

Predicting the Euro 2020 results using tournament rank probability scores from the `soccer` package

Claus Thorn Ekstrøm
UCPH Biostatistics
ekstrom@sund.ku.dk

June 19th, 2020
 @ClausEkstrom



Predicting the Euro ~~2020~~²⁰²¹ results using tournament rank probability scores from the socceR package

Claus Thorn Ekstrøm
UCPH Biostatistics
ekstrom@sund.ku.dk

June 19th, 2020
 @ClausEkstrom



EURO 2020 prediction competition rules

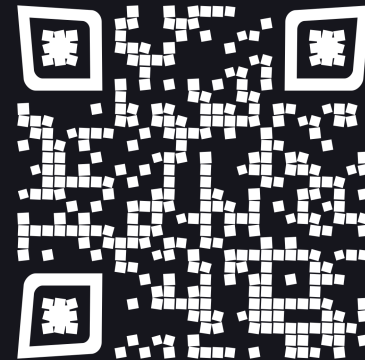
Who will be best at predicting the EURO 2020 tournament result?

You should provide

- A 7×24 matrix. Columns are teams, rows are ranks: 1st, 2nd, 3rd, 4th, 5th-8th, 9th-16th, 17th-24th.
- Deadline: **Sunday June 6th, 2021**

For full info + upload prediction:

github.com/ekstroem/euro2020



Example prediction

TEAMS (24 COLUMNS)

RANKS

##		Austria	Belgium	Croatia	Czech Rep.	Denmark	England	Finland	France	G
##	[1,]	0.00004	0.00336	0.00000	0.00546	0.00000	0.00000	0.01103	0.11020	0
##	[2,]	0.00036	0.00971	0.00001	0.01311	0.00000	0.00003	0.02964	0.07291	0
##	[3,]	0.00100	0.01756	0.00001	0.02110	0.00003	0.00004	0.03935	0.14017	0
##	[4,]	0.00545	0.04279	0.00009	0.04519	0.00040	0.00040	0.08714	0.08193	0
##	[5,]	0.04448	0.21546	0.00277	0.23179	0.00964	0.01689	0.40411	0.33635	0
##	[6,]	0.21338	0.55319	0.02717	0.54652	0.03778	0.06732	0.31851	0.23613	0
##	[7,]	0.73529	0.15793	0.96995	0.13683	0.95215	0.91532	0.11022	0.02231	0
##		Italy	Netherlands	Play-off A						
##	[1,]	0.00111	0.09852	0.00014						
##	[2,]	0.00385	0.12623	0.00078						
##	[3,]	0.00835	0.14964	0.00165						
##	[4,]	0.03284	0.13852	0.01101						
##	[5,]	0.10272	0.14247	0.04741						
##	[6,]	0.38553	0.32957	0.21897						
##	[7,]	0.46560	0.01505	0.72004						

PROBABILITIES: DENMARK

WILL END IN THE GROUP
STAGE WITH PROB 95%

COLUMNS MUST SUM TO 1

ROWS MUST SUM TO #TEAMS

Find winner: tournament rank probability score

T teams, R rank groups, X prediction matrix, O result indicator matrix

$$\text{TRPS}(O, X) = \frac{1}{T} \sum_{t=1}^T \frac{1}{R-1} \sum_{r=1}^{R-1} (\mathcal{O}_{rt} - \mathcal{X}_{rt})^2.$$

$\mathcal{O}$ and $\mathcal{X}$ are *cumulative* versions of O and X . **Smaller number = better**

- Cumulative differences (obtain at least rank r)
- Take **all** predictions into account. Not just the winner.
- Non-local and **sensitive to distance**
- Penalize **confident** classifications that are **incorrect**.
- See paper <https://arxiv.org/abs/1912.07364>

Examples

$$X^1 = \begin{bmatrix} 0.75 & 0.25 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 \\ 0 & 0 & 0.75 & 0.25 \\ 0 & 0 & 0.25 & 0.75 \end{bmatrix} \quad X^2 = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}$$

Tournament outcome: 1, 2, 3, 4: TRPS becomes 0.0208 and 0.2083.
(Log-loss = 1.15 and 5.55)

Tournament outcome: 2, 1, 4, 3: TRPS becomes 0.188 and 0.2083
(Log-loss = 5.55 and 5.55)

```
# install.packages("socceR") # To install package
library("socceR")
m1 <- matrix(c(1, 0, 0, 0, 0, 1, 0, 0, 0, 0, .5, .5, 0, 0, .5, .5), 4)
m1 # Prediction where certain on the top 2 ranks
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    0 0.0  0.0
## [2,]    0    1 0.0  0.0
## [3,]    0    0 0.5  0.5
## [4,]    0    0 0.5  0.5
```

```
# Compare prediction matrix and final ranks
# columns in m1 must match length of vector (containing ranks)
trps(m1, c(1, 2, 3, 4))
```

```
## [1] 0.04166667
```

Optimize weights

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    0 0.0  0.0
## [2,]    0    1 0.0  0.0
## [3,]    0    0 0.5  0.5
## [4,]    0    0 0.5  0.5
```

```
##      [,1] [,2] [,3] [,4]
## [1,] 0.5  0.5    0    0
## [2,] 0.5  0.5    0    0
## [3,] 0.0  0.0    1    0
## [4,] 0.0  0.0    0    1
```

```
##      [,1] [,2] [,3] [,4]
## [1,] 0.5  0.5  0.0  0.0
## [2,] 0.5  0.5  0.0  0.0
## [3,] 0.0  0.0  0.5  0.5
## [4,] 0.0  0.0  0.5  0.5
```

```
##      [,1] [,2] [,3] [,4]
## [1,] 0.25 0.25 0.25 0.25
## [2,] 0.25 0.25 0.25 0.25
## [3,] 0.25 0.25 0.25 0.25
## [4,] 0.25 0.25 0.25 0.25
```

```
optimize_weights(list(m1, m2, m3, m4), 1:4)
```

```
## [1] 5.000451e-01 4.999549e-01 5.405964e-09 2.103913e-22
```

Let's find the real winner!



Deadline: **Sunday June 6th, 2021**

github.com/ekstroem/euro2020

Lightning session at eRum 2022?

