A negatív binomiális eloszlás paramétereinek becslése

Szakdolgozat

Írta: Királyné Maczkó Renáta
alkalmazott matematikus MSc hallgató

Témavezető: Arató Miklós
egyetemi docens
Eötvös Loránd Tudományegyetem
Természettudományi Kar
Matematikai Intézet
Valószínűségelméleti és Statisztika Tanszék

Budapest, 2013. január 7.

Tartalomjegyzék

Táblázatok jegyzéke **6**

1. Bevezetés **7**

2. A negatív binomiális eloszlás **8**

2.1. Momentum módszer 10

2.2. Maximum likelihood módszer 11

2.3. Maximum kvázi-likelihood módszer 13

2.3.1. Kiterjesztett maximum kvázi-likelihood módszer 14

2.3.2. Duplán kiterjesztett kvázi likelihood módszer 15

2.4. Szimulációk és eredmények 16

2.4.1. Momentum módszer 17

2.4.2. Maximum likelihood módszer 25

2.4.3. Kiterjesztett maximum kvázi-likelihood módszer 29

2.4.4. Duplán kiterjesztett maximum kvázi-likelihood módszer 33

3. A változó időtartamú megfigyelések esete **38**

3.1. Momentum módszer 38

3.2. Maximum likelihood módszer 39

3.3. Kiterjesztett maximum kvázi-likelihood módszer 40

3.4. Duplán kiterjesztett maximum kvázi-likelihood módszer 41

3.5. Szimulációk és eredmények 42

3.5.1. Momentum módszer 42

3.5.2. Maximum likelihood módszer 47

3.5.3. Kiterjesztett maximum kvázi-likelihood módszer 50

3.5.4. Duplán kiterjesztett maximum kvázi-likelihood módszer 54

4. Összefoglalás **59**

Táblázatok jegyzéke

2.1. 10000 szimulációs momentum módszeres becslésem n=50 minta elemszám esetén 18

2.2. Clark és Perry [4] 10000 szimulációs momentum módszeres becslése n=50 minta elemszám esetén . . . 18

2.3. 10000 szimulációs (a harmadik momentum használatával kapott) momentum módszeres becslésem n=50 minta elemszám esetén . . . 19

2.4. 10000 szimulációs momentum módszeres becslésem kiegészítése n=50 minta elemszám esetén . . . 20

2.5. 10000 szimulációs (a harmadik momentum használatával kapott) momentum módszeres becslésem kiegészítése n=50 minta elemszám esetén . . . 20

2.6. 1000 szimulációs momentum módszeres becslésem . . . 21

2.7. Clark és Perry [4] 1000 szimulációs momentum módszeres becslése . . . 22

2.8. 1000 szimulációs (a harmadik momentum használatával kapott) momentum módszeres becslésem . . . 23

2.9. 1000 szimulációs momentum módszeres becslésem kiegészítése . . . 24

2.10. 1000 szimulációs (a harmadik momentum használatával kapott) momentum módszeres becslésem kiegészítése . . . 24

2.11. 10000 szimulációs maximum likelihood becslésem n=50 minta elemszám esetén . 25

2.12. Piegorsch [8] 10000 szimulációs maximum likelihood becslése n=50 minta elemszám esetén . . . 26

2.13. 10000 szimulációs maximum likelihood becslésem kiegészítése n=50 minta elemszám esetén . . . 26

2.14. 1000 szimulációs maximum likelihood becslésem . . . 27

2.15. Piegorsch [8] 1000 szimulációs maximum likelihood becslése . . . 28

2.16. 1000 szimulációs maximum likelihood becslésem kiegészítése . . . 28

2.17. 10000 szimulációs kiterjesztett maximum kvázi-likelihood becslésem n=50 minta elemszám esetén . . . 29

2.18. Clark és Perry [4] 10000 szimulációs kiterjesztett maximum kvázi-likelihood becslése n=50 minta elemszám esetén . . . 30
2.19. 10000 szimulációs kiterjesztett maximum kvázi-likelihood becslésem kiegészítése n=50 minta elemszám esetén . . . 31
2.20. 1000 szimulációs kiterjesztett maximum kvázi-likelihood becslésem . . . 31
2.21. Clark és Perry [4] 1000 szimulációs kiterjesztett maximum kvázi-likelihood becslése 32
2.22. 1000 szimulációs kiterjesztett maximum kvázi-likelihood becslésem kiegészítése . 33
2.23. 10000 szimulációs duplán kiterjesztett maximum kvázi-likelihood becslésem n=50 minta elemszám esetén . . . 34
2.24. 10000 szimulációs duplán kiterjesztett maximum kvázi-likelihood becslésem kiegészítése n=50 minta elemszám esetén . . . 35
2.25. 1000 szimulációs duplán kiterjesztett maximum kvázi-likelihood becslésem . . . 36
2.26. 1000 szimulációs duplán kiterjesztett maximum kvázi-likelihood becslésem kiegészítése . . . 37

3.1. 10000 szimulációs változó időtartamú momentum módszeres becslésem n=50 minta elemszám esetén . . . 42
3.2. 10000 szimulációs változó időtartamú momentum módszeres becslésem kiegészítése n=50 minta elemszám esetén . . . 43
3.3. 1000 szimulációs változó időtartamú momentum módszeres becslésem . . . 44
3.4. 1000 szimulációs változó időtartamú momentum módszeres becslésem kiegészítése 45
3.5. 10000 szimulációs változó időtartamú momentum módszeres becslésem (a másodszorra ismertetett módszerrel) n=50 minta elemszám esetén . . . 45
3.6. 10000 szimulációs változó időtartamú momentum módszeres becslésem (a másodszorra ismertetett módszerrel) kiegészítése n=50 minta elemszám esetén . . . 46
3.7. 1000 szimulációs változó időtartamú momentum módszeres becslésem (a másodszorra ismertetett módszerrel) . . . 46
3.8. 1000 szimulációs változó időtartamú momentum módszeres becslésem (a másodszorra ismertetett módszerrel) kiegészítése . . . 47
3.9. 10000 szimulációs változó időtartamú maximum likelihood becslésem n=50 minta elemszám esetén . . . 48
3.10. 10000 szimulációs változó időtartamú maximum likelihood becslésem kiegészítése n=50 minta elemszám esetén . . . 49
3.11. 1000 szimulációs változó időtartamú maximum likelihood becslésem . . . 50
3.12. 1000 szimulációs változó időtartamú maximum likelihood becslésem kiegészítése . 51

3.13. 10000 szimulációs változó időtartamú kiterjesztett maximum kvázi-likelihood becslésem n=50 minta elemszám esetén 52
3.14. 10000 szimulációs változó időtartamú kiterjesztett maximum kvázi-likelihood becslésem kiegészítése n=50 minta elemszám esetén 52
3.15. 1000 szimulációs változó időtartamú kiterjesztett maximum kvázi-likelihood becslésem 53
3.16. 1000 szimulációs változó időtartamú kiterjesztett maximum kvázi-likelihood becslésem kiegészítése 54
3.17. 10000 szimulációs változó időtartamú duplán kiterjesztett maximum kvázi-likelihood becslésem n=50 minta elemszám esetén 55
3.18. 10000 szimulációs változó időtartamú duplán kiterjesztett maximum kvázi-likelihood becslésem kiegészítése n=50 minta elemszám esetén 55
3.19. 1000 szimulációs változó időtartamú duplán kiterjesztett maximum kvázi-likelihood becslésem 56
3.20. 1000 szimulációs változó időtartamú duplán kiterjesztett maximum kvázi-likelihood becslésem kiegészítése 57

1. fejezet

Bevezetés

A dolgozatban, mint ahogy a cím is sugallja a negatív binomiális eloszlást fogom körüljárni. Ez az eloszlás azért fontos, mert több alkalmazásnál (például a biztosítási kárszámoknál) megfigyelték, hogy a negatív binomiális eloszlás jól illeszkedik az adatokra, gyakran lényegesen jobban, mint a természetesen adódó Poisson eloszlás. A jó illeszkedés egyik lehetséges magyarázata az, hogy a negatív binomiális eloszlás egyben keverék Poisson eloszlás is (gamma keverővel).

A szakdolgozatomban ismertetem a momentum módszert, a maximum likelihood módszert, a kiterjesztett maximum kvázi-likelihood módszert és a duplán kiterjesztett maximum kvázi-likelihood módszert. Ezeket a módszereket alkalmazom a negatív binomiális eloszlás paramétereinek a becslésére, majd generált adatokat véve tesztelem ezeket a módszereket, végül elemzem és összehasonlítom a kapott eredményeimet a cikkekben látott eredményekkel.

A szimuláláshoz az R statisztikai programot hívtam segítségül, melyben generáltam adott paraméterekre egyre nagyobb elemszámú mintákat a negatív binomiális eloszlásból, majd ezekre az adatokra lefuttattam a fent említett módszereket, és az eredmények különböző statisztikai mutatóit táblázatokba rendeztem.

A 2. fejezetben az állandó megfigyelésű idejű, a 3. fejezetben a változó megfigyelésű esetet vizsgálom. A 2. fejezet elején az egyetemen tanultak alapján és Arató [2] jegyzete alapján összefoglalom a negatív binomiális eloszlás fontosabb tulajdonságait, és [2] alapján megmutatom, hogy keverék Poisson eloszlás és speciális $(a, b, 0)$ eloszlású. Az ezt követő szakaszokban kiszámítom a különböző módszerek által kapott becsléseket, illetve a nem tanult kvázi-likelihood módszereknél McCullach és Nelder [5] könyve, valamint több cikk alapján a módszereket is kicsit részletesebben ismertetem. A megfelelő szakaszokban a 3. fejezetben is kiszámolom a módszerekből kapott becsléseket. Mindkét fejezet a szimulációs eredményeket tartalmazó szakasszal zárul (2.4. és 3.5. szakaszok). Ezekben összehasonlítom eredményeimet egymással, valamint a vizsgált cikkek eredményeivel. A 4. fejezetben röviden összefoglalom a kapott eredményeket.

2. fejezet

A negatív binomiális eloszlás

Először definiáljuk a negatív binomiális kicsit általánosabban a megszokottnál, ahogy a feldolgozott cikkekben is szerepel. Erre azért van szükség, mert ezt az általánosabb változatot lehet sikeresen alkalmazni több különféle modellben (például biztosításban, pénzügyi matematikában, demográfiában), továbbá így az én eredményeimet is jobban össze lehet hasonlítani a cikkekben lévő táblázatokkal. Az általánosabb definíciónak is több változata létezik, viszont ezek csak paraméterezésükben különböznek egymástól. Mivel nem akarjuk, hogy az átparaméterezés miatt adódjanak különbségek az eredmények között, a feldolgozott cikkekben használt paraméterezést fogjuk használni.

A definíció után megadom néhány olyan ismert tulajdonságát a negatív binomiális eloszlásnak, amelyre szükségünk lesz a továbbiakban és megmutatom, hogy a negatív binomiális eloszlás miért lesz keverék Poisson eloszlás.

2.1. Definíció. $Y \sim NB(\mu, \alpha)$ eloszlású, ha

$$P(Y = y) = \frac{\Gamma\left(y + \alpha^{-1}\right)}{y!\,\Gamma\left(\alpha^{-1}\right)} \left(\frac{\mu}{\mu + \alpha^{-1}}\right)^{y} (1 + \alpha\mu)^{-\frac{1}{\alpha}}, \tag{2.1}$$

ahol $y = 0,1,2...$, $\alpha > 0$, $\mu > 0$.

Ekkor $\mu = E(Y)$ és $D^2(Y) = \mu + \alpha\mu^2$.

2.2. Definíció. A gamma függvényt a következő képlettel lehet definiálni:

$$\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt.$$

Parciális integrálással a gamma függvény alábbi számunkra igen fontos tulajdonságához jutunk:

$$\Gamma(s) = s\Gamma(s-1),$$

ha $s > 1$. Ezenkívül ismert, hogy $\Gamma(1) = 1$. Ebből a két tulajdonságból következik, hogy $\Gamma(n) = (n-1)!$, tehát a gamma függvény a faktoriális függvénynek egy folytonos kiterjesztése.

Ha a 2.1. Definíció által adott eloszlást átparaméterezzük úgy, hogy $r = \frac{1}{\alpha}$, $q = \frac{\mu}{\mu+\alpha^{-1}}$, akkor az eloszlás azon paraméterezését kapjuk, ami sokkal jobban hasonlít a „hagyományos" változathoz:

2.3. Definíció. $X \sim NB(r, q)$ eloszlású, ha

$$P(X = n) = \frac{\Gamma(r+n)}{\Gamma(r)\, n!} (1-q)^r q^n, \tag{2.2}$$

ahol $n = 0,1,\ldots$, $r > 0$, $0 \leqslant q < 1$.

$E(X) = \frac{r}{1-q}$, $D^2(X) = \frac{rq}{(1-q)^2}$.

2.4. Megjegyzés. Amennyiben r pozitív egész, akkor $\eta = X + r$ eloszlása a „hagyományos" negatív binomiális lesz, hiszen ekkor

$$P(\eta = n) = \binom{n-1}{r-1} (1-q)^r q^{n-r},$$

ahol $n=r,r+1,\ldots$

A továbbiakban a 2.1. Definíciónak megfelelően jelöljük az eloszlásokat.

Felsoroljuk az eloszlás legfontosabb jellemzőit

- generátorfüggvénye: $G(z) = E\left(z^Y\right) = \left(\frac{1}{1+\alpha\mu(1-z)}\right)^{\frac{1}{\alpha}}$, $|z| < 1 + \frac{1}{\alpha\mu}$
- momentumgeneráló függvénye: $M(t) = E\left(e^{tY}\right) = \left(\frac{1}{1+\alpha\mu(1-e^t)}\right)^{\frac{1}{\alpha}}$, $t < -\log\frac{\mu}{\mu+\alpha^{-1}}$
- karakterisztikus függvénye: $\varphi(t) = E\left(e^{itY}\right) = \left(\frac{1}{1+\alpha\mu(1-e^{\mathrm{i}t})}\right)^{\frac{1}{\alpha}}$, $t \in \mathbb{R}$
- második momentuma: $E\left(Y^2\right) = \mu + \mu^2(1+\alpha)$
- ferdesége: $\beta_1(Y) = \frac{E(Y-E(Y))^3}{\left(E(Y-E(Y))^2\right)^{\frac{3}{2}}} = \frac{2\alpha\mu+1}{\sqrt{\alpha\mu^2+\mu}}$
- lapultsága: $\beta_2(Y) = \frac{E(Y-E(Y))^4}{\left(E(Y-E(Y))^2\right)^2} - 3 = 6\alpha + \frac{1}{\alpha\mu^2+1}$

Most megmutatjuk azt a jól ismert tényt, hogy a negatív binomiális eloszlást megkaphatjuk, mint keverék Poisson eloszlást.

Legyen ν egy $\left(\lambda, \frac{1}{\alpha}\right)$ paraméterű gamma eloszlás. Továbbá tegyük fel, hogy $\eta(t)$ rögzített $\nu = v$ melletti feltételes eloszlása egy vt paraméterű Poisson eloszlás. Ekkor $\eta(t)$ generátorfüggvénye:

$$G_{\eta(t)}(z) = E\left(E\left(z^{\eta(t)}\middle|\nu\right)\right) = E\left(e^{\nu t(z-1)}\right) = L_\nu(t(1-z)) =$$

$$= \left(\frac{\lambda}{\lambda + t(1-z)}\right)^{\frac{1}{\alpha}} = \left(\frac{1}{1 + \alpha\frac{t}{\alpha\lambda}(1-z)}\right)^{\frac{1}{\alpha}}. \tag{2.3}$$

Itt kihasználtuk, hogy a Poisson eloszlás generátorfüggvénye

$$G_{\eta(t)|v}(z) = E\left(z^{\eta(t)}\middle|v\right) = e^{vt(z-1)},$$

a gamma elosztás Laplace transzformáltja pedig

$$L_\nu(s) = \left(\frac{\lambda}{\lambda + s}\right)^{\frac{1}{\alpha}}.$$

Azt vesszük észre, hogy ez nem más, mint a $\left(\frac{t}{\alpha\lambda}, \alpha\right)$ paraméterű negatív binomiális eloszlás generátorfüggvénye. Ezzel beláttuk, hogy a negatív binomiális eloszlás egyben egy gamma keverővel ellátott keverék Poisson eloszlás.

Általában (numerikus szempontból) nagyon jól használható a diszkrét eloszlások következő családja.

2.5. Definíció. A nem-negatív egész értékű ξ valószínűségi változó $(a, b, 0)$ eloszlású, ha

$$P(\xi = n) = \left(a + \frac{b}{n}\right) P(\xi = n - 1),$$

ahol $n = 1,2, \ldots$

2.6. Megjegyzés. A fenti eloszláscsaládot Katz eloszláscsaládnak is szokták nevezni.

A negatív binomiális eloszlás az $(a, b, 0)$ eloszlásnak egy speciális esete, ahol $a = \frac{\alpha\mu}{\alpha\mu+1}$ és $b = \frac{\mu-\alpha\mu}{\alpha\mu+1}$.

2.1. Momentum módszer

Ezt a becslési eljárást általában akkor használjuk, ha a minta eloszlásának sok ismeretlen paramétere van. Mi most az eloszlásunk mindkét paraméterét szeretnénk becsülni és bízunk a jó közelítésben.

Az első két momentumot ismerjük. $E(Y) = \mu$, $E\left(Y^2\right) = \mu + \mu^2 + \alpha\mu^2$. Két egyenletünk van két ismeretlennel, így a két becslésre a következőt kapjuk, ha M_i-vel jelöljük az i-edik tapasztalati momentumot:

$$\hat{\mu} = M_1,$$
$$\hat{\alpha} = \frac{M_2 - M_1 - M_1^2}{M_1^2}.$$

Megjegyezzük, hogy sokszor, így Clark és Perry [4] cikkében is, a második momentumot használó momentum módszeres becsléseket a szórásból vezetik le a második momentum helyett, és a szórást a korrigált tapasztalati szórással becsülik. Később szimulációs eredményeinkből látni fogjuk, hogy ezzel a módosítással valóban pontosabb becslést kapunk, főleg kis minta elemszám esetén. A teljesség kedvéért megjegyezzük, hogy amennyiben D^2 jelöli a korrigált tapasztalati szórásnégyzetet, az így kapott becslés α-ra:

$$\hat{\alpha} = \frac{D^2 - M_1}{M_1^2}.$$

Most nézzük, hogy hogyan változik a becslésünk, ha az első és a harmadik momentumot vesszük. A harmadik momentum kiszámításához a momentumgeneráló függvényt használtam, mivel ha $M(t)$ véges a $t = 0$ pont egy környezetében, akkor ott akárhányszor deriválható és ezek a deriváltak a $t = 0$ pontban az Y momentumait adják.

$$E\left(Y^3\right) = M'''(0) = \mu^3 + 3\mu^2 + \mu + 2\alpha^2\mu^3 + 3\alpha\mu^3 + 3\alpha\mu^2.$$

Az első momentumból megkaptuk a μ becslését, ezért a harmadik momentumnál α-ra egy másodfokú egyenletet kell megoldanunk. Így α-ra két becslésünk is lehetne, de mivel α pozitív, ezért csak az egyik gyök lehet alkalmas.

$$\hat{\alpha}_2 = \frac{-3M_1^2 - 3M_1 + \sqrt{8M_3M_1 + M_1^2 - 6M_1^3 + M_1^4}}{4M_1^2},$$
$$\hat{\mu} = M_1.$$

2.2. Maximum likelihood módszer

A paraméterek becslését ebben a módszerben az adott minta likelihood függvény maximumhelyének vagy a likelihood függvény logaritmusának a maximumhelyének megkeresésével határozzuk meg. A maximum likelihood módszer a paraméterbecslés egyik leggyakrabban alkalmazott eljárása, mivel torzítatlan és aszimptotikusan hatásos. Ennek a módszernek sajnos az a fő hátránya, hogy bizonyos esetekben a likelihood egyenletek nehezen oldhatóak meg, így numerikus közelítést kell alkalmaznunk, ami kissé pontatlanná teszi a becslésünket (és a futásidőt is növeli).

