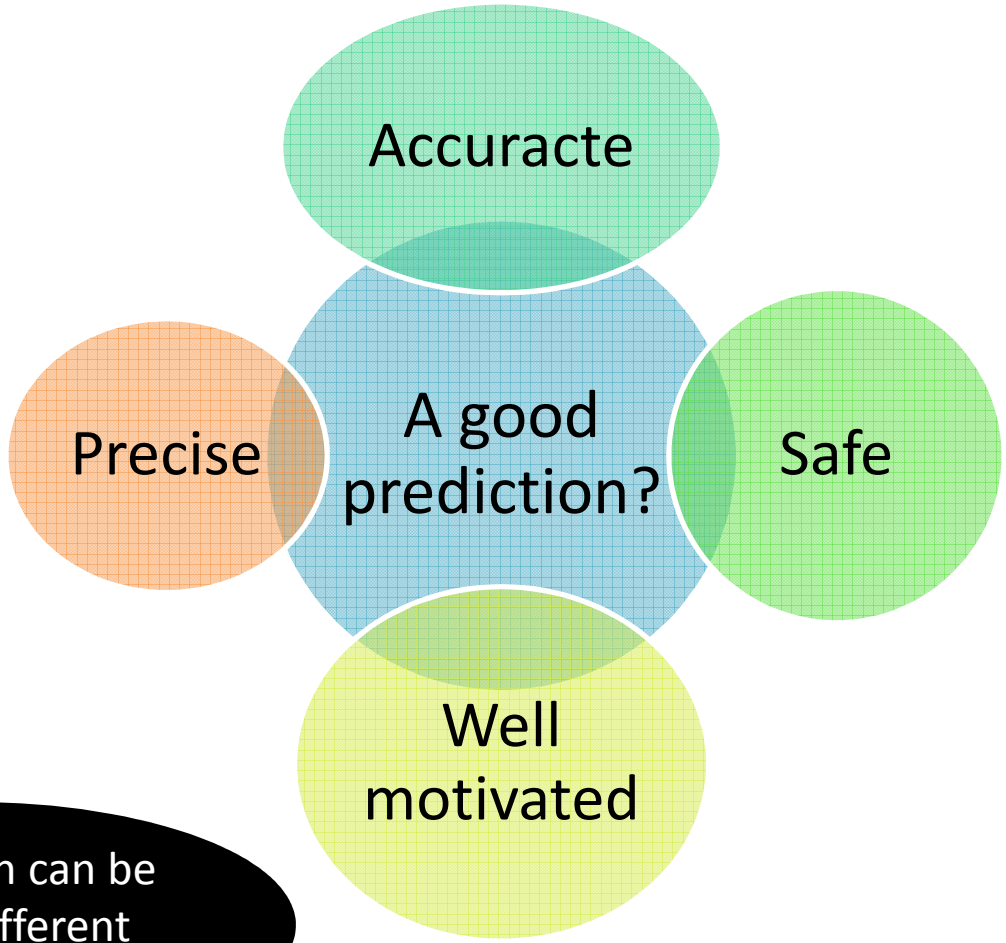
Lundaloppet Predictive Challenge

Ullrika Sahlin PhD

