Fitting a simulation model for Fifa World Cup 2010

José Mario Martínez

www.ime.unicamp.br/∼martinez

Department of Applied Mathematics, University of Campinas, Brazil. This is a joint work with Julián Martínez, Department of Chemical and Food Engineering, Federal University of Santa Catarina, Brazil

August 2, 2011

Outline

1. Simulation model for Fifa 2010 is introduced.
2. Output of simulation: Probabilities of "everything".
3. Parameters: Probabilities of individual games.
4. Goal: fitting the model (individual probabilities) to global beliefs on the final results of the Cup.
5. Tools: Derivative-free optimization. Bobyqa and others.

Native probabilities

To each individual game (A versus B) a probability is given for the victory of A, victory of B or tie (in the first phase).
Native probabilities are computed according to the published FIFA ranking of all the national teams.
If the difference in the FIFA ranking is the maximal one (31) the native probability of victory of the best of the two teams is 0.98. If it is the minimal one (1) the native probabilit of victory of the best is 0.35.
Probabilities of ties are computed using a secret formula.

Simulation with native probabilities

Output of Simulation using native (pairwise) probabilities:

	Champion	Vice	Third	Fourth	5th-8th	9th-16th	Not-class
HOLAN	14.2	10.8	8.3	5.2	31.1	22.4	8.0
ESPAN	14.1	8.9	7.3	3.8	21.4	37.6	6.8
BRAS	13.1	9.4	7.2	4.2	21.9	30.1	14.0
ITAL	10.4	8.5	7.3	4.6	30.0	36.5	2.6
ALEM	8.4	8.1	10.6	7.8	23.1	30.8	11.0
PORT	7.6	7.2	5.5	4.1	20.1	33.1	22.3
FRAN	6.3	6.7	8.2	7.0	19.1	30.9	21.9
ARGEN	5.7	6.4	8.2	7.3	21.2	34.5	16.7
INGL	4.9	6.1	7.5	7.4	22.2	32.1	19.7

Motivation for fitting pairwise probabilities

Brazilian people “know” that the probability of Brazil being Champion is around 0.95, but, at the same time, they believe that the probability of beating Argentina is not bigger than 0.50. When Brazil fails to win the Cup, brazilians atribute the fact to lack of patriotism, determination, enthusiasm, courage and other virtues that they suppose are very common among argentine players. (They don’t realise that Argentina rarely wins a World Cup.)

The question addressed here is: Is it possible to obtain reasonable modifications of pairwise probabilities in order to get “desired” global results?

Reasonable modifications of pairwise probabilities?

Assume that the native probabilities of Brazil beating the other teams are: $a_1, \ldots, a_m$.
Let $t_{Bra} \in [-1, 1]$ a control parameter (to be adjusted). The modified probabilities of Brazil beating team i under the control t will be:

$$a_i + t_{Bra}(1 - a_i) \quad \text{if} \quad t \geq 1$$

and

$$a_i + t_{Bra}(-a_i) \quad \text{if} \quad t \leq 1.$$

The parameter t_{Bra} will be called *Enhance parameter* associated with Brazil.

Fitting Enhance parameters

We will use, as unknowns, the Enhance parameters relative to Brazil, Argentina, Spain and Holland.
The user will give, for example, the probabilities of Brazil, Argentina, Spain and Holland winning the Cup.
The optimization code will find (if possible) the enhance parameters that fit these given probabilities.

Type of optimization problem

Each function evaluation involves running the simulation model "many" times in order to compute the model probabilities.

