

Modelling sport results with extreme value methods

Csáfordi József András
Applied Mathematician MSc

Supervisor: Dr. Zempléni András

Contents

- 1 Previous work and changes
- 2 Generalized Pareto Distribution
- 3 Algorithm
- 4 Different plots and figures
- 5 Analysis
- 6 Goals during the next semesters

Previous work and changes

Previous work

- Berlin marathon
- Block maxima method
- Annual winning times
- Change in time
- It was possible to work with less data

Current work

- Boston marathon
- Threshold method
- The data should be below the limit
- Each year separately
- It requires more data

Generalized Pareto Distribution

- Generalized Pareto Distribution (GPD)
- Cumulative distribution function:

$$F(x) = 1 - \left(1 + \alpha \frac{x - \beta}{\gamma}\right)^{-1/\alpha},$$

where γ is the scale parameter, β is the location parameter, and α is the shape parameter

- The location parameter β in the formulas is set to 0
- The maximum estimator, where μ is the threshold :

$$\mu - \frac{\gamma}{\alpha}$$

Algorithm

- 1 Data processing and cleaning
- 2 For loop each year and each quantile
 - `gpd.fit` and `ad.test`
 - Upload parameter table
- 3 Appropriate quantile in a for loop
- 4 Diagnostic of this parameters
- 5 Different plots and analysis

Parameter table

	Year	Quantile(%)	Data size	Threshold	Estimation	Estimated Scale par	Estimated Shape par	P value
1	1975	50	927	-10918.16	-7738.516	1418.4108	-0.4456972	7.546909e-03
2	1975	51	908	-10884.00	-7781.782	1396.9488	-0.4434493	8.744642e-10
3	1990	60	2674	-12066.61	-7200.626	1372.0213	-0.2820389	1.074408e-05
4	1990	61	2608	-12019.00	-7530.687	1339.9466	-0.2750227	2.116735e-22
5	2000	80	2142	-11350.58	-7687.141	853.9196	-0.1826570	5.967873e-15
6	2000	81	2040	-11295.60	-7534.899	823.7002	-0.1722140	3.970602e-13
7	2019	92	1318	-10380.00	-7571.126	635.2455	-0.1331953	7.957052e-02
8	2019	93	1174	-10304.12	-7500.268	621.9485	-0.1292419	2.938754e-02

Diagnostic plot

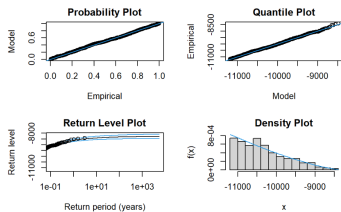


Figure: 1976 diag.

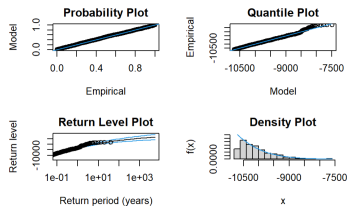
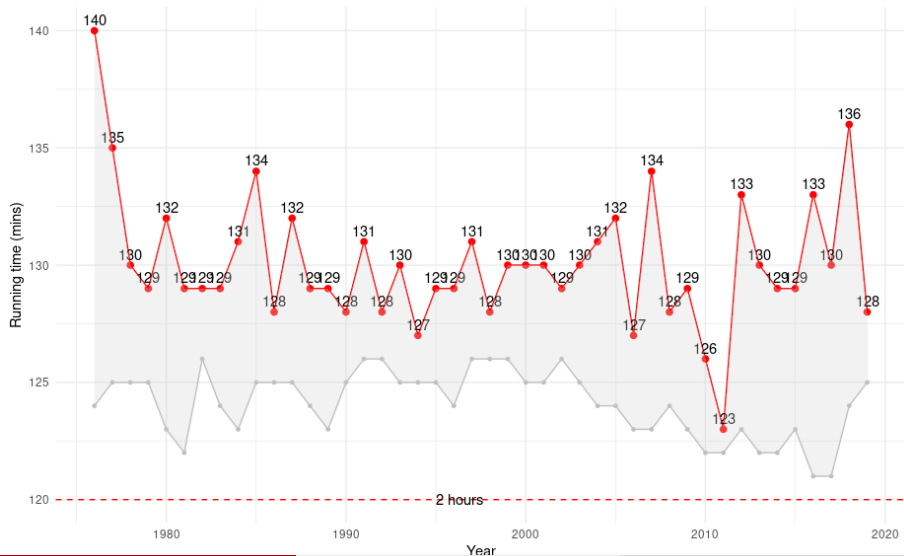


Figure: 2019 diag.

Plot

Best results for each years and the estimated range (male)



- Appropriate quantile from covariance matrix
- Sample generation from normal distribution with respect the 95% quantile
- Estimated range from the generated sample and the minimum of this range

Goals during the next semester

- Berlin - Boston relationship
- Multidimensional models:
 - Berlin-Boston marathon
 - Female-Male comparison
 - Boston with other races
- Additional variables (if possible)

Thank you for your attention!