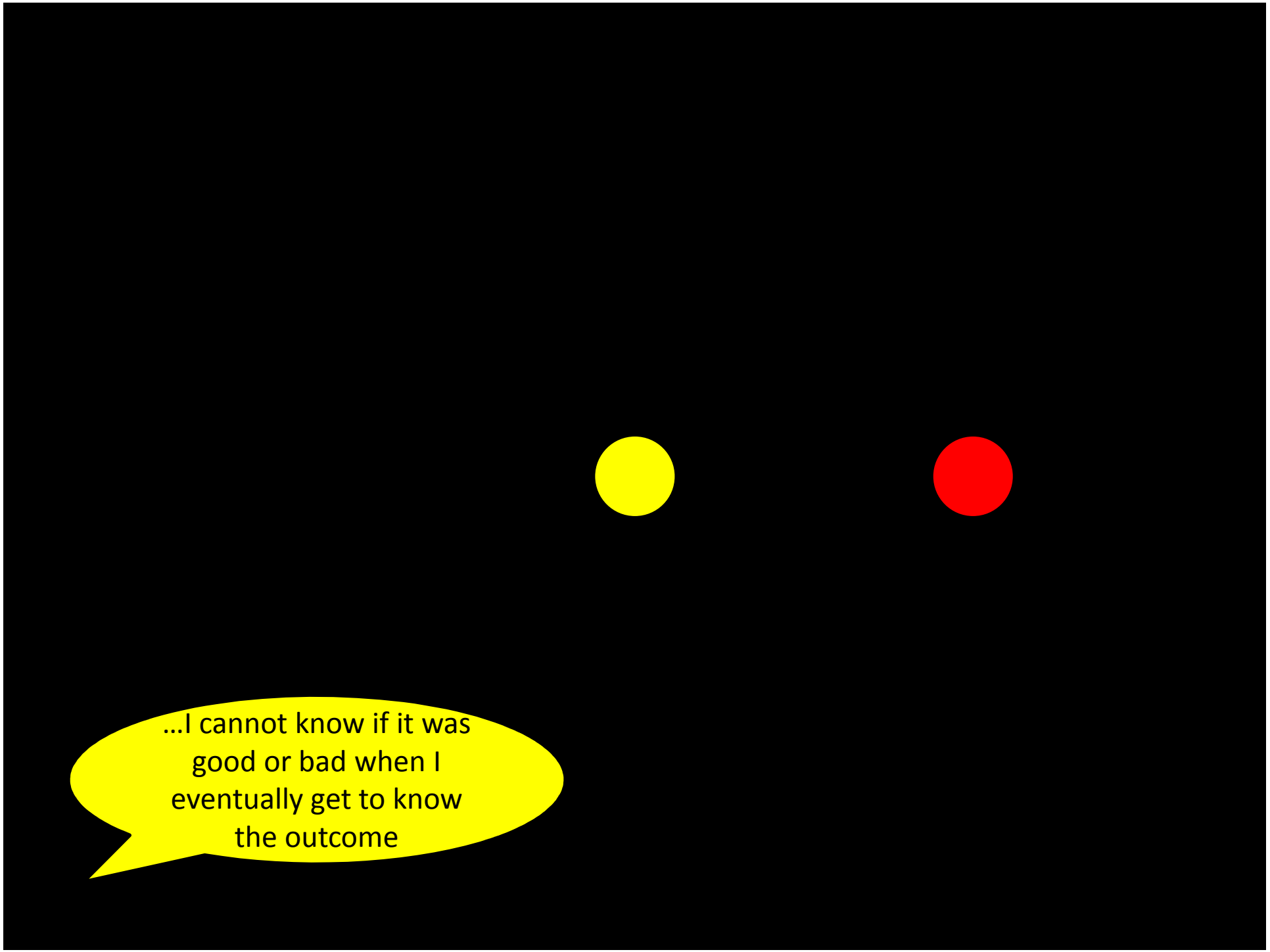
Lund University Centre of Environmental and Climate Research

