An implementation of: A Simple Random Walk Model for Predicting Track and Field World Records[1]

Aritz Lizoain

Computational statistics

Université de Neuchâtel

Aim

Build a **prediction model** for the men's 100m world record.

Applications:

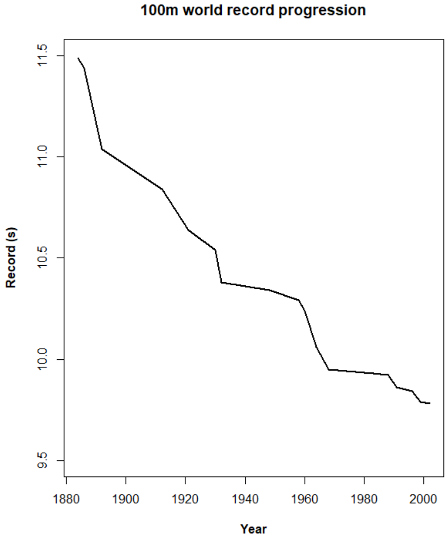
- Estimate human performance limits. How fast can humans run?
- Help coaches and athletes setting up goals, training plans, diets
- Potentially identify athletes using performance enhancing drugs

Model (1/5)

- The record can either decrease or stay the same
- Time series approach is unusable (non-stationary process)
- **One-sided random walk** model

$$X_t = X_{t-1} - \epsilon_t$$

- $\epsilon_t \geq 0$ and independent



Model (2/5)

The **error term distribution** is a mixture of a discrete and continuous RVs

$$f_t(\epsilon_t) = (1 - p(t))\delta_0(\epsilon_t) + p(t)h_t(\epsilon_t)$$

where $\delta_0(\epsilon_t)$ is a degenerate distribution at zero (i.e. $\mathrm{P}(\epsilon_t = 0) = 1$)

Probability that a record is broken in year $t \sim Bernoulli(p(t))$

$$p(t) = \frac{exp(\beta_{11}+\beta_{12}\cdot t)}{1+exp(\beta_{11}+\beta_{12}\cdot t)}$$

Amount by which a record is broken $\sim Exp(\lambda(t))$

$$\lambda(t) = exp(-\beta_{21} - \beta_{22} \cdot t)$$

Model (3/5)

The **error term distribution** is a mixture of a discrete and continuous RVs

$$f_t(\epsilon_t) = (1 - p(t))\delta_0(\epsilon_t) + p(t)h_t(\epsilon_t)$$

where $\delta_0(\epsilon_t)$ is a degenerate distribution at zero (i.e. $\mathrm{P}(\epsilon_t = 0) = 1$)

Probability that a record is broken in year $t \sim Bernoulli(p(t))$

$$p(t) = \frac{exp(\beta_{11}+\beta_{12}\cdot t)}{1+exp(\beta_{11}+\beta_{12}\cdot t)}$$

Amount by which a record is broken $\sim Exp(\lambda(t))$

$$\lambda(t) = exp(-\beta_{21} - \beta_{22} \cdot t)$$

Model (4/5)

Probability that a record is broken in year $t \sim$ $Bernoulli(p(t))$

$$p(t) = \frac{exp(\beta_{11}+\beta_{12}\cdot t)}{1+exp(\beta_{11}+\beta_{12}\cdot t)}$$

$\boldsymbol{\beta}_1$ estimated via logistic regression

Amount by which a record is broken $\sim Exp(\lambda(t))$

$$\lambda(t) = exp(-\beta_{21} - \beta_{22} \cdot t)$$

$\boldsymbol{\beta}_2$ estimated via max. likelihood[2]

$$L(\boldsymbol{\beta}_2) = \prod_{t=1}^{n}[\lambda exp(-\lambda\epsilon_t)]^{I(\epsilon_t>0)}$$

Parameters	$\hat{\beta}_{21}$	$\hat{\beta}_{12}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
Paper	-2.68	1.45	-1.02	-2.18
Obtained	-2.33	1.04	-1.00	-2.21

Model (5/5)

Probability that a record is broken in year $t \sim Bernoulli(p(t))$

$$p(t) = \frac{exp(\beta_{11}+\beta_{12}\cdot t)}{1+exp(\beta_{11}+\beta_{12}\cdot t)}$$