2.7. Definíció. Legyen $\mathcal{P}_\Theta$ a Θ paramétercsaláddal paraméterezett eloszláscsalád, melyben a $\theta \in \Theta$ paraméterű eloszlás sűrűségfüggvénye f_θ. Ekkor az $Y = (y_1, y_2, \ldots, y_n)$ független, azonos ($\mathcal{P}_\Theta$-ból) származó eloszlású minta $L_Y : \Theta \to \mathbb{R}$ likelihood függvénye:

$$L_Y(\theta) = \prod_{i=1}^{n} f_\theta(y_i), \theta \in \Theta.$$

A maximalizálás szempontjából mindegy, hogy a likelihood függvényt vagy annak a logaritmusát maximalizáljuk, ezért vegyük a negatív binomiális eloszlást és írjuk fel a log-likelihood függvényét, majd próbáljuk megbecsülni az α paramétert.

$$\widetilde{l}(\alpha, \mu) = \sum_{i=1}^{n} \log\left(\frac{\Gamma\left(y_i + \alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right)}\right) - \sum_{i=1}^{n} \log y_i! + \sum_{i=1}^{n} y_i \log(\mu) + \sum_{i=1}^{n} y_i \log(\alpha) -$$

$$- \sum_{i=1}^{n} \left(y_i + \alpha^{-1}\right) \log(1 + \alpha\mu) \tag{2.4}$$

A likelihood függvény maximalizálása szempontjából a $\sum_{i=1}^{n} \log y_i!$ tag egyáltalán nem játszik szerepet, ezért ez a tag elhagyható. Az sem változtat a becslésünkön, ha leosztunk n-nel. Így a következő kifejezéshez jutunk, amit szintén l-el jelölünk:

$$l(\alpha, \mu) = \frac{1}{n} \sum_{i=1}^{n} \log\left(\frac{\Gamma\left(y_i + \alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right)}\right) + \bar{y} \log(\mu) - \left(\bar{y} + \alpha^{-1}\right) \log(1 + \alpha\mu) + \bar{y} \log(\alpha) \tag{2.5}$$

Lawless [6] 1987-ben megmutatta, hogy a két gamma függvény hányadosának a logaritmusa egy olyan összeggel helyettesíthető, amiben már nincs szükség a gamma függvényre, így a gamma hányados egy összegképlettel helyettesíthető.

$$\Gamma\left(\left(y_i + \alpha^{-1}\right) / \Gamma\left(\alpha^{-1}\right)\right. = \alpha^{-1}\left(1 + \alpha^{-1}\right) \cdots \left(y_i - 1 + \alpha^{-1}\right)$$

$$\log\left(\frac{\Gamma\left(y_i + \alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right)}\right) = \sum_{\nu=0}^{y_i - 1} \log\left(\frac{(1 + \alpha\nu)}{\alpha}\right) = \sum_{\nu=0}^{y_i - 1} \log(1 + \alpha\nu) - y_i \log(\alpha)$$

Ebből azt kapjuk, hogy

$$l(\alpha, \mu) = \frac{1}{n} \sum_{i=1}^{n} \sum_{\nu=0}^{y_i - 1} \log(1 + \alpha\nu) + \bar{y} \log(\mu) - \left(\bar{y} + \alpha^{-1}\right) \log(1 + \alpha\mu). \tag{2.6}$$

Deriválás után pedig a következőket kapjuk:

$$\frac{\partial l(\alpha, \mu)}{\partial \mu} = \frac{\bar{y}}{\mu} - \frac{1 + \alpha\bar{y}}{1 + \alpha\mu}, \tag{2.7}$$

$$\frac{\partial l\left(\alpha,\mu\right)}{\partial\alpha}=\frac{1}{n}\sum_{i=1}^{n}\sum_{\nu=0}^{y_i-1}\frac{\nu}{1+\alpha\nu}+\frac{1}{\alpha^2}\log\left(1+\alpha\mu\right)-\frac{\mu\left(\bar{y}+\alpha^{-1}\right)}{1+\alpha\mu}. \tag{2.8}$$

Ha a (2.7) kifejezést egyenlővé tesszük nullával, akkor rögtön adódik, hogy a $\hat{\mu}=\bar{y}$ egy jó választás lesz. A (2.2) kifejezés zérushelyének egzakt meghatározása bonyolult, ezért helyette numerikus közelítést alkalmazunk. Mi a szimulációink során az R programmal a (2.6) képlet által adott log-likelihood függvényt maximalizáljuk az α változóra a $\mu=\bar{y}$ helyen.

2.3. Maximum kvázi-likelihood módszer

Ebben a szakaszban McCullagh és Nelder [5] könyve alapján ismertetjük a maximum kvázi-likelihood módszert, majd megadjuk a negatív binomiális eloszlás kvázi-likelihood függvényét.

A maximum kvázi-likelihood módszer hasonló a maximum likelihood módszerhez, csak annyiban különbözik tőle, hogy ebben az esetben nem a likelihood függvényt maximalizáljuk, hanem egy hozzá közeli függvényt. Azt gondolnánk, hogy ebben az esetben rosszabb becslést fogunk kapni a paramétereinkre, viszont a kvázi-likelihood függvény maximumát könnyebben határozzuk meg, így előfordulhat, hogy legalább olyan jó becslést kapunk.

Legyen $\underline{Y}^T=(Y_1,\ldots,Y_N)$ független mintára $E(\underline{Y})=\underline{\mu}$, $\Sigma\left(\underline{Y}\right)=\sigma^2V\left(\underline{\mu}\right)$, $\underline{\mu}=\underline{\mu}\left(\underline{\beta}\right)$.

2.8. Definíció. Az $\widetilde{l}$-ot kvázi-likelihoodnak nevezzük, ha

$$\frac{\partial\widetilde{l}\left(\underline{\mu},\underline{y}\right)}{\partial\underline{\mu}}=V^{-1}\left(\underline{\mu}\right)\left(\underline{y}-\underline{\mu}\right). \tag{2.9}$$

2.9. Megjegyzés. Általában ezt az egyenletet szükségtelen megoldanunk, elég, ha

$$\frac{\partial\widetilde{l}\left(\underline{\beta},\underline{y}\right)}{\partial\underline{\beta}}=0,$$

azaz

$$D^TV^{-1}\left(\underline{y}-\underline{\mu}\left(\underline{\beta}\right)\right)=0$$

egyenlet gyökeit határozzuk meg, ahol $D=\frac{\partial\underline{\mu}}{\partial\underline{\beta}}$.

Negatív binomiális esetben a szórásnégyzet $V(\mu)=\mu+\alpha\mu^2$, és a kvázi-likelihood függvény $\widetilde{l}(\mu,y)=y\log\left(\frac{\mu}{\mu+\alpha^{-1}}\right)-\frac{1}{\alpha}\log(1+\alpha\mu)$ alakú.

2.3.1. Kiterjesztett maximum kvázi-likelihood módszer

Ebben a szakaszban ismét McCullagh és Nelder [5] könyve alapján ismerkedünk meg a kiterjesztett maximum kvázi-likelihood módszerrel.

2.10. Definíció. Az l_2-t kiterjesztett kvázi-likelihood függvénynek nevezzük, ha

$$l_2 = \frac{1}{\phi}\left(Q - Q_0\right) - \frac{1}{2}\log\left(2\pi\phi V(y)\right). \tag{2.10}$$

Ahol Q jelöli a (2.9) differenciálegyenlet megoldását, azaz

$$Q = \int \frac{y-\mu}{V(\mu)} d\mu,$$

Q_0 pedig az $\mu = y$ esetet, ϕ a szórás paraméter, míg a $V(y)$ ismét az y szórásmátrixát jelöli.

A negatív binomiális eloszlás esetében általában a $V(\mu) = \phi\mu\left(1+\alpha\mu\right)$, de a mi esetünkben $\phi = 1$, ezért a $V(\mu) = \mu + \alpha\mu^2$-t írjuk be a Q-ba és integráljuk ki.

$$Q = \int \frac{y-\mu}{\mu+\alpha\mu^2} d\mu = \int \left(\frac{y}{\mu} - \frac{1+\alpha y}{1+\alpha\mu}\right) d\mu = y\log\mu - \left(\frac{1+\alpha y}{\alpha}\right)\log(1+\alpha\mu).$$

Emiatt $Q_0 = y\log y - \left(\frac{1+\alpha y}{\alpha}\right)\log(1+\alpha y)$ lesz.

A kiterjesztett kvázi-likelihood függvényt a negatív binomiális eloszlásra felírva azt kapjuk, hogy

$$l_2(\mu,\alpha) = y\log\mu - \left(\frac{1+\alpha y}{\alpha}\right)\log(1+\alpha\mu) - y\log y + \left(\frac{1+\alpha y}{\alpha}\right)\log(1+\alpha y) - \frac{1}{2}\log\left(2\pi(y+\alpha y^2)\right).$$

2.11. Megjegyzés. [5] szerint a kiterjesztett kvázi-likelihood függvényhez úgy is eljuthattunk volna, ha a negatív binomiális eloszlás sűrűségfüggvényében szereplő faktoriálist a Stirling-formulával helyettesítjük, ami $k! \approx \sqrt{2\pi k}k^k e^{-k}$, továbbá a gamma függvényt is közelítjük a Stirling-formulával.

A jobb közelítés érdekében viszont most a Stirling-formulának egy eltoltját használjuk. Mégpedig $k! \approx \sqrt{2\pi\left(k+c\right)}k^k e^{-k}$, ahol $c = \frac{1}{6}$. Ez az eredeti felírásunkban azzal ekvivalens, ha a $V(Y)$-t helyettesíthetjük $V_2\left(Y\right)$-vel, ahol

$$V_2\left(Y\right) = \frac{\left(y+\nu\right)^2\left(y+c\right)\left(\nu+c\right)}{\nu^2\left(y+\nu+c\right)}. \tag{2.11}$$

2.12. Megjegyzés. Ha a $V_2(Y)$ helyébe beírjuk a $c = 0$-t és $\nu = \frac{1}{\alpha}$-t, akkor pont a $V(Y)$-t kapjuk vissza, mivel

$$\mu + \alpha\mu^2 = \frac{\mu\left(\mu+\nu\right)}{\nu},$$

amiből a $\nu = \frac{1}{\alpha}$-t kapunk.

A fent megadott ν és c értéket beírva:

$$V_2(Y) = \frac{\left(y+\frac{1}{\alpha}\right)^2\left(y+\frac{1}{6}\right)\left(\frac{1}{\alpha}+\frac{1}{6}\right)}{\left(\frac{1}{\alpha}\right)^2\left(y+\frac{1}{a}+\frac{1}{6}\right)} = \frac{\alpha^2\left(y+\frac{1}{\alpha}\right)^2\left(y+\frac{1}{6}\right)\left(1+\frac{\alpha}{6}\right)}{\left(\alpha y+1+\frac{\alpha}{6}\right)}.$$

Ha $V_2(Y)$-nal írjuk fel a kiterjesztett kvázi-likelihood függvényünk az $(y_1, ..., y_n)$ mintára (összegezve), n-nel osztva, akkor a következőt kapjuk:

$$\widetilde{l_2}(\mu,\alpha) = \sum_{i=1}^{n}\left[y_i\log\left(\frac{\mu}{y_i}\right) - \left(\frac{1+\alpha y_i}{\alpha}\right)\log\left(\frac{1+\alpha\mu}{1+\alpha y_i}\right) - \frac{1}{2}\log(2\pi) - \log(1+\alpha y_i) - \right.$$

$$\left. -\frac{1}{2}\log\left(y_i+\frac{1}{6}\right) - \frac{1}{2}\log\left(1+\frac{\alpha}{6}\right) + \frac{1}{2}\log\left(\alpha y_i+1+\frac{\alpha}{6}\right)\right]. \tag{2.12}$$

Ezen kiterjesztett kvázi-likelihood függvény deriváltjai a következők:

$$\frac{\partial\widetilde{l_2}(\mu,\alpha)}{\partial\mu} = \sum_{i=1}^{n}\left[\frac{y_i}{\mu} - \frac{1+\alpha y_i}{1+\alpha\mu}\right], \tag{2.13}$$

$$\frac{\partial\widetilde{l_2}(\mu,\alpha)}{\partial\alpha} = \sum_{i=1}^{n}\left[\frac{1}{\alpha^2}\log\left(\frac{1+\alpha\mu}{1+\alpha y_i}\right) - \frac{\mu(1+\alpha y_i)}{\alpha(1+\alpha\mu)} + \frac{y_i}{\alpha} - \frac{y_i}{1+\alpha y_i} - \right.$$

$$\left. -\frac{1}{12\left(1+\frac{\alpha}{6}\right)} + \frac{y_i+\frac{1}{6}}{2\left(\alpha y_i+1+\frac{\alpha}{6}\right)}\right]. \tag{2.14}$$

A (2.13) kifejezés nullhelyéből most is rögtön adódik a $\bar{y} = \hat{\mu}$ választás. A (2.14) egzakt megoldása ismét bonyolult, ezért helyette ismét numerikus közelítést alkalmazunk. Mi az R programmal az $\widetilde{l_2}$ függvény α szerinti maximumhelyét kerestük a $\mu = \bar{y}$ helyen. A maximalizálás szempontjából most is van olyan tag, amiket elhagyhatunk, mert nem befolyásolják a maximum helyét. Ilyen tagok az α-t nem tartalmazó tagok, például a $-\frac{1}{2}\log(2\pi)$ vagy az $y_i\log\left(\frac{\mu}{y_i}\right)$.

2.3.2. Duplán kiterjesztett kvázi likelihood módszer

Lee és Nelder [7] 2001-ben bevezette a duplán kiterjesztett kvázi likelihood módszert. A lényeg, hogy úgy gondolták, hogy a Stirling-formula nem ad elég jó közelítést, ezért a gamma függvény logaritmusára adnak egy újabb közelítést. A módszer alábbi leírását Saha és Paul [9] cikkéből vettem.

A következő közelítést használjuk:

$$\log\Gamma(z) \approx \left(z-\frac{1}{2}\right)\log(z) + \frac{1}{2}\log(2\pi) - z + \frac{1}{12z}. \tag{2.15}$$

Írjuk be ezt a közelítést a negatív binomiális eloszlás log-likelihood függvényébe a gamma függvény logaritmusa helyébe és a faktoriális helyébe.

A log-likelihood függvényünk így néz ki a közelítés után:

$$l_3(\mu,\alpha) = \sum_{i=1}^{n} \left[y_i \log(\mu) + \left(y_i + \frac{1}{\alpha}\right) \log\left(\frac{1+\alpha y_i}{1+\alpha\mu}\right) - \frac{1}{2}\log(1+\alpha y_i) - \right.$$

$$\left. - \left(y_i - \frac{1}{2}\right)\log(y_i) + \frac{\alpha}{12(1+\alpha y_i)} - \frac{\alpha}{12} - \frac{1}{12y_i} - \frac{1}{2}\log(2\pi) \right]. \tag{2.16}$$

Nézzük ennek a parciális deriváltjait:

$$\frac{\partial l_3(\mu,\alpha)}{\partial \mu} = \sum_{i=1}^{n} \left(\frac{y_i}{\mu(1+\alpha\mu)} - \left(\frac{1}{1+\alpha\mu}\right)\right) \tag{2.17}$$

$$\frac{\partial l_3(\mu,\alpha)}{\partial \alpha} = \left[\frac{1}{\alpha^2}\log\left(\frac{1+\alpha\mu}{1+\alpha y_i}\right) + \frac{y_i-\mu}{\alpha(1+\alpha\mu)} - \frac{2\alpha y_i + 2 - \alpha}{2\alpha^2(1+\alpha y_i)} + \frac{2-\alpha}{2\alpha^2} - \frac{\alpha y_i(2+\alpha y_i)}{12(1+\alpha y_i)^2} \right]. \tag{2.18}$$

A (2.17) derivált megegyezik a maximum likelihood módszerben kapott μ szerinti (2.7) derivált n-szeresével, így $\hat{\mu} = \bar{y}$ megfelelő választás lesz most is. A (2.18) kifejezés ismét túl bonyolult, így numerikus közelítéssel oldjuk meg a maximalizálási feladatot. Mi az R programmal az l_3 függvény α szerinti maximumhelyét kerestük a $\mu = \bar{y}$ helyen.

A maximalizálás szempontjából itt is vannak olyan tagok, amiket elhagyhatunk, mert nem befolyásolják a maximum hely keresését. Ezek a következők: $y_i \log(\mu)$, $-\frac{1}{2}\log(2\pi)$ és $-\frac{1}{12y_i}$.

2.4. Szimulációk és eredmények

A szimulációkat az R statisztikai programmal végeztük. Először generáltunk n elemszámú α és μ paraméterű negatív binomiális eloszlású mintákat, majd ezekre futtattunk le az előbbiekben tárgyalt módszereket. A táblázatok tartalmazzák a lefuttatott „eredeti” paramétereket, a becslésként kapott $\hat{\alpha}$ értéknek a statisztikai mutatóit (átlagát, mediánját, standard hibáját és alsó illetve felső kvartilisét) és azt, hogy hány értek lesz negatív a szimulációk során.

Minden esetben kétféle szimulációt használtunk. Az egyik esetben 10000 szimulációt futtattunk állandó méretű elemszám ($n = 50$) esetén úgy, hogy közben változtattuk a μ és az α paramétereket. Az α és a μ paramétereit és a szimuláció számát is a feldolgozott cikkekben lévő paraméterekből vettük, mert így az eredményeink összehasonlíthatóak a feldolgozott cikkekben

található eredményekkel. Készítettünk egy kiegészítést is, ami tartalmazza a $\alpha = 0{,}1$ és $\alpha = 0{,}05$ paramétereket. Itt azt figyelhetjük meg, hogy kis szórás esetén mennyire jó közelítéseket kapunk.

A másik esetben 1000 szimulációt futtattunk állandó μ és α paraméterekre, csak az elemszámot változtattuk. Itt is a cikkek voltak a paraméterekben a mérvadók. Ebben az esetben is készítettünk egy kiegészítést, amiben az elemszámot vettük nagyra. Ekkor azt tudjuk megfigyelni, hogy mennyire nagy elemszám szükséges a jó közelítéshez, illetve nagy elemszám esetén van-e különbség a különböző becsléseink között.

Mielőtt az eredmények ismertetésébe fognánk megjegyezzük, hogy amennyiben a becsléshez függvény maximalizálására volt szükség, úgy a maximalizálást az R `optimise` nevű függvényével végeztük, és általában a maximum helyét a $(-1/\max\{y_1, ..., y_n\}, 100)$ intervallumon kerestük.

2.4.1. Momentum módszer

Az első szimulációban a momentum módszer esetén azt várnánk, hogy mivel nagy szimuláció számmal dolgozunk, ezért az $\hat{\alpha}$ becslések átlaga elég jól fogja közelíteni az α-t. Sajnos ez a becsléseinkből nem látszik, sőt minél kisebb α-t becsülünk, annál nagyobb eltérés mutatkozik a két szám közt, viszont a μ növelésével egyre jobb becsléseket kapunk. Az eltérés oka, hogy a minta elemszáma, ami most 50, nem túl nagy, így pontos becslést nem várhatunk. A μ növekedésével tapasztalható javulás oka az, hogy α a szórásra jellemző paraméter és ha a μ-t növeljük, akkor a szórás nő, így α becslése is javulni fog. Már itt megjegyezzük, hogy az eltérések azt mutatják, hogy a becslésünk torzít. Erre a felfedezésre később adunk magyarázatot.

