázatból leolvasható, hogy az normális közelítés (lásd:(1)) ebben a példában pontatlannak bizonyult, lényegesen alulbecsülte a valódi értékeket.

Érdemes grafikusan is megjeleníteni a valódi eloszlásfüggvényt és a normális közelítés alapján kapott eloszlásfüggvényt.



1. ábra. Bevezető példa: a valódi eloszlásfüggvény és a normális közelítés alapján kapott eloszlásfüggvény kapcsolata

Ahogyan a korábbi számításaink jelezték, alátámasztja az ábra is: ezen feladat esetében a normális közelítés eredményeivel meglehetősen óvatosan kell bánnunk.

Sokszor, amikor valószínűségi változók konvolúcióját vizsgáljuk, mindenképpen fontos, hogy annak eloszlását pontosan ismerjük, ne pusztán közelítőleg adjuk meg. A további fejezetekben arra fogunk példákat látni, hogy generátorfüggvények segítségével milyen módszerekkel tudjuk meghatározni egyes valószínűségi változók konvolúciójának eloszlását.

3. Független, azonos eloszlású, nemnegatív egész értékű valószínűségi változók konvolúciója

3.1. Bevezetés

Legyenek $X_1, X_2, \ldots X_n$ független, azonos eloszlású, nemnegatív egész értékű valószínűségi változók. Jelölje a közös eloszlásukat $p_X(x) = P(X = x)$, s tegyük fel, hogy $p_X(0) > 0$.

Írjuk fel az eloszláshoz tartozó generátorfüggvényt:

$$G_X(u) = \sum_{x=0}^{\infty} p_X(x) u^x \tag{2}$$

Legyen $S = X_1 + X_2 + \ldots + X_n$. Ekkor a konvolúció generátorfüggvénye:

$$G_S(u) = \big(G_X(u)\big)^n \tag{3}$$

Keressük $p_S(s)$ eloszlást. Legyen $p_X^{*k}(s) = P(X_1 + X_2 + \ldots + X_k = s)$. Ekkor:

$$p_S(s) = p_X^{*n}(s) \tag{4}$$

Hogyan kaphatjuk meg $p_S(s)$ értékeit? Könnyen látható, hogy p_X^{*k+1}-et egyértelműen meghatározzák $p_X^{*1}, p_X^{*2}, \ldots p_X^{*k}$ eloszlások. Ezt felhasználva felírható az alábbi összefüggés:

$$p_X^{*k+1}(y) = \sum_{x=0}^{y} p_X(x) p_X^{*k}(y - x), \text{ ahol: } y = 0, 1 \ldots s \tag{5}$$

Ezen megállapítás elvezet p_X^{*n} rekurziós meghatározásához. Írjuk fel tehát sorban a megfelelő $n - 1$ összefüggést a $k = 1, 2 \ldots n - 1$ értékekre, s végezetül megkapjuk a keresett $p_S(s)$ eloszlást.

Az itt tárgyalt rekurziós elv problémáját az okozza, hogy rendkívül műveletigényes nagy n értékek esetén, főleg abban az esetben, amikor X sokféle különböző értéket is felvehet. Felmerül az igény egy ennél hatékonyabb, jóval kevesebb művelet elvégzését igénylő eljárás kidolgozására. A következő alfejezetben egy ilyen lehetséges módszer kerül bemutatásra.

3.2. A rekurzió

3.2.1. Tétel. (De Pril, 1985, [2]) *Legyenek* $X_1, X_2, \dots X_n$ *független, azonos eloszlású, nemnegatív egész értékű valószínűségi változók. Jelölje a közös eloszlásukat* $p_X(x) = P(X = x)$, *s tegyük fel, hogy* $p_X(0) > 0$. *Ekkor igazak rájuk az alábbi összefüggések:*

$$p_S(0) = \big(p_X(0)\big)^n \tag{6}$$

$$p_S(s) = \frac{1}{p_X(0)} \sum_{x=1}^{s} \left(\frac{n+1}{s} x - 1 \right) p_X(x) p_S(s-x) \qquad s = 1, 2, \dots \tag{7}$$

Bizonyítás. A nemnegativitásból könnyen adódik az (6)-os összefüggés, ugyanis $p_S(0) = 0$ pontosan akkor teljesül, ha valamennyi i esetén $X_i = 0$, s ennek valószínűsége a függetlenség miatt $\big(p_X(0)\big)^n$.

A (7)-es egyenlet igazolásához használjuk fel a (3)-as összefüggést! Lederiválva kapjuk, hogy:

$$G'_S(u) = n \cdot G'_X(u) \big(G_X(u)\big)^{n-1}$$

Mindkét oldalt megszorozva $G_X(u)$-val:

$$G_X(u) G'_S(u) = n \cdot G'_X(u) \big(G_X(u)\big)^{n}$$

A jobb oldali $\big(G_X(u)\big)^n$ helyére $G_S(u)$-t helyettesítve:

$$G_X(u) G'_S(u) = n \cdot G'_X(u) G_S(u)$$

A Leibniz-formula segítségével írjuk fel mindkét oldal $(s-1)$-edik deriváltját:

$$G_X(u) G_S^{(s)}(u) + \sum_{x=1}^{s-1} \binom{s-1}{x} G_X^{(x)}(u) G_S^{(s-x)}(u) =$$
$$= n \sum_{x=0}^{s-1} \binom{s-1}{x} G_X^{(x+1)}(u) G_S^{(s-1-x)}(u)$$

Használjuk fel, hogy tetszőleges generátorfüggvény esetén igaz az alábbi összefüggés:

$$p_\epsilon(k) = \frac{G_\epsilon^{(k)}(0)}{k!}$$

Ebből $u = 0$ helyettesítéssel adódik:

$$p_X(0) \cdot s! \cdot p_S(s) + \sum_{x=1}^{s-1} \binom{s-1}{x} x! \cdot (s-x)! \cdot p_X(x) \cdot p_S(s-x) =$$
$$= n \sum_{x=0}^{s-1} \binom{s-1}{x} (x+1)! \cdot (s-1-x)! \cdot p_X(x+1) \cdot p_S(s-1-x)$$

Ezt átrendezve kapjuk az (7)-es összefüggést. $\square$

Megjegyzés

A tétel második felének bizonyítására ismeretes egy másik, különösen elegáns bizonyítás is:

Legyen X_{n+1} egy olyan valószínűségi változó, mely független $X_1, X_2, \ldots X_n$ mindegyikétől, s ezekkel azonos eloszlású. Vizsgáljuk az alábbi feltételes várható értéket $s > 0$ esetén:

$$E\left[\frac{n+1}{s} X_{n+1} - 1 \,\middle|\, S + X_{n+1} = s\right] = 0$$

Ebből következik:

$$\sum_{x=0}^{s} \left(\frac{n+1}{s} x - 1\right) p_X(x) p_S(s-x) = 0$$

Amelyből rögtön adódik a tétel második fele. $\square$

3.3. Panjer-rekurzió

3.3.1. Definíció. Az η nemnegatív egész értékű valószínűségi változó $(a, b, 0)$ eloszlású, ha léteznek hozzá olyan $a, b \in \mathbb{R}$ paraméterek, hogy eloszlására teljesül az alábbi rekurzió:

$$P\big(\eta = n\big) = \left(a + \frac{b}{n}\right) \cdot P\big(\eta = n-1\big) \qquad \text{n=1,2, \dots}$$