$\boldsymbol{\beta}_1$ estimated via logistic regression

Amount by which a record is broken $\sim Exp(\lambda(t))$

$$\lambda(t) = exp(-\beta_{21} - \beta_{22} \cdot t)$$

$\boldsymbol{\beta}_2$ estimated via max. likelihood[2]

$$L(\boldsymbol{\beta}_2) = \prod_{t=1}^{n}[\lambda exp(-\lambda\epsilon_t)]^{I(\epsilon_t>0)}$$

Parameters	$\hat{\beta}_{21}$	$\hat{\beta}_{12}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
Paper	-2.68	1.45	-1.02	-2.18
Obtained	-2.33	1.04	-1.00	-2.21

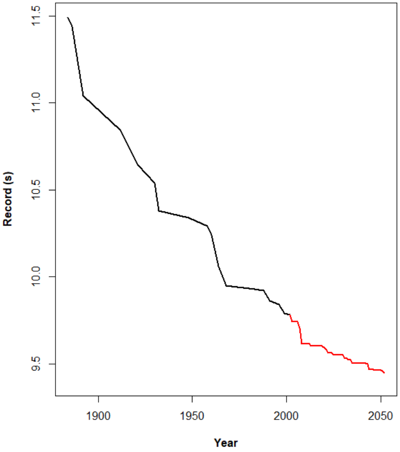
Monte Carlo simulation algorithm

1. Randomly generate
   $a_t \sim Exp(\hat{\lambda}(t))$
   $b_t \sim Bernoulli(\hat{p}(t))$

2. $\epsilon_t = a_t \cdot b_t$

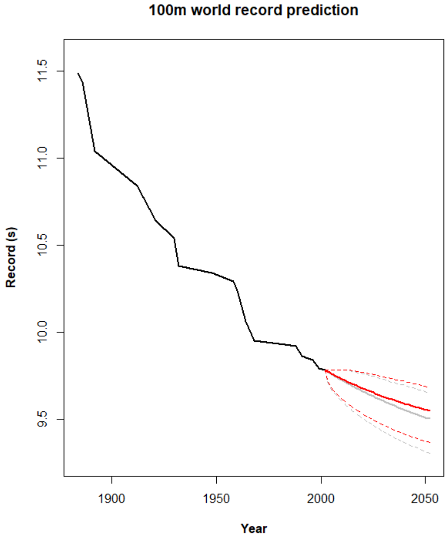
3. $X_t = X_{t-1} - \epsilon_t$

4. Repeat steps 1-3 B times



Predictions - 50 years

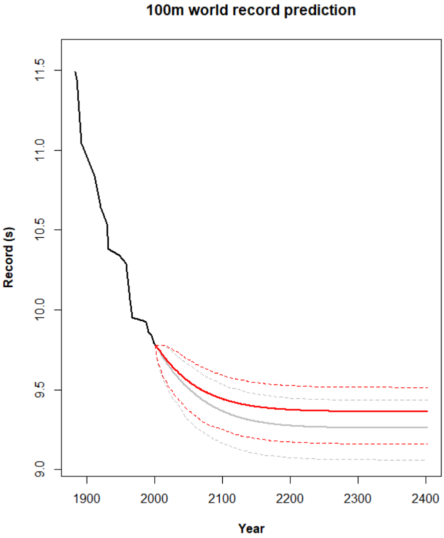
- B = 100,000
- The **mean** for each year is used **as a point estimate**
- **95% confidence intervals** are constructed taking the **2.5% and 97.5% percentiles**



Predictions - 400 years

How fast can humans run?

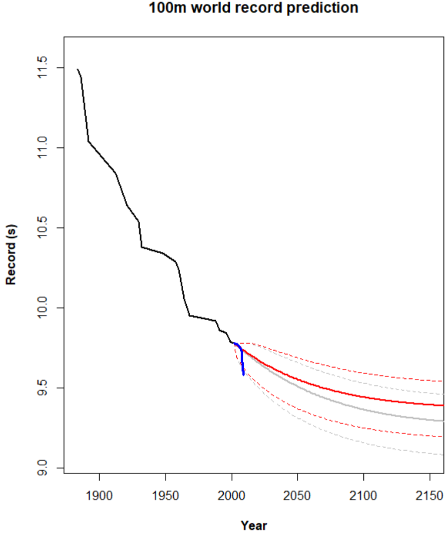
X_t converges to 9.36(9.26)s



World records after 2002 (1/2)