A 10000 szimulációs momentum módszeres eredményeink, ami a 2.1. táblázatban látható, szinte minden esetben rosszabbak, mint a Clark és Perry [4] cikkben szereplő eredmények, amiket a 2.2. táblázat mutat. Ez látható mind az átlagnál, mind a mediánnál, mind az alsó kvartilisek esetében. Ezzel szemben a felső kvartilisek mindegyike közelebb van a mi megfigyeléseinknél a becsült α-hoz. Az alsó kvartilisek rosszabb eredménye mellett több negatív értéket is kaptunk a becslésünk során. (Ezek mind a becslésünk torzítottságát támasztják alá.) Például $\mu = 1$ és $\alpha = 0{,}2$ esetén a 2.2. táblázat megfelelő sora szerint 2311 negatív érték adódott Clark és Perry szimulációja során és az alsó kvartilis 0,009 lett míg a mi eredményeinket összefoglaló 2.1. táblázat alapján a szimulációnk során 3074 negatív érték adódott, így az alsó kvartilis is negatív $-0{,}035$. Bár eredményeink rosszabbak, a standard hiba a mi szimulációink során szinte minden esetben kisebb, ami kisebb szórást mutat. Az eltérések és a becslésünk torzítottságának oka az, hogy mi a második tapasztalati momentumot használtuk, míg Clark és Perry [4] a *korrigált* tapasztalati szórásnégyzetet. Ez magyarázza a kisebb szórást is, mivel a két becslés közti eltérést a korrigált tapasztalati szórásnégyzet alkalmazása okozza, ami (a becslés $-M_1$-es tagjáról eltekintve) egy 50/49-es szorzót jelent.

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	0,911	0,529	98	0,833	0,550	1,182
	0,333	0,299	0,307	1557	0,259	0,086	0,467
	0,2	0,170	0,258	2759	0,136	-0,008	0,317
3	1	0,949	0,348	0	0,897	0,708	1,134
	0,333	0,314	0,148	42	0,299	0,211	0,400
	0,2	0,188	0,115	301	0,176	0,107	0,257
5	1	0,952	0,316	0	0,904	0,737	1,113
	0,333	0,319	0,120	1	0,305	0,235	0,390
	0,2	0,191	0,087	17	0,182	0,130	0,243
10	1	0,957	0,287	0	0,914	0,757	1,105
	0,333	0,323	0,098	0	0,313	0,254	0,381
	0,2	0,193	0,063	0	0,187	0,149	0,232
15	1	0,955	0,279	0	0,913	0,762	1,100
	0,333	0,322	0,089	0	0,312	0,259	0,374
	0,2	0,194	0,056	0	0,189	0,154	0,228
20	1	0,963	0,277	0	0,919	0,771	1,107
	0,333	0,322	0,085	0	0,314	0,262	0,371
	0,2	0,195	0,054	0	0,190	0,157	0,228

2.1. táblázat. 10000 szimulációs momentum módszeres becslésem n=50 minta elemszám esetén

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	0,949	0,539	81	0,870	0,579	1,226
	0,333	0,320	0,306	1287	0,286	0,108	0,496
	0,2	0,192	0,268	2311	0,162	0,009	0,348
3	1	0,968	0,353	0	0,916	0,722	1,157
	0,333	0,327	0,152	27	0,312	0,220	0,416
	0,2	0,198	0,117	249	0,186	0,116	0,268
5	1	0,977	0,320	0	0,930	0,756	1,144
	0,333	0,329	0,122	1	0,316	0,244	0,399
	0,2	0,198	0,089	12	0,189	0,135	0,251
10	1	0,973	0,289	0	0,927	0,768	1,130
	0,333	0,330	0,098	0	0,320	0,261	0,389
	0,2	0,198	0,066	0	0,192	0,152	0,238
15	1	0,979	0,283	0	0,939	0,782	1,125
	0,333	0,330	0,090	0	0,319	0,267	0,383
	0,2	0,199	0,059	0	0,193	0,158	0,234
20	1	0,978	0,277	0	0,935	0,784	1,123
	0,333	0,330	0,086	0	0,321	0,268	0,380
	0,2	0,200	0,055	0	0,194	0,161	0,233

2.2. táblázat. Clark és Perry [4] 10000 szimulációs momentum módszeres becslése n=50 minta elemszám esetén

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	0,843	0,547	152	0,751	0,468	1,113
	0,333	0,269	0,309	1737	0,228	0,059	0,428
	0,2	0,153	0,258	2888	0,118	-0,024	0,295
3	1	0,901	0,377	0	0,832	0,641	1,082
	0,333	0,304	0,160	40	0,281	0,193	0,389
	0,2	0,179	0,122	388	0,164	0,097	0,247
5	1	0,908	0,353	0	0,840	0,671	1,067
	0,333	0,310	0,133	0	0,291	0,218	0,378
	0,2	0,184	0,091	33	0,171	0,119	0,235
10	1	0,916	0,327	0	0,853	0,695	1,064
	0,333	0,316	0,108	0	0,300	0,241	0,374
	0,2	0,191	0,071	0	0,182	0,141	0,230
15	1	0,923	0,314	0	0,864	0,708	1,071
	0,333	0,317	0,099	0	0,304	0,248	0,371
	0,2	0,190	0,062	0	0,183	0,147	0,224
20	1	0,918	0,306	0	0,863	0,710	1,062
	0,333	0,318	0,094	0	0,305	0,252	0,369
	0,2	0,192	0,058	0	0,185	0,151	0,225

2.3. táblázat. 10000 szimulációs (a harmadik momentum használatával kapott) momentum módszeres becslésem n=50 minta elemszám esetén

A 2.3. táblázat mutatja az első és a harmadik momentumból kapott becslésünket. Ebben az esetben nem meglepő, hogy szinte minden értékben rosszabb eredményeink születtek a második momentumra épülő becslésünknél. Ezt vártuk is, mivel a harmadik tapasztalati momentumot egy-egy nagy érték a mintában nagyon meg tudja növelni. Tehát ez a becslésünk kicsit jobban torzított és (általában) picit nagyobb standard hibájú, mint a második momentumból kapott momentum módszeres becslés (2.1. táblázat). Érdemes megfigyelni, hogy μ értékének növelésekor ebben az esetben is jobb becslést kapunk.

A 2.4. és a 2.5. táblázatok mutatják azon szimulációkat, melyekben a momentum módszeres becsléseinket még kisebb α paraméterű negatív binomiális eloszlásokra alkalmaztuk. Szinte ugyanazt kaptuk, mint az előbb, ismét látszik a torzítás, a becslés javulása μ növelésekor, és hogy a harmadik momentumból kapott becslésünk kicsit rosszabb. Viszont picit meglepő, hogy a két becsléssel majdnem ugyanazt az eredményt kaptuk $\alpha = 0{,}05$-re a $\mu = 20$ esetben, és még $\mu = 15$ is csak egy-egy ezrednyi eltérés van a kapott értékek között.

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	0,1	0,070	0,228	4240	0,041	-0,084	0,200
	0,05	0,027	0,214	4993	0,000	-0,116	0,148
3	0,1	0,092	0,091	1474	0,084	0,027	0,148
	0,05	0,041	0,078	3255	0,034	-0,014	0,089
5	0,1	0,093	0,062	478	0,089	0,050	0,131
	0,05	0,044	0,050	1942	0,040	0,008	0,075
10	0,1	0,096	0,042	23	0,093	0,067	0,121
	0,05	0,046	0,030	456	0,043	0,025	0,065
15	0,1	0,097	0,034	2	0,094	0,073	0,118
	0,05	0,047	0,023	115	0,046	0,031	0,062
20	0,1	0,097	0,031	0	0,094	0,076	0,116
	0,05	0,048	0,020	20	0,046	0,034	0,060

2.4. táblázat. 10000 szimulációs momentum módszeres becslésem kiegészítése n=50 minta elemszám esetén

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	0,1	0,056	0,230	4465	0,032	-0,102	0,180
	0,05	0,013	0,216	5249	-0,009	-0,133	0,132
3	0,1	0,086	0,093	1722	0,076	0,019	0,141
	0,05	0,038	0,080	3355	0,030	-0,017	0,085
5	0,1	0,091	0,066	578	0,083	0,045	0,129
	0,05	0,042	0,052	2086	0,038	0,006	0,073
10	0,1	0,095	0,044	19	0,090	0,064	0,121
	0,05	0,046	0,032	478	0,043	0,024	0,066
15	0,1	0,096	0,036	1	0,092	0,070	0,117
	0,05	0,047	0,024	99	0,045	0,030	0,062
20	0,1	0,096	0,032	0	0,093	0,073	0,115
	0,05	0,048	0,021	22	0,046	0,033	0,060

2.5. táblázat. 10000 szimulációs (a harmadik momentum használatával kapott) momentum módszeres becslésem kiegészítése n=50 minta elemszám esetén

A fentiek tükrében már nem meglepő, hogy most az eredményeink 1000 szimulációnál (2.6. táblázat), mind az átlagnál, mind a mediánnál és mind az alsó kvartilisnél sokkal rosszabb értéket mutatnak, mint a Clark és Perry [4] cikkben szereplő értékek (2.7. táblázat), hiszen erre ismét magyarázatot ad a becslésünk torzítottsága. Most főleg az az érdekes, hogy sem az átlagnál, sem a mediánnál nem jutottunk elég közel az α értékéhez. Míg Clark és Perry [4] eredményeiben csak az alsó kvartiliseknél láthatunk 6 negatív értéket, addig a mi eredményeinkben az alsó kvartiliseknél 8 negatív, és az egész táblázatban (a $\mu = 1$ és $\alpha = 0{,}1$ és $\alpha = 0{,}05$ esetekhez tartozó értékeket nem számolva) 9 negatív eredmény szerepel, mivel egy esetben a medián is

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	0,574	0,958	249	0,400	0,000	0,987
		20	0,816	0,811	88	0,666	0,300	1,182
		30	0,823	0,617	42	0,701	0,401	1,109
		50	0,909	0,545	15	0,814	0,539	1,188
1	0,333	10	0,140	0,651	460	0,065	-0,259	0,400
		20	0,225	0,459	342	0,179	-0,093	0,482
		30	0,229	0,373	285	0,181	-0,012	0,436
		50	0,304	0,319	153	0,259	0,088	0,458
1	0,2	10	0,028	0,575	558	-0,012	-0,306	0,250
		20	0,126	0,428	459	0,059	-0,167	0,330
		30	0,139	0,323	365	0,111	-0,074	0,331
		50	0,157	0,244	279	0,132	-0,008	0,282
1	0,1	10	-0,005	0,563	587	-0,063	-0,339	0,235
		20	0,053	0,395	511	-0,003	-0,184	0,240
		30	0,059	0,287	473	0,032	-0,136	0,213
		50	0,074	0,228	422	0,063	-0,084	0,222
1	0,05	10	-0,088	0,501	653	-0,172	-0,400	0,183
		20	-0,002	0,342	574	-0,031	-0,224	0,181
		30	0,000	0,280	554	-0,020	-0,185	0,154
		50	0,020	0,215	516	-0,001	-0,120	0,130
3	0,2	10	0,144	0,249	288	0,097	-0,023	0,262
		20	0,175	0,179	134	0,142	0,048	0,269
		30	0,174	0,146	87	0,154	0,071	0,258
		50	0,184	0,118	39	0,170	0,108	0,245

2.6. táblázat. 1000 szimulációs momentum módszeres becslésem

negatív (ha beleszámítjuk a $\mu = 1$ és $\alpha = 0{,}1$ és $\alpha = 0{,}05$ eseteket is, akkor 26 negatív értéket számolhatunk össze). Vegyük észre, hogy minél nagyobb elemszámot választunk, annál jobb lesz a becslésünk, és annál jobban közelít Clark és Perry [4] eredményeihez, mivel annál jobban közelít a tapasztalati szórás (ami megfeleltethető a tapasztalati második momentumnak) és a korrigált tapasztalati szórásnégyzet. Viszont az $n = 10$ elemszám annyira kicsi, hogy nagyon rossz eredményeket mutat, sokkal rosszabbat, mint a Clark és Perry [4] eredményei, mivel ekkor már a korrigált tapasztalati szórásnégyzet alkalmazása egy 10/9-es szorzót jelent.

A 2.6. és a 2.8. táblázatok az 1000 szimulációs momentum módszeres becsléseinket tartalmazzák. Az első két momentumból származó becsléseink ismét jobb eredményeket adnak minden vizsgált esetben, de ezen már nem lepődünk meg. Érdemes megfigyelni, hogy $\alpha = 1$ és $n = 10$ esetén még a felső kvartilis sem éri el a becsült értéket, sőt a harmadik momentumot használó becslésünknél még csak meg sem közelíti. Ez a kevés elemszámból adódó nagyobb torzítás miatt lehetséges.

μ	α	n	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	0,774	1,091	188	0,583	0,098	1,262
		20	0,862	0,787	82	0,725	0,367	1,229
		30	0,937	0,636	33	0,825	0,493	1,253
		50	0,978	0,545	9	0,892	0,616	1,259
1	0,333	10	0,319	0,747	355	0,128	-0,166	0,638
		20	0,293	0,494	263	0,228	-0,028	0,513
		30	0,330	0,395	198	0,272	0,031	0,539
		50	0,321	0,307	135	0,276	0,113	0,501
1	0,2	10	0,178	0,644	411	0,072	-0,213	0,419
		20	0,192	0,454	358	0,109	-0,113	0,422
		30	0,186	0,346	314	0,163	-0,058	0,375
		50	0,215	0,267	218	0,189	0,023	0,373
3	0,2	10	0,191	0,268	266	0,149	-0,008	0,349
		20	0,201	0,185	104	0,169	0,076	0,292
		30	0,196	0,157	71	0,175	0,089	0,274
		50	0,197	0,112	26	0,187	0,122	0,260

2.7. táblázat. Clark és Perry [4] 1000 szimulációs momentum módszeres becslése

A 2.9. és a 2.10. táblázatokban azt láthatjuk, hogy az n elemszámot nagyon megnöveltük, és eddig hiába voltak becsléseink kevésbé pontosak immár az átlag, a medián és a kvartilisek is elég jól közelítik a becsült paramétert. A második momentumból kapott módszernél az átlag mindhárom minta elemszámra nagyon közel van a becsült paraméterhez. A harmadik momentum segítségével kapott becslésünk esetében még ezres minta esetén tapasztalható némi eltérés, főként nagyobb α esetén, viszont megállapítható, hogy a két becslés is egyre jobban közelít egymáshoz, sok jellemző paraméterük megegyezik, sőt néhány kivételes esetben egy-egy paramétere a harmadik momentumos becslésünknek jobb.

Összességében azt a következtetést vonhatjuk le, hogy a második momentum segítségével kapott becslés jobb, mint a harmadik momentum segítségével kapott, viszont Clark és Perry [4] korrigált tapasztalati szórásnégyzetet használó becslése még jobb, főként kis minta esetén. A továbbiakban csak a második momentum segítségével kapott becslésünket hasonlítjuk a többiekhez, mivel ez jobb a másik momentum módszeres becslésünknél, és több szimulációs eredményt ismerünk rá, mint amit Clark és Perry [4] ismertetett.

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	0,428	0,770	286	0,319	-0,056	0,819
		20	0,679	0,728	121	0,531	0,234	1,009
		30	0,735	0,600	69	0,627	0,318	1,036
		50	0,849	0,554	18	0,753	0,478	1,104
1	0,333	10	0,055	0,548	530	-0,009	-0,265	0,310
		20	0,175	0,455	390	0,094	-0,121	0,384
		30	0,222	0,377	302	0,158	-0,034	0,415
		50	0,263	0,317	197	0,207	0,046	0,435
1	0,2	10	-0,035	0,503	600	-0,068	-0,328	0,187
		20	0,075	0,388	484	0,013	-0,181	0,282
		30	0,121	0,337	426	0,052	-0,106	0,303
		50	0,149	0,268	310	0,112	-0,042	0,291
1	0,1	10	-0,091	0,434	669	-0,153	-0,328	0,116
		20	0,024	0,361	532	-0,020	-0,233	0,201
		30	0,027	0,289	511	-0,004	-0,173	0,178
		50	0,053	0,227	460	0,026	-0,104	0,181
1	0,05	10	-0,134	0,410	702	-0,177	-0,405	0,082
		20	-0,046	0,320	642	-0,099	-0,249	0,091
		30	0,001	0,284	558	-0,035	-0,191	0,153
		50	0,013	0,214	549	-0,016	-0,127	0,137
3	0,2	10	0,112	0,224	330	0,077	-0,037	0,218
		20	0,139	0,169	201	0,118	0,023	0,233
		30	0,163	0,142	100	0,146	0,061	0,248
		50	0,180	0,121	33	0,167	0,096	0,250

2.8. táblázat. 1000 szimulációs (a harmadik momentum használatával kapott) momentum módszeres becslésem

μ	α	n	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1000	1,001	0,124	0	0,992	0,913	1,076
		10000	0,999	0,040	0	0,999	0,973	1,026
		50000	1,000	0,017	0	1,000	0,988	1,011
1	0,333	1000	0,330	0,071	0	0,330	0,279	0,375
		10000	0,333	0,022	0	0,332	0,317	0,347
		50000	0,333	0,010	0	0,333	0,326	0,339
1	0,2	1000	0,197	0,059	0	0,195	0,156	0,235
		10000	0,200	0,018	0	0,199	0,188	0,212
		50000	0,200	0,008	0	0,200	0,195	0,205
1	0,1	1000	0,100	0,054	27	0,097	0,063	0,135
		10000	0,100	0,016	0	0,100	0,090	0,110
		50000	0,100	0,007	0	0,100	0,095	0,105
1	0,05	1000	0,049	0,049	162	0,048	0,014	0,083
		10000	0,049	0,015	1	0,048	0,040	0,059
		50000	0,050	0,007	0	0,050	0,045	0,054
3	0,2	1000	0,200	0,026	0	0,200	0,183	0,217
		10000	0,200	0,008	0	0,199	0,194	0,205
		50000	0,200	0,004	0	0,200	0,197	0,203

2.9. táblázat. 1000 szimulációs momentum módszeres becslésem kiegészítése

μ	α	n	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1000	0,992	0,162	0	0,981	0,883	1,088
		10000	1,002	0,052	0	0,999	0,964	1,036
		50000	0,999	0,024	0	0,998	0,982	1,013
1	0,333	1000	0,329	0,077	0	0,324	0,273	0,380
		10000	0,333	0,024	0	0,333	0,317	0,349
		50000	0,333	0,011	0	0,333	0,326	0,341
1	0,2	1000	0,199	0,067	0	0,196	0,152	0,237
		10000	0,199	0,021	0	0,199	0,186	0,212
		50000	0,200	0,009	0	0,200	0,194	0,206
1	0,1	1000	0,095	0,055	40	0,093	0,058	0,134
		10000	0,100	0,018	0	0,101	0,088	0,113
		50000	0,100	0,008	0	0,100	0,094	0,105
1	0,05	1000	0,047	0,050	162	0,045	0,015	0,079
		10000	0,051	0,016	1	0,050	0,040	0,061
		50000	0,050	0,007	0	0,050	0,045	0,054
3	0,2	1000	0,200	0,029	0	0,198	0,180	0,218
		10000	0,200	0,009	0	0,200	0,194	0,206
		50000	0,200	0,004	0	0,200	0,197	0,203

2.10. táblázat. 1000 szimulációs (a harmadik momentum használatával kapott) momentum módszeres becslésem kiegészítése

2.4.2. Maximum likelihood módszer

Ebben a fejezetben először Piegorsch [8] cikkbeli eredményeit fogom összehasonlítani a kapott eredményeinkkel, azután összevetjük az előző fejezetben kapott első és második momentumból származó momentum módszeres eredményeinkkel

μ	α	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	0,985	0,522	88	0,924	0,618	1,285
	0.333	0,311	0,311	1502	0,280	0,097	0,488
	0.2	0,178	0,276	2800	0,152	-0,008	0,341
3	1	0,985	0,302	0	0,954	0,772	1,167
	0.333	0,325	0,146	38	0,312	0,222	0,412
	0.2	0,192	0,114	279	0,183	0,112	0,259
5	1	0,988	0,260	0	0,963	0,803	1,148
	0.333	0,324	0,112	0	0,315	0,245	0,392
	0.2	0,193	0,083	25	0,186	0,135	0,245
10	1	0,986	0,218	0	0,968	0,833	1,120
	0.333	0,326	0,087	0	0,320	0,264	0,381
	0.2	0,195	0,061	0	0,190	0,152	0,233
15	1	0,989	0,211	0	0,974	0,839	1,116
	0.333	0,327	0,081	0	0,321	0,270	0,378
	0.2	0,195	0,054	0	0,191	0,157	0,229
20	1	0,988	0,203	0	0,976	0,846	1,114
	0.333	0,327	0,075	0	0,321	0,273	0,374
	0.2	0,196	0,050	0	0,192	0,160	0,227

2.11. táblázat. 10000 szimulációs maximum likelihood becslésem n=50 minta elemszám esetén

Ha megnézzük a 2.11. és a 2.12. táblázatokat, akkor azt vesszük észre, hogy eredményeink közelítenek a Piegorsch [8] cikkben található eredményekhez. Ebben az esetben nem lehet jobbnak mondani egyik szimulációt sem, sem az átlagban, sem a mediánban, sem az alsó kvartilisben, sem a felső kvartilisben. A mi becsléseinknek nagyobb kicsit a szórása, ezt a standard hibából tudjuk, viszont az eltérés okát nem látjuk. Lehet, hogy a véletlenszám generátor okozza a különbséget.