3.3.2. Állítás. *η $(a, b, 0)$ eloszlású pontosan akkor, ha eloszlása binomiális, negatív binomiális vagy Poisson.*

A megfelelő paraméterértékek:

	a és b értékei	
$Bin(n, p)$	$a = -\frac{p}{1-p}$	$b = \frac{(n+1)p}{1-p}$
$NegBin(r, q)$	$a = q$	$b = (r-1)q$
$Poi(\lambda)$	$a = 0$	$b = \lambda$

2. táblázat. $(a, b, 0)$ típusú eloszlások

Panjer-rekurzió

3.3.3. Tétel. (Panjer) *Legyen η $(a, b, 0)$ eloszlású, továbbá legyenek $Z_1, Z_2, \dots$ független, azonos eloszlású, pozitív egész értékű valószínűségi változók. Tetszőleges ϵ egész értékű valószínűségi változóra legyen $p_\epsilon(n) = P(\epsilon = n)$.*

Ekkor $S = Z_1 + Z_2 + \dots Z_\eta$-ra teljesül az alábbi rekurzív azonosság:

$$p_S(0) = p_\eta(0) \tag{8}$$

$$p_S(n) = \sum_{y=1}^{n} \big(a + b\frac{y}{n}\big) p_{Z_1}(y) p_S(n-y) \qquad n = 1, 2, \dots \tag{9}$$

3.4. A korábban tárgyalt rekurzió interpretációja, illetve kapcsolata a Panjer-rekurzióval

([2]): Legyen K nemnegatív egész értékű valószínűségi változó $p_K(k)$ eloszlással, $G_K(u)$ generátorfüggvénnyel. Legyenek $Y_1, Y_2, \dots$ független, azonos eloszlású, pozitív egész értékű valószínűségi változók $p_Y(y)$ eloszlással, $G_Y(u)$ generátorfüggvénnyel. Tegyük fel, hogy valamennyi Y_i K-tól is független.

Legyen

$$R = Y_1 + Y_2 + \dots Y_K,$$

illetve megállapodás szerint $R = 0$ abban az esetben, amikor $K = 0$.

Ekkor az R valószínűségi változó eloszlására teljesülnek az alábbiak:

$$p_R(0) = p_K(0) \tag{10}$$

$$p_R(r) = \sum_{k=1}^{r} p_K(k) p_Y^{*k}(r) \qquad r = 1, 2, \dots \tag{11}$$

A generátorfüggvényére:

$$G_R(u) = G_K[G_Y(u)] \tag{12}$$

A következőek alapján egy újabb lehetőség nyílik a korábban tárgyalt S valószínűségi változó eloszlásának meghatározására. Az ötlet lényege, hogy K és Y megfelelő megválasztásával elérhető, hogy R eloszlása megegyezzen S-ével, így elég az előbbi megadása.

3.4.1. Tétel. *Legyenek*

$$p_K(k) = \binom{n}{k} \big(1 - p_X(0)\big)^k \big(p_X(0)\big)^{n-k}, \qquad k = 0, 1, \dots, n \tag{13}$$

$$p_Y(y) = \frac{p_X(y)}{1 - p_X(0)} \tag{14}$$

Ekkor S és R valószínűségi változók azonos eloszlásúak.

Bizonyítás. (13) alapján:

$$G_K(u) = \sum_{k=0}^{n} p_K(k)u^k = \sum_{k=0}^{n} \binom{n}{k} \big(1 - p_X(0)\big)^k \big(p_X(0)\big)^{n-k} u^k =$$
$$= \sum_{k=0}^{n} \binom{n}{k} \big[\big(1 - p_X(0)\big)u\big]^k \big(p_X(0)\big)^{n-k} = \big[p_X(0) + \big(1 - p_X(0)\big)u\big]^n$$

(14) alapján:

$$G_Y(u) = \sum_{k=1}^{\infty} p_Y(k)u^k = \sum_{k=1}^{\infty} \frac{p_X(k)}{1 - p_X(0)} \cdot u^k = \frac{1}{1 - p_X(0)} \cdot \sum_{k=1}^{\infty} p_X(k)u^k =$$
$$= \frac{1}{1 - p_X(0)} \cdot \big(G_X(u) - p_X(0)\big) = \frac{G_X(u) - p_X(0)}{1 - p_X(0)}$$

Továbbá a (12)-es és (3)-as összefüggéseket felhasználva:

$$G_R(u) = \left[p_X(0) + \left((1 - p_X(0)) \cdot \frac{G_X(u) - p_X(0)}{1 - p_X(0)}\right)\right]^n =$$
$$= \big(G_X(u)\big)^n = G_S(u)$$

Tehát azt kapjuk, hogy R és S valószínűségi változók generátorfüggvényei megegyeznek. Abból adódóan, hogy a generátorfüggények egyértelműen meghatározzák az eloszlást: R és S azonos eloszlásúak. $\square$

A tétel mögött megbúvó ötlet az, hogy az összkár kiszámításakor nem egyesével vizsgáljuk a különböző szerződésekhez tartozó kifizetéseket (amelyeknek értéke sok esetben természetesen 0), hanem káronként összegzünk. Ekkor valamennyi összeadandó tag pozitív lesz. A tételben használt jelölésekkel:

- K: bekövetkezett káresetek száma, eloszlása binomiális n és $1 - p_X(0)$ paraméterekkel
- Y_i: az i-edik bekövetkezett káresethez tartozó kifizetés

Így S eloszlása:

$$p_S(0) = [p_X(0)]^n$$
$$p_S(s) = \sum_{k=1}^{s} \binom{n}{k} [1 - p_X(0)]^k [p_X(0)]^{n-k} p_Y^{*k}(s), \qquad s = 1, 2, \ldots$$

Az előbbi eredményből az is látszik, hogy a 3.2.1. tételben felírt rekurziós összefüggés speciális esete az előző alfejezetben ismertetett Panjer-rekurziónak. A paraméterértékek megfelelő megválasztásával:

$$a = -\frac{p}{1-p} = -\frac{1-p_X(0)}{p_X(0)} \qquad b = \frac{(n+1)p}{1-p} = \frac{(n+1)(1-p_X(0))}{p_X(0)}$$

Ekkor:

$$\begin{aligned} p_S(s) &= \sum_{y=1}^{s} \big(a + b\frac{y}{s}\big) p_Y(y) p_S(s-y) = \\ &= \sum_{y=1}^{s} \Bigg(-\frac{1-p_X(0)}{p_X(0)} + \frac{(n+1)(1-p_X(0))}{p_X(0)} \cdot \frac{y}{s} \Bigg) p_Y(y) p_S(s-y) \end{aligned}$$

Használjuk fel, hogy $y > 0$ esetén teljesül az alábbi:

$$\begin{aligned} p_Y(y) = P(Y=y) = P(X=y|X>0) &= \frac{P(X=y, X>0)}{1-P(X=0)} = \\ = \frac{P(X=y)}{1-P(X=0)} &= \frac{p_X(y)}{1-p_X(0)} \end{aligned}$$

Ekkor folytatván az egyenlőséget:

$$\begin{aligned} p_S(s) &= \sum_{y=1}^{s} \Bigg(-\frac{1}{p_X(0)} + \frac{(n+1)}{p_X(0)} \cdot \frac{y}{s} \Bigg) p_X(y) p_S(s-y) = \\ &= \frac{1}{p_X(0)} \sum_{y=1}^{s} \Bigg(\frac{n+1}{s} y - 1 \Bigg) p_X(y) p_S(s-y) \end{aligned}$$

Ami szintén a korábbiakban (3.2.1. tétel) megkapott eredményt adja. $\square$

3.5. Alkalmazás

3.5.1. 1. példa

A Központi Statisztikai Hivatal adatai alapján 2014-ben Magyarország népessége 9.877.365 fő, valamilyen közúti közlekedési balesetben súlyos sérülést szenvedettek száma 5.331 fő, közúti közlekedési balesetben elhunytak száma 626 fő. [3]

Egy biztosítónál biztosítás köthető közúti közlekedési baleset bekövetkezésének esetére: a kifizetés súlyos sérülés esetén 1 millió forint, halál esetén 3 millió forint.