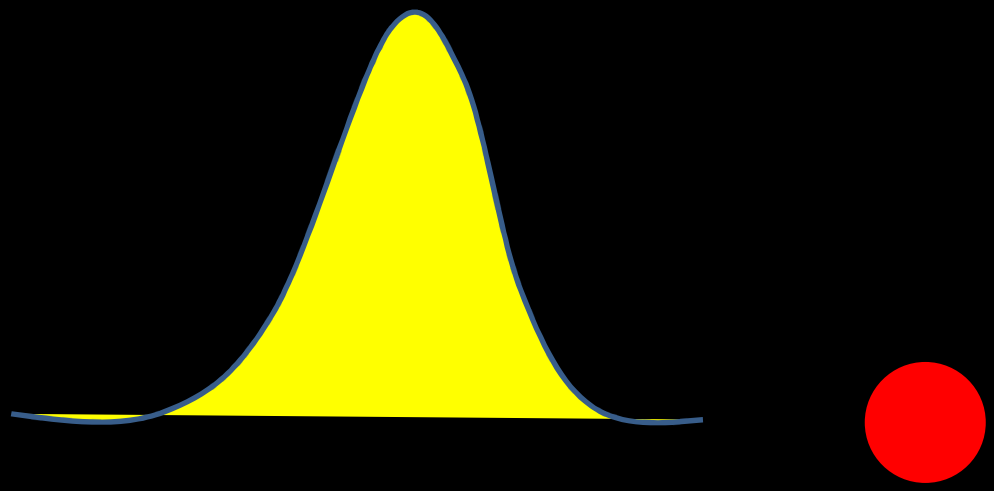
May 2014



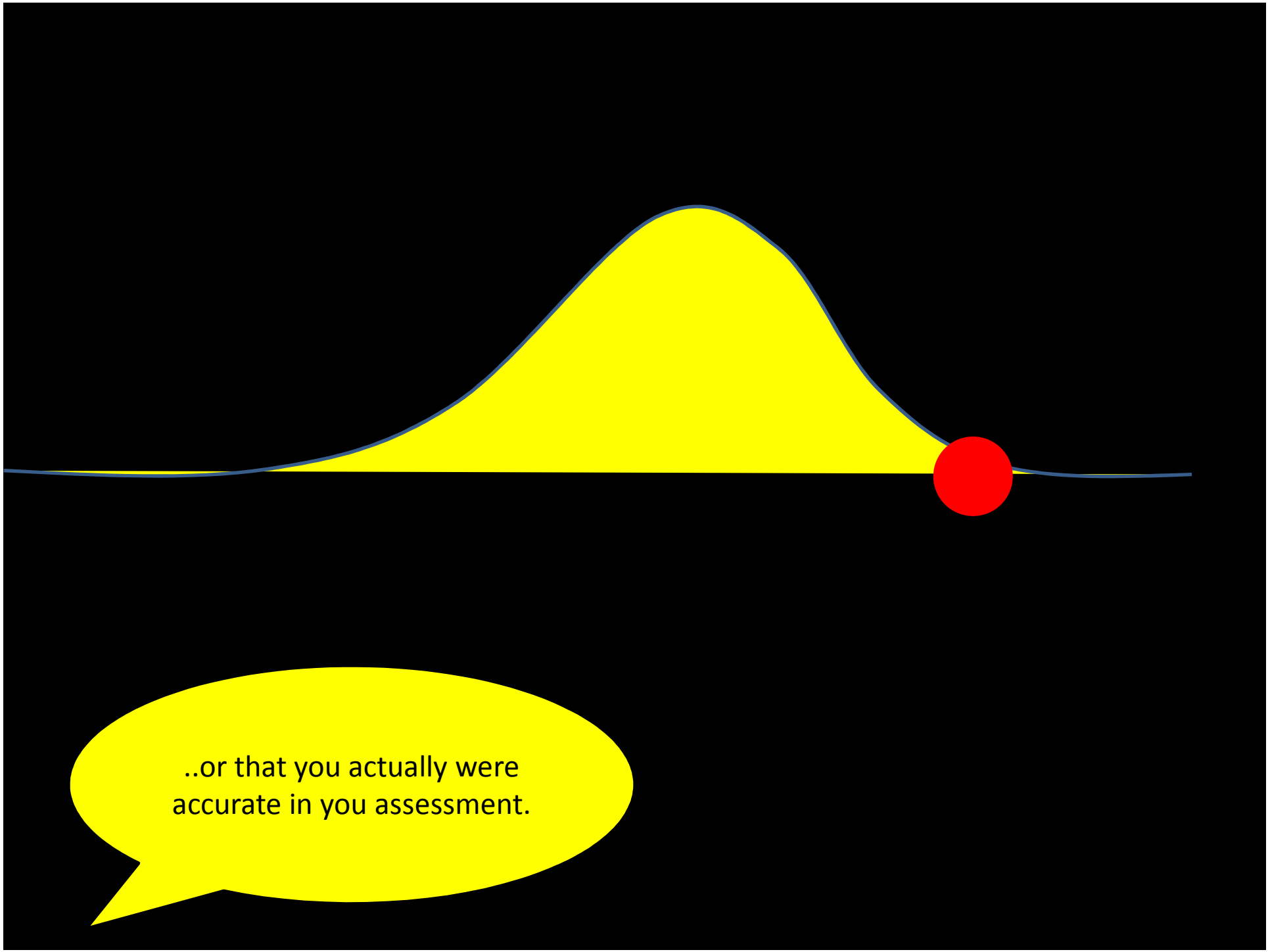
A prediction can be good in different ways

If you make a prediction as a single value…


…I cannot know if it was good or bad when I eventually get to know the outcome



By stating the uncertanty in you prediction I can now that it missed including the outcome that acutally ocurred..


..or that you actually were
accurate in you assessment.

