

## A modern bayesi elemzések eszköztára és alkalmazása\*

---

### **Kehl Dániel**

PhD, a Pécsi  
Tudományegyetem adjunktusa  
E-mail: kehd@ktk.pte.hu

### **Várpalotai Viktor**

PhD, a Nemzetgazdasági  
Minisztérium főosztályvezető-  
helyettese, a Pécsi Tudomány-  
egyetem tudományos  
főmunkatársa  
E-mail:  
viktor.varpalotai@ngm.gov.hu

A tanulmány a modern bayesi elemzések eszköztárának egyik leggyakrabban alkalmazott elemét, a Gibbs-mintavételt mutatja be. Az elméleti alapok, majd az algoritmus bemutatása után rövid számszerű példa illusztrálja az eljárást, valamint a konvergencia gyorsaságát. A gyakorlati felhasználást egy klasszikus, jól ismert feladat bayesi eszköztár segítségével történő megoldása mutatja be, amely egyben lehetőséget teremt a klasszikus szemlélettel közös és attól eltérő eredmények, illetve előnyök és hátrányok bemutatására is.

TÁRGYSZÓ:  
Bayes.  
Ökonometria.  
MCMC.

\* A tanulmány a TÁMOP 4.2.2.C-11/1/KONV-2012-0005 sz. („Jól-lét az információs társadalomban” című) pályázat támogatásával készült. Írásunk korábbi változataihoz fűzött számos értékes észrevételéért köszönetet mondunk *Hunyadi László* professzornak és a *Statisztikai Szemle* lektorának.

A bayesi statisztika és ökonometria az utóbbi évtizedekben széles körben elterjedt, egyre intenzívebben alkalmazott,<sup>1</sup> hatékony elemzési eszköztárrá gyarapodott. A bayesi statisztika és ökonometria növekvő alkalmazásának több oka van. Az elméleti egyszerűség mellett egyfelől számos kiváló kézikönyv<sup>2</sup> jelent meg az utóbbi évtizedben, amelyek a bayesi elemzések gyakorlati alkalmazásához szükséges ismereteket gyűjtik össze, másfelől a számítógépes véletlenszám-generálás területén bekövetkezett módszertani fejlődés ledöntötte a bayesi eszköztár gyakorlati alkalmazhatóságának korlátait.

A bayesi elemzések egyre növekvő száma, valamint az ezzel járó ismeretterjesztő hatás azzal is párosult, hogy a hagyományos (más néven frekventista vagy klasszikus) és bayesi statisztika szembenállása napjainkra enyhült. A legtöbb kutató véleménye, hogy vannak problémák, melyek esetén a klasszikus vagy a bayesi megközelítés célravezetőbb, így optimális az lenne, ha a statisztikával foglalkozó kutatók mindkét megközelítéssel tisztában lennének (*Casella [2007]*, *Hunyadi [2011]*). Sajnos egyelőre kevés olyan szakember van, aki mindkét módszertanban elmélyült ismeretekkel rendelkezik, cikkünk ezen is próbál változtatni, a bayesi szemlélet népszerűsítésével.

A bayesi elemzések egy rendkívül egyszerű összefüggésen, a Bayes-tételre alapulnak, a megközelítés filozófiájáról, alapfogalmairól a témakörben nem elmélyült Olvasó *Hunyadi [2011]* művében és az irodalomjegyzék tételeiben talál további értékes információt. A bayesi módszertanban a prior és a likelihood segítségével a (kvázi) poszterior előállítás nem okoz problémát, hiszen ehhez csupán egy szorzást kell elvégezni. Az igazi problémát a poszteriorban (sokváltozós együttes sűrűségfüggvényben) rejlő (marginális) információ kinyerése jelenti. Ezzel el is érkeztünk a módszer egyik fő sajátosságához, annak számításigényességéhez, hisz az ehhez szükséges integrálás gyakran analitikusan nem, csak numerikusan végezhető el. A felhasznált numerikus módszerek ugyan ismertek voltak, legalábbis alapjaikban már korábban is (*Metropolis et al. [1953]*), nagy lendületet azonban a számítógépek, valamint az egyszerűbb programozási nyelvek elterjedése adott a területnek a 80-as évek második felében és a 90-es évek elején.<sup>3</sup> Jelen cikkünk első részében

<sup>1</sup> A bayesi statisztika növekvő népszerűségét a bayesi tanulmányok számának növekedése is jól mutatja a nemzetközi szakirodalomban (*Várpalotai [2008]*).

<sup>2</sup> A kézikönyvek közül kiemelünk néhányat, melyeket jelen tanulmány elkészítése és egyéb bayesi kötődésű munkáink során felhasználtunk (*Albert [2009]*, *Congdon [2005]*, *Gelman et al. [2004]*, *Geweke [2005]*, *Koop [2003]*, *Koop et al. [2007]*, *Robert-Casella [2004]*).

<sup>3</sup> Ma a leggyakrabban alkalmazott, bayesi statisztikát (is) támogató szoftverek a MATLAB, az R, valamint a BUGS (Bayesian inference Using Gibbs Sampling) különböző verziói, mint a WinBUGS és az OpenBUGS.

az ún. Markov-lánc Monte-Carlo (Markov Chain Monte Carlo – MCMC) forradalom egyik „zászlóshajóját”, a Gibbs-mintavételt, valamint konvergenciájának intuícióját mutatjuk be röviden, majd a második részben egy olyan példát, amelynél a klasszikus és bayesi ökonometriai elemzés jellegét tekintve hasonlít egymásra, ugyanakkor míg a klasszikus elemzéssel a paraméterbecslés egzakt bizonytalansága csak körülményesen határozható meg, addig a bayesi elemzésben ez egyszerűen kiszámítható.

## 1. Markov-lánc Monte-Carlo-módszerek

Az MCMC-módszerek napjainkra a legfontosabb, leggyakrabban használt algoritmusok közé kerültek (Hunyadi [2011]), sikertörténetükről, fejlődésükről rövid összefoglalót ad Casella és Berger [2011],<sup>4</sup> illetve megemlítjük Metropolis et al. [1953], Hastings [1970], Geman és Geman [1984], valamint Gelfand és Smith [1990] alapvető jelentőségű tanulmányait.

A MCMC-módszerek célja, hogy mintát tudjunk venni egy (jellemzően összetett, többdimenziós) sűrűségfüggvénnyel adott, akár ismeretlen valószínűség-eloszlásból. Szakítva a független azonos eloszlású véletlen értékeket generáló algoritmusokkal, az MCMC-technikák közös jellemzője, hogy olyan Markov-lánc(ka)t állítanak fel, melyek egyensúlyi eloszlása megegyezik a kívánt eloszlással. Ezután minden lépés utáni állapotot a céleloszlásból származó mintaelemnek tekintünk, amik azonban a Markov-tulajdonság miatt nem lesznek függetlenek. A Markov-lánc konstruálása jellemzően nem okoz különösebb nehézséget, a gyakorlati alkalmazások esetén a probléma inkább a konvergencia megállapításában rejlik.

A Gibbs-algoritmus lehetővé teszi a bonyolult, sokdimenziós problémák lebontását kisebb, egyszerűbb feladatokra, Markov-láncok felhasználásával. A megoldandó probléma egy együttes eloszlás (a poszterior) marginális eloszlásainak (az egyes paraméterek), jellemzőinek meghatározása. A legkézenfekvőbb eljárás az együttes eloszlás integrálása lenne, ez azonban sok esetben analitikusan nem oldható meg. Szintén lehetséges numerikus integrálási módszereket alkalmazni, magas dimenziószám-ban, ez azonban nehézkes és lassú. Ilyen esetekben nyújthat segítséget a Gibbs mintavételi technika, ami lehetőséget ad a kívánt együttes eloszlásból való mintavételre, méghozzá indirekt módon, a feltételes eloszlások segítségével. A módszert leggyakrabban a bayesi megközelítés használja, de összetett likelihoodokkal kapcsolatos számítások esetén a klasszikus statisztikában is alkalmazható.

<sup>4</sup> Magyar nyelvű ismertetését lásd a *Statisztikai Szemle* hasábjain (Kehl [2012]).

