

# Sporteredmények modellezése extrém érték modellekkel

*Csáfordi József András*

Az önálló projektem első félévében a Berlieni maraton győztes eredményeit próbáltam modellezni extrém-érték modellek segítségével. Az extrém-elméleti modelleket nagyon ritkán előforduló események bekövetkezésének kockázatára használjuk, mint az időjárásban (legmelegebb nap) vagy a pénzügyi területeken (legnagyobb tőzsdei összeomlás). Sporteredményekben a rekordok létfontosságú események a fejlődés jellemzésében, azonban ezek a legszélsőségesebb eredmények is előfordulásuk idején, így természetes összefüggés van a szélsőérték-elmélet és a rekord elmélet között. A modern atlétikában számos tényező szerepet játszhat egy új rekord beállítása közben. Mivel a sportolók leginkább az olimpiai évre hozzák magukat csúcsformába, ezért feltételezhetjük, hogy ebben az évben nagyobb a valószínűsége egy új rekord beállításának, de a maratoni futásnál egyéb más szempontokat is figyelembe lehetne venni például, hogy milyen erősen és irányból fúj a szél vagy, hogy mennyire van meleg. Mivel az olimpiát minden negyedik évben rendezik meg, ezért nem lenne célszerű az olimpiai aranyérmesek eredményeire támaszkodni a modellünk során, hiszen ez nem szolgál nekünk sok adattal. Ezért vizsgáljuk a leghíresebb maratonok legjobb eredményeit. Ilyen híres maraton a Berlieni, Fukuokai (Japán) vagy a Bostoni. A GEV számításokhoz az R programnyelvet használtam.

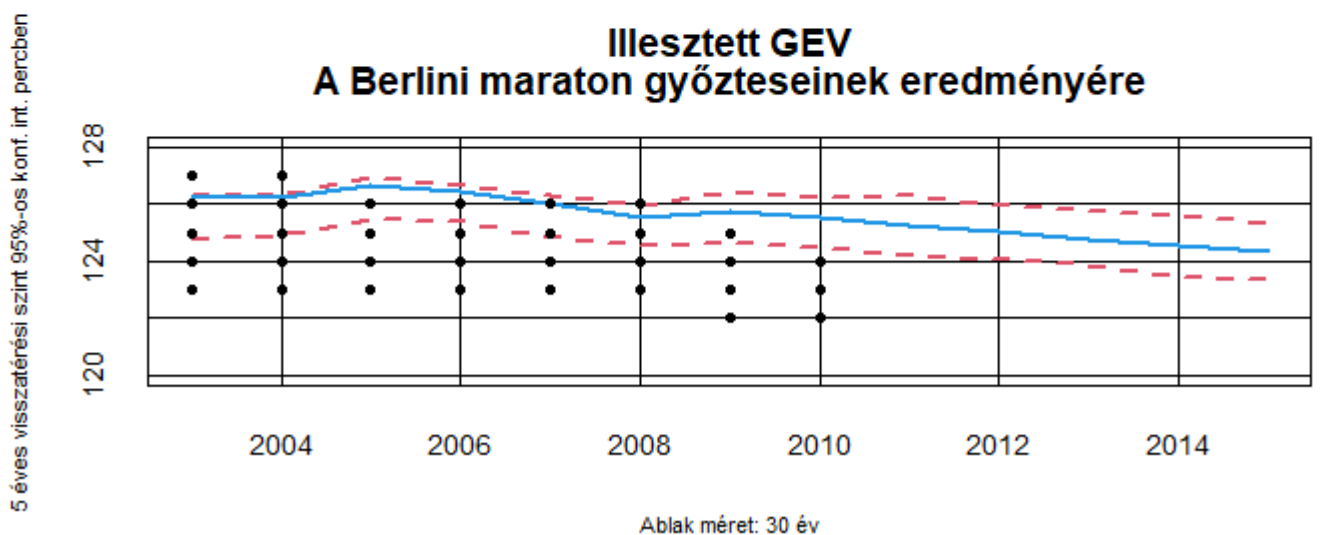
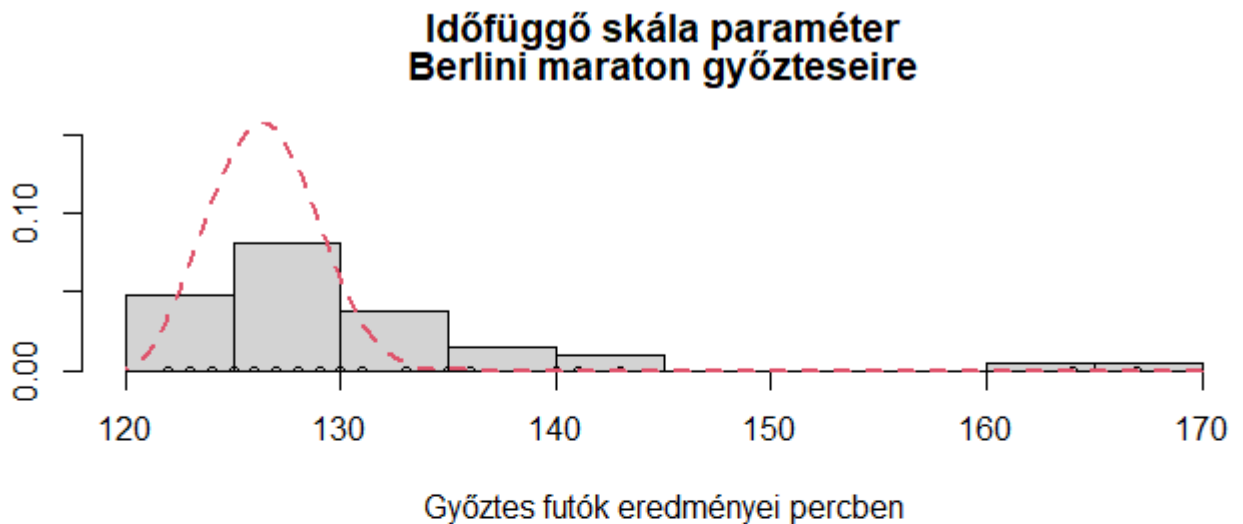
Az idősor szélsőértékeinek modellezésére jól használható a blokk-maximum módszer. A modellem során csak a győztesek eredményeit használom, így nem szükséges erre ezt a blokkos módszert alkalmazni, mert már megvannak a maximum értékeink. Általánosságban a stacionárius idősorokhoz jól használható a GEV-eloszlás. A paraméterek kiszámításánál figyelembe kell venni, hogy a győztes idők nem stacionáriusak.