A 2.1. és a 2.11. táblázatok alapján elmondhatjuk, hogy a maximum likelihood módszeres becslésünk esetében szinte minden esetben jobb becslést kaptunk, mint a momentum módszeres becslésünk esetében a μ paraméter növelése esetén. Mi ezt vártuk a két becsléstől, mivel a maximum likelihood módszerről tudjuk, hogy aszimptotikus értelemben hatásos becslést ad.

A 2.4 és a 2.13. táblázatok azt mutatják, hogy nagy μ és kis α paraméterek esetén, a momentum módszerrel és maximum likelihood módszerrel végzett szimulációk az alsó kvartilisek sokszor már megegyeznek egymással. Ettől még továbbra is igaz az, hogy a maximum likelihood becsléseink szinte minden esetben jobbak vagy legalább olyan jók, mint a momentum módszeres

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	0,978	0,518	100	0,914	0,619	1,266
	0,333	0,314	0,314	1512	0,278	0,096	0,495
	0,2	0,179	0,281	2845	0,152	-0,013	0,347
3	1	0,986	0,303	0	0,959	0,773	1,166
	0,333	0,322	0,145	34	0,308	0,221	0,412
	0,2	0,191	0,112	302	0,181	0,112	0,261
5	1	0,987	0,258	0	0,963	0,806	1,141
	0,333	0,326	0,113	0	0,318	0,246	0,395
	0,2	0,194	0,084	25	0,187	0,135	0,245
10	1	0,986	0,222	0	0,967	0,831	1,119
	0,333	0,327	0,088	0	0,318	0,264	0,380
	0,2	0,195	0,061	0	0,191	0,152	0,233
15	1	0,989	0,210	0	0,973	0,839	1,121
	0,333	0,327	0,080	0	0,321	0,270	0,377
	0,2	0,195	0,054	0	0,192	0,157	0,229
20	1	0,986	0,202	0	0,972	0,842	1,111
	0,333	0,326	0,076	0	0,320	0,272	0,374
	0,2	0,196	0,050	0	0,192	0,161	0,226

2.12. táblázat. Piegorsch [8] 10000 szimulációs maximum likelihood becslése n=50 minta elemszám esetén

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	0,1	0,079	0,246	4170	0,053	-0,093	0,221
	0,05	0,027	0,228	5037	0,000	-0,133	0,160
3	0,1	0,093	0,091	1522	0,087	0,029	0,149
	0,05	0,043	0,078	3065	0,038	-0,011	0,091
5	0,1	0,094	0,063	483	0,089	0,050	0,132
	0,05	0,045	0,051	1934	0,041	0,008	0,076
10	0,1	0,096	0,040	19	0,093	0,067	0,122
	0,05	0,048	0,031	447	0,045	0,026	0,067
15	0,1	0,097	0,033	1	0,094	0,073	0,118
	0,05	0,048	0,024	96	0,046	0,031	0,063
20	0,1	0,097	0,030	0	0,095	0,076	0,116
	0,05	0,048	0,020	17	0,047	0,034	0,061

2.13. táblázat. 10000 szimulációs maximum likelihood becslésem kiegészítése n=50 minta elemszám esetén

becslések, bár ez a kép immár kissé árnyaltabb.

A 2.6 és a 2.14. táblázatokban az 1000 szimulációs momentum módszeres becsléseink és maximum likelihood becsléseink találhatóak. A maximum likelihood becsléseink jobb eredményeket mutatnak az átlagnál és a mediánnál, viszont az alsó kvartiliseknél nagyjából együtt mozog a két

μ	n	α	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	0,883	1,466	250	0,569	-0,001	1,414
		20	0,925	0,860	102	0,783	0,366	1,365
		30	0,973	0,725	54	0,880	0,475	1,356
		50	0,986	0,528	12	0,913	0,615	1,292
1	0,333	10	0,286	1,194	472	0,041	-0,300	0,598
		20	0,290	0,583	344	0,190	-0,100	0,565
		30	0,301	0,427	255	0,241	-0,001	0,512
		50	0,299	0,306	162	0,272	0,086	0,476
1	0,2	10	0,148	0,839	532	-0,014	-0,333	0,378
		20	0,152	0,447	427	0,094	-0,157	0,377
		30	0,161	0,370	377	0,113	-0,086	0,360
		50	0,192	0,272	257	0,166	0,000	0,353
1	0,1	10	0,062	0,755	594	-0,120	-0,486	0,316
		20	0,096	0,455	493	0,014	-0,203	0,309
		30	0,078	0,337	468	0,043	-0,154	0,261
		50	0,081	0,246	412	0,057	-0,079	0,225
1	0,05	10	-0,033	0,587	648	-0,159	-0,500	0,213
		20	0,012	0,420	577	-0,033	-0,264	0,210
		30	0,006	0,305	551	-0,014	-0,212	0,176
		50	0,029	0,231	496	0,010	-0,139	0,165
3	0,2	10	0,167	0,271	305	0,118	-0,027	0,303
		20	0,178	0,188	159	0,159	0,045	0,276
		30	0,183	0,156	98	0,168	0,074	0,264
		50	0,190	0,115	37	0,183	0,109	0,266

2.14. táblázat. 1000 szimulációs maximum likelihood becslésem

becslés, ami a kis elemszám miatt lehetséges. De a felső kvartiliseknél minden eredmény rosszabb, mint a momentum módszeres becsléseinknél. Viszont ezen nem lepődünk meg, hiszen emlékszünk, hogy a momentum módszernél az alacsony felső kvartilist a minta torzítottsága adta. Az is észrevehető, hogy a standard hiba a maximum likelihood módszeres becsléseinknél nagyobb, főként kis mintára ($n = 10{,}20$ esetén.). Ezért szokás kis minta esetén inkább a momentum módszert alkalmazni (persze inkább a Clark és Perry [4] által vizsgált korrigált változatot).

A 2.14 és a 2.15. táblázatok most is azt mutatják – mint a 10000 szimulációnál –, hogy a becslési értékeink nagyjából egyformán mozognak Piegorsch [8] cikkben lévő eredményeivel. Nem lehet egyik becslést sem jobbnak mondani. Érdekes tény, hogy a mi becsléseink között $\mu = 1$ esetben több negatív érték volt, viszont $\mu = 3$ esetben nekünk volt kevesebb negatív értékünk. Kicsit meglepő dolog, hogy a Piegorsch [8] cikkében többször tapasztalható, hogy a standard hiba kisebb, viszont pont a legnagyobb standard hibájú helyeken a mi eredményeink a jobbak. A különbségek itt is a véletlenszám generátorból adódhatnak vagy a kis minta elemszám miatt.

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	0,900	1,592	269	0,540	-0,014	1,355
		20	0,991	0,926	100	0,814	0,365	1,467
		30	0,950	0,660	47	0,836	0,512	1,320
		50	0,993	0,517	9	0,938	0,634	1,261
1	0,333	10	0,290	0,908	461	0,127	-0,274	0,656
		20	0,270	0,528	363	0,207	-0,114	0,574
		30	0,294	0,423	251	0,256	$-1{,}0 \cdot 10^{-6}$	0,518
		50	0,310	0,317	147	0,277	0,102	0,494
1	0,2	10	0,127	0,798	570	-0,014	-0,333	0,378
		20	0,152	0,476	425	0,089	-0,157	0,370
		30	0,176	0,361	343	0,146	-0,078	0,375
		50	0,185	0,285	267	0,147	-0,007	0,350
3	0,2	10	0,147	0,263	321	0,106	-0,004	0,281
		20	0,173	0,197	183	0,146	0,036	0,267
		30	0,187	0,154	89	0,169	0,076	0,281
		50	0,192	0,111	25	0,180	0,117	0,262

2.15. táblázat. Piegorsch [8] 1000 szimulációs maximum likelihood becslése

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1000	0,995	0,108	0	0,992	0,925	1,067
		10000	1,002	0,033	0	1,001	0,979	1,023
		50000	1,000	0,015	0	1,000	0,990	1,010
1	0,333	1000	0,329	0,066	0	0,327	0,283	0,375
		10000	0,333	0,021	0	0,333	0,319	0,346
		50000	0,333	0,010	0	0,333	0,327	0,340
1	0,2	1000	0,199	0,055	0	0,199	0,161	0,236
		10000	0,200	0,018	0	0,199	0,188	0,212
		50000	0,200	0,008	0	0,200	0,195	0,206
1	0,1	1000	0,101	0,050	19	0,099	0,068	0,132
		10000	0,099	0,016	0	0,099	0,088	0,110
		50000	0,100	0,007	0	0,100	0,095	0,105
1	0,05	1000	0,048	0,049	161	0,047	0,016	0,080
		10000	0,051	0,015	0	0,050	0,041	0,060
		50000	0,050	0,007	0	0,050	0,045	0,055
3	0,2	1000	0,200	0,025	0	0,201	0,182	0,218
		10000	0,200	0,008	0	0,200	0,194	0,205
		50000	0,200	0,004	0	0,200	0,198	0,203

2.16. táblázat. 1000 szimulációs maximum likelihood becslésem kiegészítése

A 2.9. és a 2.16. táblázatok azt mutatják, hogy a nagy elemszám miatt a momentum módszeres becsléseink és a maximum likelihood módszeres becsléseink már kevésbé térnek el egymástól mind az átlagnál, mind a felső kvartilis esetében. A medián és az alsó kvartilis esetében még azért

a maximum likelihood módszeres becsléseink adnak jobb becslést, de észrevehető az is, hogy a momentum módszeres becsléseink néhány esetben vannak olyan jók vagy jobbak, mint a maximum likelihood módszeres becsléseinkben. A standard hiba immár mindenütt jobb a momentum módszeres becslésre, mint azt a korábbi javulásból is vártuk.

2.4.3. Kiterjesztett maximum kvázi-likelihood módszer

Ebben a fejezetben a kiterjesztett maximum kvázi-likelihood módszeres becsléseinket fogjuk összehasonlítani a Clark és Perry [4] cikkében található eredményekkel, majd megvizsgáljuk, hogy valóban rosszabb becslést kapunk-e, mint a maximum likelihood becsléseink esetében, és megnézzük, hogy ezek a becslések hogyan viszonyulnak a momentum módszeres becsléseinkhez.

A kiterjesztett maximum kvázi-likelihood módszer becslése esetében az R statisztikai programmal végrehajtott lefutás már kevesebb időt vett igénybe, mint amikor a maximum likelihood módszer becslését futtattuk, hiszen a maximalizálandó függvényünkön egyszerűsítést hajtottunk végre. A lefutási idő főleg nagy elemszámú minta esetén fontos. (Itt jegyezzük meg, hogy a momentum módszernél jobb futásidő sehol sem várható, hiszen ott maximalizálást nem végzünk.)

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1,020	0,545	90	0,957	0,641	1,322
	0,333	0,327	0,326	1498	0,289	0,094	0,516
	0,2	0,187	0,280	2745	0,154	-0,008	0,347
3	1	1,006	0,309	0	0,974	0,788	1,189
	0,333	0,325	0,147	41	0,311	0,223	0,412
	0,2	0,194	0,116	292	0,185	0,112	0,267
5	1	1,001	0,265	0	0,977	0,814	1,161
	0,333	0,328	0,113	1	0,319	0,248	0,398
	0,2	0,192	0,084	32	0,186	0,133	0,244
10	1	0,994	0,225	0	0,974	0,833	1,132
	0,333	0,329	0,088	0	0,323	0,267	0,384
	0,2	0,196	0,062	0	0,191	0,152	0,235
15	1	1,002	0,215	0	0,988	0,851	1,135
	0,333	0,327	0,080	0	0,322	0,272	0,378
	0,2	0,195	0,054	0	0,192	0,157	0,229
20	1	0,995	0,207	0	0,978	0,851	1,125
	0,333	0,328	0,076	0	0,323	0,275	0,376
	0,2	0,195	0,050	0	0,192	0,160	0,227

2.17. táblázat. 10000 szimulációs kiterjesztett maximum kvázi-likelihood becslésem n=50 minta elemszám esetén

A kiterjesztett maximum kvázi-likelihood módszeres becsléseink, amit a 2.17. táblázat mutat,

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1,012	0,550	100	0,944	0,625	1,321
	0,333	0,319	0,321	1567	0,284	0,094	0,506
	0,2	0,185	0,280	2727	0,150	-0,012	0,349
3	1	1,001	0,311	0	0,971	0,781	1,192
	0,333	0,325	0,147	34	0,312	0,222	0,415
	0,2	0,193	0,116	293	0,181	0,112	0,264
5	1	1,005	0,266	0	0,979	0,816	1,167
	0,333	0,326	0,115	1	0,317	0,246	0,396
	0,2	0,193	0,085	19	0,188	0,132	0,246
10	1	0,997	0,228	1	0,978	0,836	1,136
	0,333	0,329	0,089	0	0,322	0,267	0,385
	0,2	0,196	0,061	0	0,192	0,153	0,234
15	1	0,999	0,214	0	0,983	0,849	1,132
	0,333	0,328	0,079	1	0,322	0,272	0,376
	0,2	0,197	0,054	0	0,194	0,159	0,231
20	1	0,995	0,207	0	0,978	0,849	1,125
	0,333	0,328	0,076	0	0,324	0,274	0,376
	0,2	0,196	0,050	0	0,193	0,161	0,228

2.18. táblázat. Clark és Perry [4] 10000 szimulációs kiterjesztett maximum kvázi-likelihood becslése n=50 minta elemszám esetén

nagyon jól közelítik Clark és Perry [4] cikkbeli eredményeit az összes vizsgált esetben, ami a 2.18. táblázatban található. Csak elég kevés eltérés tapasztalható, ami a véletlenszám generátor miatt adódhatott.

Érdekes, hogy a 2.17. táblázat értékei azt mutatják, hogy a maximum kvázi-likelihood módszer becslései jobbak az átlagot, a mediánt az alsó kvartilisek értékét és sok esetben a standard hibát tekintve is, mint a maximum likelihood módszer becsléséi (2.11. táblázat). Ez azért nem annyira meglepő, mert az egyszerűbb függvénynél az optimalizálás könnyebben elvégezhető.

A 2.19. táblázat azt mutatja, hogy jobb becslést kaptunk most is a maximum kvázi-likelihood módszeres esetben, mint a maximum likelihood módszer esetében (2.13. táblázat). Viszont az $\alpha = 0{,}1$ és $\alpha = 0{,}05$ esetekben egyre többször fordul elő az, hogy a maximum likelihood becslésünk eléri illetve jobb becslést ad, főként a mediánt tekintetében. Ez főleg azért érdekes számunkra, mert az α paramétert csökkentve egyre rosszabb becslésre számíthatunk a maximum kvázi-likelihood esetében.

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	0,1	0,086	0,243	4130	0,052	-0,089	0,224
	0,05	0,038	0,229	5025	0,000	-0,131	0,167
3	0,1	0,091	0,091	1527	0,084	0,027	0,148
	0,05	0,042	0,080	3215	0,035	-0,014	0,091
5	0,1	0,095	0,063	444	0,090	0,050	0,134
	0,05	0,046	0,051	1834	0,042	0,010	0,078
10	0,1	0,097	0,041	13	0,094	0,068	0,122
	0,05	0,047	0,030	454	0,045	0,026	0,066
15	0,1	0,097	0,033	0	0,095	0,074	0,118
	0,05	0,048	0,024	96	0,046	0,031	0,063
20	0,1	0,097	0,030	0	0,095	0,076	0,116
	0,05	0,048	0,020	16	0,047	0,034	0,061

2.19. táblázat. 10000 szimulációs kiterjesztett maximum kvázi-likelihood becslésem kiegészítése n=50 minta elemszám esetén

μ	α	n	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	1,081	1,742	259	0,635	0,000	1,584
		20	1,030	0,947	108	0,840	0,381	1,511
		30	1,028	0,741	43	0,886	0,538	1,403
		50	1,034	0,573	9	0,952	0,618	1,374
1	0,333	10	0,356	1,116	490	0,064	-0,257	0,659
		20	0,308	0,542	337	0,210	-0,088	0,556
		30	0,295	0,427	263	0,221	-0,003	0,509
		50	0,333	0,326	150	0,289	0,098	0,535
1	0,2	10	0,172	0,924	574	-0,020	-0,288	0,365
		20	0,173	0,450	454	0,079	-0,166	0,395
		30	0,182	0,374	379	0,134	-0,102	0,387
		50	0,182	0,267	268	0,145	-0,004	0,339
1	0,1	10	0,086	0,744	653	-0,114	-0,360	0,288
		20	0,068	0,415	540	-0,011	-0,235	0,233
		30	0,072	0,312	496	0,000	-0,160	0,244
		50	0,089	0,238	400	0,063	-0,085	0,233
1	0,05	10	0,031	0,642	673	-0,160	-0,360	0,214
		20	0,017	0,362	596	-0,073	-0,257	0,202
		30	0,037	0,284	548	-0,013	-0,178	0,183
		50	0,031	0,216	521	-0,002	-0,130	0,149
3	0,2	10	0,165	0,297	301	0,120	-0,024	0,291
		20	0,173	0,189	168	0,145	0,041	0,271
		30	0,183	0,149	91	0,164	0,084	0,268
		50	0,193	0,114	30	0,178	0,115	0,264

2.20. táblázat. 1000 szimulációs kiterjesztett maximum kvázi-likelihood becslésem

Meglepő, hogy kis α paraméter és egyre nagyobb μ paraméter esetén már a momentum módszeres becsléseink (2.4. táblázat) is nagyon közel vannak a becsléseinkhez, sőt el is érik azokat. Ez azt mutatja, hogy ha valamilyen előzetes kalkulációból tudjuk, hogy ilyen esetben vagyunk, akkor elég a momentum módszeres becslésünkre hagyatkoznunk, nem feltétlenül érdemes a maximum kvázi-likelihood módszert használnunk.