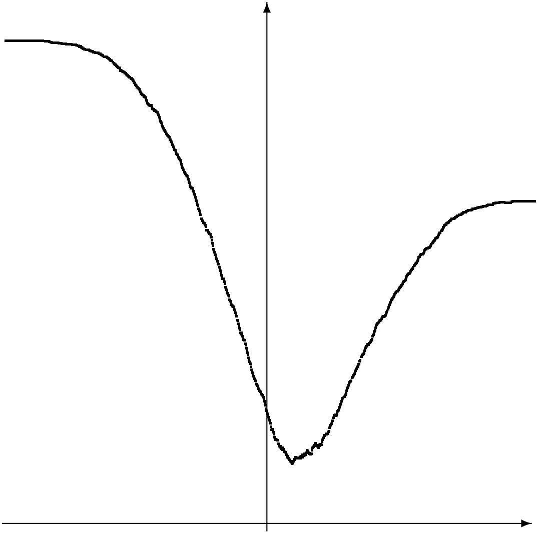
We don't have the derivatives of this objective function, therefore this is a genuine non-derivative problem, for which it is mandatory to use a derivative free method.

Bounds on the variables: Enhance parameters are between -1 and 1.

We use, as objective function, the sum of absolute values of differences between model probabilities and given probabilities.

The objective function is "quasi-convex with small oscilations". It is constant near the boundary of the box. (Typical: constant 60 from -1 to -0.85. Constant 40 from 0.91 to 1. Minimum 7.3 in $t = 0.08$.

Profile of the objective function



Algorithm

We use Bobyqa, Fortran subroutine released by M. J. D. Powell in 2009.

Bobyqa uses successive interpolating quadratics using selected past iterates.

At each iteration Bobyqa minimizes the interpolated quadratic in the box.

No convergence theorem has been proved yet.

We also tried the classical method of Nelder and Mead.

Example 1 (15, 15, 15, 15)

At the return from BOBYQA Number of function values = 63
Least value of F = 3.000001907348633D-01 The corresponding X is: 8.064884D-02 1.810350D-01 1.056468D-01 4.124307D-02
Results:
Enhance parameter for Brazil: 8.06488350820740302E-002
Enhance parameter for Argentina: 0.18103502507393154
Enhance parameter for Spain: 0.10564680709879409
Enhance parameter for Holland: 4.12430744679619515E-002
100 x Probability(Brazil Champion) = 14.899999618530273
100 x Probability(Argentina Champion) = 15.000000000000000
100 x Probability(Spain Champion) = 15.199999809265137
100 x Probability(Holland Champion) = 15.000000000000000

Example 2 (95, 0, 0, 0)

At the return from BOBYQA Number of function values = 70
Least value of F = 2.049999952316284D+00 The corresponding X is: 1.000000D+00 -4.754130D-01 -5.417783D-01 -6.631729D-01
Results:
Enhance parameter for Brazil: 1.0000000000000000
Enhance parameter for Argentina: -0.47541297198836702
Enhance parameter for Spain: -0.54177832823479755
Enhance parameter for Holland: -0.66317294247330338
100 x Probability(Brazil Champion) = 95.000000000000000
100 x Probability(Argentina Champion) = 0.25000000000000000
100 x Probability(Spain Champion) = 0.50000000000000000
100 x Probability(Holland Champion) = 0.20000000298023224

Example 3 (50, 50, 0, 0)

At the return from BOBYQA Number of function values = 73
Least value of F = 4.249997615814209D+00 The corresponding X is: 1.000000D+00 9.866129D-01 -6.511769D-02 -6.654745D-01

Results:
Enhance parameter for Brazil: 1.0000000000000000
Enhance parameter for Argentina: 0.98661287145966914
Enhance parameter for Spain: -6.5117686138675773E-002
Enhance parameter for Holland: -0.66547445976551878
100 x Probability(Brazil Champion) = 48.700000762939453
100 x Probability(Argentina Champion) = 49.150001525878906
100 x Probability(Spain Champion) = 2.0999999046325684
100 x Probability(Holland Champion) = 1.0000000317107685E-030

Conclusions

1. We introduced a simulation model for the Fifa World Cup that may be used for fun, or for realising that common sense does not deal well with probabilities.
2. Many constrained and unconstrained genuine non-derivative problems in which the objective function is the simulation may be defined, providing sensible tests for derivative-free optimization.
3. In all the problems Nelder-Mead uses twice the number of evaluations of Bobyqa.