

Syllabus

Analiză Bayesiană*

Iunie 2014

Instructor: Constantin Manuel Bosâncianu
Afilieră: Universitatea Centrală Europeană, Budapesta
Email: Bosancianu_Constantin@ceu-budapest.edu
Website: <https://manuelbosancianu.github.io>

Introducere

Analiza Bayesiană reprezintă un corp de metode statistice relativ noi în științele sociale, dar bazat pe principii care sunt vechi de câteva secole. Deși metodele parcurse în acest modul au fost elaborate cu decenii în urmă, implementarea lor în analizele empirice a fost întârziată de cerințele computaționale înalte pe care aceste metode le au. În ultimul deceniu, însă, putem observa o propagare din ce în ce mai alertă a metodelor Bayesiene în științele sociale. Parțial, această popularitate poate fi pusă pe seama disponibilității mai largi a tehnologiei de calcul avansate.¹ Cu toate acestea, o explicație bună a fenomenului trebuie să includă proprietățile specifice ale metodelor Bayesiene. Acestea oferă posibilitatea de a estima parametri cu asumții minime privind distribuția acestora, în același timp încorporând o gamă mult mai variată de forme funcționale ale modelelor statistice decât ar fi disponibile în paradigma frecventistă. Abilitatea de a estima modele complexe, cu asumții distribuționale minime, precum și de a folosi în estimare informația conținută în eșantionul cu care lucrăm dar și cunoștințele prealabile privind subiectul investigat, au făcut aceste metode extrem de populare în științele sociale.

Datorită nevoii de a păstra modului la un nivel introductiv, majoritatea materialelor obligatorii oferite sunt din două manuale al căror public țintă sunt cercetătorii din științele sociale. Chiar și cu această restricție conținutul matematic al modului este, probabil, cel mai mare dintre toate cursurile oferite la această școală de vară. Recomand participanților să aloce timpul corespunzător capitolelor din cartea lui Lynch (2007) și Kruschke (2011). În mod special capitolele introductive din ambele cărți vor forma fundația fără de care capitolele mai avansate matematic (referitoare la metodele *Markov Chain Monte Carlo*) nu vor putea fi înțelese. Intenția nu este ca participanții să aibă abilitatea de a reproduce formule matematice după cele 8 săptămâni alocate modului, ci de a dobândi o înțelegere intuitivă a modalității în care aceste metode funcționează.

La fel ca în cazul modului de Modele Ierarhice Liniare, am alocat unei săptămâni o serie de articole empirice bazate pe metode Bayesiene, pentru a pune în perspectivă informația teoretică asimilată în celelalte săptămâni. În cazul în care unii participanți preferă articole din propriul domeniu de activitate, iar nu din științele politice (domeniu cu care eu sunt mai familiar), v-aș ruga să mă contactați. În această situație vă pot pune la dispoziție câteva articole din aria dumneavoastră de expertiză.

*Modul pregătit pentru Școala de Vară de Metode Aplicate, ediția 2014 (2 Iunie - 26 Iulie), Facultatea de Științe Politice, Administrative și ale Comunicării, Universitatea Babeș-Bolyai, Cluj Napoca, România.

¹Serverele cu memorie RAM de 1000 GB încep să devină comune în centrele de cercetare statistică.

Bibliografie

Săptămâna 1 (2–8 Iunie): Probabilitate în paradigma frecventistă și Bayesiană

Probabil cel mai dificil pas pentru un începător în metodele Bayesiene este de a fi sensibil la diferențele dintre noțiunea frecventistă și cea Bayesiană de *probabilitate*. Această săptămână este alocată acestui subiect, precum și a implicațiilor acestor diferențe pentru analizele noastre.

Obligatoriu

- Kruschke, John K. (2011). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Burlington, MA: Academic Press. Capitolul III (pp. 23–47). URL: <https://faculty.washington.edu/jmiyamot/p548/kruschkejk%20bayesian%20data%20analysis.ch01-06.pdf>.

Opțional

- Lynch, Scott M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer. Capitolul II (pp. 9–44). URL: http://faculty.ksu.edu.sa/69424/us_B00k/Introduction%20to%20Applied%20Bayesian%20Statistics.pdf.
- Berger, James O., și Donald A. Berry. (1988). Statistical Analysis and the Illusion of Objectivity. *American Scientist*, **76**, 159–165. URL: <http://www.med.mcgill.ca/epidemiology/joseph/courses/EPIB-675/Berger.Berry.pdf>.

Săptămâna 2 (9–15 Iunie): Concepte de bază în analiza Bayesiană

Săptămâna aceasta vom examina principiul fundamental pe care se bazează analiza Bayesiană (Teorema lui Bayes), precum și o serie de termeni de bază folosiți în acest tip de analiză (informație anterioară–*prior*–și posterioară–*posterior*).

Obligatoriu

- Lynch, Scott M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer. Capitolul III (pp. 47–74).

Opțional

- Kruschke, John K. (2011). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Burlington, MA: Academic Press. Capitolul IV (pp. 51–69).
- Albert, Jim. (2009). *Bayesian Computation with R*. New York: Springer. Capitolul II (pp. 19–35).
- Goldstein, Michael. (2006). Subjective Bayesian Analysis: Principles and Practice. *Bayesian Analysis*, **1**(3), 403–420. URL: http://www.stat.cmu.edu/~rsteorts/btheory/goldstein_subjective_2006.pdf.
- Western, Bruce. (1999). Bayesian Analysis for Sociologists: An Introduction. *Sociological Methods & Research*, **28**(1), 7–34. URL: http://www.tc.umn.edu/~alonso/Western_SocMetRes_1999.pdf.