Felhasználva a 2014-es magyarországi adatokat 100.000-es szerződésszám mellett meghatározandó az összkár eloszlása, valamint azon legkisebb z_1 és z_2 számok értéke, amelyek 99%-os, illetve 99,5%-os valószínűséggel elegek lesznek az összkár kifizetésére. (Azaz keressük azon legkisebb z_1 és z_2 értékeket, melyekre $P(S \leq z_1) \geq 0{,}99$, illetve $P(S \leq z_2) \geq 0{,}995$.)

Megoldás

Modellezzük a feladatot!

Legyen X_i az i. biztosított számára történő kifizetést jelentő valószínűségi változó, $i = 1, 2, \ldots 100.000$. Az összkárt adó S valószínűségi változó ezen X_i-k konvolúciójaként áll elő: $S = \sum\limits_{i=1}^{100.000} X_i$.

A károkat reprezentáló valószínűségi változókat függetleneknek és azonos eloszlásúaknak tekintve az összkár eloszlásának kiszámításakor alkalmazható a 3.2.1. tételben látott rekurzió.

A tétel jelöléseivel, a 2014-es adatokból származtatott tapasztalati valószínűségekkel dolgozván:

- $p_X(1) = \frac{5.331}{9.877.365}$
- $p_X(3) = \frac{626}{9.877.365}$
- $p_X(0) = 1 - p_X(1) - p_X(3)$
- $p_X(x) = 0$, ha $x \neq 0, 1, 3$

Továbbá a szerződésszám: $n = 100.000$.

A rekurzió felhasználásával, Matlab-programot készítvén megadható az S konvolúció eloszlása: a következő oldalakon található táblázatokban öt tizedesjegyre kerekítve szerepelnek $P(S = s)$, valamint $P(S \leq s)$ értékei. Alaposan áttanulmányozva a táblázatokat több fontos adatot is kinyerhetünk belőlük az összkár eloszlásának jellemzése céljából.

A táblázat adatainak elemzése

Egyrészt megadható az S konvolúció mediánja: $m = 73$ millió forint. Ez összehasonlítva a várható értékkel (millió forintban számolva):

$$
\begin{aligned}
E(S) = E\left(\sum_{i=1}^{100.000} X_i\right) &= 100.000 \cdot E(X_1) = \\
&= 100.000 \cdot \left(1 \cdot \frac{5.331}{9.877.365} + 3 \cdot \frac{626}{9.877.365}\right) = \\
&= 72{,}985 \approx 73
\end{aligned}
$$

Vagyis azt kapjuk, hogy a mediánérték lényegében megegyezik az összkár várható értékével.

Másrészt leolvashatóak a táblázatból azon keresett forintértékek –z_1 és z_2–, melyek 99%-os, illetve 99,5%-os valószínűséggel elegek lesznek az összkár kifizetésére:

- $z_1 = 99$ millió forint
- $z_2 = 102$ millió forint

Érdemes lehet összevetni ezen z-értékeket az összkár maximumával.
$S_{\max} = 3 \cdot 100.000 = 300.000$ (millió forintban számolva), azaz az összkár akár a 300 milliárd forintot is elérheti. Látható tehát, hogy a biztosító által maximálisan kifizetendő összegnek a töredéke is elég lesz arra, hogy az 99%-os (illetve 99,5%-os) valószínűséggel fedezze a biztosító összkifizetését.

Ennek az a magyarázata, hogy a káresetek bekövetkezési valószínűsége már önmagában sem nagy ($p_X(1) = \frac{5.331}{9.877.365}$, $p_X(3) = \frac{626}{9.877.365}$), ennek következtében annak a valószínűsége, hogy az adott évben relatíve sok kár fog együttesen bekövetkezni, elhanyagolhatóan kicsi lesz. A táblázat adataiból az is leolvasható, hogy már 126 millió forint esetében is igaz lesz az, hogy annál többet csak lényegében 0 valószínűséggel fog kifizetni a biztosító.

s (millió forintban)	$P(S=s)$	$P(S \leq s)$	s (millió forintban)	$P(S=s)$	$P(S \leq s)$
0	≈ 0	≈ 0	35	$0,00001$	$0,00002$
1	≈ 0	≈ 0	36	$0,00002$	$0,00004$
2	≈ 0	≈ 0	37	$0,00003$	$0,00007$
3	≈ 0	≈ 0	38	$0,00004$	$0,00011$
4	≈ 0	≈ 0	39	$0,00007$	$0,00018$
5	≈ 0	≈ 0	40	$0,00011$	$0,00029$
6	≈ 0	≈ 0	41	$0,00016$	$0,00045$
7	≈ 0	≈ 0	42	$0,00024$	$0,00069$
8	≈ 0	≈ 0	43	$0,00035$	$0,00104$
9	≈ 0	≈ 0	44	$0,00050$	$0,00154$
10	≈ 0	≈ 0	45	$0,00070$	$0,00224$
11	≈ 0	≈ 0	46	$0,00096$	$0,00320$
12	≈ 0	≈ 0	47	$0,00131$	$0,00451$
13	≈ 0	≈ 0	48	$0,00175$	$0,00626$
14	≈ 0	≈ 0	49	$0,00230$	$0,00855$
15	≈ 0	≈ 0	50	$0,00298$	$0,01153$
16	≈ 0	≈ 0	51	$0,00381$	$0,01534$
17	≈ 0	≈ 0	52	$0,00479$	$0,02013$
18	≈ 0	≈ 0	53	$0,00595$	$0,02608$
19	≈ 0	≈ 0	54	$0,00729$	$0,03337$
20	≈ 0	≈ 0	55	$0,00881$	$0,04218$
21	≈ 0	≈ 0	56	$0,01051$	$0,05269$
22	≈ 0	≈ 0	57	$0,01238$	$0,06507$
23	≈ 0	≈ 0	58	$0,01441$	$0,07949$
24	≈ 0	≈ 0	59	$0,01657$	$0,09606$
25	≈ 0	≈ 0	60	$0,01884$	$0,11490$
26	≈ 0	≈ 0	61	$0,02116$	$0,13606$
27	≈ 0	≈ 0	62	$0,02350$	$0,15956$
28	≈ 0	≈ 0	63	$0,02582$	$0,18539$
29	≈ 0	≈ 0	64	$0,02806$	$0,21345$
30	≈ 0	≈ 0	65	$0,03018$	$0,24363$
31	≈ 0	≈ 0	66	$0,03212$	$0,27575$
32	≈ 0	≈ 0	67	$0,03384$	$0,30959$
33	≈ 0	$0,00001$	68	$0,03530$	$0,34489$
34	$0,00001$	$0,00001$	69	$0,03646$	$0,38136$