### 1.1. Markov-láncok néhány fontos tulajdonsága, jelölések

Elsőként röviden tekintjük át a Markov-láncok azon jellemzőit, melyek az MCMC-módszerek szempontjából jelentőséggel bírnak. A véges ( $k$ ) állapotterű diszkrét Markov-lánc egy speciális sztochasztikus folyamat  $t \geq 0$  indexszel, amely  $X_0$  indulási érték, kezdeti állapot után  $X_1, X_2, \dots, X_t, \dots$  állapotokba kerül, ha

$$P(X_{t+1} = a | \mathbf{X} = \mathbf{b}) = P(X_{t+1} = a | X_t = b_t)$$

minden  $(a, \mathbf{b})$  párra és  $t \geq 0$ -ra, ahol  $\mathbf{X} = (X_0 \ X_1 \ \dots \ X_t)$  a megelőző és a jelen állapotok vektora és  $\mathbf{b} = (b_0 \ b_1 \ \dots \ b_t)$ . Mindez azt jelenti, hogy a következő állapot alakulása csak a jelenlegi állapoton múlik, a múltbeli állapotok nem befolyásolják azt. Ha ezek a feltételes valószínűségek időben állandók, azokat gyakran jelöljük a  $p_{ba} = P(X_{t+1} = a | X_t = b)$  módon, melyeket kézenfekvő egy ún. átmenetmátrixba rendezni:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \vdots & \vdots & & \vdots \\ p_{k1} & p_{k2} & \dots & p_{kk} \end{bmatrix}.$$

Az átmenetmátrix definíciójából következik néhány tulajdonsága: négyzetes, minden eleme nemnegatív, illetve minden sora egy feltételes eloszlás, azaz sorösszegei egyet adnak,<sup>5</sup> a  $b$ . sor  $a$ . oszlopa a  $b$  állapotból  $a$  állapotba kerülés valószínűségét mutatja meg.

Az átmenetmátrix segítségével meghatározható az egylépéses átmenetek mellett a többlépéses átmenetek valószínűsége is, méghozzá  $m$  lépés esetén  $\mathbf{P}^m$  módon, amely mátrix sorai  $X_{t+m}$  feltételes eloszlásait adják meg adott  $X_t$  állapotok mellett.

Az átmenetek valószínűségei mellett szólnunk kell a kezdeti állapotok valószínűségét leíró vektorról ( $\mathbf{v}$ ), kezdeti eloszlásról is, hiszen a Markov-lánc együttes eloszlásának meghatározásához az átmenetmátrixon kívül erre is szükségünk van. Az együttes eloszlásból pedig meghatározható a  $t$ . időpont marginális eloszlása, méghozzá  $\mathbf{v}\mathbf{P}^t$  módon. Markov-láncok esetén létezik olyan speciális kezdeti eloszlás, mely teljesíti a  $\mathbf{v}\mathbf{P} = \mathbf{v}$  egyenlőséget, az ilyen vektorokat a Markov-lánc egyensúlyi eloszlásának hívjuk.

<sup>5</sup> Az ilyen mátrixokat sztochasztikus mátrixoknak nevezzük.

Bizonyos esetekben ez a speciális kezdeti eloszlás egyedi és fontos tulajdonságokkal rendelkezik (*DeGroot–Schervish* [2012]): ha létezik olyan  $m$ , melyre  $\mathbf{P}^m$  valamennyi eleme szigorúan pozitív,<sup>6</sup> akkor

- a Markov-lánc egyetlen  $\mathbf{v}$  egyensúlyi eloszlással rendelkezik,
- $\lim_{t \rightarrow \infty} \mathbf{P}^t$  egy olyan mátrix, melynek minden sora  $\mathbf{v}$ , és
- függetlenül attól, hogy a Markov-lánc milyen kezdeti eloszlásból indul,  $t$  lépés után az eloszlása  $\mathbf{v}$ -hez tart, ahogy  $t \rightarrow \infty$ .

A harmadik pont különösen fontos, hiszen azt mondja, hogy bárhonnan indítva a láncot, azt elegendően hosszú ideig futtatva a  $t$ . lépésben kapott érték tulajdonképp egy  $\mathbf{v}$ -ből származó véletlen változó. Mindezek végtelen állapotterrel rendelkező Markov-lánccokra is igazak, a Gibbs-mintavétel pedig tulajdonképp ezt használja ki úgy, hogy olyan Markov-láncot állít fel, melynek egyensúlyi eloszlása épp a generálni kívánt eloszlás.

A módszert és a konvergencia intuícióját előbb egy könnyen átlátható diszkrét példán mutatjuk be, majd ezután az általános algoritmust adjuk meg.

## 1.2. Kétváltozós Gibbs-mintavétel

Tekintsünk elsőként egy egyszerű, kétváltozós esetet, az  $(X, Y)$  együttes eloszlást. A Gibbs-mintavétel  $X$  marginális eloszlásból úgy vesz mintát, hogy magát a marginális eloszlást nem, csupán az  $X|Y$  és az  $Y|X$  feltételes eloszlásokat használja, még hozzá a következő módon. Adott  $y_0$  kezdeti érték segítségével váltakozva generálunk az

$$\begin{aligned} X_i &\sim f_1(x|Y_i = y_i), \\ Y_{i+1} &\sim f_2(y|X_i = x_i) \end{aligned} \quad /1/$$

eloszlásokból, ahol  $f_1(\cdot)$  és  $f_2(\cdot)$  a megfelelő feltételes eloszlások sűrűségfüggvényei. Ebben a véletlen számokból álló „Gibbs-sorozatban” elég nagy  $k$  esetén  $X_k = x_k$  egy  $f(x)$ -ből származó mintaelemnek tekinthető.

A legegyszerűbb esetben  $X$  és  $Y$  is bináris valószínűségi változók a következő együttes eloszlással (a példa *Casella–George* [1992] tanulmánya alapján készült):

<sup>6</sup> Amennyiben a Markov-lánc irreducibilis (az állapotok egymásból kölcsönösen elérhetők, azaz kommunikálnak egymással), aperiodikus, véges állapotterű, akkor létezik ilyen  $m$ . Az ilyen Markov-lánccokat gyakran ergodikusanak nevezik.

$X/Y$	0	1
0	$p_{00}$	$p_{01}$
1	$p_{10}$	$p_{11}$

ahol a valószínűségek egyre összegződnek. Természetesen a marginális eloszlások ebben az egyszerű esetben triviálisan adódnak, azt kívánjuk illusztrálni, hogy csak a feltételes eloszlások segítségével is generálhatók olyan véletlen értékek, melyek eloszlása pontosan a kívánt peremeloszlás. A feltételes eloszlásokat sztochasztikus mátrixokkal írhatjuk le:

$$A_{y|x} = \begin{bmatrix} \frac{p_{00}}{p_{00} + p_{10}} & \frac{p_{10}}{p_{00} + p_{10}} \\ \frac{p_{01}}{p_{01} + p_{11}} & \frac{p_{11}}{p_{01} + p_{11}} \end{bmatrix}, \text{ illetve } A_{x|y} = \begin{bmatrix} \frac{p_{00}}{p_{00} + p_{01}} & \frac{p_{01}}{p_{00} + p_{01}} \\ \frac{p_{10}}{p_{10} + p_{11}} & \frac{p_{11}}{p_{10} + p_{11}} \end{bmatrix}. \quad /2/$$

A két mátrix egy-egy Markov-lánc átmenetmátrixa, melyek azt mutatják meg, hogy adott  $x$  állapotból milyen valószínűséggel jutunk adott  $y$  állapotba. Jellemzően azonban nem ezekre a valószínűségekre, hanem adott  $x$  állapotból egy újabb  $x$  állapotba kerülés valószínűségére vagyunk kíváncsiak vagy épp ugyanerre az  $y$ -ra vonatkozóan. Ezek a lépések nem közvetlenül, hanem a másik változón keresztül történnek meg, de könnyen meghatározhatjuk az egy lépéses átmenet-valószínűségeket, méghozzá

$$A_{x|x} = A_{y|x} A_{x|y}, \text{ illetve } A_{y|y} = A_{x|y} A_{y|x} \quad /3/$$

formában. A többlépéses átmenetmátrixok pedig a /3/-ban meghatározott mátrixok megfelelő hatványaiként állíthatók elő. A Markov-láncok tulajdonságainál említett tétel szerint pedig ahogy  $k \rightarrow \infty$  a  $k$ . állapot eloszlása épp a marginális eloszlás lesz. Könnyen beláthatóan a marginális eloszlás kielégíti a  $\mathbf{vP} = \mathbf{v}$  ( $f_x A_{x|x} = f_x A_{y|x} A_{x|y} = f_x$ ) feltételt, azaz a Markov-lánc egyensúlyi eloszlása:

$$\begin{bmatrix} p_{00} + p_{10} & p_{01} + p_{11} \end{bmatrix} \begin{bmatrix} \frac{p_{00}}{p_{00} + p_{10}} & \frac{p_{10}}{p_{00} + p_{10}} \\ \frac{p_{01}}{p_{01} + p_{11}} & \frac{p_{11}}{p_{01} + p_{11}} \end{bmatrix} \begin{bmatrix} \frac{p_{00}}{p_{00} + p_{01}} & \frac{p_{01}}{p_{00} + p_{01}} \\ \frac{p_{10}}{p_{10} + p_{11}} & \frac{p_{11}}{p_{10} + p_{11}} \end{bmatrix} = \\ = \begin{bmatrix} p_{00} + p_{10} & p_{01} + p_{11} \end{bmatrix}.$$