μ	α	n	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	0,992	1,693	270	0,554	-0,002	1,458
		20	0,981	0,925	117	0,801	0,365	1,411
		30	1,017	0,723	40	0,913	0,512	1,396
		50	1,038	0,551	11	0,978	0,661	1,348
1	0,333	10	0,360	1,036	475	0,080	-0,252	0,653
		20	0,288	0,534	339	0,200	-0,108	0,557
		30	0,319	0,421	249	0,262	-0,002	0,539
		50	0,319	0,318	157	0,273	0,098	0,509
1	0,2	10	0,185	0,819	553	-0,023	-0,295	0,355
		20	0,177	0,485	457	0,069	-0,175	0,399
		30	0,177	0,365	377	0,131	-0,098	0,377
		50	0,213	0,290	252	0,177	0,002	0,376
3	0,2	10	0,161	0,284	335	0,115	-0,047	0,323
		20	0,190	0,192	124	0,157	0,063	0,276
		30	0,183	0,150	94	0,164	0,080	0,268
		50	0,191	0,111	27	0,181	0,116	0,258

2.21. táblázat. Clark és Perry [4] 1000 szimulációs kiterjesztett maximum kvázi-likelihood becslése

A 2.20. táblázatban található 1000 szimulációs eredményeink nagyon jól közelítik Clark és Perry [4] a 2.21. táblázatban szereplő eredményeit mind az átlag, mind a medián és mind a negatív értékek számának esetében. Kicsit eltér a két táblázat a standard hibánál, az alsó kvartilis és a felső kvartilisek értékeinél, de $n < 50$ esetén ez még nem jelent nagy eltérést. Picit meglepő viszont az, hogy $\mu = 1$ és $n = 10$ esetben az $\hat{\alpha}$ átlaga nagyon jól közelíti az α értékét minkét esetben. A 2.20. táblázatban megfigyelhető, hogy az α értékét csökkentve most sok negatív értékünk lett a medián és az alsó kvartilis értékeinél is.

Az elemszámot változtatva az előbbi észrevételünk most is fennáll, hiszen a maximum kvázi-likelihood módszeres becsléseink jobbnak mondhatóak a maximum likelihood módszeres becsléseinknél (2.14. táblázat) mind az átlag és mind az alsó kvartilis esetében, ezzel szemben a standard hiba minden esetben nagyobb lesz. A medián értékei egyre kisebb α-t véve jobban közelítenek a maximum likelihood módszeres becsléseinkhez.

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1000	1,031	0,112	0	1,030	0,953	1,103
		10000	1,032	0,036	0	1,030	1,007	1,057
		50000	1,032	0,017	0	1,032	1,021	1,043
1	0,333	1000	0,336	0,067	0	0,334	0,288	0,379
		10000	0,339	0,022	0	0,338	0,324	0,354
		50000	0,340	0,010	0	0,340	0,333	0,345
1	0,2	1000	0,203	0,058	0	0,200	0,163	0,241
		10000	0,203	0,018	0	0,203	0,191	0,214
		50000	0,203	0,008	0	0,203	0,197	0,209
1	0,1	1000	0,102	0,051	18	0,101	0,066	0,135
		10000	0,100	0,016	0	0,100	0,089	0,111
		50000	0,100	0,007	0	0,100	0,095	0,105
1	0,05	1000	0,050	0,048	154	0,049	0,015	0,082
		10000	0,049	0,015	0	0,049	0,038	0,060
		50000	0,050	0,007	0	0,050	0,045	0,055
3	0,2	1000	0,203	0,026	0	0,201	0,185	0,220
		10000	0,201	0,008	0	0,201	0,196	0,207
		50000	0,201	0,004	0	0,201	0,198	0,203

2.22. táblázat. 1000 szimulációs kiterjesztett maximum kvázi-likelihood becslésem kiegészítése

A 2.22. táblázat azt mutatja, hogy nem kapunk annyira pontos becslést nagy elemszám esetén, mint a maximum likelihood módszer (2.16. táblázat) és a momentum módszer - (2.9. táblázat) esetében szinte semelyik jellemzőt tekintve sem. Ezt az eredményt vártuk is, hiszen a maximum likelihood függvényünk eltolása miatt eddig is kicsit rosszabb becslésre számítottunk. E miatt a kiterjesztett maximum kvázi-likelihood módszert főként kis elemszámú mintákra érdemes alkalmazni.

2.4.4. Duplán kiterjesztett maximum kvázi-likelihood módszer

Ebben a fejezetben főként azt vizsgálom, hogy a duplán kiterjesztett maximum kvázi-likelihood módszer jobb vagy rosszabb becslést ad-e, mint a kiterjesztett maximum kvázi-likelihood módszer. Ezután összevetjük az eredményeinket a maximum likelihood módszeres becsléseinkkel és a momentum módszeres becsléseinkkel is. Saha és Paul [9] cikkében sajnos az eredményeik nem táblázatban vannak, hanem grafikusan ábrázolták őket, így nem tudjuk elég pontosan összevetni az eredményeinket az ő eredményeikkel és ők csak az 1000 szimulációs becsléseket hajtották végre kis minta elemszám esetén.

A 2.23. táblázat azt mutatja, hogy rosszabb becslést kaptunk szinte az összes átlag, az összes medián és az összes alsó kvartilis esetén, mint a 2.17. táblázatban található maximum kvázi-likelihood becsléseinknél. Megfigyelhető, hogy $\alpha = 0{,}2$ paraméternél a μ növelése mellett az át-

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	0,932	0,469	73	0,883	0,605	1,213
	0,333	0,312	0,307	1554	0,286	0,096	0,497
	0,2	0,170	0,266	2862	0,145	-0,013	0,328
3	1	0,967	0,279	0	0,944	0,771	1,139
	0,333	0,322	0,144	34	0,310	0,220	0,411
	0,2	0,190	0,113	272	0,182	0,111	0,258
5	1	0,971	0,244	0	0,947	0,800	1,119
	0,333	0,324	0,113	2	0,314	0,246	0,394
	0,2	0,193	0,083	32	0,187	0,135	0,245
10	1	0,974	0,211	0	0,960	0,824	1,108
	0,333	0,327	0,088	0	0,320	0,265	0,382
	0,2	0,195	0,061	1	0,192	0,152	0,234
15	1	0,975	0,200	0	0,962	0,832	1,100
	0,333	0,326	0,080	0	0,321	0,270	0,377
	0,2	0,195	0,053	0	0,191	0,157	0,229
20	1	0,973	0,191	0	0,960	0,841	1,093
	0,333	0,325	0,075	0	0,320	0,272	0,373
	0,2	0,196	0,050	0	0,192	0,161	0,228

2.23. táblázat. 10000 szimulációs duplán kiterjesztett maximum kvázi-likelihood becslésem n=50 minta elemszám esetén

lagban egyre jobban közelíti jelen becslésünk a kiterjesztett maximum kvázi-likelihood módszeres becsléseinket. Összességében viszont a μ paraméter növelésével sem jutunk jobb eredményhez, mint a maximum kvázi-likelihood becslésnél. A negatív számok nagyjából egyformán mozognak, viszont jobb eredményt kaptunk a standard hiba és a felső kvartilis esetében, tehát bár az átlag rosszabb, a körül kevésbé szórnak eredményeink.

De még a maximum likelihood módszeres becsléseinknél (2.11. táblázat) sem kapunk jobb eredményt, kivéve a standard hiba és a felső kvartilis esetében, mert ott jobbnak mondható a duplán kiterjesztett maximum kvázi-likelihood módszeres becslésünk (2.23. táblázat). A negatív számoknál csak kis eltérés tapasztalható. Sajnos hiába vártunk jobb becslés a maximum likelihood módszeres becsléseinknél, a μ paraméter növelése sem hozta meg a várt hatást, bár megjegyzendő, hogy a kapott becslés a rosszabb átlag körül kevésbé szór.

Az α paramétert tovább csökkentve az előbbi észrevételünk helyesnek bizonyul, mivel a 2.24. táblázatban a duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseink átlagai a 2.19. táblázatban látható kiterjesztett maximum kvázi-likelihood becslések átlagát először csak közelítik, és legalább $\mu = 15$ paraméternél már megegyeznek az értékek. Az előbb rosszabb becslést kaptunk a mediánra, de most kis μ esetén jobbat kaptunk, ugyanez mondható el az alsó

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	0,1	0,080	0,238	4194	0,054	-0,088	0,220
	0,05	0,036	0,224	4949	0,006	-0,127	0,170
3	0,1	0,094	0,092	1492	0,086	0,029	0,149
	0,05	0,042	0,079	3180	0,036	-0,013	0,091
5	0,1	0,093	0,062	500	0,089	0,050	0,132
	0,05	0,045	0,051	1920	0,041	0,009	0,077
10	0,1	0,096	0,040	22	0,093	0,067	0,121
	0,05	0,047	0,030	430	0,045	0,026	0,066
15	0,1	0,097	0,033	0	0,095	0,073	0,118
	0,05	0,048	0,024	90	0,047	0,031	0,063
20	0,1	0,097	0,030	0	0,095	0,076	0,116
	0,05	0,048	0,020	17	0,047	0,034	0,061

2.24. táblázat. 10000 szimulációs duplán kiterjesztett maximum kvázi-likelihood becslésem kiegészítése n=50 minta elemszám esetén

kvartilis esetén is. Cserébe viszont nagyon közeli eredményeket kaptunk a felső kvartilis értékeire és a standard hibára.

A duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseink a maximum likelihood módszeres becsléseinkhez (2.13. táblázat) képest már $\mu = 5$ paraméternél megegyeznek az átlag értékei, $\mu > 5$ paraméternél már az alsó és nagyjából a felső kvartilisek is megegyeznek. A medián értékei is nagyon közel vannak egymáshoz. Levonhatjuk azt a következtetést, hogy kis α paraméter és növekvő μ paraméter esetén a duplán kiterjesztett maximum kvázi-likelihood módszer a becsléseink alapján ad olyan jó becslést, mint a maximum likelihood módszer, de rosszabb becslést ad, mint a kiterjesztett maximum kvázi-likelihood becslés.

Az 1000 szimulációt elvégezve azt tapasztalhatjuk, hogy a 2.25. táblázatban most is általában rosszabb becslést kaptunk az átlag, a medián és az alsó kvartilis tekintetében, mint a maximum kvázi-likelihood módszeres becsléseink esetében, amit a 2.20. táblázat mutat. Néha előfordul egy-egy jobb becslés is, de az főleg a kis elemszámnak köszönhető. A standard hiba viszont sokkal jobb becslést ad, emiatt kisebb lesz a minta szórása. A felső kvartilis esetében $\alpha = 0{,}2$ paraméterig jobb becslést kapunk, majd kisebb α-t véve rosszabb, de azért közeli eredményeket kapunk.

Meglepő eredményeket kapunk, ha megnézzük a duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseinket (2.25. táblázat) és a maximum likelihood módszeres becsléseinket (2.14. táblázat). Ha az alsó kvartilis esetét vizsgálva $\mu = 1$ és $\alpha = 0{,}2$ paramétereknél jobb becslést kaptunk a duplán kiterjesztett maximum kvázi-likelihood módszeres becslések esetén, míg a becsléseink többi esetében rosszabbat. A medián értékeinél $\alpha = 0{,}2$ paraméternél jobb becsléseket értünk el, és az α értékét csökkentve ugyanúgy jobb eredményeket kaptunk, mint a

μ	α	n	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	0,863	1,229	246	0,643	0,000	1,437
		20	0,918	0,832	109	0,806	0,367	1,305
		30	0,930	0,636	45	0,830	0,497	1,284
		50	0,943	0,482	6	0,899	0,608	1,226
1	0,333	10	0,227	0,770	477	0,000	-0,285	0,462
		20	0,265	0,470	339	0,198	-0,082	0,531
		30	0,312	0,411	236	0,252	0,036	0,536
		50	0,311	0,308	152	0,281	0,102	0,510
1	0,2	10	0,102	0,691	569	-0,073	-0,372	0,357
		20	0,132	0,434	453	0,079	-0,199	0,367
		30	0,166	0,350	384	0,112	-0,097	0,366
		50	0,177	0,269	286	0,152	-0,016	0,346
1	0,1	10	0,052	0,611	609	-0,126	-0,399	0,314
		20	0,068	0,395	523	-0,003	-0,246	0,296
		30	0,065	0,304	496	0,010	-0,156	0,246
		50	0,090	0,245	398	0,067	-0,077	0,221
1	0,05	10	-0,001	0,549	638	-0,194	-0,399	0,210
		20	0,025	0,380	588	-0,067	-0,255	0,211
		30	0,023	0,284	544	-0,018	-0,193	0,192
		50	0,029	0,226	513	-0,002	-0,132	0,164
3	0,2	10	0,161	0,267	292	0,118	-0,018	0,290
		20	0,181	0,185	155	0,159	0,050	0,289
		30	0,184	0,152	98	0,165	0,078	0,273
		50	0,191	0,112	27	0,186	0,105	0,265

2.25. táblázat. 1000 szimulációs duplán kiterjesztett maximum kvázi-likelihood becslésem

maximum likelihood módszernél. Az átlag és a negatív számok is hasonlóan mozognak mindkét táblázatban, viszont a standard hiba jelen esetben ismét jobb.

Nagy elemszámú minta esetén $\alpha = 1$ esetben a duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseink (2.26. táblázat) nem érték el a felső kvartilis értékénél az 1 értéket, így természetesen ebben az esetben nem is lehet jobb a becslés a kiterjesztett maximum kvázi-likelihood módszeres becslésnél (2.22. táblázat), mivel a becslésünk ebben az esetben erősen torzított. $\alpha > 1$ esetén jobb becslést kaptunk mind az átlag, mind a medián és mind az alsó kvartilis értékeit tekintve.

A duplán kiterjesztett maximum likelihood módszeres becsléseinket tartalmazó 2.26. táblázat $\alpha < 0{,}1$ esetén jobb becslést kaptunk az átlag, a medián és az alsó kvartilis esetében, viszont $\alpha = 0{,}1$ paraméternél már rosszabbak lettek a becsléseink, mint a maximum likelihood módszeres becsléseink esetében, amit a 2.16. táblázat mutat. A standard hiba ismét picivel jobb, de az eltérés már nem olyan jelentős.

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1000	0,964	0,100	0	0,961	0,896	1,031
		10000	0,968	0,032	0	0,968	0,946	0,990
		50000	0,967	0,014	0	0,966	0,957	0,977
1	0,333	1000	0,330	0,064	0	0,325	0,286	0,372
		10000	0,332	0,021	0	0,332	0,318	0,346
		50000	0,332	0,009	0	0,332	0,325	0,338
1	0,2	1000	0,197	0,059	0	0,195	0,157	0,238
		10000	0,199	0,018	0	0,199	0,187	0,211
		50000	0,199	0,008	0	0,200	0,194	0,205
1	0,1	1000	0,099	0,052	23	0,099	0,062	0,135
		10000	0,100	0,016	0	0,100	0,089	0,111
		50000	0,100	0,008	0	0,100	0,095	0,105
1	0,05	1000	0,049	0,046	142	0,049	0,019	0,080
		10000	0,050	0,015	0	0,050	0,040	0,060
		50000	0,050	0,007	0	0,050	0,045	0,055
3	0,2	1000	0,200	0,025	0	0,199	0,183	0,216
		10000	0,200	0,008	0	0,200	0,195	0,205
		50000	0,200	0,004	0	0,200	0,197	0,202

2.26. táblázat. 1000 szimulációs duplán kiterjesztett maximum kvázi-likelihood becslésem kiegészítése

Itt is megállapíthatjuk, hogy a duplán kiterjesztett maximum kvázi-likelihood, ugyanúgy, mint a kiterjesztett, kis mintákra ad elfogadható becslést, és ereje leginkább abban mutatkozik meg, hogy becsléseink közül a legkisebb standard hibájú. Ezért viszont azt az árat kell fizetnünk, hogy átlagban rosszabb közelítést ad.

3. fejezet

A változó időtartamú megfigyelések esete

Ebben a fejezetben is a negatív binomiális eloszlásokkal fogunk foglalkozni, viszont most az eseményeket t_i ideig figyeljük meg, ahol $t_i \sim E(0{,}2)$ egyenletes eloszlású valószínűségi változó, így az események $NB(t_i\mu, \alpha)$ eloszlásúak. Ugyanazokkal a módszerekkel foglalkozunk most is, mint amiket már az előző fejezetben ismertettünk. Arra keressük a választ, hogy ha a mintavételkor csak egy bizonyos ideig figyeljük meg az eseményeket, akkor jobb vagy rosszabb becslést kapunk-e α-ra. Ez a modell jobban közelíti a valóságos helyzetet, hiszen egy adott megfigyelési időszakban a megfigyelt egyedek megfigyelési időtartamai a legritkább esetben egyenlőek, mivel sok belépés és kilépés is történhet.

A változó időtartalmú megfigyelésekkel azért foglalkozunk, mert ez egy természetes paraméterezés, ami abból ered, hogy a negatív binomiális eloszlás egyben keverék Poisson eloszlás, ahol a Poisson eloszlás paramétere vt, ahol t megfelel a fenti idő paraméternek.

A különböző alfejezetekben megnézzük, hogy a módszerek miben változnak a megfigyelések során. Majd ezekre lefuttatjuk a szimulációkat és ezeket az eredményeket összevetjük az előző fejezetek eredményeivel.

3.1. Momentum módszer

Ebben az esetben csak az első és a második momentummal foglalkozunk, mivel láttuk, hogy az első és a harmadik momentum nem ad jobb becslést. Mivel a kapott eredményeket saját szimulációs eredményekkel szeretnénk összehasonlítani a tapasztalati szórásból eredő becsléssel nem foglalkozunk, továbbá, mivel a harmadik momentumra rosszabb becslést kaptunk már korábban, most azzal sem foglalkozunk.

Legyen $X \sim NB(t_i\mu, \alpha)$ valószínűségi változó, ahol a $t_i \sim E(0{,}2)$ valószínűségi változó. Ekkor $E(X) = t_i\mu$, $D^2(X) = t_i\mu + t_i^2\alpha\mu^2$ és $E(X^2) = t_i\mu + t_i^2\mu^2(1+\alpha)$. Ebben az esetben a megfigyeléseink nem azonos eloszlásúak.

$$E\left(\frac{X_1 + ... + X_n}{n}\right) = \frac{\sum_{i=1}^n t_i\mu}{n},$$

$$E\left(\frac{X_1^2 + ... + X_n^2}{n}\right) = \frac{(\sum t_i)\mu}{n} + \frac{(\sum t_i^2)\mu^2(1+\alpha)}{n}.$$

A momentum módszerrel kapott becsléseink ezzel az átparaméterezett negatív binomiális eloszlású változóra a következők. a $\bar{t} = \sum_{i=1}^n \frac{t_i}{n}$ jelöléssel:

$$\hat{\mu} = \bar{X}/\bar{t},$$

$$\hat{\alpha} = \frac{(\bar{t})^2\bar{X^2} - (\bar{t})^2\bar{X}}{\bar{t^2}(\bar{x})^2} - 1.$$

A momentum módszer kiszámításakor felmerülhet még egy olyan lehetőség, hogy abból indulunk ki, hogy rögzített t_i esetén ha $X_i \sim NB(t_i\mu, \alpha)$, akkor

$$E\left(\frac{X_i}{t_i}\right) = \mu,$$

$$E\left(\frac{X_i^2}{t_i^2}\right) = \frac{\mu}{t_i} + \mu^2(1+\alpha).$$

Ezekből kapjuk, hogy

$$\hat{\mu_2} = \frac{1}{n}\sum_{i=1}^n \frac{X_i}{t_i},$$

$$\hat{\alpha_2} = \frac{\frac{1}{n}\sum_{i=1}^n \frac{X_i}{t_i} - \frac{1}{n^2}\sum_{i=1}^n \frac{X_i}{t_i}\sum_{i=1}^n \frac{1}{t_i}}{\left(\frac{1}{n}\sum_{i=1}^n \frac{X_i}{t_i}\right)^2}.$$

Nem fűzünk túl nagy reményeket ehhez a becsléshez, mert ha a t_i valószínűségi változó nagyon kicsi, akkor bár valószínűleg az X_i valószínűségi változónk és így $\frac{X_i}{t_i}$ nagy valószínűséggel 0 lesz, míg a $\frac{1}{t_i}$ tag nagyon nagy lesz. Ekkor a számláló elszállhat és nem kapunk pontos becsléseket.