3. táblázat. 3.5.1. példa: az S konvolúció eloszlása 0-tól 69 millió forintig

s (millió forintban)	$P(S = s)$	$P(S \leq s)$	s (millió forintban)	$P(S = s)$	$P(S \leq s)$
70	0,03731	0,41866	105	0,00060	0,99789
71	0,03781	0,45648	106	0,00048	0,99837
72	0,03797	0,49445	107	0,00037	0,99875
73	0,03779	0,53224	108	0,00029	0,99904
74	0,03728	0,56952	109	0,00023	0,99927
75	0,03645	0,60598	110	0,00018	0,99944
76	0,03534	0,64132	111	0,00014	0,99958
77	0,03398	0,67530	112	0,00010	0,99968
78	0,03240	0,70770	113	0,00008	0,99976
79	0,03064	0,73833	114	0,00006	0,99982
80	0,02874	0,76708	115	0,00005	0,99987
81	0,02676	0,79383	116	0,00003	0,99990
82	0,02471	0,81855	117	0,00003	0,99993
83	0,02265	0,84120	118	0,00002	0,99995
84	0,02061	0,86181	119	0,00001	0,99996
85	0,01861	0,88042	120	0,00001	0,99997
86	0,01669	0,89711	121	0,00001	0,99998
87	0,01486	0,91197	122	0,00001	0,99999
88	0,01313	0,92510	123	≈ 0	0,99999
89	0,01153	0,93662	124	≈ 0	0,99999
90	0,01005	0,94667	125	≈ 0	0,99999
91	0,00870	0,95538	126	≈ 0	≈ 1
92	0,00749	0,96286	127	≈ 0	≈ 1
93	0,00640	0,96926	128	≈ 0	≈ 1
94	0,00543	0,97470	129	≈ 0	≈ 1
95	0,00459	0,97928	130	≈ 0	≈ 1
96	0,00384	0,98313	131	≈ 0	≈ 1
97	0,00320	0,98633	132	≈ 0	≈ 1
98	0,00265	0,98898	133	≈ 0	≈ 1
99	0,00218	0,99117	134	≈ 0	≈ 1
100	0,00179	0,99296	135	≈ 0	≈ 1
101	0,00145	0,99441	136	≈ 0	≈ 1
102	0,00118	0,99559	137	≈ 0	≈ 1
103	0,00095	0,99654	138	≈ 0	≈ 1
104	0,00076	0,99729	139	≈ 0	≈ 1

4. táblázat. 3.5.1. példa: az S konvolúció eloszlása 70-től 139 millió forintig

A példa további vizsgálata: mit kaptunk volna a normális eloszlásból?

Mivel a károkat reprezentáló X_i valószínűségi változókat függetleneknek és azonos eloszlásúaknak tekintettük, így konvolúciójuk eloszlását közelíthetjük megfelelő paraméterű normális eloszlással.

Legyen $m = E(X_i)$, $\sigma^2 = D^2(X_i)$.

Ekkor:

$$m = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 3 \cdot p_X(3) = 1 \cdot \frac{5.331}{9.877.365} + 3 \cdot \frac{626}{9.877.365} = \\ = \frac{7.209}{9.877.365} \approx 0{,}00073$$

$$\sigma^2 = E(X_i^2) - E^2(X_i) = \big(0^2 \cdot p_X(0) + 1^2 \cdot p_X(1) + 3^2 \cdot p_X(3)\big) - m^2 = \\ = \left(1 \cdot \frac{5.331}{9.877.365} + 9 \cdot \frac{626}{9.877.365}\right) - \left(\frac{7.209}{9.877.365}\right)^2 \approx 0{,}00111$$

A centrális határeloszlás-tétel alapján a standardizált összegre teljesül az alábbi:

$$F(z) = P\left(\frac{S - 100.000 \cdot m}{\sqrt{100.000 \cdot \sigma^2}} < z\right) \approx \phi(z)$$

Ezen összefüggést felhasználva becsülhető z_1 és z_2:

$$\hat{z}_1 = 98 \text{ (millió forintban)}$$
$$\hat{z}_2 = 101 \text{ (millió forintban)}$$

Összehasonlítva ezeket a valódi értékekkel: kijelenthető, hogy a normális közelítésből kapott $\hat{z}$-értékek nagyon jó becsléseknek bizonyultak.

	Pontos érték (millió forintban)	Normális közelítésből kapott érték (millió forintban)	A kettő eltérése (millió forintban)
z_1	99	98	1
z_2	102	101	1

5. táblázat. 3.5.1. példa: a pontos értékek összevetése a normális közelítésből kapottakkal

Hibabecslés

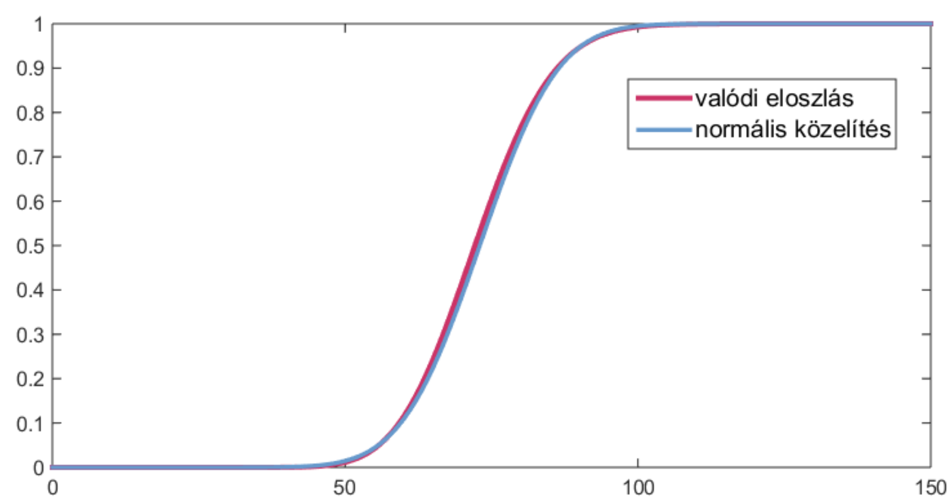
A normális közelítés hibáját független, azonos eloszlású valószínűségi változók esetében a Berry-Esséen-tétel segítségével becsülhetjük. A jelenlegi legjobb becslés a tételben szereplő konstansra Tyurin 2010-es eredménye alapján: 0,4785. ([4]) Így:

$$\sup_z |F(z) - \phi(z)| \le 0{,}4785 \cdot \frac{E|X_1 - m|^3}{\sigma^3 \cdot \sqrt{n}} =$$
$$= 0{,}4785 \cdot \left(\frac{\left(|0 - \frac{7.209}{9.877.365}|^3 \cdot p_X(0)\right) + \left(|1 - \frac{7.209}{9.877.365}|^3 \cdot p_X(1)\right) + \left(|3 - \frac{7.209}{9.877.365}|^3 \cdot p_X(3)\right)}{0{,}00111^{\frac{3}{2}} \cdot \sqrt{100.000}} \right)$$

Behelyettesítve a megfelelő értékeket:

$$\sup_z |F(z) - \phi(z)| \le 0{,}09$$

Megállapítható tehát, hogy példánkban a lehetséges eloszlásfüggvénybeli eltérés jóval kisebb a bevezető feladatban szereplőnél. Folytatván a vizsgálódást azt kapjuk, hogy a tényleges eltérés még a hibabecslésből adódó értéknek is csak a töredéke. Érdemes grafikusan is megjeleníteni $F(z)$ és $\phi(z)$ kapcsolatát. Látható, hogy ebben a példában olyannyira jó becslésnek bizonyult a normális közelítés, hogy annak eloszlásfüggvénye szinte rásimul a valódi eloszlásfüggvényre:



2. ábra. 3.5.1. példa: a valódi eloszlásfüggvény és a normális közelítés alapján kapott eloszlásfüggvény kapcsolata

3.5.2. 2. példa

A gyakorlatban az tapasztalható, hogy a valamely biztosítónál balesetbiztosítást kötő személyek esetében a közúti közlekedési balesetek bekövetkezések aránya relatíve alacsonyabb, mint a teljes lakosság körében. Érdemes tehát a tapasztalati valószínűségeket egy $0 < \alpha < 1$ szorzóval megszorozni, és ezen valószínűségekkel továbbdolgozni.