A  $2 \times 2$ -es mátrixhoz hasonlóan írható fel az általánosabb, de továbbra is csupán két változót tartalmazó,  $n \times m$ -es eset, ahol egy rövid számpéldán keresztül a konvergenciát mutatjuk be. Legyen az együttes eloszlás például:

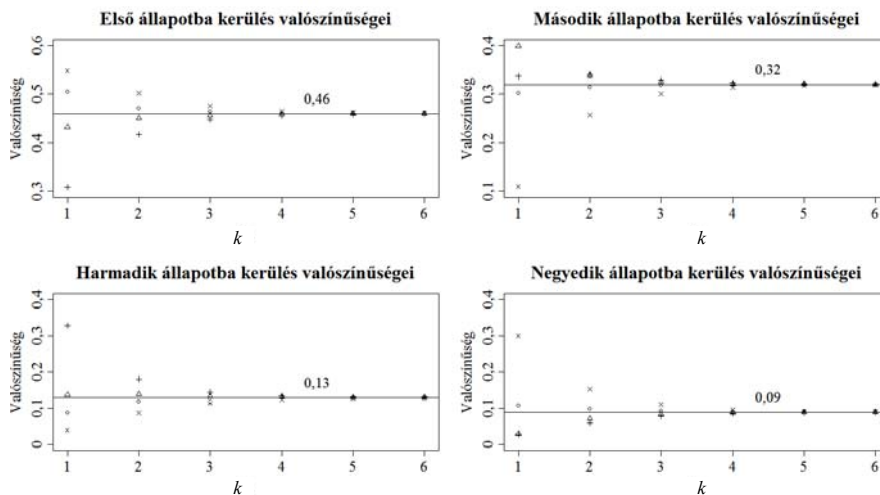
$$A_{x|x} = \begin{bmatrix} 0,10 & 0,15 & 0,05 & 0 \\ 0,20 & 0,10 & 0 & 0,02 \\ 0,10 & 0,01 & 0,01 & 0,07 \\ 0,01 & 0,01 & 0,04 & 0 \\ 0,05 & 0,05 & 0,03 & 0 \end{bmatrix}, \quad /4/$$

amiből a /2/ és /3/ képlettel analóg módon képezhető az  $A_{x|x}$  egylépéses átmenetmátrix:

$$A_{x|x} = \begin{bmatrix} 0,504 & 0,301 & 0,087 & 0,107 \\ 0,433 & 0,399 & 0,137 & 0,031 \\ 0,309 & 0,336 & 0,327 & 0,028 \\ 0,548 & 0,110 & 0,041 & 0,300 \end{bmatrix}.$$

Az átmenetmátrix hatványainak elemeit az 1. ábra mutatja be  $k$  függvényében, a vízszintes vonalak a marginális valószínűségeket reprezentálják, a különböző jelzések a különböző induló állapotokból adott állapotba jutás valószínűségeit mutatják.

1. ábra. Az egyes állapotokba kerülés valószínűsége  $k$  lépés után az egyes kezdeti állapotokból



Az ábra alapján az látható, hogy néhány lépés után, bárhol is indítjuk útjára a Markov-láncot, annak a valószínűsége, hogy egy adott állapotba jutunk, éppen az adott állapot marginális valószínűsége, illetve az, hogy ebben az egyszerű példában a konvergencia rendkívül gyors. Természetesen ez a példa csupán az intuíciót kívánja bemutatni. Bizonyítani nem kívánjuk, csupán megemlíjük, hogy „folytonos állapot-tér esetén a Markov-láncok matematikája sokkal összetettebb, de hasonlóan kell elképzelni a folyamatot egy végtelen átmenetmátrixszal” (Casella–George [1992]).

### 1.3. Többváltozós Gibbs-mintavétel

Tegyük fel, hogy  $\mathbf{X} = (X_1, X_2, \dots, X_p)$  véletlen vektorváltozó, ahol az  $X_j$ -k egy- vagy többdimenziós komponensek (blokkok), valamint azt is, hogy képesek vagyunk a következő  $f_1, f_2, \dots, f_p$  feltételes sűrűségfüggvényekkel adott eloszlásokból véletlen számo(ka)t generálni, ismerjük, azaz ismert eloszlásként azonosítani tudjuk a

$$f_j(x_j | x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_p), \quad /5/$$

ún. teljes feltételes (full conditional) eloszlásokat minden  $j = 1, 2, \dots, p$ -re.

Ekkor a Gibbs-algoritmus:<sup>7</sup>

1. Válasszunk  $\mathbf{X}^{(0)} = \mathbf{x}^{(0)}$  kezdőértékeket ( $m = 0$ ).

2. Ismételjük a következő lépéseket, amíg a lánc az egyensúlyi eloszlásához nem konvergál:

a)  $X_1^{(m+1)} \sim f_1(x_1 | x_2^{(m)}, x_3^{(m)}, \dots, x_p^{(m)})$ .

b)  $X_2^{(m+1)} \sim f_2(x_2 | x_1^{(m+1)}, x_3^{(m)}, \dots, x_p^{(m)})$ .

c)  $X_3^{(m+1)} \sim f_3(x_3 | x_1^{(m+1)}, x_2^{(m+1)}, \dots, x_p^{(m)})$ .

d) ...

e)  $X_p^{(m+1)} \sim f_p(x_p | x_1^{(m+1)}, x_2^{(m+1)}, \dots, x_{p-1}^{(m+1)})$ .

f) Növeljük  $m$  értékét.

3. A konvergencia előtti értékeket (burn in period) levágva a lánc elejéről megkapjuk a kívánt eloszlásból származó véletlen mintát.

<sup>7</sup> Érdemes megjegyezni, hogy az algoritmus nagyon hasonlít az egyenletrendszerek megoldására alkalmas Gauss–Siedel iterációs eljárásához.



Az átláthatóság érdekében  $m$  felsőindexként szerepel, azaz  $x_j^{(m)}$  a  $j$ . komponens  $m$ . lépésben (iterációban) felvett értékét jelöli. Az algoritmus jelentősége rögtön szembetűnő, amennyiben azt az  $f(\theta_1, \theta_2, \dots, \theta_p | y)$  poszterior sűrűségfüggvényre írjuk fel. Bár első látásra az összes feltételes eloszlás ismerete erős feltételezésnek tűnik, de megfelelő (konjugált) priorok választása esetén a gyakorlatban használt ökonometriai modellek túlnyomó részénél (például lineáris regresszió, vektor autoregresszív modellek, látens változós modellek) az összes feltételes eloszlás beazonosítható, azok könnyen generálhatók.

Miután a lánc konvergált (az ehhez szükséges gyenge feltételeket lásd *Gelfand–Smith* [1990]) az  $m^*$  lépésben, egy újabb lépés a kívánt együttes eloszlásból származó véletlen értéként tekinthető. A szükséges számú ( $M$ ) véletlen érték generálására több eljárás létezik. Az első lehetőség, hogy az algoritmust  $m^* + 1$ . lépésig futtatjuk  $M$  alkalommal, minden esetben megtartva az utolsó értéket. Sokkal gyakrabban alkalmazott technika, hogy a láncot hagyjuk futni  $m^* + M$  lépésig, majd a konvergenciáig szükséges iterációk eredményeit elhagyva kapjuk a szükséges számú mintaelemet. Az első eljárás hátránya, hogy lassú, hiszen az összes iteráció száma magas, a második pedig az, hogy a véletlen értékek autokorreláltak lesznek. Ennek kivédésére szokás a lánc csak minden  $r$ . értékét megtartani, ezzel csökkentve ezt a negatív hatást, ekkor  $m^* + rM$  iteráció szükséges. Ezen eljárás szakirodalmi elnevezése *thinning* vagy *ritkítés*.