Először a GEV eloszlás segítségével szeretnénk előrejelzést adni a jövőbeni győztes időkre. Mivel egy esetleges Berlieni maratonton elért új rekordot szeretnénk becsülni, és a modellünk maximumot számol, ezért az adatainkat  $-1$ -el kell szoroznunk, hogy az extrém érték keretébe kerüljenek. Ezzel a modellel arra a kérdésre kapunk választ, hogy mekkora a legnagyobb érték, ami várhatóan bekövetkezik egy adott időintervallumban.

Az első ábrán egy időfüggő hisztogram látható az időszak végére számolt sűrűségfüggvénnyel. Ez egy 4 paraméteres függvény lesz, aminek a paramétereit egy iteratív becslési függvény segítségével határozzuk meg. Minden becslésünkönél egy helyparaméter, lépték-

paraméter, alakparaméter és időparaméter meghatározása szükséges a modellünk felépítéséhez.

A második ábrán láthatjuk, hogy 95%-os konfidencia intervallummal számolva 5 éven belül a visszatérési szint majdnem eléri a 124 percet az utolsó 30 év eredményei alapján, ami azt jelenti, hogy a következő 5 éven belül várható közel 124 perces győztes eredmény.



A következő félévben érdemes lenne megvizsgálni más maratonok esetében a kapott értékeket, hogyan alakulnak, illetve megvizsgálni a kapcsolatot, akár más hosszútávú futószámok eredményei és a maraton eredményei alapján, többdimenziós modelleket alkalmazva.

# Irodalomjegyzék

- [1] Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. Modelling extremal events: for insurance and finance. Vol. 33. Springer Science Business Media, 2013.
- [2] [https://hu.wikipedia.org/wiki/Berlini\\_maraton](https://hu.wikipedia.org/wiki/Berlini_maraton)
- [3] Coles, Stuart, et al. An introduction to statistical modeling of extreme values. Vol. 208. London: Springer, 2001.