Használjuk továbbra is a Központi Statisztikai Hivatal 2014-es adatait! Vizsgáljuk meg, hogyan alakul az S konvolúció eloszlása $\alpha_1 = 0{,}9$ és $\alpha_2 = 0{,}8$ esetén, valamint melyek lesznek azon legkisebb forintértékek, melyek 99%-os, illetve 99,5%-os valószínűséggel fedezni fogják az összkár értékét! (Az α_1-hez tartozó értékek: z_1' és z_2', az α_2-höz tartozó értékek: z_1'' és z_2''.)

Megoldás

Akárcsak a korábbi feladatban, most is a 3.2.1. tételben látott rekurziót fogjuk alkalmazni a különböző valószínűségértékekhez tartozó konvolúció eloszlásának meghatározására.

Az $\alpha_1 = 0{,}9$-es szorzóhoz tartozó S' konvolúcióra:

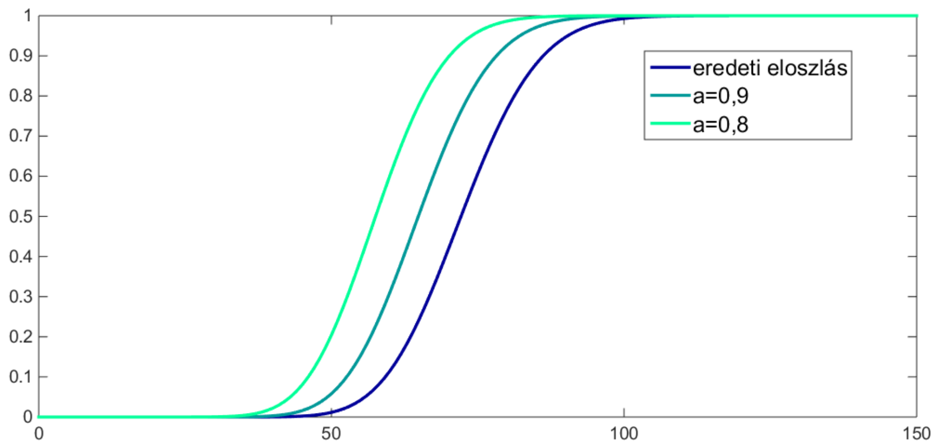
- $p'_X(1) = 0{,}9 \cdot \frac{5.331}{9.877.365}$
- $p'_X(3) = 0{,}9 \cdot \frac{626}{9.877.365}$
- $p'_X(0) = 1 - p'_X(1) - p'_X(3)$
- $p'_X(x) = 0$, ha $x \neq 0, 1, 3$

Az $\alpha_2 = 0{,}8$-es szorzóhoz tartozó S'' konvolúcióra:

- $p''_X(1) = 0{,}8 \cdot \frac{5.331}{9.877.365}$
- $p''_X(3) = 0{,}8 \cdot \frac{626}{9.877.365}$
- $p''_X(0) = 1 - p''_X(1) - p''_X(3)$
- $p''_X(x) = 0$, ha $x \neq 0, 1, 3$

Matlab-program segítségével meghatározható az S', illetve az S'' konvolúció eloszlása, továbbá megadhatóak a keresett z-forintértékek.

Kirajzolván az eredeti S, valamint a valószínűségértékek módosításával kapott S', illetve S'' konvolúciók eloszlásfüggvényét az alábbi ábrát kapjuk:



3. ábra. Az eredeti eloszlásfüggvény és az α_1, illetve α_2 szorzókkal vett módosításainak kapcsolata

Ha tehát nem a "tiszta" tapasztalati valószínűségekkel dolgozunk, hanem azoknak a gyakorlati tapasztalataink szerint vett lehetséges módosításaival, akkor az S konvolúció eloszlása lényegesen változik. Ennek egyik megnyilvánulása látható az ábrán is: az eloszlásfüggvény balra tolódik. Ezzel összefüggésben például a keresett z-forintértékek is jelentősen csökkenni fognak, amelynek gyakorlati következménye, hogy a biztosító számára jóval kevesebb pénzösszeg is elég lesz arra, hogy az 99%-os, illetve 99,5%-os valószínűséggel fedezze az összkifizetést.

Nézzük meg, hogy egészen pontosan mennyivel módosultak a keresett z-értékek! Leolvasható, hogy az α_1-es szorzó 9 millió forinttal, az α_2-es szorzó 17 millió forinttal csökkentette az eredeti z_1 és z_2 értékeit.

	Az eredeti eloszlásból kapott érték	$\alpha_1 = 0{,}9$-es módosítás esetén	$\alpha_2 = 0{,}8$-es módosítás esetén
z_1	99	90	82
z_2	102	93	85

6. táblázat. Az eredeti eloszlás alapján kapott értékek összehasonlítása a módosítások után kapottakkal

4. Független, nemnegatív egész értékű valószínűségi változók konvolúciója

4.1. Bevezetés

A korábbi fejezetben azon speciális esettel foglalkoztunk, melyben a valószínűségi változók függetlenek és azonos eloszáslúak. Most tegyük fel, hogy nem teljesül rájuk az utóbbi tulajdonság. Példaként vizsgáljuk a következő biztosítási modellt, melyben a biztosító baleseti halál esetén valamilyen fix összeget fizet, mely szerződésenként eltérhet, továbbá feltesszük, hogy ismert valamennyi biztosított esetében a baleseti halál egy bizonyos időintervallumon, például egy éven belüli bekövetkezési valószínűsége.

Soroljuk osztályokba a szerződéseket a baleseti halál esetén kifizetendő összeg és a bekövetkezési valószínűség szerint, majd az osztályok alapján foglaljuk táblázatba az adatokat!

Az i-edik oszlopban kifizetendő összeg i, a j-edik sorban a bekövetkezési valószínűség q_j. Kerüljön a táblázat i-edik oszlopának j-edik sorába az adott osztályba tartozó szerződések száma: c_{ij}.

	1	2	i	a
q_1				
q_2				
q_j			c_{ij}	
q_b				

7. táblázat. Segédtáblázat a nem azonos eloszlású modellhez

Használjuk az alábbi jelöléseket:

$p_j = 1 - q_j$, azaz a "túlélési" valószínűség az adott időintervallumon

$c_j = \sum\limits_{i=1}^{a} c_{ij}$, tehát a q_j bekövetkezési valószínűséghez tartozó szerződések száma

$c = \sum\limits_{j=1}^{b} c_j$, azaz az összes szerződés száma

$m = \sum\limits_{i=1}^{a} \sum\limits_{j=1}^{b} i \cdot c_j$, tehát az összes szerződésre vonatkozó maximálisan kifizetendő összeg

Jelölje S az összes szerződésre vonatkozó kifizetendő összeghez tartozó valószínűségi változót, eloszlása: $p_S(s) = P(S = s)$. A feladat ezen S valószínűségi változó eloszlásának meghatározása.