A konvergencia megállapítása nem egyszerű feladat. *Geweke* [1992] egyetlen Markov-láncon alapuló, idősor-elemzési eszköztárra támaszkodó diagnosztikai módszert javasol. Alapötlete, hogy a sorozat elejének (például az értékek első 10 százalékának) és végének (például az értékek 50 százalékának) átlagait hasonlítja össze. *Gelman és Rubin* [1992] módszere több lánc különböző kezdő értékekről való indításával, majd láncokon belüli és láncok közötti varianciák összehasonlításával dolgozik. Szintén gyakori a láncok egymás utáni értékeinek ábrázolása (*trace*), kumulatív módon számolt átlagok állandóságának, valamint a generált véletlen értékek alapján becsült sűrűségfüggvények vizuális vizsgálata. *Cowles és Carlin* [1996] a '90-es évek nagy MCMC hullámának 13 diagnosztikai eszközét elemzi jellegük, az igényelt láncok száma, elméleti hátterük, alkalmazhatóságuk és összetettségük szerint, következtetésükben pedig arra jutnak, hogy minden eszköznek vannak hátrányai, így érdemes több diagnosztikai eljárást alkalmazni. Mivel ezek a diagnosztikai eszközök nem tévedhetetlenek, így soha nem lehetünk biztosak benne, hogy a lánc ténylegesen konvergált-e a kívánt eloszláshoz, a gyakran használt ökonometriai modellek esetén azonban ez nem szokott problémát okozni, főként, ha a konvergenciát vizsgáló eszközök nem jeleznek problémát.

## 2. Szívkoszorúér-megbetegedések miatti halálozás modellezése klasszikus és bayesi MCMC-eljáráson alapuló ökonometriai eszközökkel

Ebben a részben bemutatjuk, hogy az előző fejezetben ismertetett MCMC-módszer miként használható az empirikus elemzésekben. A következő példa elsősorban a Gibbs mintageneráló algoritmus illusztrálására szolgál, de emellett arra is rámutat, hogy adott esetben a klasszikus ökonometria elemzési eszköze igen hasonló a bayesi elemzéseknél használt Gibbs mintavételi eljáráshoz. Míg azonban a klasszikus ökonometriai elemzés elsődlegesen pontbecsléseket szolgáltat, addig a bayesi ökonometriai eszköztárral a paraméterbecslés bizonytalansága is közvetlenül meghatározható.<sup>8</sup>

Választott példánk *Ramanathan* [2003] könyvének 4-7 számmal jelölt adatállományát használja, mely a szívkoszorúér-megbetegedések miatti halálozási rátát és annak lehetséges magyarázóváltozóit tartalmazza az 1947–1980 időszakra (34 megfigyelés).<sup>9</sup> Azért választottuk ezt az adatállományt, mert a klasszikus és ökonometriai eszközök által szolgáltatott eredmények összevetésekor hivatkozni tudunk *Ramanathan* klasszikus ökonometriai számításaira. Az adatállomány változói (lásd *Ramanathan* [2003] 664. old.):

- CHD: 100 000 főre jutó szívkoszorúér-megbetegedés miatt elhunytak száma,
- CAL: egy főre jutó napi kalciumfogyasztás grammban,
- UNEMP: munkanélküliek a 16 éves és idősebb munkavállalók százalékában,
- CIG: egy főre jutó cigarettafogyasztás a 18 évesek és idősebbek körében (font),
- EDFAT: egy főre jutó étkezési zsír- és olajfogyasztás (font),
- MEAT: egy főre jutó húsfogyasztás (font),
- SPIRITS: egy főre jutó égetettszesz-fogyasztás a 18 évesek és idősebbek körében (gallon),
- BEER: egy főre jutó sörfogyasztás a 18 évesek és idősebbek körében (gallon),
- WINE: egy főre jutó borfogyasztás a 18 évesek és idősebbek körében (gallon).

<sup>8</sup> A bayesi elemzéshez felhasznált MATLAB-kódokat megkeresés esetén szívesen rendelkezésre bocsátjuk.

<sup>9</sup> Az adatok letölthetők a <http://econweb.ucsd.edu/~rramanathan/XLDATA/DATA4-7.XLS> címről.

## 2.1. Autokorrelált hibatagú lineáris regressziós modell elemzése klasszikus ökonometriai eszközökkel

A rendelkezésre álló összes változó felhasználásával felírt lineáris regressziós modellből a nem szignifikáns változók elhagyása után a következő, az információs kritériumok által is preferált modellváltozat adódott (*Ramanathan* [2003] 221. old.):

$$CHD_t = \beta_0 + \beta_1 \cdot CIG_t + \beta_2 \cdot EDFAT_t + \beta_3 \cdot SPIRITS_t + \beta_4 \cdot BEER_t + u_t, \quad /6/$$

ahol  $\beta_0$  a konstans,  $\beta_i$  az egyes magyarázóváltozókhoz tartozó együttható,  $u_t$  a modell hibatagja. A becslés eredményei a táblázatban találhatóak.

A lineáris regressziós modellben az eltérésváltozók a Lagrange multiplikátor próba alapján autokorreláltak bizonyultak (*Ramanathan* [2003] 408. old.). A hibatagok autokorreláltságának következménye (lásd például *Ramanathan* [2003] 404. old.), hogy a legkisebb négyzetek módszerével becsült együtthatók bár továbbra is torzítatlannak és konzisztensek, de nem lesznek hatásosak. Továbbá az együtthatók becsült varianciái torzítottak és inkonzisztensek lesznek, így a hipotézisvizsgálatok is érvényüket veszítik. Ezek a következmények egyrészt azt jelentik, hogy a /6/ regresszió alapján tett megállapítás a becsült együtthatók nullától szignifikánsan különböző voltára önmagában érvénytelen. Másrészt célszerű olyan becslési eljárást alkalmaznunk, mint a Cochrane–Orcutt-féle iteratív eljárás,<sup>10</sup> amely a hibatagok autokorreláltságának megfelelő figyelembe vételével a korábbi negatív következményeket kiküszöböli.

Így a példában szereplő lineáris regresszió a hibatagok autokorreláltságát is szem előtt tartva a következőképpen írható fel:

$$CHD_t = \beta_0 + \beta_1 \cdot CIG_t + \beta_2 \cdot EDFAT_t + \beta_3 \cdot SPIRITS_t + \beta_4 \cdot BEER_t + u_t, \quad /7/$$

$$u_t = \rho_1 u_{t-1} + \varepsilon_t, \quad /8/$$

ahol  $\varepsilon_t$  hibatagról már feltehető, hogy autokorrelálatlan.<sup>11</sup>

A /7/–/8/ modell becslése a Cochrane–Orcutt-féle iteratív eljárással a következő:

1. Becsüljük meg a /7/ modellt a legkisebb négyzetek módszerével.
2. A becsült  $\hat{\beta}_i$  együtthatók segítségével számítsuk ki az  $\hat{u}_t$  reziduumokat az

<sup>10</sup> A módszer leírását lásd például *Dufour et al.* [1980] vagy *Ramanathan* [2003] 469. old.

<sup>11</sup> A hibatag autokorreláltságának figyelembe vételéhez elegendő egy késleltetést szerepeltetni, mivel a tesztek szerint a becsült  $\varepsilon_t$  hibatag már autokorrelálatlan (*Ramanathan* [2003] 413. old.)

$\hat{u}_i = CHD_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot CIG_i - \hat{\beta}_2 \cdot EDFAT_i - \hat{\beta}_3 \cdot SPIRITS_i - \hat{\beta}_4 \cdot BEER_i$   
kifejezés felhasználásával.

3. Az  $\hat{u}_i$  reziduumokra legkisebb négyzetek módszerével illesszük a /8/ modellt.

4. Az előző lépésben becsült  $\hat{\rho}$  együttható segítségével definiáljuk  $CHD_i^* = CHD_i - \hat{\rho} \cdot CHD_{i-1}$  változót, illetve ezzel analóg módon a  $CIG_i^*$ ,  $EDFAT_i^*$ ,  $SPIRITS_i^*$  és  $BEER_i^*$  idősorokat. Ezt követően, a  $\hat{\rho}$  együtthatót adottnak véve, a legkisebb négyzetek módszerével becsüljük meg /7/ modellt, de az eredeti változók helyett mindenütt a csillagozott változókat használva:

$$CHD_i^* = \beta_0 \cdot (1 - \hat{\rho}) + \beta_1 \cdot CIG_i^* + \beta_2 \cdot EDFAT_i^* + \beta_3 \cdot SPIRITS_i^* + \beta_4 \cdot BEER_i^* + u_i.$$

5. Menjünk vissza a második lépéshez mindaddig, amíg a becsült együtthatók nem konvergálnak.

Az iteráció eredményeként kapott  $\hat{\beta}_i$  és  $\hat{\rho}$  együtthatók becslése továbbra is torzítatlan és konzisztens, illetve az együtthatók varianciái is konzisztensen meghatározhatók. A Cochrane–Orcutt-féle együtthatóbecslés eredményeit a táblázat második blokkja mutatja.<sup>12</sup> Mint látható, a reziduumok autokorreláltságának figyelembe vétele érdemben megváltoztatta a pontbecsléseket miközben a standard hibák lényegesen nem módosultak. Az eredmények alapján a korábban szignifikánsnak tűnő együtthatók egy kivétellel (BEER) inszignifikánsná váltak.