3.2. Maximum likelihood módszer

A maximum likelihood módszer esetében ugyanúgy a log-likelihood függvényt számoljuk ki és azt maximalizáljuk. Az átparaméterezett negatív binomiális eloszlású változóra a következő log-likelihood függvényt kapjuk:

$$\widetilde{l}(\alpha,\mu)=\sum_{i=1}^{n}\log\left(\frac{\Gamma\left(y_i+\alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right)}\right)-\sum_{i=1}^{n}\log y_i!+\sum_{i=1}^{n}y_i\log\left(t_i\mu\right)-$$

$$-\sum_{i=1}^{n}\left(y_i+\alpha^{-1}\right)\log\left(1+t_i\alpha\mu\right)+\sum_{i=1}^{n}y_i\log\left(\alpha\right). \tag{3.1}$$

A $\sum_{i=1}^{n} y_i \log(t_i\mu)$ tag felírható $\sum_{i=1}^{n} y_i \log(t_i) + \sum_{i=1}^{n} y_i \log(\mu)$ alakban, ezért ennek az összegnek az első tagja elhagyható a maximalizáláskor. Az előző fejezetben a maximalizáláskor elhagyott tagok most is elhagyhatóak és az átalakítások most is elvégezhetőek. A likelihood függvényünket osszuk le szintén n-el.

$$l(\alpha,\mu)=\frac{1}{n}\sum_{i=1}^{n}\sum_{\nu=0}^{y_i-1}\log\left(1+\alpha\nu\right)+\frac{1}{n}\sum_{i=1}^{n}y_i\log\left(\mu\right)-\frac{1}{n}\sum_{i=1}^{n}\left(y_i+\alpha^{-1}\right)\log\left(1+t_i\alpha\mu\right). \tag{3.2}$$

Deriválás után pedig a következőket kapjuk:

$$\frac{\partial l(\alpha,\mu)}{\partial\mu}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\mu(1+t_i\alpha\mu)}-\frac{t_i\mu}{\mu(1+t_i\alpha\mu)}\right), \tag{3.3}$$

$$\frac{\partial l(\alpha,\mu)}{\partial\alpha}=\frac{1}{n}\sum_{i=1}^{n}\sum_{\nu=0}^{y_i-1}\frac{\nu}{1+\alpha\nu}+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\alpha^2}\log\left(1+t_i\alpha\mu\right)-\frac{t_i\mu\left(y_i+\alpha^{-1}\right)}{1+t_i\alpha\mu}\right). \tag{3.4}$$

A (3.3) derivált nullhelyéből rögtön adódik, hogy a $\hat{\mu}=\bar{y}/\bar{t}$ becslés jó lesz. A (3.4) kifejezés egzakt nullhelyének meghatározása bonyolult, ezért ismét numerikus közelítést kell alkalmaznunk. Mi a korábbiakhoz hasonlóan az R programmal az l függvény maximumhelyét kerestük α-ra a $\mu=\bar{y}/\bar{t}$ helyen. Az α-tól független tagokat ez esetben is elhagyhatjuk.

3.3. Kiterjesztett maximum kvázi-likelihood módszer

Hasonlóképpen számoljuk ki a log-likelihood függvényünket az átparaméterezett negatív binomiális eloszlás esetében is a kiterjesztett maximum kvázi-likelihood módszernél. A következőt kapjuk:

$$l_2(\mu,\alpha)=\sum_{i=1}^{n}\left[y_i\log\left(\frac{t_i\mu}{y_i}\right)-\left(\frac{1+\alpha y_i}{\alpha}\right)\log\left(\frac{1+t_i\alpha\mu}{1+\alpha y_i}\right)-\frac{1}{2}\log\left(2\pi\right)-\log\left(1+\alpha y_i\right)-\right.$$

$$\left.-\frac{1}{2}\log\left(y_i+\frac{1}{6}\right)-\frac{1}{2}\log\left(1+\frac{\alpha}{6}\right)+\frac{1}{2}\log\left(\alpha y_i+1+\frac{\alpha}{6}\right)\right]. \tag{3.5}$$

A log-likelihood függvény deriváltjai a következők:

$$\frac{\partial l_2(\mu,\alpha)}{\partial \mu} = \sum_{i=1}^{n}\left[\frac{y_i}{\mu} - \frac{1+\alpha y_i}{1+t_i\alpha\mu}\right], \tag{3.6}$$

$$\frac{\partial l_2(\mu,\alpha)}{\partial \alpha} = \sum_{i=1}^{n}\left[\frac{1}{\alpha^2}\log\left(\frac{1+t_i\alpha\mu}{1+\alpha y_i}\right) - \frac{t_i\mu(1+\alpha y_i)}{\alpha(1+t_i\alpha\mu)} + \frac{y_i}{\alpha} - \frac{y_i}{1+\alpha y_i} - \right.$$

$$\left. - \frac{1}{12\left(1+\frac{\alpha}{6}\right)} + \frac{y_i+\frac{1}{6}}{2\left(\alpha y_i+1+\frac{\alpha}{6}\right)}\right]. \tag{3.7}$$

A (3.6) derivált nullhelyéből rögtön adódik, hogy ugyanaz a becslésünk jó lesz, mint a maximum likelihood módszernél. Tehát $\hat{\mu} = \bar{y}/\bar{t}$. A (3.6) kifejezés egzakt nullhelyének meghatározása bonyolult, ezért ismét numerikus közelítést kell alkalmaznunk. Mi a korábbiakhoz hasonlóan az R programmal az l_2 függvény maximumhelyét kerestük α-ra a $\mu = \bar{y}/\bar{t}$ helyen. Az α-tól független tagok természetesen ismét elhagyhatóak.

3.4. Duplán kiterjesztett maximum kvázi-likelihood módszer

A log-likelihood függvényünk így néz ki az átparaméterezés után:

$$l_3(\mu,\alpha) = \sum_{i=1}^{n}\left[y_i\log(t_i\mu) + \left(y_i+\frac{1}{\alpha}\right)\log\left(\frac{1+\alpha y_i}{1+t_i\alpha\mu}\right) - \frac{1}{2}\log(1+\alpha y_i) - \right.$$

$$\left. - \left(y_i-\frac{1}{2}\right)\log(y_i) + \frac{\alpha}{12(1+\alpha y_i)} - \frac{\alpha}{12} - \frac{1}{12y_i} - \frac{1}{2}\log(2\pi)\right]. \tag{3.8}$$

Nézzük ennek a deriváltjait:

$$\frac{\partial l_3(\mu,\alpha)}{\partial \mu} = \sum_{i=1}^{n}\left(\frac{y_i}{\mu(1+t_i\alpha\mu)} - \left(\frac{1}{1+t_i\alpha\mu}\right)\right), \tag{3.9}$$

$$\frac{\partial l_3(\mu,\alpha)}{\partial \alpha} = \sum_{i=1}^{n}\left[\frac{1}{\alpha^2}\log\left(\frac{1+t_i\alpha\mu}{1+\alpha y_i}\right) + \frac{y_i-t_i\mu}{\alpha(1+t_i\alpha\mu)} - \frac{2\alpha y_i+2-\alpha}{2\alpha^2(1+\alpha y_i)} + \frac{2-\alpha}{2\alpha^2} - \frac{\alpha y_i(2+\alpha y_i)}{12(1+\alpha y_i)^2}\right]. \tag{3.10}$$

A (3.9) derivált nullhelyéből ismét adódik, hogy ugyanaz a becslésünk jó lesz, mint a maximum likelihood módszernél. Tehát $\hat{\mu} = \bar{y}/\bar{t}$. A (3.9) kifejezés egzakt nullhelyének meghatározása bonyolult, ezért ismét numerikus közelítést kell alkalmaznunk. Mi a korábbiakhoz hasonlóan az R programmal az l_3 függvény maximumhelyét kerestük α-ra a $\mu = \bar{y}/\bar{t}$ helyen. Az α-tól független tagok természetesen ismét elhagyhatóak.

3.5. Szimulációk és eredmények

Ugyanazokat a szimulációkat hajtottuk végre ugyanazokkal a módszerekkel, mint az előző fejezetben, viszont a momentum módszer esetében a harmadik és az első momentum által adott becslést már nem szimuláltuk. Először a 10000 szimulációs eredményeinket vizsgáljuk, majd az 1000 szimulációs becsléseket.

3.5.1. Momentum módszer

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	0,886	0,605	196	0,785	0,472	1,171
	0,333	0,292	0,323	1629	0,249	0,073	0,463
	0,2	0,163	0,276	2894	0,130	-0,028	0,313
3	1	0,908	0,449	1	0,825	0,608	1,115
	0,333	0,305	0,190	212	0,280	0,171	0,409
	0,2	0,182	0,144	759	0,166	0,082	0,263
5	1	0,915	0,407	0	0,844	0,637	1,108
	0,333	0,310	0,164	74	0,289	0,195	0,402
	0,2	0,187	0,121	317	0,172	0,103	0,256
10	1	0,929	0,399	0	0,855	0,660	1,113
	0,333	0,313	0,144	12	0,294	0,215	0,390
	0,2	0,187	0,099	79	0,176	0,119	0,242
15	1	0,925	0,378	0	0,859	0,664	1,103
	0,333	0,314	0,136	2	0,296	0,219	0,388
	0,2	0,190	0,093	47	0,180	0,126	0,242
20	1	0,921	0,380	0	0,850	0,662	1,097
	0,333	0,314	0,132	0	0,297	0,224	0,385
	0,2	0,188	0,088	44	0,178	0,126	0,238

3.1. táblázat. 10000 szimulációs változó időtartamú momentum módszeres becslésem n=50 minta elemszám esetén

Azt vehetjük észre, hogy a 3.1. táblázat a várt eredményeket mutatja. A változó időtartamú megfigyeléseknél rosszabb átlagot, standard hibát, mediánt és alsó kvartilist kaptunk, mint ami a 2.1. táblázatban kaptunk. A hiba növekedését mutatja az is, hogy a felső kvartilis esetében megfigyelhető, hogy bár a $\mu = 1$ esetén a változó időtartalmú megfigyelések esetében a becsült α-hoz közelebbi értéket kaptunk, viszont ez a mediántól távolabb van, és a μ növekedésével a becsült α-tól is jobban eltávolodik, mint az időben nem változó esetben, de elég közel marad a 2.1. táblázat felső kvartilis értékeihez. Egy kicsit meglepő, hogy még $\mu > 10$ esetén is ilyen sok negatív értékünk van a szimuláció során.

Vegyes képet kapunk mostani eredményeink és az első és a harmadik momentum becsléseiből adódott 2.3. táblázat eredményeinek összehasonlításakor, mivel átlagban jobb lett a közelítésünk, viszont standard hibája már nagyobb, és a negatív értékek számát, a mediánt illetve a kvartiliseket tekintve is inkább a 2.3. táblázat hozható ki győztesnek.

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	0,1	0,071	0,234	4126	0,047	-0,092	0,203
	0,05	0,028	0,221	4965	0,002	-0,124	0,153
3	0,1	0,090	0,112	2143	0,080	0,012	0,156
	0,05	0,041	0,097	3498	0,034	-0,026	0,100
5	0,1	0,092	0,086	1308	0,085	0,032	0,145
	0,05	0,044	0,072	2829	0,038	-0,006	0,088
10	0,1	0,093	0,067	647	0,088	0,048	0,135
	0,05	0,045	0,051	1886	0,041	0,009	0,076
15	0,1	0,094	0,060	405	0,089	0,053	0,131
	0,05	0,046	0,043	1334	0,044	0,017	0,073
20	0,1	0,094	0,056	277	0,091	0,055	0,129
	0,05	0,047	0,040	1032	0,045	0,020	0,071

3.2. táblázat. 10000 szimulációs változó időtartamú momentum módszeres becslésem kiegészítése n=50 minta elemszám esetén

Picit meglepő, hogy $\mu = 1$-re és kis α-ra kicsivel talán jobb becslést ad jelen módszerünk a 3.2. táblázatot nézve, mint a 2.4. táblázatban vizsgált. Viszont a nagyobb μ esetén továbbra is a fix paraméterű becslés a jobb. Itt is megfigyelhető, hogy sokkal több a negatív becslés, ahogy a μ értéke növekszik, például a 2.4. táblázat szerint $\mu = 20$ és $\alpha = 0{,}05$ esetén csak 20 negatív szám van, míg a 3.2. táblázatban ebben az esetben 1032 negatív szám szerepel. Viszont a felső és az alsó kvartilis is mind rosszabb eredményt mutat.

Ha összevetjük a 3.2. táblázat eredményeit a 2.5. táblázat eredményeivel, akkor azt vehetjük észre, hogy itt is jobb becslést kapunk $\mu = 1$ esetében, viszont nagyobb μ esetén már tapasztalható a nagyobb α-ra már látott tendencia, hogy a fix paraméteres becslés itt is kicsit jobb becslést ad.

A 3.3. táblázat eredményeiben az látható, hogy kis minta esetén már átlagban sem kapunk jobbat, mint a 2.6. táblázat eredménye, csak néhány kivételes esetben. Észrevehető, hogy sok helyen alsó kvartilisnek is negatív az értéket adnak, ami összefügg azzal, hogy viszonylag sok negatív értéket kaptunk a szimuláció során.

A nagy elemszám miatt már nincs nagy különbség a 3.4. táblázat eredményei és a 2.9. táblázat eredményei között sem az átlag, sem a medián sem pedig a felső kvartilis értékei között, viszont rosszabb eredményeket kaptunk az alsó kvartilis és a standard hiba értékeire.

μ	α	n	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	0,495	0,928	302	0,353	-0,089	0,893
		20	0,759	0,826	137	0,599	0,201	1,134
		30	0,828	0,787	79	0,661	0,310	1,152
		50	0,883	0,615	20	0,777	0,447	1,198
1	0,333	10	0,086	0,637	522	-0,027	-0,308	0,362
		20	0,243	0,492	338	0,162	-0,083	0,477
		30	0,280	0,421	260	0,209	-0,009	0,501
		50	0,287	0,320	172	0,251	0,070	0,444
1	0,2	10	-0,004	0,564	573	-0,074	-0,332	0,242
		20	0,132	0,446	425	0,072	-0,175	0,347
		30	0,165	0,356	351	0,112	-0,068	0,340
		50	0,173	0,276	261	0,137	-0,007	0,323
1	0,1	10	-0,061	0,506	613	-0,121	-0,357	0,188
		20	0,044	0,376	489	0,009	-0,207	0,230
		30	0,055	0,320	479	0,015	-0,161	0,212
		50	0,066	0,225	404	0,047	-0,094	0,198
1	0,05	10	-0,096	0,472	662	-0,144	-0,388	0,143
		20	-0,002	0,355	546	-0,031	-0,237	0,172
		30	-0,003	0,265	546	-0,035	-0,188	0,147
		50	0,008	0,204	507	-0,005	-0,137	0,134
3	0,2	10	0,101	0,269	398	0,050	-0,078	0,237
		20	0,159	0,225	241	0,128	0,002	0,277
		30	0,165	0,185	184	0,147	0,037	0,263
		50	0,186	0,140	73	0,170	0,090	0,267

3.3. táblázat. 1000 szimulációs változó időtartamú momentum módszeres becslésem

Az viszont kicsit meglepő, hogy a 3.4. táblázatban látható eredményeink szerint a 2.10. táblázat eredményeihez képest most jobb a becslésünk főleg nagyobb α-ra.

A 3.5. táblázatban láthatjuk, hogy a vártnál sokkal rosszabb becsléseket kaptunk az előző momentum módszeres esetünknél, (amit a 3.1. táblázat foglal össze). Megfigyelhető, hogy a μ paraméter növelésével egyre jobb közelítéseket kaptunk, de már $\mu \geq 5$ esetén észrevehető, hogy legjobban a medián közelíti a becsült paramétert. A standard hiba is és a negatív értékek száma is csökken a μ növelésekor, $\mu = 20$ és $\alpha = 1$ paraméterek esetén már csak 287 negatív számunk van, míg $\mu = 1$ és $\alpha = 1$ paraméterek esetén 5967 negatív számunk volt. A kvartilisekből és a mediánból látszik, hogy egy-egy kiugró érték rontja csak le az átlagot és a standard hibát, ami valószínűleg annak felel meg, hogy egy-egy mintába nagyon kis t_i érték került, és így a $\frac{1}{n}\sum\frac{1}{t_i}$ érték nagyon nagy lett.

A 3.6. táblázat azt mutatja, hogy kisebb α értékeket nézve most is megfigyelhető a javuló tendencia a becsléseinknél és már nem szóródnak olyan nagyon.