4.2. A rekurzió

4.2.1. Tétel. (De Pril, 1986, [5])

Legyenek

$$a(j,k) = \left(\frac{q_j}{p_j}\right)^k$$

$$A(i,k) = (-1)^{k+1} i \sum_{j=1}^{b} c_{ij} a(j,k)$$

Ekkor igazak az alábbi összefüggések:

$$p_S(0) = \prod_{j=1}^{b} (p_j)^{c_j} \tag{15}$$

$$s \cdot p_S(s) = \sum_{i=1}^{min(a,s)} \sum_{k=1}^{[s/i]} A(i,k) p_S(s - ki) \qquad s = 1, 2, \ldots, m \tag{16}$$

Bizonyítás. Minthogy S független valószínűségi változók konvolúciójaként áll elő, így generátorfüggvénye azok szorzata:

$$G_S(u) = \prod_{i=1}^{a} \prod_{j=1}^{b} (p_j + q_j u^i)^{c_{ij}}$$

Ebből $u = 0$ helyettesítéssel rögtön adódik a (15)-ös állítás.

A tétel második felének belátásához vegyük mindkét oldal logaritmusát:

$$\log(G_S(u)) = \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} \log(p_j + q_j u^i)$$

Ezt u szerint deriválva kapjuk:

$$\frac{G'_S(u)}{G_S(u)} = \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} \frac{q_j i u^{i-1}}{p_j + q_j u^i}$$

Átszorozva $G_S(u)$-val, majd továbbalakítva:

$$G'_S(u) = G_S(u) \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} i u^{i-1} \cdot \frac{q_j}{p_j} \cdot \frac{1}{1 + \frac{q_j}{p_j} u^i}$$

$\frac{1}{1+\frac{q_j}{p_j}u^i}$ -t hatványsorba fejtve:

$$G'_S(u) = G_S(u) \cdot \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} i u^{i-1} \frac{q_j}{p_j} \sum_{k=0}^{\infty} (-1)^k \left(\frac{q_j}{p_j} u^i \right)^k$$

Ezt továbbalakítva, majd felhasználva $a(j,k)$ és $A(i,k)$ definícióját:

$$\begin{aligned} G'_S(u) &= G_S(u) \cdot \sum_{i=1}^{a} \sum_{k=0}^{\infty} \left[(-1)^k (-1)^2 i \sum_{j=1}^{b} a(j, k+1) c_{ij} \right] u^{(k+1)i-1} = \\ &= G_S(u) \cdot \sum_{i=1}^{a} \sum_{k=0}^{\infty} A(i, k+1) u^{(k+1)i-1} = \\ &= G_S(u) \cdot \sum_{i=1}^{a} \sum_{k=1}^{\infty} A(i,k) u^{ki-1} \end{aligned}$$

A bizonyítás befejezése innentől kezdve hasonló az azonos eloszlású esetre felírt rekurziónál (3.2.1. tétel) látotthoz: a Leibniz-formulát felhasználva mindkét oldal $(s-1)$-edik deriváltját vesszük, majd $u = 0$ helyettesítéssel megkapjuk a (16)-os összefüggést. □

4.3. Egy speciális eset

Legyen $a = 1$, azaz vizsgáljuk azt a speciális esetet, mikor a biztosító kifizetése valamennyi káreset bekövetkezésekor ugyanannyi. Legyen N az összkárigény valószínűségi változója, ennek eloszlását keressük.

Legyen

$$A(k) = (-1)^{k+1} \sum_{j=1}^{b} c_j \left(\frac{q_j}{p_j} \right)^k$$

Ekkor az előzőek alapján N eloszlására teljesülnek az alábbiak:

$$p_N(0) = \prod_{j=1}^{b} (p_j)^{c_j} \tag{17}$$

$$n \cdot p_N(n) = \sum_{k=1}^{n} A(k) p_N(n-k) \qquad n = 1, 2, \ldots, c. \tag{18}$$

4.4. Alkalmazás

4.4.1. 1. példa

Tekintsük az alábbi egyszerűsített biztosítási modellt: egy biztosítónál biztosítás köthető közúti közlekedési balesetben bekövetkező halál esetére. A biztosító kifizetése minden haláleset bekövetkezésekor ugyanannyi: 3.000.000 forint.

A biztosító a halálesetek bekövetkezési valószínűségeit az alábbi módszerrel becsüli meg: rendelkezésére állnak az Országos Területfejlesztési és Területrendezési Információs Rendszer 2013-as, megyei bontású adatai (az állandó lakosság száma, valamint a közlekedési balesetben életüket vesztettek száma), forrás: *www.teir.hu*. Ezen adatok alapján a biztosító tapasztalati valószínűségekkel becsüli azt, hogy egy adott megyéből származó ember mekkora valószínűséggel fogja életét veszíteni közúti közlekedési balesetben egyéves időintervallumon belül.

Tegyük fel, hogy a biztosítónál 100.000 személy köt biztosítást, s ezen biztosítottak megyénkénti eloszlása megegyezik az ország teljes lakosságának megyénkénti eloszlásával.

Meghatározandó a rendelkezésre álló, 2013-as adatok alapján az összkár eloszlása, valamint azon legkisebb z_1 és z_2 számok értéke, amelyek 99%-os, illetve 99,5%-os valószínűséggel elegek lesznek az összkár kifizetésére. (Azaz keressük azon legkisebb z_1 és z_2 értékeket, melyekre $P(S \leq z_1) \geq 0{,}99$, illetve $P(S \leq z_2) \geq 0{,}995$.)

Megoldás

A feladat feltételei alapján a biztosító kifizetése egyféle lehet. Függetleneknek tekintve a reprezentáló valószínűségi változókat: teljesülni fognak a 4.2.1.-es tétel feltételei, azaz az összkár eloszlásának meghatározásakor alkalmazható a tételben felírt rekurzió.

Foglaljuk táblázatba az említett adatokat!

Megye	Bekövetkezett halálesetek száma	Állandó lakosság száma	Becsült bekövetkezési valószínűség	Szerződésszám
Bács-Kiskun	47	532822	$0,0000882$	5301
Baranya	18	392917	$0,0000458$	3909
Békés	17	366002	$0,0000464$	3641

8. táblázat. I. segédtáblázat a 4.4.1.-es példához

Megye	Bekövetkezett halálesetek száma	Állandó lakosság száma	Becsült bekövetkezési valószínűség	Szerződésszám
Borsod-Abaúj-Zemplén	24	700482	$0,0000343$	6969
Budapest	49	1698406	$0,0000289$	16897
Csongrád	25	419796	$0,0000596$	4176
Fejér	27	430136	$0,0000628$	4279
Győr-Moson-Sopron	40	448500	$0,0000892$	4462
Hajdú-Bihar	32	549515	$0,0000582$	5467
Heves	14	309351	$0,0000453$	3078
Jász-Nagykun-Szolnok	35	393312	$0,0000890$	3913
Komárom-Esztergom	24	312026	$0,0000769$	3104
Nógrád	8	203811	$0,0000393$	2028
Pest	63	1253772	$0,0000502$	12474
Somogy	19	325243	$0,0000584$	3236
Szabolcs-Szatmár-Bereg	27	578004	$0,0000467$	5750
Tolna	14	234202	$0,0000598$	2330
Vas	14	257718	$0,0000543$	2564
Veszprém	30	359103	$0,0000835$	3573
Zala	13	286331	$0,0000454$	2849