## 2.2. Autokorrelált hibatagú lineáris regressziós modell elemzése bayesi ökonometriai eszközökkel

A klasszikus ökonometria eredményeit követően a bayesi ökonometria MCMC családjába tartozó Gibbs mintavételi eljárással becsüljük meg a /7/–/8/ egyenletekkel adott modellspecifikáció együtthatóit.<sup>13</sup> A tanulmány első részében bemutatott Gibbs-mintavétel alkalmazásához a priorok, a modell likelihoodja és a poszterior felírását követően meg kell határoznunk az együtthatók feltételes poszterior eloszlásait.

Amennyiben – a klasszikus ökonometriai megközelítéshez hasonlóan – feltételezzük, hogy az  $\varepsilon_i$  hibatagok normális eloszlásúak, akkor az ismeretlen együtthatók konjugált prior eloszlásai a következők.<sup>14</sup>

<sup>12</sup> A közölt eredmények megegyeznek Ramanathan [2003] 413. oldalon leírt eredményével.

<sup>13</sup> Autoregresszív hibatagú lineáris modellek bayesi becsléséről lásd Chib [1993].

<sup>14</sup> Lásd például Koop [2003] 134. old.

$$\beta \sim N(B_0, V_0), \quad /9/$$

$$\rho \sim N(R_0, W_0), \quad /10/$$

$$\sigma^2 \sim IG(S_0, T_0), \quad /11/$$

ahol  $\beta = [\beta_0 \ \beta_1 \ \beta_2 \ \beta_3 \ \beta_4]'$ ,  $\sigma^2 = \text{Var}(u_i)$ . Továbbá  $N(R_0, W_0)$  (illetve  $N(B_0, V_0)$ ) jelöli az  $R_0$  ( $B_0$ ) várható értékű és  $W_0$  ( $V_0$ ) variancia-kovarianciamátrixú (többváltozós) normális,  $IG(S_0, T_0)$  pedig az  $S_0$  lokációs paraméterű és  $T_0$  szabadságfokú inverz gamma eloszlást.

A /7/-/8/ egyenletekkel adott modell likelihoodja, feltételezve, hogy  $\varepsilon_i$  független azonos normális eloszlású:

$$f(y, X | \beta, \rho, \sigma^2) = (\sigma\sqrt{2\pi})^{-T} \exp\left(-\frac{1}{2\sigma^2} (y^*(\rho) - X^*(\rho)\beta)' (y^*(\rho) - X^*(\rho)\beta)\right), \quad /12/$$

ahol  $T$  a megfigyelések száma,<sup>15</sup>  $y^*(\rho)$  a  $CHD_i^*(\rho) = CHD_i - \rho \cdot CHD_{i-1}$  megfigyelésekből képzett oszlopvektor.  $y^*(\rho)$ -vel analóg módon definiáljuk  $CIG^*(\rho)$ ,  $EDFAT^*(\rho)$ ,  $SPIRITS^*(\rho)$  és  $BEER^*(\rho)$  oszlopvektorokat a  $CIG_i$ ,  $EDFAT_i$ ,  $SPIRITS_i$  és  $BEER_i$  idősorokból. Ezek felhasználásával legyen  $X^*(\rho)$  a  $X_i^*(\rho) = [1 \ CIG_i^*(\rho) \ EDFAT_i^*(\rho) \ SPIRITS_i^*(\rho) \ BEER_i^*(\rho)]$  megfigyelésekből képzett  $T \times 4$ -es mátrix.

Bayes tételét használva a poszterior a /9/-/11/ priorok és a /12/ likelihood felhasználásával a következő:

$$f(\beta, \rho, \sigma^2 | y, X) \propto f(y, X | \beta, \rho, \sigma^2) f(\beta) f(\rho) f(\sigma^2). \quad /13/$$

A tanulmány függeléke alapján meghatározhatók a Gibbs mintavételhez szükséges feltételes poszterior eloszlások  $\beta | \rho, \sigma^2, y, X \sim N(B_1, V_1)$ ,  $\rho | \beta, \sigma^2, y, X \sim N(R_1, W_1)$ ,  $\sigma^2 | \beta, \rho, y, X \sim IG(S_1, T_1)$ , ahol  $B_1, V_1, R_1, W_1, S_1$  és  $T_1$  értékeit a Függelék /F8/, /F9/, /F12/, /F13/, /F15/ és /F16/ képletei határozzák meg.<sup>16</sup>

<sup>15</sup> Esetünkben a késleltetések szerepeltetése miatt szükséges mintakorrekció után  $T = 33$ .

<sup>16</sup> Vegyük észre, hogy a feltételes poszterior eloszlások azonos típusúak a megfelelő prior eloszlásokkal, azaz ténylegesen konjugált priorokkal dolgozunk.

A Gibbs-algoritmus alkalmazása során a kezdeti értékek megadása után a bemutatott feltételes eloszlásokból kell ismételtelen véletlen mintát generálnunk.<sup>17</sup> Az algoritmus lépései esetünkben a következők:

1. A kezdeti értékek legyenek  $\rho^{(0)} = 0$  és  $\sigma^{2(0)} = 1$ . Legyen  $m = 0$ .
2. Legyenek  $y^*(\rho^{(m)})$  és  $X^*(\rho^{(m)})$  elemei a következők:

$$CHD_t^*(\rho^{(m)}) = CHD_t - \rho_1^{(m)} CHD_{t-1}$$

$$X_t^*(\rho^{(m)}) = \begin{bmatrix} 1 - \rho^{(m)} \\ CIG_t - \rho^{(m)} \cdot CIG_{t-1} \\ EDFAT_t - \rho^{(m)} \cdot EDFAT_{t-1} \\ SPIRITS_t - \rho^{(m)} \cdot SPIRITS_{t-1} \\ BEER_t - \rho^{(m)} \cdot BEER_{t-1} \end{bmatrix}'.$$

3. Generáljunk egy  $\beta^{(m+1)}$  véletlen vektort a  $\beta | \rho^{(m)}, \sigma^{2(m)} \sim N(B_1, V_1)$  feltételes eloszlásból, ahol:

$$B_1 = \left( V_0^{-1} + \frac{1}{\sigma^{2(m)}} X^*(\rho^{(m)})' X^*(\rho^{(m)}) \right)^{-1} \left( V_0^{-1} B_0 + \frac{1}{\sigma^{2(m)}} X^*(\rho^{(m)})' y^*(\rho^{(m)}) \right),$$

$$V_1 = \left( V_0^{-1} + \frac{1}{\sigma^{2(m)}} X^*(\rho^{(m)})' X^*(\rho^{(m)}) \right)^{-1}.$$

4. Legyenek  $u(\beta^{(m+1)})$  és  $U(\beta^{(m+1)})$  elemei a következők:

$$u_t(\beta^{(m+1)}) = CHD_t - \beta_0^{(m+1)} - \beta_1^{(m+1)} \cdot CIG_t - \beta_2^{(m+1)} \cdot EDFAT_t - \beta_3^{(m+1)} \cdot SPIRITS_t - \beta_4^{(m+1)} \cdot BEER_t,$$

$$U_t(\beta^{(m+1)}) = u_{t-1}(\beta^{(m+1)}).$$

5. Generáljunk egy  $\sigma^{2(m+1)}$  véletlen számot a  $\sigma^2 | \beta^{(m+1)}, \rho^{(m)} \sim IG(S_1, T_1)$  feltételes eloszlásból, ahol:

<sup>17</sup> A normális és inverz gamma eloszlásokból történő mintavétel a Függelék 2. pontjában ismertetjük.

$$S_1 = S_0 + \left( u(\beta^{(m+1)}) - U(\beta^{(m+1)})\rho^{(m)} \right)' \left( u(\beta^{(m+1)}) - U(\beta^{(m+1)})\rho^{(m)} \right),$$

$$T_1 = T_0 + T.$$

6. Generáljunk egy  $\rho^{(m+1)}$  véletlen vektort a  $\rho | \beta^{(m+1)}, \sigma^{2(m+1)} N(R_1, W_1)$  feltételes eloszlásból, ahol:

$$R_1 = \left( W_0^{-1} + \frac{1}{\sigma^{2(m+1)}} U(\beta^{(m+1)})' U(\beta^{(m+1)}) \right)^{-1} \left( W_0^{-1} R_0 + \frac{1}{\sigma^{2(m+1)}} U(\beta^{(m+1)})' u(\beta^{(m+1)}) \right),$$

$$W_1 = \left( W_0^{-1} + \frac{1}{\sigma^{2(m+1)}} U(\beta^{(m+1)})' U(\beta^{(m+1)}) \right)^{-1}.$$

7. Tároljuk el a generált  $\beta^{(m+1)}$ ,  $\rho^{(m+1)}$  és  $\sigma^{2(m+1)}$  véletlen vektort és számokat, legyen  $m = m + 1$  és menjünk vissza a 3. lépéshez.