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1000	1,001	0,157	0	0,989	0,885	1,102
		10000	1,000	0,049	0	0,999	0,969	1,031
		50000	1,001	0,022	0	1,000	0,987	1,015
1	0,333	1000	0,331	0,078	0	0,327	0,278	0,381
		10000	0,334	0,024	0	0,334	0,319	0,349
		50000	0,334	0,011	0	0,334	0,327	0,341
1	0,2	1000	0,198	0,065	0	0,194	0,156	0,239
		10000	0,201	0,020	0	0,201	0,186	0,214
		50000	0,200	0,009	0	0,200	0,194	0,206
1	0,1	1000	0,098	0,052	31	0,096	0,061	0,131
		10000	0,099	0,017	0	0,099	0,088	0,111
		50000	0,100	0,007	0	0,099	0,094	0,105
1	0,05	1000	0,050	0,050	165	0,048	0,014	0,081
		10000	0,049	0,015	0	0,049	0,038	0,059
		50000	0,050	0,007	0	0,050	0,045	0,055
3	0,2	1000	0,199	0,034	0	0,198	0,176	0,219
		10000	0,200	0,011	0	0,200	0,193	0,208
		50000	0,200	0,005	0	0,200	0,197	0,203

3.4. táblázat. 1000 szimulációs változó időtartamú momentum módszeres becslésem kiegészítése

μ	α	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	-4,027	50,475	5967	-0,349	-1,977	0,579
	0,333	-10,060	475,820	7667	-0,958	-2,569	-0,063
	0,2	-33,780	2364,722	8091	-1,080	-2,762	-0,236
3	1	-0,698	32,011	2416	0,545	0,027	0,954
	0,333	-0,811	10,804	5070	-0,008	-0,496	0,262
	0,2	-2,425	120,943	6169	-0,122	-0,616	0,133
5	1	0,089	10,863	1338	0,709	0,390	1,019
	0,333	-0,551	16,137	3308	0,145	-0,126	0,325
	0,2	-0,689	11,459	4626	0,024	-0,253	0,182
10	1	0,596	3,978	637	0,821	0,592	1,065
	0,333	-0,220	15,449	1635	0,236	0,107	0,351
	0,2	-0,148	4,210	2589	0,118	-0,008	0,207
15	1	0,733	2,803	415	0,847	0,657	1,071
	0,333	-0,159	15,616	1114	0,265	0,168	0,359
	0,2	-0,011	3,488	1708	0,145	0,064	0,216
20	1	0,641	11,961	287	0,865	0,689	1,086
	0,333	0,167	2,234	782	0,280	0,196	0,364
	0,2	-0,051	5,389	1318	0,156	0,089	0,217

3.5. táblázat. 10000 szimulációs változó időtartamú momentum módszeres becslésem (a másodszorra ismertetett módszerrel) n=50 minta elemszám esetén

μ	α	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	0,1	-4,352	30,597	8319	-1,177	-2,762	-0,318
	0,05	-4,564	60,625	8360	-1,206	-2,829	-0,354
3	0,1	-1,127	9,603	7073	-0,212	-0,692	0,044
	0,05	-1,687	26,302	7571	-0,268	-0,792	-0,008
5	0,1	-1,644	69,925	6072	-0,064	-0,347	0,084
	0,05	-14,410	1363,444	6940	-0,106	-0,391	0,036
10	0,1	-0,217	2,646	4059	0,031	-0,093	0,106
	0,05	-0,246	3,003	5536	-0,015	-0,140	0,057
15	0,1	-0,462	19,902	2936	0,053	-0,022	0,108
	0,05	-0,254	5,945	4605	0,008	-0,065	0,055
20	0,1	-0,057	2,568	2211	0,068	0,013	0,111
	0,05	-0,193	3,775	3629	0,021	-0,030	0,057

3.6. táblázat. 10000 szimulációs változó időtartamú momentum módszeres becslésem (a másodszorra ismertetett módszerrel) kiegészítése n=50 minta elemszám esetén

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	-4,261	25,303	574	-0,240	-2,389	0,560
		20	-6,466	78,738	575	-0,251	-1,858	0,570
		30	-3,419	21,909	584	-0,358	-2,331	0,593
		50	-943,500	29698,673	580	-0,336	-1,992	0,575
1	0,333	10	-6,464	53,388	750	-0,698	-2,630	0,000
		20	-4,723	31,651	728	-0,757	-2,370	0,063
		30	-4,231	27,526	754	-0,861	-2,454	-0,017
		50	-2,841	11,427	786	-0,928	-2,536	-0,111
1	0,2	10	-4,310	15,701	771	-0,760	-2,696	-0,073
		20	-6,318	59,663	770	-0,945	-2,648	-0,070
		30	-5,712	52,157	792	-1,012	-2,746	-0,140
		50	-4,102	33,160	788	-0,967	-2,632	-0,150
1	0,1	10	-5,997	51,219	788	-0,804	-2,928	-0,085
		20	-7,203	69,371	807	-0,973	-2,700	-0,215
		30	-6,513	91,613	795	-1,077	-3,001	-0,184
		50	-12,020	260,384	828	-1,208	-3,151	-0,306
1	0,05	10	-10,290	200,636	810	-0,785	-2,645	-0,164
		20	-5,672	35,917	806	-1,039	-2,923	-0,218
		30	-6,672	78,564	818	-1,091	-2,974	-0,267
		50	-3,997	25,521	836	-1,188	-2,788	-0,386
3	0,2	10	-1,566	18,423	597	-0,085	-0,478	0,150
		20	-2,288	30,029	583	-0,084	-0,488	0,166
		30	-2,072	28,854	607	-0,108	-0,683	0,150
		50	-1,012	8,343	629	-0,136	-0,617	0,125

3.7. táblázat. 1000 szimulációs változó időtartamú momentum módszeres becslésem (a másodszorra ismertetett módszerrel)

A 3.7. táblázatunkban látható, hogy a kis minta elemszám miatt nagyon nagy eltérések lehet az átlag és a standard hiba értékei között is. Ez mutatja, hogy nagyon nagy kiugró értékek is előfordulhatnak a becslésben. A medián és a kvartilisek egymástól való eltérései nem túl nagyok. Láthatjuk, hogy $\mu = 3$ esetben a felső kvartilisek nem is érik el a becsülni kívánt α értékét sem.

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1000	-2,435	14,508	770	-1,002	-2,483	-0,081
		10000	-4,742	46,058	839	-1,581	-3,233	-0,580
		50000	-2,648	61,411	864	-1,856	-3,331	-0,829
1	0,333	1000	-4,228	23,773	857	-1,618	-3,207	-0,662
		10000	-7,065	85,357	907	-2,257	-3,703	-1,332
		50000	-15,280	340,982	910	-2,694	-4,085	-1,718
1	0,2	1000	-5,261	36,540	877	-1,808	-3,331	-0,878
		10000	-7,443	82,727	908	-2,425	-3,879	-1,453
		50000	-5,740	50,936	916	-2,790	-4,269	-1,880
1	0,1	1000	-2,854	13,936	880	-1,772	-3,363	-0,887
		10000	-3,853	11,104	915	-2,560	-4,021	-1,562
		50000	-5,017	31,930	918	-2,792	-4,357	-1,823
1	0,05	1000	-13,870	210,142	894	-1,914	-3,593	-0,955
		10000	-3,925	16,191	903	-2,516	-4,013	-1,374
		50000	-3,894	10,180	922	-2,991	-4,418	-2,049
3	0,2	1000	-0,678	4,555	790	-0,373	-0,838	-0,068
		10000	-15,850	453,817	862	-0,610	-1,149	-0,266
		50000	-0,666	9,599	896	-0,703	-1,155	-0,385

3.8. táblázat. 1000 szimulációs változó időtartamú momentum módszeres becslésem (a másodszorra ismertetett módszerrel) kiegészítése

Nem túl meglepő módon a 3.8. táblázatunk azt mutatja, hogy a minta elemszám növelésével nemhogy jobb, hanem sokkal rosszabb eredményeket kaptunk, mivel egy nagyobb mintában még inkább előfordulhat egy-egy nagyon kis t_i érték. Negatív értékkel szerepel mind az átlag, mind a medián, mind az alsó és felső kvartilis is. Sajnos az is észrevehető, hogy az α paraméter csökkenésével abszolút nem mutat még közeli eredményt sem az átlag. A standard hiba nagyságából látszik, hogy nagyon sok kiugró értékünk van, ami a becslést nagyon elrontja.

3.5.2. Maximum likelihood módszer

Most megnézzük, hogy a változó időtartamú megfigyelések milyen mértékben változnak a maximum likelihood módszeres becsléseinkhez képest.

A változó időtartamú maximum likelihood becsléseink, ami a 3.9. táblázatban található, nagyon jól közelítik alulról az átlagot, az alsó és a felső kvartiliseket és a standard hibát is, sőt időnként még el is érik a fix paraméterű maximum likelihood becsléseinknél tapasztalt értékeket,

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	0,976	0,536	84	0,902	0,605	1,255
	0.333	0,313	0,305	1313	0,273	0,107	0,483
	0.2	0,176	0,259	2476	0,145	0,001	0,319
3	1	0,988	0,325	0	0,956	0,755	1,178
	0.333	0,323	0,149	36	0,308	0,218	0,411
	0.2	0,190	0,112	251	0,180	0,112	0,258
5	1	0,985	0,275	0	0,960	0,792	1,147
	0.333	0,322	0,119	0	0,313	0,238	0,393
	0.2	0,193	0,086	36	0,186	0,133	0,245
10	1	0,987	0,237	0	0,971	0,817	1,131
	0.333	0,326	0,095	0	0,318	0,258	0,385
	0.2	0,194	0,065	0	0,189	0,148	0,234
15	1	0,990	0,222	0	0,973	0,834	1,130
	0.333	0,328	0,085	0	0,321	0,267	0,380
	0.2	0,195	0,058	0	0,191	0,155	0,231
20	1	0,992	0,212	0	0,974	0,843	1,125
	0.333	0,328	0,082	0	0,323	0,270	0,380
	0.2	0,195	0,052	0	0,191	0,158	0,227

3.9. táblázat. 10000 szimulációs változó időtartamú maximum likelihood becslésem n=50 minta elemszám esetén

ami a 2.11. táblázatban található. Kicsivel kevesebb negatív szám található, ennek köszönhető, hogy az alsó kvartilisek között nincs negatív értékünk.

Mivel ennyire közeli eredményeket kaptunk a maximum likelihood módszeres becslésekhez, ezért nem meglepő, hogy az átlag és a medián sokkal jobb, az alsó kvartilis értékei pedig általában jobb eredményeket mutatnak a változó időtartamú becsléseinknél (3.9. táblázat), mint a fix paraméteres momentum módszeres becsléseink esetében (2.1. táblázat), (és mint az fix paraméterűnél rosszabb értékeket mutató változó paraméteres momentum módszeres becsléseink esetében). Nagyon hasonló eredményeket kaptunk a standard hibára. Érdekes viszont az a tény, hogy a felső kvartilis $\alpha = 1$ paraméter esetén most rosszabb, de a többi α paraméterre már nagyon közeli.

Meglepő, hogy kisebb α paraméter esetén már nem lehet éles különbséget tenni a változó időtartamú (3.10. táblázat) és fix paraméteres (2.13. táblázat) maximum likelihood módszeres becsléseink között az átlag, a medián, az alsó és felső kvartilis értékeinél. Kis μ esetén a változó paraméterű esetben találhatóak kicsit jobb értékek, míg nagyobb μ esetén a fix paraméterű esetben, de az eltérés nem jelentős.

Mivel ennyire közeli eredményeket kaptunk a két maximum likelihood módszeres becsléseink

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	0.1	0,080	0,225	3908	0,052	-0,073	0,206
	0.05	0,029	0,206	4832	0,008	-0,115	0,146
3	0.1	0,091	0,088	1385	0,082	0,029	0,143
	0.05	0,041	0,073	3080	0,034	-0,011	0,086
5	0.1	0,094	0,062	440	0,088	0,050	0,132
	0.05	0,045	0,049	1830	0,041	0,010	0,074
10	0.1	0,095	0,042	24	0,092	0,066	0,121
	0.05	0,046	0,030	462	0,044	0,024	0,065
15	0.1	0,096	0,036	2	0,094	0,071	0,118
	0.05	0,048	0,024	99	0,046	0,031	0,062
20	0.1	0,096	0,032	0	0,094	0,073	0,116
	0.05	0,048	0,021	21	0,046	0,033	0,060

3.10. táblázat. 10000 szimulációs változó időtartamú maximum likelihood becslésem kiegészítése n=50 minta elemszám esetén

között, ezért már azt is tudjuk, hogy a változó időtartamú maximum likelihood módszeres becsléseink jobb eredményt adnak a momentum módszeres eredményeknél, hiszen a fix paraméterű maximum likelihood módszeres eredményeink is jobb becslést adtak és a momentum módszernél láttuk, hogy a fix paraméteres momentum módszer jobb becslést adott, mint a változó időtartamú momentum módszer.

$\alpha = 1$ és $\mu = 1$ esetén kis minta elemszámra is nagyon jól közelít az átlagban a 3.11. táblázatban található változó időtartamú maximum likelihood módszeres becslésünk és még a medián értékei is jobbak, mint a 2.14. táblázatban található fix paraméterű maximum likelihood módszeres becslésem. Kisebb α paraméter esetén hiába kapunk jobb értéket szinte az összes alsó és felső kvartilisre, sem a medián, sem az átlag nem jobb. $\mu = 3$ esetben is ugyanúgy rosszabb átlagot és mediánt kaptunk. Érdekes viszont, hogy a standard hiba szinte minden esetben alacsonyabb a változó paraméteres esetben.

Azt várnánk, hogy kevés minta elemszám esetén a 3.3. momentum módszeres táblázat jobb eredményt fog adni, mint a maximum likelihood módszeres becsléseink. Tényleg jobb eredményeket kapunk szinte az összes kvartilis esetében és sokkal kisebb standard hibát észlelünk, mégis sem az átlag, sem a medián nem adott jobb becslés ebben az esetben. Ez következik a momentum módszeres becslésnél megfigyelt torzításból.

Az előző észrevételeink, amit az 1000 szimulációs maximum likelihood módszeres becsléseink között tettünk a 3.12. táblázatot vizsgálva látható, hogy nagy minta elemszámra is hasonlóakat mondhatunk el. Természetesen az átlag és a medián értékei is már alig térnek el a becsülni kívánt α értékétől a 3.12. táblázat eredményeiben a változó időtartamú maximum likelihood

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	1,036	2,133	256	0,570	-0,017	1,364
		20	0,979	1,007	107	0,790	0,354	1,369
		30	1,005	0,761	39	0,864	0,480	1,361
		50	0,953	0,530	9	0,884	0,600	1,232
1	0,333	10	0,258	1,096	479	0,029	-0,333	0,526
		20	0,284	0,554	306	0,206	-0,072	0,509
		30	0,288	0,406	237	0,225	0,012	0,497
		50	0,324	0,320	141	0,297	0,100	0,493
1	0,2	10	0,139	0,801	531	-0,028	-0,333	0,336
		20	0,148	0,450	403	0,079	-0,154	0,355
		30	0,166	0,392	342	0,104	-0,069	0,313
		50	0,180	0,245	237	0,152	0,007	0,339
1	0,1	10	0,007	0,676	618	-0,158	-0,333	0,224
		20	0,055	0,367	517	-0,012	-0,198	0,243
		30	0,062	0,327	472	0,016	-0,139	0,200
		50	0,080	0,227	399	0,050	-0,069	0,203
1	0,05	10	-0,039	0,649	657	-0,174	-0,333	0,135
		20	0,006	0,348	576	-0,053	-0,234	0,166
		30	0,013	0,272	546	-0,034	-0,175	0,154
		50	0,027	0,209	493	0,002	-0,119	0,142
3	0,2	10	0,139	0,270	318	0,090	-0,039	0,261
		20	0,169	0,189	185	0,144	0,033	0,263
		30	0,184	0,152	85	0,160	0,079	0,267
		50	0,188	0,112	24	0,173	0,108	0,262

3.11. táblázat. 1000 szimulációs változó időtartamú maximum likelihood becslésem

módszeres becslésünk esetében. Viszont most is megfigyelhető, hogy a $\mu = 1$ és $\alpha = 1$ paraméterek esetét kivéve jobban közelítenek a kvartilisek, mint a 2.16. táblázat eredményei, amiben a fix paraméteres maximum likelihood módszeres becsléseink találhatóak, és a standard hiba is kisebb.

Pontosan azokat az eredményeket kaptunk, mint amiket vártunk nagy minta elemszám esetén. A 3.4. táblázat $n = 1000$ paraméter esetén rosszabb értékeket mutat, mint a 3.12. táblázatban található maximum likelihood módszeres becsléseink. Nagyon n paraméter esetén már szinte megegyeznek a két táblázatban az átlag és a medián értékei, viszont szinte minden esetben kevesebb standard hibát kaptunk a maximum likelihood módszeres becslésünk során.

3.5.3. Kiterjesztett maximum kvázi-likelihood módszer

Ebben az esetben is kicsit rosszabb becslést vártunk. Meglepő módon a 3.13. táblázatunk azt mutatja, hogy a változó időtartamú maximum kvázi-likelihood módszeres becsléseink $\mu = 1$ esetben jobbat adnak a kvartilisek és a standard hiba esetében, mindemellett a negatív számok

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1000	0,994	0,114	0	0,988	0,916	1,066
		10000	1,000	0,036	0	0,999	0,977	1,024
		50000	0,999	0,016	0	0,999	0,988	1,010
1	0,333	1000	0,332	0,065	0	0,329	0,288	0,375
		10000	0,334	0,021	0	0,334	0,320	0,347
		50000	0,333	0,009	0	0,333	0,328	0,340
1	0,2	1000	0,200	0,053	0	0,199	0,161	0,235
		10000	0,201	0,017	0	0,200	0,189	0,212
		50000	0,200	0,007	0	0,200	0,195	0,205
1	0,1	1000	0,098	0,045	11	0,099	0,066	0,129
		10000	0,100	0,015	0	0,099	0,089	0,109
		50000	0,100	0,007	0	0,100	0,095	0,104
1	0,05	1000	0,050	0,043	117	0,049	0,022	0,079
		10000	0,050	0,014	0	0,050	0,041	0,059
		50000	0,050	0,006	0	0,050	0,046	0,054
3	0,2	1000	0,199	0,025	0	0,198	0,183	0,216
		10000	0,200	0,008	0	0,200	0,195	0,206
		50000	0,200	0,004	0	0,200	0,198	0,202

3.12. táblázat. 1000 szimulációs változó időtartamú maximum likelihood becslésem kiegészítése

is kevesebben vannak. Az átlag is már $\alpha = 0{,}2$ esetben már jobbat ad, mint a 2.17. táblázatban található maximum kvázi-likelihood módszeres becsléseink. De a μ paraméter növelésével már rosszabb becsléseket kapunk az összes megfigyelés esetén, viszont nagyon sokszor tapasztalhatjuk, hogy a megfigyeléseink nagyon jól közelítik a fix paraméteres maximum kvázi-likelihood módszeres becsléseinkre kapott értékeket.

Ha megnézzük a változó időtartamú maximum likelihood (3.9. táblázat) és kiterjesztett maximum kvázi-likelihood (3.13. táblázat) módszeres becsléseinket, akkor azt tapasztalhatjuk, hogy az alsó és a felső kvartilisek esetében viszonylag kevés eltérés mutatkozik. A μ paraméter növelésével az átlag is jobbnak mondható a standard hibával együtt a kiterjesztett maximum kvázi-likelihood módszeres becsléseink esetén.

A korábbi tapasztalataink fényében picit meglepő módon, de a 3.14. táblázatunkban már kis α paraméter esetén egyre többször tapasztalhatjuk, hogy legalább olyan jó vagy még jobb becslést kaptunk, mint a 2.19. táblázatban lévő maximum kvázi-likelihood módszeres becsléseink.

A maximum likelihood módszeres becsléseink (3.10. táblázat) $\mu \geq 5$ paraméter esetén szinte megegyeznek az átlagban és $\mu \geq 10$ paraméter esetén pedig a standard hiba is megegyezik a kiterjesztett maximum kvázi-likelihood módszeres 3.14. táblázatunk eredményeivel.

A 3.15. és a 2.20. táblázatok azt mutatják, hogy a változó időtartamú becsléseinknél kevesebb negatív szám jelenik meg. Ezzel összefüggésben az alsó kvartilisek is közelebb vannak a becsült

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1,016	0,561	70	0,943	0,625	1,315
	0,333	0,317	0,305	1368	0,274	0,098	0,493
	0,2	0,194	0,259	2419	0,154	0,005	0,334
3	1	1,009	0,336	0	0,972	0,774	1,199
	0,333	0,324	0,151	43	0,311	0,217	0,415
	0,2	0,189	0,114	278	0,179	0,109	0,258
5	1	0,999	0,281	0	0,971	0,801	1,168
	0,333	0,327	0,121	0	0,315	0,241	0,400
	0,2	0,193	0,087	23	0,185	0,132	0,246
10	1	1,002	0,243	0	0,983	0,832	1,151
	0,333	0,326	0,094	0	0,320	0,260	0,384
	0,2	0,195	0,065	0	0,189	0,150	0,234
15	1	1,002	0,229	0	0,982	0,841	1,141
	0,333	0,328	0,087	0	0,322	0,267	0,382
	0,2	0,195	0,057	0	0,190	0,156	0,230
20	1	0,998	0,215	0	0,982	0,846	1,131
	0,333	0,326	0,081	0	0,321	0,270	0,377
	0,2	0,196	0,053	0	0,192	0,158	0,230

3.13. táblázat. 10000 szimulációs változó időtartamú kiterjesztett maximum kvázi-likelihood becslésem n=50 minta elemszám esetén

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	0.1	0,094	0,209	3902	0,057	-0,073	0,213
	0.05	0,049	0,188	4817	0,009	-0,100	0,147
3	0.1	0,094	0,087	1328	0,085	0,033	0,147
	0.05	0,042	0,072	3115	0,034	-0,011	0,084
5	0.1	0,094	0,061	441	0,088	0,050	0,131
	0.05	0,045	0,048	1798	0,041	0,011	0,074
10	0.1	0,095	0,042	27	0,091	0,066	0,121
	0.05	0,046	0,030	462	0,044	0,025	0,065
15	0.1	0,097	0,036	1	0,094	0,072	0,118
	0.05	0,048	0,024	86	0,046	0,031	0,062
20	0.1	0,097	0,032	0	0,095	0,074	0,118
	0.05	0,048	0,021	20	0,046	0,033	0,061

3.14. táblázat. 10000 szimulációs változó időtartamú kiterjesztett maximum kvázi-likelihood becslésem kiegészítése n=50 minta elemszám esetén

paraméterhez, bár nem mindig, hiszen $\alpha = 1$ és $\mu = 1$ esetben és a $\mu = 3$ esetben rosszabb értékeket kapunk. Az átlag jobbnak mondható, de főleg $n = 10$ esetén kapunk más eredményt,

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	1,119	2,122	256	0,559	-0,021	1,488
		20	0,995	0,995	103	0,797	0,320	1,401
		30	1,015	0,744	31	0,881	0,514	1,357
		50	1,007	0,539	8	0,942	0,627	1,308
1	0,333	10	0,350	1,283	469	0,039	-0,167	0,471
		20	0,329	0,551	328	0,192	-0,073	0,551
		30	0,336	0,434	216	0,254	0,039	0,507
		50	0,330	0,322	123	0,273	0,097	0,491
1	0,2	10	0,241	1,194	534	-0,039	-0,167	0,407
		20	0,193	0,461	415	0,071	-0,125	0,349
		30	0,181	0,333	349	0,103	-0,075	0,337
		50	0,191	0,250	234	0,151	0,011	0,333
1	0,1	10	0,118	0,589	614	-0,115	-0,167	0,234
		20	0,100	0,349	508	-0,008	-0,157	0,233
		30	0,095	0,302	460	0,028	-0,125	0,222
		50	0,105	0,220	386	0,066	-0,073	0,226
1	0,05	10	0,088	0,635	636	-0,125	-0,167	0,163
		20	0,046	0,351	610	-0,084	-0,167	0,149
		30	0,057	0,255	519	-0,010	-0,125	0,173
		50	0,047	0,184	476	0,009	-0,101	0,147
3	0,2	10	0,148	0,247	333	0,086	-0,041	0,255
		20	0,179	0,193	167	0,144	0,038	0,289
		30	0,184	0,149	80	0,168	0,085	0,262
		50	0,183	0,117	34	0,170	0,101	0,249

3.15. táblázat. 1000 szimulációs változó időtartamú kiterjesztett maximum kvázi-likelihood becslésem

ami a kevés minta elemszámból adódhatott. Érdekes módon a medián és a standard hiba $\alpha > 0{,}2$ esetén már jobbá válik, mint a fix paraméteres maximum kvázi-likelihood módszeres becslésem esetében.