9. táblázat. II. segédtáblázat a 4.4.1.-es példához

Jelölje az összkifizetés valószínűségi változóját S. Az adatok felhasználásával, Matlab-programot írván az S konvolúció eloszlására a következő oldalon található táblázatot kapjuk. A táblázatból leolvasható a keresett z_1 és z_2 értéke:

- $z_1 = 33$ millió forint
- $z_2 = 36$ millió forint

Továbbá megadható S mediánja: $m = 15$ millió forint, valamint megmondható, hogy az az érték, aminél többet lényegében 0 valószínűséggel fog kifizetni a biztosító: 54 millió forint.

s (millió forint)	$P(S=s)$	$P(S \leq s)$	s (millió forint)	$P(S=s)$	$P(S \leq s)$
0	$0,00465$	$0,00465$	30	$0,02561$	$0,97823$
3	$0,02496$	$0,02960$	33	$0,01250$	$0,99073$
6	$0,06704$	$0,09664$	36	$0,00560$	$0,99633$
9	$0,12003$	$0,21667$	39	$0,00231$	$0,99864$
12	$0,16120$	$0,37787$	42	$0,00089$	$0,99953$
15	$0,17318$	$0,55105$	45	$0,00032$	$0,99985$
18	$0,15504$	$0,70609$	48	$0,00011$	$0,99995$
21	$0,11897$	$0,82506$	51	$0,00003$	$0,99999$
24	$0,07988$	$0,90494$	54	$0,00001$	≈ 1
27	$0,04768$	$0,95262$	57	≈ 0	≈ 1

10. táblázat. 4.4.1. példa: az S konvolúció eloszlása

A példa további vizsgálata: mit kaptunk volna, ha összesített adatokkal dolgozunk azok megyei bontása helyett?

Érdemes megvizsgálni, hogy hogyan alakul a konvolúció eloszlása abban az esetben, amikor nem területi bontású adatokkal dolgozunk, hanem azok összesítésével. A 8. és 9. táblázatok alapján 2013-ban az összlakosság száma: 10.051.449, közúti közlekedési balesetben elhunytak száma: 540. Ezen adatok felhasználásával kapjuk, hogy annak a tapasztalati valószínűsége, hogy egy személy egy adott évben közúti közlekedési balesetben életét veszti: $\frac{540}{10.051.449}$.

Jelölje X'_i az i. személyre vonatkozó kifizetést. X'_i-ket független és azonos eloszlású valószínűségi változóknak tekintve:

- $P(X'=3) = \frac{540}{10.051.449}$
- $P(X'=0) = 1 - \frac{540}{10.051.449}$
- $P(X'=k) = 0$, ha $k \neq 0, 3$

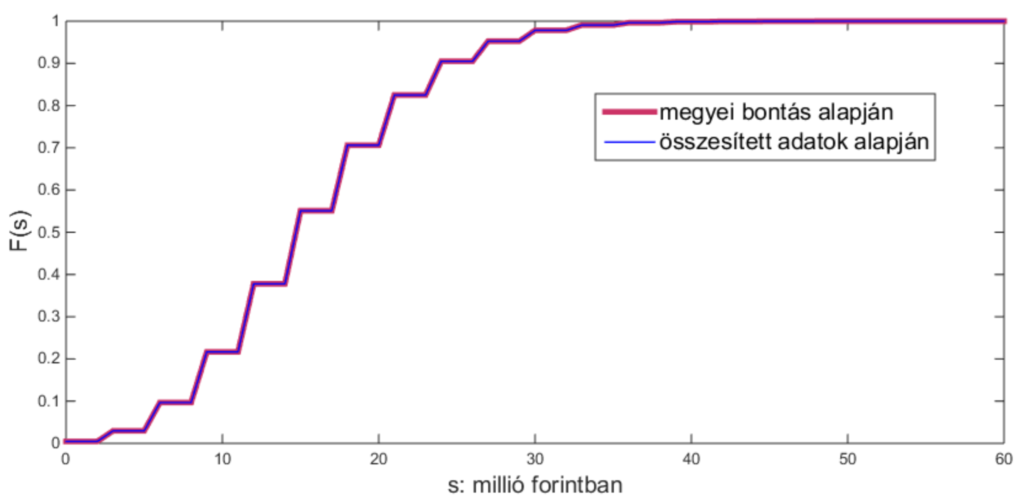
Ekkor az összkár: $S' = \sum_{i=1}^{100.000} X'_i$.

Meghatározandó S' eloszlása. Egy alkalmas módszer: az előzőekben is használt rekurzió, amely alapján a következő táblázatot kapjuk:

s (millió forint)	$P(S = s)$	$P(S \leq s)$	s (millió forint)	$P(S = s)$	$P(S \leq s)$
0	$0,00464$	$0,00465$	30	$0,02563$	$0,97821$
3	$0,02494$	$0,02958$	33	$0,01251$	$0,99072$
6	$0,06700$	$0,09659$	36	$0,00560$	$0,99632$
9	$0,11999$	$0,21658$	39	$0,00232$	$0,99864$
12	$0,16116$	$0,37774$	42	$0,00089$	$0,99953$
15	$0,17317$	$0,55091$	45	$0,00032$	$0,99985$
18	$0,15505$	$0,70597$	48	$0,00011$	$0,99995$
21	$0,11900$	$0,82497$	51	$0,00003$	$0,99999$
24	$0,07991$	$0,90488$	54	$0,00001$	≈ 1
27	$0,04770$	$0,95258$	57	≈ 0	≈ 1

11. táblázat. 4.4.1. példa további vizsgálata: az S' konvolúció eloszlása

Ezt összevetve a 10. táblázattal: kijelenthető, hogy szignifikáns különbségek nincsenek az S és az S' konvolúciók eloszlásai között, tehát az adatok megyei bontásával dolgozván lényegében hasonló eredményre jutunk, mintha azok összesítésével számolnánk. Ennek egy lehetséges magyarázata, hogy a különböző megyékből származó emberek esetében kicsik a különbségek a bekövetkezési valószínűségek között. Megjegyezhető tehát, hogy nem mindig érdemes egy bonyolult modellel számolni, mert elképzelhető, hogy az nem ad sokkal pontosabb eredményt egy egyszerű megközelítésnél.