A bayesi becslés megvalósításához az eredményeket alig vagy egyáltalán nem befolyásoló (nem informatív) priorokat választottunk:<sup>18</sup>

$$B_0 = [0 \ 0 \ 0 \ 0 \ 0]', V_0 = 100I_5, R_0 = 0, W_0 = 100, S_0 = 0 \text{ és } T_0 = 0.$$

Az iterációt a rendelkezésünkre álló adatokon 101 000-szer ismételtük, eredményeink az első 1 000 minta elhagyásával kapott 100 000-es mintán alapulnak. Az együtthatók poszterior eloszlását a 2. ábra hisztogramjai szemléltetik, illetve a poszterior eloszlás jellemző értékeiről a táblázat harmadik blokkja tartalmaz további információt.

Az eredmények értékelése előtt a klasszikus és bayesi elemzés módszertanát vetjük össze. A Cochrane–Orcutt-féle iteratív eljárás és a Gibbs-algoritmus jellegét tekintve igen hasonló egymáshoz. Mindkét eljárás az együtthatók egy halmazát adottnak feltételezve határozza meg a többi együtthatót úgy, hogy folyamatosan felcseréli az adottnak feltételezett és a meghatározandó együtthatókat. A jellegében hasonló eljárásokban ugyanakkor lényeges különbségek is vannak. A Cochrane–Orcutt-féle eljárás – lineáris modellek esetén – tulajdonképpen a feltételes poszterior móduszokat adja becslésül és egyetlen pontbecsléshez konvergál, addig a Gibbs-eljárás a feltételes poszterior módusz körül választ megfelelő értékeket úgy, hogy az ismétlések révén a paraméterek együttes poszterior eloszlása bontakozzon ki. A Gibbs-eljárásnak ezen felül az is előnye, hogy nem egyetlen fixpont értékeket keres meg, amelyek

<sup>18</sup> Ez  $\beta$  paraméterre vonatkozó prior esetén úgy érthető el, hogy a  $V_0$  és  $W_0$  kovarianciák főátlóiban szereplő értékeket kellően nagyok választjuk.

meghatározása főleg sok együtthatós, nemlineáris modellek esetén okozhat numerikus problémát.<sup>19</sup>

A poszterior eloszlásokat szemlélve feltűnő, hogy a bayesi megközelítés a /7/ egyenletben szereplő késleltetett endogén változóhoz tartozó együtthatóra ferde poszterior peremeloszlást eredményezett. Ez első látásra meglepő lehet, hiszen a Gibbs mintavételi eljárásban a /7/ egyenletben szereplő együtthatók feltételes eloszlása szimmetrikus. Ugyanakkor tudjuk, hogy a klasszikus ökonometria – némileg hosszadalmas levezetést igénylő – eredménye is hasonló: a késleltetett endogén változóhoz tartozó együttható legkisebb négyzetek elvével történő becslése nem a szokásos  $t$ -eloszlást követi. Valójában a bayesi becslés során ezt az eredményt látjuk viszont a levezetések bonyodalmai nélkül.

*Becslési eredmények: klasszikus és bayesi ökonometriai eszközökkel*

Módszer	OLS /6/		Cochrane–Orcutt /7/–/8/		Bayes /7/–/8/	
	Pontbecslés	CI 95 százalék	Pontbecslés	CI 95 százalék	Pontbecslés	HPD 95 százalék
$\beta_0$ (konstans)	139,678	–18,548 (a)	341,120	170,182 (a)	309,925	140,452 (a)
	(77,944)	297,904 (f)	(84,206)	512,059 (f)	(85,065)	474,504 (f)
$\beta_1$ (CIG)	10,706	1,388 (a)	2,902	–6,724 (a)	4,523	–4,7736 (a)
	(4,590)	20,024 (f)	(4,742)	12,529 (f)	(4,759)	13,973 (f)
$\beta_2$ (EDFAT)	3,380	1,417 (a)	0,371	–1,748 (a)	0,732	–1,339 (a)
	(0,967)	5,343 (f)	(1,044)	2,491 (f)	(1,079)	2,901 (f)
$\beta_3$ (SPIRITS)	26,749	12,464 (a)	12,005	–4,081 (a)	12,837	–3,771 (a)
	(7,037)	41,034 (f)	(7,924)	28,092 (f)	(8,895)	31,133 (f)
$\beta_4$ (BEER)	–4,132	–5,884 (a)	–2,202	–4,143 (a)	–2,289	–4,468 (a)
	(0,863)	–2,380 (f)	(0,956)	–0,261 (f)	(1,080)	–0,222 (f)
$\rho (u_{t-1})$			0,614	0,333 (a)	0,509	0,197 (a)
			(0,138)	0,895 (f)	(0,171)	0,876 (f)

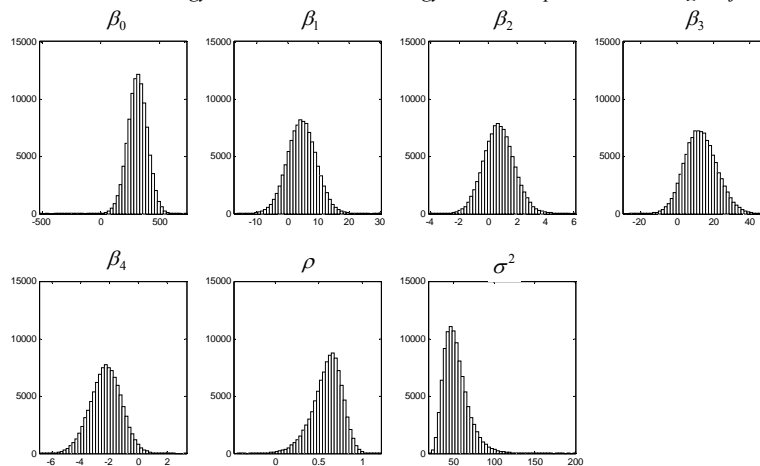
*Megjegyzés.* A klasszikus becsléseknél a CI a konfidencia intervallumot, a bayesi becsléseknél a pontbecslés a poszterior várható értéket, a HPD (highest posterior density) pedig a legnagyobb valószínűségi intervallumot jelöli, vagyis azt a legszűkebb intervallumot, ahová a poszterior eloszlás adott százaléka esik. Az (a) és (f) az intervallumok alsó és felső értékeire utal. A pontbecslés alatt zárójelben az együttható szórása szerepel.

<sup>19</sup> Az MCMC-módszereknek ez általános előnye minden pontbecslési, így például a maximum likelihood eljárással szemben: egyetlen maximumhely megkeresése helyett, mely összetett, nemlineáris ökonometriai modellek esetén numerikusan igen nehéz feladat lehet, a teljes poszterior eloszlást szimulálja, amiből a paraméterek jellemző értékei már könnyen meghatározhatók.



A /7–/8/ egyenletek együtthatóinak becslt értékeit tekintve elmondható, hogy a klasszikus és a bayesi becslés numerikusan hasonló eredményekhez vezetett, ami természetes, hiszen a bayesi becsléshez alacsony információ tartalmú (praktikusan nem informatív is tekinthető) priorokat használtunk. A numerikus különbségeket alapvetően az okozza, hogy a bayesi becslésben a poszterior átlagot számítottuk ki, ami ferde eloszlások esetén különbözik a klasszikus megközelítés módusz becslésétől.

