Simulare Monte Carlo pentru Estimarea Câștigătoarei UEFA Champions League

Barbu David-FLorian, Filote Ionut-Alexandru

12 ianuarie 2026

Rezumat

Acest proiect își propune să estimeze probabilitățile de câștig ale echipelor participante în UEFA Champions League (sezoanele 2024-2025 si 2025-2026) folosind metoda Monte Carlo. Având în vedere complexitatea noului format competițional și natura statistică a fotbalului, nu există o soluție analitică simplă. Utilizând date reale de tip Expected Goals (xG) și un model probabilistic bazat pe distribuția Poisson, am simulat desfășurarea turneului de $N = 50.000$ de ori pentru a obține o estimare robustă a probabilităților finale.

Cuprins

1 Descrierea Problemei și Motivație

1.1 Alegerea Temei

Fotbalul este un sistem complex guvernat de aleatoriu, dar și de performanța subiacentă a echipelor. Estimarea câștigătoarei unei competiții precum UEFA Champions League este o problemă dificilă din două motive:

1. Formatul Competițional: Noul format "Swiss-style" și bracket-ul eliminatoriu introduc un număr imens de combinații posibile de meciuri.
2. Natura Datelor: Scorul unui meci nu reflectă întotdeauna fidel performanța. Metricile moderne, precum *Expected Goals (xG)*, oferă o imagine mai clară asupra forței reale a unei echipe.

Am ales această problemă deoarece combină analiza statistică sportivă (Sports Analytics) cu metode numerice avansate, având o aplicabilitate reală în predicții economice (piața de pariuri) și analiză de performanță.

2 Formularea Matematică

Pentru a simula un meci între Echipa A (gazdă) și Echipa B (oaspete), modelăm numărul de goluri marcate ca variabile aleatoare independente distribuite Poisson:

$$G_A \sim \text{Poisson}(\lambda_A), \quad G_B \sim \text{Poisson}(\lambda_B)$$

Parametrii λ reprezintă numărul așteptat de goluri și sunt calculați astfel:

$$\lambda_A = \text{MediaLigii} \times \text{Att}_A \times \text{Def}_B \times \text{HomeAdv} \tag{1}$$

$$\lambda_B = \text{MediaLigii} \times \text{Att}_B \times \text{Def}_A \times \frac{1}{\text{HomeAdv}} \tag{2}$$

Unde:

- Att_A: Coeficientul de atac al echipei A (derivat din xG creat).
- Def_B: Coeficientul de apărare al echipei B (derivat din xG permis).
- HomeAdv: Factorul de avantaj al terenului propriu (estimat la 1.1).

3 Algoritmul Monte Carlo

Metoda Monte Carlo se bazează pe Legea Numerelor Mari, aproximând valoarea așteptată a unei variabile aleatoare prin media eșantionului rezultat din simulări repetate.

3.1 Pașii Algoritmului

1. Parsarea Datelor: Calculăm rating-urile ofensive și defensive pentru fiecare echipă pe baza meciurilor din faza ligii.

2. Simularea unui Turneu:

   - Se generează scoruri pentru fiecare meci din bracket folosind distribuția Poisson.
   - În caz de egalitate în fazele eliminatorii, se alege câștigătorul aleator (simulare penalty-uri).

3. Repetiție: Procesul se repetă de N ori.

4. Agregare: Probabilitatea de câștig a echipei i este estimată prin:

$$\hat{p}_i = \frac{\text{Număr victorii echipa } i}{N}$$

4 Rigoare Teoretică: Teorema Limită Centrală

Pentru a estima numărul necesar de simulări (N) și a valida precizia rezultatelor, am utilizat **Teorema Limită Centrală (TLC)**.

Fie X_i o variabilă aleatoare Bernoulli pentru o echipă dată într-o singură simulare i:

$$X_i = \begin{cases} 1 & \text{dacă echipa câștigă turneul} \\ 0 & \text{altfel} \end{cases}$$

Estimatorul probabilității de câștig este media eșantionului: $\hat{p} = \frac{1}{N}\sum_{i=1}^{N} X_i$.

Conform Teoremei Limită Centrale, pentru un N suficient de mare, distribuția erorii de estimare converge către o distribuție normală:

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{N}\right)$$

4.1 Calculul Numărului Necesar de Simulări

Pentru a construi un interval de încredere de 70%, utilizăm scorul standard $Z_{\alpha/2} \approx 1.04$ (corespunzător pentru $\alpha = 0.30$). Marja de eroare (ϵ) este dată de formula:

$$\epsilon = Z \cdot \sqrt{\frac{p(1-p)}{N}} \tag{3}$$

Pentru a determina N necesar pentru o eroare maximă de 1% ($\epsilon = 0.01$), considerăm cazul cel mai defavorabil (varianța maximă), care are loc atunci când $p = 0.5$. Astfel, $p(1-p) = 0.25$.

Rezolvând ecuația pentru N:

$$N = \frac{Z^2 \cdot p(1-p)}{\epsilon^2} = \frac{1.04^2 \cdot 0.25}{0.01^2}$$

$$N = \frac{1.0816 \cdot 0.25}{0.0001} = \frac{0.2704}{0.0001} \approx 2.704$$

4.2 Analiza Preciziei pentru N=50.000

Deși calculul teoretic (bazat pe un nivel de încredere de 70%, unde $Z \approx 1.04$) indică faptul că aproximativ 2.704 simulări ar fi fost suficiente pentru o precizie de $\pm 1\%$, în cadrul acestui proiect am ales să rulăm $N = 50.000$ de simulări pentru a asigura o convergență robustă și pentru a reduce zgomotul statistic.

