

Proiect

Grupele 241, 242, 243 și 244

Notă privind redactarea proiectului: Raportul trebuie scris prin intermediul pachetului `Rmarkdown` din R. Informații introductive despre modul de folosire a acestui pachet pot fi găsite [aici](#) sau [aici](#) iar pentru mai multe detalii se poate consulta cartea [R Markdown: The definitive guide](#). Toate simulările, figurile și codurile folosite trebuie incluse în raport. Se va folosi doar limbajul R. Raportul trimis trebuie să conțină pe lângă fișierul generat (*HTML*, *Microsoft Word* sau *L^AT_EX*) și fișierul `.Rmd` care conține codul sursă.

Problema

Scopul acestui proiect este de a crea o aplicație web folosind pachetul `shiny` pentru R dezvoltat de `RStudio`. Prin intermediul acestui pachet, utilizatorul poate construi o aplicație web care prezintă funcționalități similare cu applet-urile Java/Javascript.


Pentru a începe, includ mai jos o serie de resurse care ar putea să vă fie de folos în procesul de documentare:

- tutorialul video de pe pagina `RStudio` <https://shiny.rstudio.com/tutorial/>
- o serie de articole despre funcționalitățile `Shiny` <https://shiny.rstudio.com/articles/> sau sumarizate într-un [cheatsheet](#)
- exemple de aplicații în `Shiny` <https://shiny.rstudio.com/gallery/>

De asemenea, mai jos regăsiți posibile teme pentru proiect dar puteți considera orice aplicație doriți¹:

- o aplicație în care să ilustrați o serie de repartiții discrete și continue (minim 15) și modul în care pot fi utilizate în calcul: astfel pentru fiecare repartiție să aveți posibilitatea să selectați valorile parametrilor care definesc repartiția, să ilustrați grafic densitatea (respectiv funcția de masă) și funcția de repartiție, să calculați probabilități de tipul $\mathbb{P}(X \leq a)$, $\mathbb{P}(X \geq b)$ sau $\mathbb{P}(a \leq X \leq b)$, pentru valori ale lui a și b date și să ilustrați grafic probabilitatea calculată prin hașurarea zonei corespunzătoare
- o aplicație care să illustreze câteva dintre paradoxurile din teoria probabilităților și baza teoretică pe care o evidențiază acestea: paradoxul lui Bertrand [link](#), paradoxul lui Simpson (ce a derutat oamenii chiar și în timpul pandemiei de Covid-19; iată cum [video](#), etc. Alte paradoxuri au fost prezentate [aici](#). Încercați să includeți simulări/grafice în R ale experimentelor derulate în cadrul problemelor abordate, acolo unde este posibil.
- o aplicație care să illustreze și evalueze acuratețea unor aproximări ale funcției de repartiție a binomialei $\mathcal{B}(n, p)$: aproximarea Poisson, aproximarea normală (rezultată prin aplicarea Teoremei Limită Centrale), aproximarea normală cu factor de corecție, aproximarea Camp-Paulson (a se vedea (Lesch and Jeske 2009)) sau aproximarea prin intermediul repartiției normale asimetrice (a se vedea (Chang et al. 2008))
- o aplicație care să trateze partea teoretică și practică (grafice/tabele) a unei probleme de interes (public) pe baza unui set de date din R, cum ar fi *the broken heart effect* de [aici](#)

¹Vă rog să ne trimiteți un email cu tema aleasă așa încât să o pot valida din punct de vedere al dificultății până pe data de 9.01.2022

- 
- o aplicație care să implementeze **Chaos Game** în două dimensiuni (mai multe detalii puteți vedea pagina de Wikipedia https://en.wikipedia.org/wiki/Chaos_game sau <https://mathworld.wolfram.com/ChaosGame.html>)
 - o aplicație care să susțină prin demonstrații și simulări ideea intrigantă (citiți pentru început caseta de la final: *on the edge at .05*) din articolul (Schilling 1990) care se găsește [aici](#)
 - o aplicație care să implementeze **Legea lui Benford** pentru șiruri (https://en.wikipedia.org/wiki/Benford%27s_law sau <https://mathworld.wolfram.com/BenfordsLaw.html>)
 - o aplicație care să ilustreze modalitatea de generare a unei variabile aleatoare repartizate normal folosind algoritmul de acceptare respingere și metoda Box-Muller (a se vedea (Rubinstein and Kroese 2016))
 - o aplicație care să prezinte componenta probabilistă din problemele de *machine learning* și *data science*, având ca sugestii bibliografice (primele capitole din) cărțile (Murphy 2012) și (Shalev-Shwartz and Ben-David 2014)

Referințe

- Chang, Ching-Hui, Jyh-Jiuan Lin, Nabendu Pal, and Miao-Chen Chiang. 2008. "A Note on Improved Approximation of the Binomial Distribution by the Skew-Normal Distribution." *The American Statistician* 62 (2): 167–70. <https://doi.org/10.1198/000313008x305359>.
- Lesch, Scott M., and Daniel R. Jeske. 2009. "Some Suggestions for Teaching about Normal Approximations to Poisson and Binomial Distribution Functions." *The American Statistician* 63 (3): 274–77. <https://doi.org/10.1198/tast.2009.08147>.
- Murphy, Kevin P. 2012. *Machine Learning: A Probabilistic Perspective*. MIT press.
- Rubinstein, Reuven Y, and Dirk P Kroese. 2016. *Simulation and the Monte Carlo Method*. Vol. 10. John Wiley & Sons.
- Schilling, Mark F. 1990. "The Longest Run of Heads." *The College Mathematics Journal* 21 (3): 196–207.
- Shalev-Shwartz, Shai, and Shai Ben-David. 2014. *Understanding Machine Learning: From Theory to Practice*. Cambridge University Press.