2. ábra. A /7–/8/ egyenletekkel adott modell együtthatóinak poszterior histogramja



### 3. Összefoglalás

A tanulmányban bemutattuk a modern bayesi ökonometriai elemzések egyik gyakran alkalmazott MCMC-módszerét, a Gibbs-mintavételt, mely lehetővé teszi, hogy a bayesi elemzés során a poszterior együttes sűrűségfüggvényben levő információkat a megszokott statisztikai fogalmakba (várható érték, módusz, szórás stb.) tömörítsük. Az MCMC-eljárások forradalmasították a bayesi elemzések eszköztárát, segítségével napjainkra olyan problémák is megoldhatóvá váltak, melyek klasszikus módszerekkel egyáltalán nem, vagy csak körülményesen kezelhetők.

A tanulmány empirikus része olyan példát mutat be, ahol a klasszikus és a bayesi elemzés módszertanilag igen hasonló. Amellett, hogy a két megközelítés numerikusan hasonló becsléseket eredményezett, a bayesi becslés azzal az előnnyel járt, hogy a késleltetett endogén változóhoz tartozó becslt együttható nem standard (ferde) eloszlására automatikusan rámutatott. A bemutatott példa tanulsága általánosítható:

míg a klasszikus megközelítésben a becslési módszerek elsődlegesen pontbecsléseket eredményeznek, amelyek bizonytalansága általában csak mély valószínűségelméleti-statisztikai tudás alapján vezethető le, addig a bayesi elemzés minden esetben az ismeretlen együtthatók együttes eloszlását határozza meg, melyből az elemző tetszőleges, az eloszlást jellemző mutatókat határozhat meg.

## Függelék

### 1. Feltételes poszterior eloszlások meghatározása a Gibbs-mintavételhez

A 2.2. alfejezetben szereplő /9–/11/ priorok sűrűségfüggvényei:

$$f(\beta) = (2\pi)^{-\frac{k_\beta}{2}} |V_0|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(B_0 - \beta)' V_0^{-1}(B_0 - \beta)\right), \quad /F1/$$

$$f(\rho) = (2\pi)^{-\frac{k_\rho}{2}} |W_0|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(R_0 - \rho)' W_0^{-1}(R_0 - \rho)\right), \quad /F2/$$

$$f(\sigma^2) = \frac{\left(\frac{S_0}{2}\right)^{\frac{T_0}{2}}}{\Gamma\left(\frac{T_0}{2}\right)} \sigma^{-(T_0+2)} \exp\left(-\frac{1}{2\sigma^2} S_0\right), \quad /F3/$$

ahol  $k_\beta$  és  $k_\rho$  a /7/ és /8/ egyenletben szereplő együtthatók száma, azaz a példában  $k_\beta = 5$  és  $k_\rho = 1$ .

A főszovegben a /13/ képlettel adott poszterior az /F1–/F3/ sűrűségfüggvények és a /12/ likelihood felhasználásával a következő:

$$\begin{aligned} f(\beta, \rho, \sigma^2 | y, X) &\propto (2\pi)^{-\frac{k_\beta}{2}} |V_0|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(B_0 - \beta)' V_0^{-1}(B_0 - \beta)\right) \times \\ &\times (2\pi)^{-\frac{k_\rho}{2}} |W_0|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(R_0 - \rho)' W_0^{-1}(R_0 - \rho)\right) \times \\ &\times \frac{\left(\frac{S_0}{2}\right)^{\frac{T_0}{2}}}{\Gamma\left(\frac{T_0}{2}\right)} \sigma^{-(T_0+2)} \exp\left(-\frac{1}{2\sigma^2} S_0\right) \times \\ &\times (\sigma\sqrt{2\pi})^{-T} \exp\left(-\frac{1}{2\sigma^2} (y^*(\rho) - X^*(\rho)\beta)' (y^*(\rho) - X^*(\rho)\beta)\right). \end{aligned} \quad /F4/$$

A Gibbs-mintavétel alkalmazásához az ismeretlen együtthatók feltételes poszterior eloszlásait a következőkben vezetjük le.

$\beta | \rho, \sigma^2, y, X$  feltételes poszterior sűrűségfüggvénye /F4/ alapján:

$$f(\beta | \rho, \sigma^2, y, X) \propto \exp\left(-\frac{1}{2}(B_0 - \beta)' V_0^{-1} (B_0 - \beta) - \frac{1}{2\sigma^2} (y^*(\rho) - X^*(\rho)\beta)' (y^*(\rho) - X^*(\rho)\beta)\right), \quad /F5/$$

ahol kihasználtuk, hogy  $\rho$  és  $\sigma^2$  értékei a feltételes eloszlásban rögzítettek.

Cholesky-felbontás segítségével  $(B_0 - \beta)' V_0^{-1} (B_0 - \beta)$  átírható  $\left(V_0^{-\frac{1}{2}} B_0 - V_0^{-\frac{1}{2}} \beta\right)' \left(V_0^{-\frac{1}{2}} B_0 - V_0^{-\frac{1}{2}} \beta\right)$  alakra, ahol  $V_0 = V_0^{-\frac{1}{2}} V_0^{-\frac{1}{2}}$ . Bevezetve a  $z = \begin{bmatrix} y^*(\rho) \\ \sigma V_0^{-\frac{1}{2}} B_0 \end{bmatrix}$  és  $Z = \begin{bmatrix} X^*(\rho) \\ \sigma V_0^{-\frac{1}{2}} \end{bmatrix}$  jelöléseket, /F5/ átírható tömörebb formára:

$$f(\beta | \rho, \sigma^2, y, X) \propto \exp\left(-\frac{1}{2\sigma^2} (z - Z\beta)' (z - Z\beta)\right). \quad /F6/$$

A dekompozíciós szabály<sup>20</sup> segítségével /F6/ így is írható:

$$f(\beta | \rho, \sigma^2, y, X) \propto \exp\left(-\frac{1}{2\sigma^2} (\beta - \hat{\beta})' Z'Z (\beta - \hat{\beta})\right), \quad /F7/$$

ami egy konstans tényezőtől eltekintve egy többváltozós normális eloszlás sűrűségfüggvénye, azaz  $\beta | \rho, \sigma^2, y, X \sim N(\hat{\beta}, \sigma^2 (Z'Z)^{-1})$ . Tehát  $\beta$  feltételes eloszlása normális  $B_1$  várható érték vektorral és  $V_1$  kovarianciamátrixszal, ahol:

$$B_1 = \left(V_0^{-1} + \frac{1}{\sigma^2} X^*(\rho)' X^*(\rho)\right)^{-1} \left(V_0^{-1} B_0 + \frac{1}{\sigma^2} X^*(\rho)' y^*(\rho)\right), \quad /F8-F9/$$

$$V_1 = \left(V_0^{-1} + \frac{1}{\sigma^2} X^*(\rho)' X^*(\rho)\right)^{-1}.$$

Hasonlóképpen  $\rho | \beta, \sigma^2, y, X$  feltételes sűrűségfüggvénye /F4/ alapján:

<sup>20</sup> Dekompozíciós szabálynak hívjuk a következő azonosságot:

$(z - Z\beta)' (z - Z\beta) = (z - Z\hat{\beta})' (z - Z\hat{\beta}) + (\beta - \hat{\beta})' Z'Z (\beta - \hat{\beta})$ , ahol  $\hat{\beta} = (Z'Z)^{-1} Z'z$ , ami nem más mint a legkisebb négyzetek (vagy maximum likelihood) becslőfüggvény.

$$f(\rho | \beta, \sigma^2, y, X) \propto \exp\left(-\frac{1}{2}(R_0 - \rho)' W_0^{-1}(R_0 - \rho)\right) \exp\left(-\frac{1}{2\sigma^2}(y^*(\rho) - X^*(\rho)\beta)'(y^*(\rho) - X^*(\rho)\beta)\right), \quad /F10/$$

ahol kihasználtuk, hogy  $\beta$  és  $\sigma^2$  értékei a feltételes eloszlásban rögzítettek. Használva  $u_t = CHD_t - \beta_0 - \beta_1 \cdot CIG_t - \beta_2 \cdot EDFAT_t - \beta_3 \cdot SPIRITS_t - \beta_4 \cdot BEER_t$  definícióját, legyen  $u(\beta)$  az  $u_t$ -ből álló oszlopvektor,  $U(\beta)$  pedig az  $u_{t-1}$  megfigyelésekből álló  $T \times 1$  méretű mátrix (vektor). Ekkor  $y^*(\rho) - X^*(\rho)\beta = u(\beta) - U(\beta)\rho$ . Ez utóbbi segítségével  $\rho | \beta, \sigma^2, y, X$  feltételes sűrűségfüggvénye:

$$f(\rho | \beta, \sigma^2, y, X) \propto \exp\left(-\frac{1}{2}(R_0 - \rho)' W_0^{-1}(R_0 - \rho)\right) \exp\left(-\frac{1}{2\sigma^2}(u(\beta) - U(\beta)\rho)'(u(\beta) - U(\beta)\rho)\right). \quad /F11/$$