Săptămâna 3 (16–22 Iunie): MCMC I–Algoritmul Gibbs

Algoritmul Gibbs reprezintă una dintre metodele de bază prin care putem simula distribuția posterioară a parametrului de interes.

Obligatoriu

- Lynch, Scott M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer. Capitolul IV (pp. 77–105).

Opțional

- Kruschke, John K. (2011). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Burlington, MA: Academic Press. Capitolul VIII (pp. 155–178).
- Albert, Jim. (2009). *Bayesian Computation with R*. New York: Springer. Capitolul VI (pp. 117–145).

Săptămâna 4 (23–29 Iunie): MCMC II–Algoritmul Metropolis–Hastings

A doua metodă de estimare a parametrului de interes prin folosirea simulațiilor statistice este algoritmul Metropolis–Hastings. Împreună cu algoritmul Gibbs, despre care ați citit săptămâna trecută, el formează nucleul metodelor *Markov Chain Monte Carlo* din analiza Bayesiană.

Obligatoriu

- Lynch, Scott M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer. Capitolul V (pp. 107–129).

Opțional

- Kruschke, John K. (2011). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Burlington, MA: Academic Press. Capitolul VII (pp. 117–146).
- Albert, Jim. (2009). *Bayesian Computation with R*. New York: Springer. Capitolul VI (pp. 117–146).

Săptămâna 5 (30 Iunie – 6 Iulie): Exemple empirice

Vom face o scurtă pauză din aspectele teoretice săptămâna aceasta pentru a trece din nou peste câteva aspecte fundamentale privind metodele Bayesiene, și a vedea cum pot fi acestea aplicate în analize empirice.

Obligatoriu

- Jackman, Simon. (2004). Bayesian Analysis for Political Research. *Annual Review of Political Science*, 7, 483–505. URL: <http://jackman.stanford.edu/mcmc/AnnualReviews.pdf>.

Opțional

- Jackman, Simon. (2001). Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Model Checking. *Political Analysis*, 9(3), 227–241. URL: <http://polmeth.wustl.edu/analysis/vol/9/PA93-227-241.pdf>.
- Western, Bruce, și Simon Jackman. (1994). Bayesian Inference for Comparative Research. *The American Political Science Review*, 88(2), 412–423. URL: <http://allman.rhon.itam.mx/~emagar/talleR/lecturas/dia2reg/western+jackmanBayes1994apsr.pdf>.

Săptămâna 6 (7–13 Iulie): Evaluarea modelelor

Deși metodele Bayesiene sunt mult mai flexibile și mai puternice decât cele frecventiste, aceasta nu înseamnă că modelele vor estima întotdeauna parametrii de interes fără probleme. În această săptămână vom parcurge modalitățile prin care putem evalua modelele noastre statistice din perspectiva performanței algoritmilor din spatele lor.

Obligativ

- Lynch, Scott M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer. Capitolul VI (pp. 131–163).

Opțional

- Cowles, Mary K., și Bradley P. Carlin. (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*, 91(434), 883–904. URL: http://www.math.pitt.edu/~cbsg/Materials/Cowles96_MCMCconvergence.pdf.
- Brooks, Stephen P., și Andrew Gelman. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455. URL: <http://www.stat.columbia.edu/~gelman/research/published/brooksgelman2.pdf>.

Săptămâna 7 (14–20 Iulie): Regresie în analiza Bayesiană

În această săptămână de final vom parcurge modul prin care parametrii dintr-o regresie liniară pot fi estimați prin metode Bayesiene, folosind algoritmi învățați în săptămânile anterioare.

Obligativ

- Lynch, Scott M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer. Capitolul VII (pp. 165–192).
- Kruschke, John K. (2011). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Burlington, MA: Academic Press. Capitolul XXIII (pp. 619–626).

Opțional

- Albert, Jim. (2009). *Bayesian Computation with R*. New York: Springer. Capitolul IX (pp. 205–217).

Examinare

Modulul nu va necesita scrierea unei lucrări la final, sau un examen scris; datorită dificultății materialelor obligatorii, o asemenea solicitare ar fi injustă. Participanții sunt rugați să scrie un mini-eseu de maximum o pagină (A4, corp de literă Georgia 11, spațiere la un rând, format Word sau PDF) în care să-și argumenteze poziția personală privind următoarea afirmație: “Metodele Bayesiene sunt mult mai bune decât cele frecventiste”. Nu sunt interesat de citate din lecturile sugerate mai sus, ci de o poziție argumentată și personală pe baza căreia pot evalua măsura în care ați înțeles conceptele discutate de-a lungul modulului. Atât timp cât acest argument este valid, poziția finală contează mai puțin (dacă este un “da”, un “nu”, sau un “depinde”). Termenul limită de trimitere a lucrării prin email este 18 Iulie, 23:59 (ora României).

PS: Mult succes cu lecturile, și sper să ne vedem în Cluj Napoca!

PPS: Dacă pot fi de ajutor cu informații privind lecturile alocate sau analiză Bayesiană până când ne vom vedea în Cluj Napoca, nu ezitați să-mi trimiteți un email!