6 new world records since 2002[3]

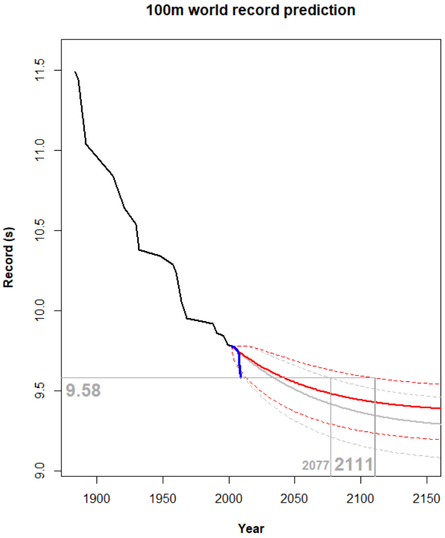
- The last record of 9.58s (2009) is out of the prediction intervals
- It will not be broken until 2111(2077)!



World records after 2002 (2/2)

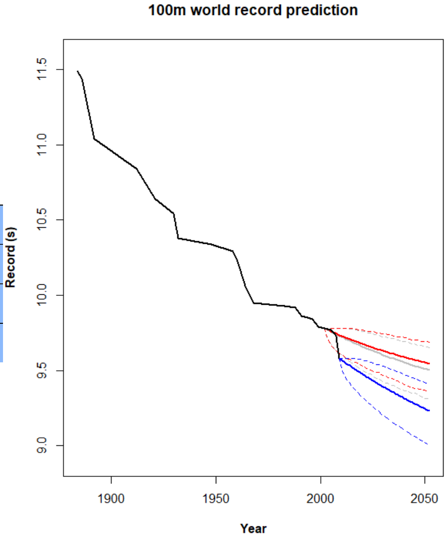
6 new world records since 2002[3]

- The last record of 9.58s (2009) is out of the prediction intervals
- It will not be broken until 2111(2077)!



Updated model predictions - 50 years

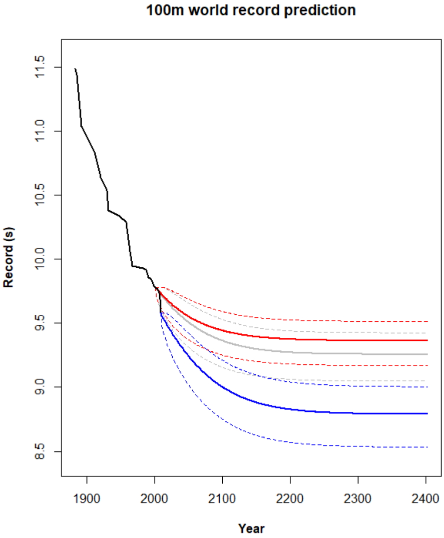
	Paper	Obtained	Updated
$\hat{\beta}_{11}$	-2.68	-2.33	-2.59
$\hat{\beta}_{12}$	1.45	1.04	1.56
$\hat{\beta}_{21}$	-1.02	-1.00	-1.18
$\hat{\beta}_{22}$	-2.18	-2.22	-1.90



Updated model predictions - 400 years

How fast can humans run?

Paper	Obtained	Updated
9.26s	9.36s	8.79s



References

[1] J. T. Terpstra and N. D. Schauer, “A simple random walk model for predicting track and field world records,” *Journal of Quantitative Analysis in Sports*, vol. 3, no. 3, 2007. DOI: doi:10.2202/1559-0410.1067. [Online]. Available: https://doi.org/10.2202/1559-0410.1067.

[2] M. Gesmann, *How to use optim in r*, 2013. [Online]. Available: https://magesblog.com/post/2013-03-12-how-to-use-optim-in-r/.

[3] I. A. of Athletics Federations, *Progression of IAAF World Records, 2015 Edition*, I. Hymans Richard; Matrahazi, Ed. 2015 (Retrieved February 24, 2018), p. 33, http://iaaf-ebooks.s3.amazonaws.com/2015/Progression-of-IAAF-World-Records-2015/projet/IAAF-WRPB-2015.pdf.

Summary

- One-sided random walk

$$X_t = X_{t-1} - \epsilon_t$$

- ϵ_t mixed distribution
  discrete + continuous
- MC simulations to obtain predictions and CI
- Human performance limits are estimated
- **A new world record can significantly change the model predictions**