Az /F11/ feltételes sűrűségfüggvény formáját tekintve megegyezik az /F5/ feltételes sűrűségfüggvénnyel, így megismételve az előzőekben leírt lépéseket kapjuk, hogy  $\rho | \beta, \sigma^2, y, X$  feltételes sűrűségfüggvénye normális eloszlású  $R_1$  várható érték vektorral és  $W_1$  kovarianciamátrixszal, ahol:

$$R_1 = \left(W_0^{-1} + \frac{1}{\sigma^2}U(\beta)'U(\beta)\right)^{-1} \left(W_0^{-1}R_0 + \frac{1}{\sigma^2}U(\beta)'u(\beta)\right), \quad /F12-13/$$

$$W_1 = \left(W_0^{-1} + \frac{1}{\sigma^2}U(\beta)'U(\beta)\right)^{-1}.$$

Végül  $\sigma^2 | \beta, \rho, y, X$  feltételes sűrűségfüggvénye /F4/ alapján:

$$f(\sigma^2 | \beta, \rho, y, X) \propto \sigma^{-(T+T_0+2)} \exp\left(-\frac{1}{2\sigma^2}S_0\right) \exp\left(-\frac{1}{2\sigma^2}(y^*(\rho) - X^*(\rho)\beta)'(y^*(\rho) - X^*(\rho)\beta)\right) =$$

$$= \sigma^{-(T+T_0+2)} \exp\left(-\frac{1}{2\sigma^2}\left[S_0 + (y^*(\rho) - X^*(\rho)\beta)'(y^*(\rho) - X^*(\rho)\beta)\right]\right) = \quad /F14/$$

$$= \sigma^{-(T+T_0+2)} \exp\left(-\frac{1}{2\sigma^2}\left[S_0 + (u(\beta) - U(\beta)\rho)'(u(\beta) - U(\beta)\rho)\right]\right).$$

ahol kihasználtuk, hogy  $\beta$  és  $\rho$  értékei a feltételes eloszlásban rögzítettek. Az /F13/ kifejezés egy konstanstól eltekintve egy  $IG(S_1, T_1)$  eloszlás sűrűségfüggvénye, ahol:

$$S_1 = S_0 + (u(\beta) - U(\beta)\rho)'(u(\beta) - U(\beta)\rho), \quad /F15-F16/$$

$$T_1 = T_0 + T.$$

## 2. Véletlen vektorok generálása adott paraméterű normál és inverz gamma eloszlásokból

### a) Véletlen vektor generálása $N(B, V)$ paraméterű normális eloszlásból

A matematikai-statisztikai programcsomagok általában rendelkeznek olyan véletlenszám-generátorral, amely képes független, standard normális eloszlást követő véletlen számokat előállítani. Jelölje  $B$  az előállítani kívánt normális eloszlás várható értékének oszlopvektorát (sorainak száma  $k$ ),  $V$  pedig a kovarianciamátrixát. Állítsunk elő egy  $k$  sorú független, standard normális eloszlást követő véletlen  $u$  vektort. Ekkor  $z = B + V^{1/2}u$  módon definiált  $z$  véletlen vektor  $B$  várható érték vektorú,  $V$  kovarianciájú normális eloszlást fog követni.

### b) Véletlen vektor generálása $IG(S, T)$ paraméterű inverz gamma eloszlásból

$S$  skála paraméter és  $T$  szabadságfokú inverz gamma eloszlást követő véletlen szám előállításához generáljunk egy  $T$  sorú véletlen  $u$  oszlopvektort a standard normális eloszlásból. Ekkor  $z = \frac{S}{u'u}$  módon definiált  $z$  véletlen változó  $S$  skála paraméterű és  $T$  szabadságfokú inverz gamma eloszlást fog követni.

## Irodalom

- ALBERT, J. H. [2009]: *Bayesian Computation with R. 2nd Edition*. Springer. New York.
- CASELLA, G. [2007]: Why (Not) Frequentist Inference (Too)? *Boletín de Estadística e Investigación Operativa*. Vol. 23. No. 1. pp. 5–6.
- CASELLA, G. – BERGER, R. L. [2011]: A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Statistical Science*. Vol. 26. No. 1. pp. 102–115.
- CASELLA, G. – GEORGE, E. I. [1992]: Explaining the Gibbs Sampler. *The American Statistician*. Vol. 46. No. 3. pp. 167–174.
- CHIB, S. [1993]: Bayes Regression with Autoregressive Errors: A Gibbs Sampling Approach. *Journal of Econometrics*. Vol. 58. Issue 3. pp. 275–294.
- CHIB, S. – GREENBERG, E. [1995]: Understanding the Metropolis-Hastings Algorithm. *The American Statistician*. Vol. 49. No. 4. pp. 327–335.
- CONGDON, P. [2005]: *Bayesian Models for Categorical Data*. Wiley. New York.
- COWLES, M. K. – CARLIN, B. P. [1996]: Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*. Vol. 91. No. 434. pp. 883–904.
- DEGROOT, M. H. – SCHERVIS, M. J. [2012]: *Probability and Statistics*. Fourth edition. Pearson. Boston.
- DUFOUR, J. M. – GAUDRY, M. – LIEM, T. C. [1980]: The Cochrane-Orcutt Procedure Numerical Example of Multiple Admissible Minima. *Economics Letters*. Vol. 6. No. 1. pp. 43–48.

- GELFAND, A. E. – SMITH, A. F. M. [1990]: Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*. Vol. 85. Issue 410. pp. 398–409.
- GELMAN, A. – CARLIN, J. B. – STERN, H. S. – RUBIN, D. B. [2004]: *Bayesian Data Analysis*. Chapman & Hall/CRC. Boca Raton.
- GELMAN, A. – RUBIN, D. B. [1992]: Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*. Vol. 7. No. 4. pp. 457–472.
- GEMAN, S. – GEMAN, D. [1984]: Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 6. Issue 6. pp. 721–741.
- GEWEKE, J. [1992]: Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In: *Bernardo, J. M. – Berger, J. O. – Dawid, A. P. – Smith, A. F. M. (eds.): Bayesian Statistics 4*. Clarendon Press. Oxford.
- GEWEKE, J. [2005]: *Contemporary Bayesian Econometrics and Statistics*. Wiley. New York.
- HASTINGS, W. K. [1970]: Monte Carlo Sampling Methods Using Markov Chains and Their Application. *Biometrika*. Vol. 57. Issue 1. pp. 97–109.
- HUNYADI L. [2011]: Bayesi gondolkodás a statisztikában. *Statisztikai Szemle*. 89. évf. 10–11. sz. 1150–1171. old.
- KEHL D. [2012]: Robert, C. – Casella, G.: Szemelvények a Markov-lánc Monte-Carlo módszerek történetéből. *Statisztikai Szemle*. 90. évf. 4. sz. 352–354. old.
- KOOP, G. [2003]: *Bayesian Econometrics*. Wiley. New York.
- KOOP, G. – POIRIER, D. J. – TOBIAS, J. L. [2007]: *Bayesian Econometric Methods*. Econometric Exercises 7. Cambridge University Press. Cambridge.
- METROPOLIS, N. – ROSENBLUTH, A. – ROSENBLUTH, M. – TELLER, A. – TELLER, E. [1953]: Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*. Vol. 21. No. 6. pp. 1087–1092.
- RAMANATHAN, R. [2003]: *Bevezetés az ökonometriába alkalmazásokkal*. Panem Kiadó. Budapest.
- ROBERT, C. P. – CASELLA, G. [2004]: *Monte Carlo Statistical Methods. 2nd Edition*. Springer. New York.
- ROBERT, C. P. – CASELLA, G. [2011]: A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Statistical Science*. Vol. 26. No. 1. pp. 102–115.
- VÁRPALOTAI V. [2008]: *Modern Bayes-i ökonometriai elemzések. Simasági priorok alkalmazása az üzleti ciklusok szinkronizációjának mérésére és az infláció előrejelzése*. PhD-értekezés. Budapesti Corvinus Egyetem. Budapest.

## Summary

This study demonstrates one of the most widely used members of modern bayesian computations, Gibbs sampling. After introducing the theoretical background and the algorithm itself, a short numerical example illustrates the procedure and the speed of convergence. The practical usefulness is proven through the solution of a well-known problem with bayesian methods. This also enables the authors to show similarities and differences of the classical frequentist theory and the bayesian framework.