4. ábra. 4.4.1.-es példa: S és S' konvolúciók eloszlása

4.4.2. 2. példa

Tekintsük egy –az előzővel rokon– egyszerűsített biztosítási modellt: egy biztosítónál biztosítás köthető közúti közlekedési balesetben bekövetkező halál esetére. A biztosító kifizetése kétféle lehet a megkötött biztosítástól függően:

- I. típusú biztosítás esetén: 5.000.000 forint
- II. típusú biztosítás esetén: 3.000.000 forint

A biztosító a halálesetek bekövetkezési valószínűségeit az előző példában (4.4.1.) bemutatott módszerrel becsüli, ugyancsak a 2013-as adatok felhasználásával. Tegyük fel, hogy most is 100.000 a biztosítottak száma, megyénkénti eloszlásuk megegyezik a teljes lakosság megyénkénti eloszlásával. Az, hogy hányan kötnek a különböző biztosítási fajtákból, az alábbi táblázatból olvasható le:

Megye	Becsült bekövetkezési valószínűség	I. típusú biztosítások száma	II. típusú biztosítások száma	Az összes szerződés száma
Bács-Kiskun	0,0000882	560	4741	5301
Baranya	0,0000458	391	3518	3909
Békés	0,0000464	401	3240	3641
Borsod-Abaúj-Zemplén	0,0000343	751	6218	6969
Budapest	0,0000289	1702	15195	16897
Csongrád	0,0000596	406	3770	4176
Fejér	0,0000628	495	3784	4279
Győr-Moson-Sopron	0,0000892	394	4068	4462
Hajdú-Bihar	0,0000582	521	4946	5467
Heves	0,0000453	310	2768	3078
Jász-Nagykun-Szolnok	0,0000890	357	3556	3913
Komárom-Esztergom	0,0000769	275	2829	3104
Nógrád	0,0000393	249	1779	2028
Pest	0,0000502	1256	11218	12474
Somogy	0,0000584	358	2878	3236

12. táblázat. I. segédtáblázat a 4.4.2.-es példához

Megye	Becsült bekövetkezési valószínűség	I. típusú biztosítások száma	II. típusú biztosítások száma	Összes szerződés száma
Szabolcs-Szatmár-Bereg	0,0000467	569	5181	5750
Tolna	0,0000598	234	2096	2330
Vas	0,0000543	277	2287	2564
Veszprém	0,0000835	388	3185	3573
Zala	0,0000454	298	2551	2849

13. táblázat. II. segédtáblázat a 4.4.2.-es példához

A feladatunk meghatározni a rendelkezésre álló adatok alapján az összkár(S) eloszlását, valamint azon legkisebb z_1 és z_2 értékeket, amelyek 99%-os, illetve 99,5%-os valószínűséggel elegek lesznek az összkár kifizetésére. (Azaz keressük azon legkisebb z_1-et és z_2-t, melyekre $P(S \leq z_1) \geq 0{,}99$, illetve $P(S \leq z_2) \geq 0{,}995$.)

Megoldás

Akárcsak az előző példában, a függetlenséget feltéve most is teljesülnek a 4.2.1-es tétel feltételei, így S eloszlásának meghatározásakor használhatjuk a tételben felírt rekurziót. Matlab-program segítségével S eloszlására a következő oldalon található táblázatot kapjuk, amiből le is olvashatóak az általunk keresett z_1 és z_2 értékei:

- $z_1 = 37$ millió forint
- $z_2 = 40$ millió forint

A táblázat adatainak elemzése

S eloszlásának meghatározásával megadható a konvolúció mediánja: $m = 17$ millió forint, valamint az az érték, aminél többet lényegében 0 valószínűséggel kell majd kifizetnie a biztosítónak: 60 millió forint.

s (millió forint)	$P(S = s)$	$P(S \leq s)$	s (millió forint)	$P(S = s)$	$P(S \leq s)$
0	$0,00464$	$0,00464$	31	$0,00836$	$0,95835$
1	0	$0,00464$	32	$0,01031$	$0,96867$
2	0	$0,00464$	33	$0,00606$	$0,97473$
3	$0,02241$	$0,02705$	34	$0,00507$	$0,97980$
4	0	$0,02705$	35	$0,00517$	$0,98496$
5	$0,00252$	$0,02958$	36	$0,00307$	$0,98803$
6	$0,05411$	$0,08369$	37	$0,00274$	$0,99077$
7	0	$0,08369$	38	$0,00240$	$0,99318$
8	$0,01219$	$0,09588$	39	$0,00149$	$0,99467$
9	$0,08710$	$0,18298$	40	$0,00134$	$0,99602$
10	$0,00069$	$0,18367$	41	$0,00105$	$0,99707$
11	$0,02943$	$0,21310$	42	$0,00069$	$0,99776$
12	$0,10516$	$0,31825$	43	$0,00060$	$0,99837$
13	$0,00331$	$0,32157$	44	$0,00044$	$0,99880$
14	$0,04737$	$0,36894$	45	$0,00030$	$0,99911$
15	$0,10168$	$0,47062$	46	$0,00025$	$0,99936$
16	$0,00800$	$0,47862$	47	$0,00018$	$0,99954$
17	$0,05719$	$0,53582$	48	$0,00013$	$0,99966$
18	$0,08234$	$0,61815$	49	$0,00010$	$0,99976$
19	$0,01288$	$0,63103$	50	$0,00007$	$0,99983$
20	$0,05525$	$0,68628$	51	$0,00005$	$0,99988$
21	$0,05783$	$0,74412$	52	$0,00004$	$0,99991$
22	$0,01555$	$0,75967$	53	$0,00002$	$0,99994$
23	$0,04454$	$0,80420$	54	$0,00002$	$0,99996$
24	$0,03637$	$0,84057$	55	$0,00001$	$0,99997$
25	$0,01502$	$0,85559$	56	$0,00001$	$0,99998$
26	$0,03086$	$0,88645$	57	$0,00001$	$0,99999$
27	$0,02108$	$0,90753$	58	≈ 0	$0,99999$
28	$0,01210$	$0,91963$	59	≈ 0	$0,99999$
29	$0,01883$	$0,93846$	60	≈ 0	≈ 1
30	$0,01154$	$0,94999$	61	≈ 0	≈ 1

14. táblázat. 4.4.2. példa: az S konvolúció eloszlása

5. Összefoglalás

A dolgozatban bemutatásra került több módszer is arra vonatkozólag, hogy generátorfüggvények segítségével miképpen tudjuk meghatározni független, nemnegatív egész értékű valószínűségi változók konvolúciójának eloszlását. A bemutatott eljárásokat konkrét példákra alkalmaztuk. Az eredmények megmutatták, hogy sok esetben óvatosan kell eljárni a normális közelítések alkalmazásaival kapcsolatban, ugyanakkor azt is láthattuk, hogy a tényleges hibák lényegesen kisebbek is lehetnek, mint az Esséen-típusú hibabecslésből adódóak.

Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani mindazoknak, akik segítették dolgozatom létrejöttét.

Mindenekelőtt hálás köszönettel tartozom konzulensemnek, Arató Miklósnak, amiért megismertetett a témával, emellett útmutatásaival, rengeteg hasznos ötletével és észrevételével mindvégig fáradhatatlanul segítette a munkámat.

Szeretném megköszönni továbbá Zempléni Andrásnak és Varga Lászlónak, hogy egyetemi éveim alatt felkeltették az érdeklődésemet a valószínűségszámítás és annak alkalmazásai iránt.

Végül, de nem utolsó sorban köszönet illeti családomat és barátaimat az elmúlt évek során nyújtott kitartó támogatásukért.

Romvári Petra

Hivatkozások

[1] Shevtsova, I. G.: *An Improvement of Convergence Rate Estimates in the Lyapunov Theorem*, Doklady Mathematics, 82(3): 862–864, 2010.

[2] De Pril, N.: *Recursions for Convolutions of Arithmetic Distributions*, ASTIN Bulletin, 15(2): 135–139, 1985.

[3] Központi Statisztikai Hivatal: Évközi adatok – Egészségügy, baleset (http://www.ksh.hu/docs/hun/xstadat/xstadat_evkozi/e_ods001.html)

[4] Tyurin, I. S.: *An improvement of upper estimates of the constants in the Lyapunov theorem*, Russian Mathematical Surveys, 65(3(393)): 201–202, 2010.

[5] De Pril, N.: *On the Exact Computation of the Aggregate Claims Distribution in the Individual Life Model*, ASTIN Bulletin, 16(2): 109–112, 1986.