Recalculând marja de eroare teoretică efectivă pentru $N = 50.000$ la acest nivel de încredere:

$$\epsilon_{final} = 1.04 \cdot \sqrt{\frac{0.25}{50.000}} = 1.04 \cdot \sqrt{0.000005} \approx 1.04 \cdot 0.00223 \approx 0.0023$$

Concluzie: Cu 50.000 de iterații, rezultatele noastre au o marjă de eroare statistică de aproximativ $\pm 0.23\%$ (la un nivel de încredere de 70%). Acest rezultat este semnificativ mai precis decât marja inițială de 1% pe care ne-am propus-o, validând soliditatea simulării.

5 Implementare și Rezultate

Am implementat simularea în Python folosind bibliotecile 'numpy' pentru generarea numerelor aleatoare și 'plotly' pentru vizualizare.

5.1 Modele Utilizate

Am testat trei scenarii pentru a observa sensibilitatea rezultatelor:

- **1. xG Clasic:** Bazat strict pe media xG din meciurile anterioare.
- **2. Hibrid:** Include ajustări bazate pe valoarea de piață a lotului.
- **3. Complet:** Introduce o varianță suplimentară ($\pm 15\%$) în parametrii λ si adaugă eficacitatea echipelor în fața porții proprii și adverse.

5.2 Rezultate Grafice

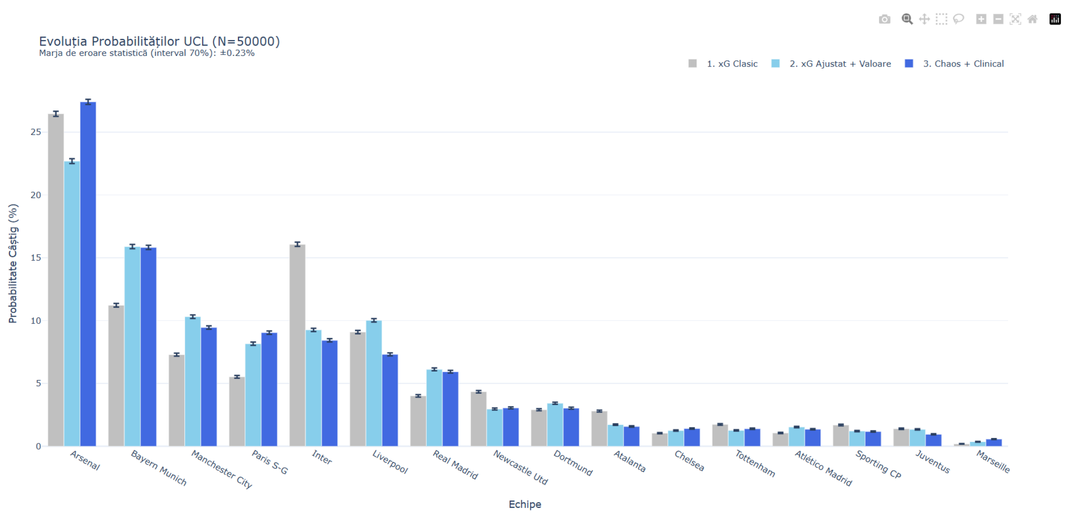

Figura 1: Probabilitățile de câștig pentru Top 16 echipe (N=50.000)

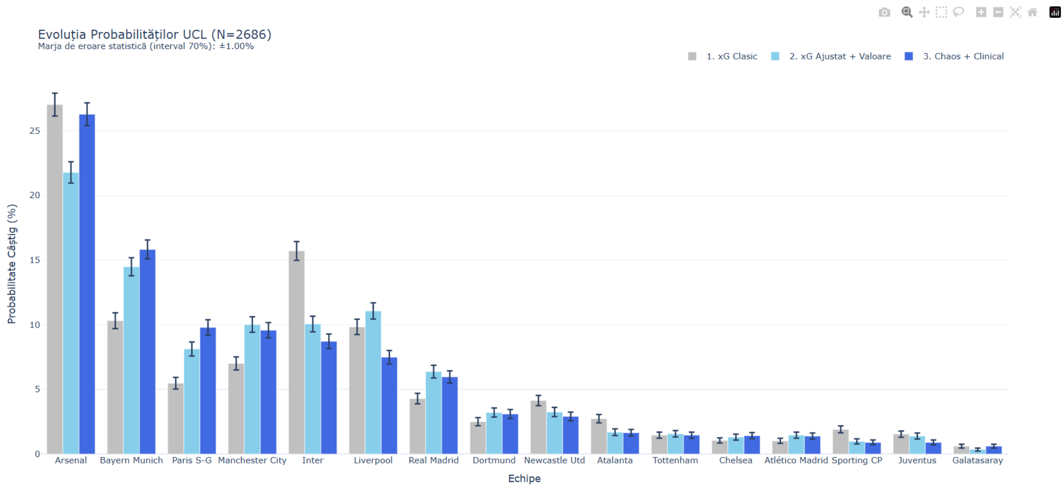

Figura 2: Probabilitățile de câștig pentru Top 16 echipe (N=2700)

Graficul evidențiază cum echipele cu un xG ridicat (ex: Arsenal, Inter) domină simulările, dar modelul "Complet" redistribuie ușor șansele către echipele pragmatice (ex: Bayern Munchen).

6 Concluzii

Metoda Monte Carlo s-a dovedit eficientă pentru a cuantifica incertitudinea în fotbal. Analiza erorilor prin Teoremă Limită Centrală confirmă validitatea statistică a celor 2700 de rulări. Rezultatele demonstrează că, deși norocul joacă un rol în meciuri individuale, pe termen lung (în mii de turnee simulate), calitatea jocului (măsurată prin xG) este determinantă.