Számítottunk arra, hogy a maximum kvázi-likelihood módszer az α paraméter csökkentésével jobb átlagban szinte paraméter értékére, mint amit a 3.11. maximum likelihood módszeres táblázatunkban lévő értékek. A kvázi-likelihood módszer esetében többször figyelhető meg kisebb standard hiba. A medián is az $\alpha = 0{,}2$ paramétert kivéve jobb közelítést adott, viszont a kvartilisek most sem adtak jobb közelítést, mint a maximum likelihood módszer esetében.

A 3.16. táblázat esetében szinte minden esetben jobb becslést kapunk, mint a 2.22. táblázat esetében, amelyekben a maximum kvázi-likelihood módszernek a változó időtartamú illetve fix paraméterű eloszlásokhoz tartozó becsléseik találhatóak. De a nagy elemszám miatt túl nagy eltérés nem tapasztalható egyik statisztikai mutató esetében sem.

$n = 1000$ minta elemszám esetén rosszabb becsléseket kaptunk az átlag értékeire a maximum

μ	α	n	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1000	1,030	0,121	0	1,026	0,951	1,102
		10000	1,028	0,036	0	1,027	1,003	1,052
		50000	1,029	0,017	0	1,028	1,017	1,040
1	0,333	1000	0,338	0,064	0	0,337	0,292	0,379
		10000	0,338	0,021	0	0,338	0,325	0,352
		50000	0,339	0,009	0	0,339	0,333	0,345
1	0,2	1000	0,206	0,055	0	0,206	0,167	0,241
		10000	0,202	0,018	0	0,202	0,189	0,214
		50000	0,202	0,008	0	0,202	0,197	0,207
1	0,1	1000	0,101	0,049	14	0,099	0,065	0,133
		10000	0,101	0,014	0	0,101	0,091	0,111
		50000	0,101	0,007	0	0,101	0,096	0,105
1	0,05	1000	0,048	0,043	131	0,045	0,018	0,074
		10000	0,050	0,013	0	0,050	0,041	0,059
		50000	0,050	0,006	0	0,051	0,046	0,054
3	0,2	1000	0,200	0,024	0	0,200	0,183	0,216
		10000	0,201	0,008	0	0,201	0,196	0,206
		50000	0,201	0,003	0	0,201	0,199	0,203

3.16. táblázat. 1000 szimulációs változó időtartamú kiterjesztett maximum kvázi-likelihood becslésem kiegészítése

kvázi-likelihood módszer esetében (3.16. táblázat), mint a maximum likelihood módszer (2.16) esetében, viszont kicsit meglepő, hogy nagyobb minta elemszám esetén már alig különböznek a statisztikai mutatók egymástól. Nagyobb α paraméter esetén kisebb standard hibája van a maximum kvázi-likelihood módszeres becsléseinknek. Mihelyt csökken az α paraméter értéke, úgy romlik a standard hiba is.

Érdekes, hogy a fix paraméteres esettel ellentétben most nagy mintákra is jól használható a kiterjesztett maximum kvázi-likelihood módszer.

3.5.4. Duplán kiterjesztett maximum kvázi-likelihood módszer

A $\mu = 1$ esetet kivéve picit rosszabb eredményeket kaptuk a 3.17. táblázatban a változó időtartamú duplán kiterjesztett maximum kvázi-likelihood módszeres becslésünknél az alsó és felső kvartilisekre és a standard hibára, mint a fix paraméteres duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseinknél (2.23. táblázat), viszont az átlag és a medián elég hasonló értékeket tükröz. Viszont, hogy a μ paramétert növelve a kezdetben a standard hiba ismét javul csakúgy, mint a fix paraméteres esetben.

Ha megnézzük a változó időtartamú duplán kiterjesztett maximum kvázi-likelihood módszeres becslésünket és a kiterjesztett maximum likelihood módszeres (3.9. táblázat) becsléseinket, akkor

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	0,933	0,473	92	0,883	0,606	1,205
	0.333	0,308	0,291	1282	0,276	0,100	0,478
	0.2	0,179	0,255	2449	0,151	0,003	0,323
3	1	0,964	0,301	0	0,939	0,754	1,148
	0.333	0,321	0,149	34	0,307	0,217	0,409
	0.2	0,189	0,113	267	0,180	0,110	0,258
5	1	0,968	0,259	0	0,949	0,785	1,128
	0.333	0,323	0,116	2	0,314	0,241	0,394
	0.2	0,192	0,086	23	0,185	0,131	0,245
10	1	0,978	0,227	0	0,963	0,818	1,119
	0.333	0,326	0,094	0	0,318	0,259	0,385
	0.2	0,193	0,066	0	0,188	0,146	0,234
15	1	0,979	0,213	0	0,964	0,830	1,113
	0.333	0,327	0,084	0	0,320	0,268	0,380
	0.2	0,195	0,057	0	0,191	0,155	0,230
20	1	0,976	0,202	0	0,963	0,835	1,103
	0.333	0,327	0,081	0	0,322	0,269	0,378
	0.2	0,195	0,053	0	0,191	0,158	0,230

3.17. táblázat. 10000 szimulációs változó időtartamú duplán kiterjesztett maximum kvázi-likelihood becslésem n=50 minta elemszám esetén

μ	α	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	0.1	0,084	0,223	3860	0,058	-0,075	0,211
	0.05	0,033	0,205	4865	0,006	-0,111	0,147
3	0.1	0,091	0,087	1364	0,082	0,031	0,143
	0.05	0,042	0,074	3028	0,035	-0,010	0,087
5	0.1	0,094	0,061	442	0,089	0,051	0,133
	0.05	0,045	0,049	1742	0,040	0,011	0,075
10	0.1	0,096	0,042	25	0,093	0,067	0,122
	0.05	0,047	0,030	434	0,044	0,025	0,065
15	0.1	0,096	0,035	1	0,093	0,071	0,117
	0.05	0,047	0,024	92	0,045	0,030	0,062
20	0.1	0,097	0,032	0	0,094	0,074	0,118
	0.05	0,048	0,021	24	0,046	0,033	0,061

3.18. táblázat. 10000 szimulációs változó időtartamú duplán kiterjesztett maximum kvázi-likelihood becslésem kiegészítése n=50 minta elemszám esetén

azt vehetjük észre, hogy míg az átlagban rosszabb becsléseink vannak, egyre kisebb α paramétert véve az átlag egyre jobban közelít az α paraméterhez. Ugyanez elmondható a medián esetében.

Ezzel szemben szinte mindenhol a standard hiba kisebb a duplán kiterjesztett maximum kvázi-likelihood módszeres becslés esetben.

A 3.18. táblázatban jól látható, hogy kisebb α paramétert véve a változó időtartamú duplán kiterjesztett maximum kvázi-likelihood módszeres becslésünk esete jobbnak mondható $\mu \leq 5$ paraméter értékénél, mint a fix paraméteres duplán kiterjesztett maximum kvázi-likelihood módszeres becslésünk (2.24. táblázat), mivel jobb eredményt kapunk a standard hibára, a kvartilisekre és majdnem olyan jókat az átlag s a medián esetében. $\mu \geq 10$ esetén pedig rosszabb eredményeket kapunk szinte az összes statisztikai mutató esetében.

Kicsit meglepő, hogy a különbséget nem tapasztalunk kis α és növekvő μ esetében a duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseink és a kiterjesztett maximum kvázi-likelihood módszeres becsléseink (3.14. táblázat) között egyik statisztikai mutató esetében sem.

μ	α	n	*Mean*($\hat{\alpha}$)	*s.e*	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	10	0,816	1,261	272	0,510	-0,058	1,295
		20	0,874	0,830	111	0,731	0,309	1,296
		30	0,900	0,646	42	0,821	0,433	1,277
		50	0,940	0,463	3	0,893	0,620	1,204
1	0,333	10	0,292	0,854	455	0,084	-0,261	0,567
		20	0,285	0,488	298	0,192	-0,060	0,530
		30	0,287	0,397	236	0,237	0,016	0,494
		50	0,321	0,301	123	0,279	0,109	0,486
1	0,2	10	0,131	0,627	525	-0,043	-0,294	0,340
		20	0,155	0,430	417	0,070	-0,138	0,360
		30	0,165	0,322	338	0,124	-0,073	0,342
		50	0,182	0,254	238	0,150	0,006	0,326
1	0,1	10	0,032	0,552	606	-0,121	-0,303	0,225
		20	0,056	0,360	499	0,001	-0,205	0,216
		30	0,072	0,284	461	0,029	-0,131	0,225
		50	0,091	0,218	373	0,064	-0,070	0,223
1	0,05	10	-0,024	0,528	667	-0,180	-0,308	0,126
		20	0,015	0,319	566	-0,047	-0,218	0,161
		30	0,026	0,270	530	-0,016	-0,178	0,161
		50	0,038	0,202	484	0,013	-0,111	0,160
3	0,2	10	0,153	0,288	326	0,092	-0,038	0,280
		20	0,182	0,195	156	0,151	0,043	0,279
		30	0,185	0,160	93	0,161	0,067	0,283
		50	0,196	0,120	29	0,181	0,112	0,269

3.19. táblázat. 1000 szimulációs változó időtartamú duplán kiterjesztett maximum kvázi-likelihood becslésem

Mivel a kiterjesztett maximum kvázi-likelihood módszeres becsléseink jobb becslést adtak, mint a maximum likelihood módszeres becsléseink, és a duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseink ugyanolyan jónak mondható, mint a kiterjesztett maximum kvázi-likelihood módszeres becsléseink, ezért emiatt tudjuk, hogy jobb eredményeket kapunk.

$\mu = 1$ és $\alpha < 1$ paraméterek esetében megállapítható, hogy szinte az összes átlag és kvartilisek becslése, valamint a standard hiba értéke jobb a duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseinknél (3.19. táblázat), mint a fix paraméteres duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseink (2.23. táblázat). $\mu = 3$ esetben rosszabbá válik a standard hiba. $\alpha = 1$ paraméter esetén a két módszer becsléseiben a standard hiba szinte megegyezik, viszont a fix paraméteres táblázatunk jobb átlagot produkál.

Nem kapunk jobb becslést a duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseink esetében, mint a kiterjesztett maximum kvázi-likelihood módszeres becsléseink (3.14. táblázat) esetében sem az átlag, sem a medián és sem a kvartilisek esetében, cserébe ekkor jobb a standard hiba.

μ	α	n	$Mean(\hat{\alpha})$	$s.e$	*No. of neg. values*	*Median*	*Lower quartile*	*Upper quartile*
1	1	1000	0,972	0,107	0	0,971	0,898	1,037
		10000	0,970	0,032	0	0,969	0,949	0,991
		50000	0,970	0,015	0	0,969	0,961	0,979
1	0,333	1000	0,333	0,062	0	0,328	0,289	0,377
		10000	0,332	0,019	0	0,331	0,319	0,345
		50000	0,332	0,009	0	0,332	0,326	0,338
1	0,2	1000	0,198	0,055	0	0,198	0,160	0,233
		10000	0,200	0,016	0	0,200	0,189	0,212
		50000	0,200	0,007	0	0,200	0,195	0,204
1	0,1	1000	0,099	0,049	18	0,098	0,065	0,129
		10000	0,099	0,014	0	0,100	0,090	0,109
		50000	0,100	0,006	0	0,100	0,096	0,105
1	0,05	1000	0,048	0,042	118	0,049	0,019	0,076
		10000	0,049	0,013	0	0,049	0,041	0,058
		50000	0,050	0,006	0	0,050	0,046	0,054
3	0,2	1000	0,201	0,025	0	0,201	0,184	0,218
		10000	0,200	0,008	0	0,200	0,194	0,205
		50000	0,200	0,004	0	0,200	0,197	0,202

3.20. táblázat. 1000 szimulációs változó időtartamú duplán kiterjesztett maximum kvázi-likelihood becslésem kiegészítése

Nagy minta elemszám esetén is elmondható, hogy a változó időtartamú duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseink (3.20. táblázat) jobbak a standard hiba értékeiben és a felső kvartilis értékeiben, és $\alpha > 0{,}1$ paraméter esetében az átlag is jobbat ad, mint

a fix paraméteres duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseink (2.24). $\mu = 3$ esetében a standard hibák megegyezik a két módszer között, viszont rosszabb az átlag a változó időtartamú duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseinknél.

A duplán kiterjesztett maximum kvázi-likelihood módszeres becsléseink esetében kisebb a standard hiba, mint a kiterjesztett maximum kvázi-likelihood módszeres becsléseinknél (3.16. táblázat), és nagyobb α esetén is jobbat kapunk az átlag becslésére, de az α paraméter csökkentésével már az átlagok közt nem mutatkozik különbség.

4. fejezet

Összefoglalás

A dolgozatban megismerkedtünk a negatív binomiális eloszlás paramétereinek lehetséges becsléseivel, több momentum módszeres becsléssel, a maximum likelihood becsléssel és a kiterjesztett illetve duplán kiterjesztett maximum kvázi-likelihood becslésekkel. Utóbbi kettő alapjaival kicsit bővebben foglalkoztunk, mivel ezek a módszerek az egyetemi tananyagban nem szerepelnek.

A kapott becsléseinket az R statisztikai programmal szimuláltuk, és a szimulációk eredményeit táblázatokba rendeztük. Ezen táblázatokat egymással és a korábbi cikkek eredményeivel hasonlítottuk össze a 2.4. szakaszban. A momentum módszer esetében azt kaptuk, hogy a legjobb becslést kis minták esetén Clark és Perry [4] korrigált tapasztalati szórásnégyzetet használó eredménye adja, mivel a mi becsléseink kicsit torzítottak. Azt is tapasztaltuk, hogy a harmadik momentumból származó becslésünk kicsit rosszabb, mint a második momentumból származók. Ezt követően megállapítottuk, hogy az átlag tekintetében a legjobb becslést a maximum likelihood módszer adja, viszont a standard hibája jobb a kiterjesztett és a duplán kiterjesztett maximum kvázi-likelihood módszereknek, így sokszor (főleg kis minták esetén) érdemesebb ezeket használni.

Ezt követően áttértünk a változó idejű megfigyelések esetére. A dolgozatnak ez a része új vizsgálatnak tekinthető, mivel a szakirodalomban ez az eset egyáltalán nem volt kutatva. Erre is felírtuk a vizsgált módszerek becsléseit, majd a kapott becsléseket szimuláltuk. A momentum módszer kicsit rosszabb becslést adott, mint a fix paraméteres eset. Ebben az esetben azt is megvizsgáltuk, hogy mi a helyzet, ha az $\frac{X_i}{t_i}$ momentumaiból indulunk ki. Ebben az esetben nagyon rossz becsléseket kaptunk, mivel volt néhány nagyon kiugró érték. A maximum likelihood módszer már hasonló eredményeket produkált, mint fix paraméteres esetben. Meglepetést okozott viszont, hogy a kiterjesztett maximum kvázi-likelihood módszer ebben az esetben többször jobb eredményt adott, mint fix paraméterre, és immár nagy mintára is jól használható az ebből kapott becslés, így itt még inkább ezt a becslést érdemes alkalmazni, a maximum likelihood módszerrel szemben. A duplán kiterjesztett maximum kvázi-likelihood módszerre azt kaptuk,

hogy standard hibája kicsit jobb a kiterjesztett maximum kvázi-likelihood módszerénél, viszont picit jobban torzít. Így időben változó megfigyelések esetén, amennyiben a torzítatlanság a cél inkább a kiterjesztett maximum kvázi-likelihood módszert érdemes alkalmazni, míg ha kisebb standard hibát szeretnénk, akkor érdemesebb a duplán kiterjesztett maximum kvázi-likelihood módszert alkalmazni. Annak is örülhetünk, hogy ezen módszerek mutatói nem állnak távol a fix paraméteres változataik mutatóitól, sőt több esetben egyértelműen jobbak.

Köszönetnyilvánítás:

Köszönetemet szeretném kifejezni Arató Miklós Tanár Úrnak, aki idejét nem sajnálva segítette a szakdolgozatom létrejöttét a rendszeres konzultáció mellett a tanácsaival és a támogatásával.

Ezenkívül szeretném férjemnek megköszönni a LaTeX-ben nyújtott rengeteg segítségét és a töretlen támogatását.

Irodalomjegyzék

[1] F. J. Anscombe: Sampling theory of the negative binomial and logarithmic series distributions *Biometrika* **36**, 358-382, 1950

[2] Arató M.: *Általános biztosításmatematika*, ELTE, Eötvös kiadó, 1997

[3] Bolla M. és Krámli A.: *Statisztikai következtetések elmélete*, Typotex Kiadó, 2005

[4] S. J. Clark and J. N. Perry: Estimation of the Negative Binomial Parameter κ by Maximum Quasi -Likelihood, *Biometrics* **45**, 309–316, 1989

[5] P. McCullagh and J. A. Nelder: *Generalized Linear Models*, London: Chapman and Hall, 1983

[6] J. F. Lawless: Negative binomial and mixed Poisson regression *Canadian Journal of Statistics* **15**, 209–225, 1987

[7] Y. Lee and J.A. Nelder: Hierarchical generalised linear models: a synthesis of generalised linear models, random effect models and structured dispersions. *Biometrika*, **88**, 987 – 1006, 2001

[8] W. W. Piegorsch: Maximum Likelihood Estimation for the Negative Binomial Dispersion Parameter, *Biometrics* **46**, 863–867, 1990

[9] K. Saha and S. Paul: Bias-Corrected Maximum Likelihood Estimator of the Negative Binomial Dispersion Parameter, *Biometrics* **61**, 179–185, 2005