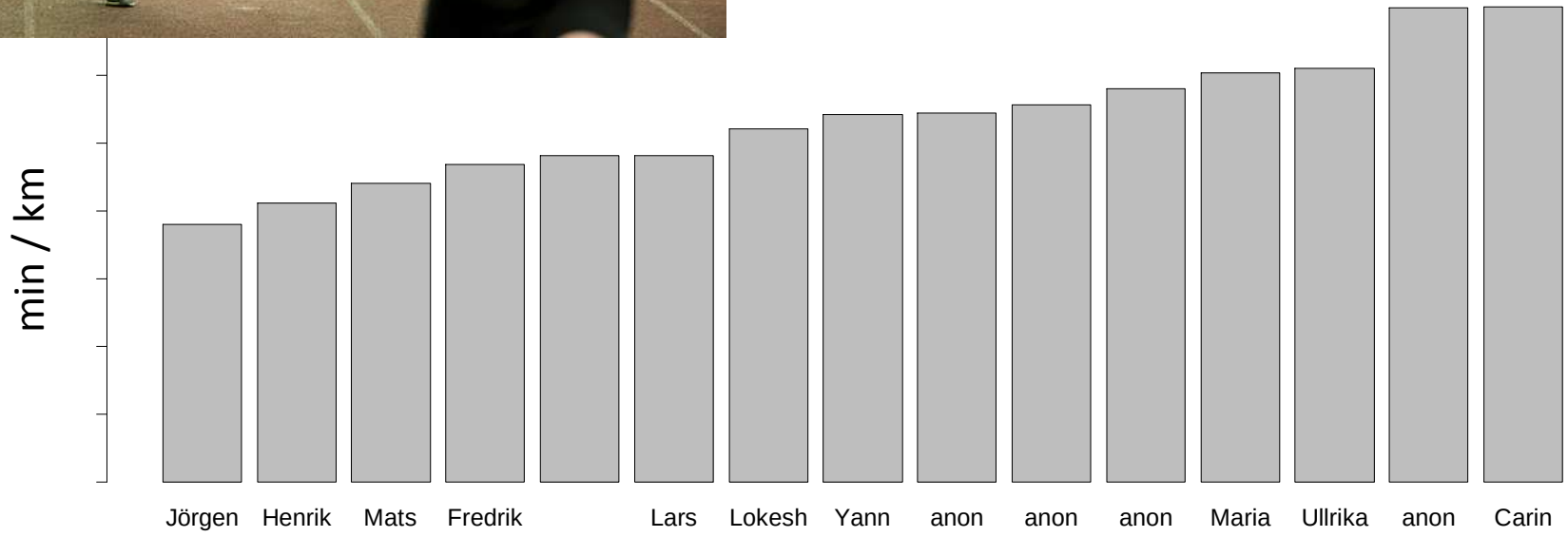
The participants were given this task:

You can use any kind of way to express your "guesstimate". For example

- an interval (e.g. between 20 and 30 minutes with 90% confidence),
- a Normal distribution (e.g. time will be around a mean of 25 with a standard deviation of 5),
- a sample of times that you think are possible (e.g. 20, 15,22, 30) or
- (for those of you who are unsure if you will complete the race) a mixing distribution (say there is a 20% chance that I will not take part and if I do, I will run for between 15 and 25 minutes).

Results

17 respondents



The fastest runner is not the winner in this challenge…

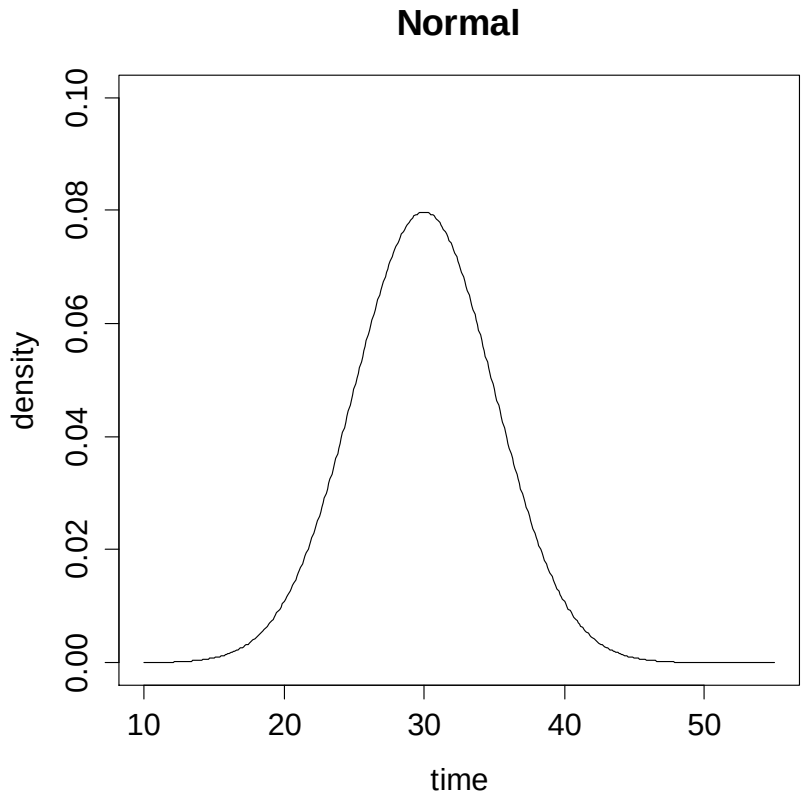
Results in the predictive challenge

- Accurate or not accurate prediction
- Most accurate prediction
- Most precise prediction
- Most safe prediction
- Most pessimistic and optimistic predictions
- Most unexpected failure

…instead there are different ways to win.

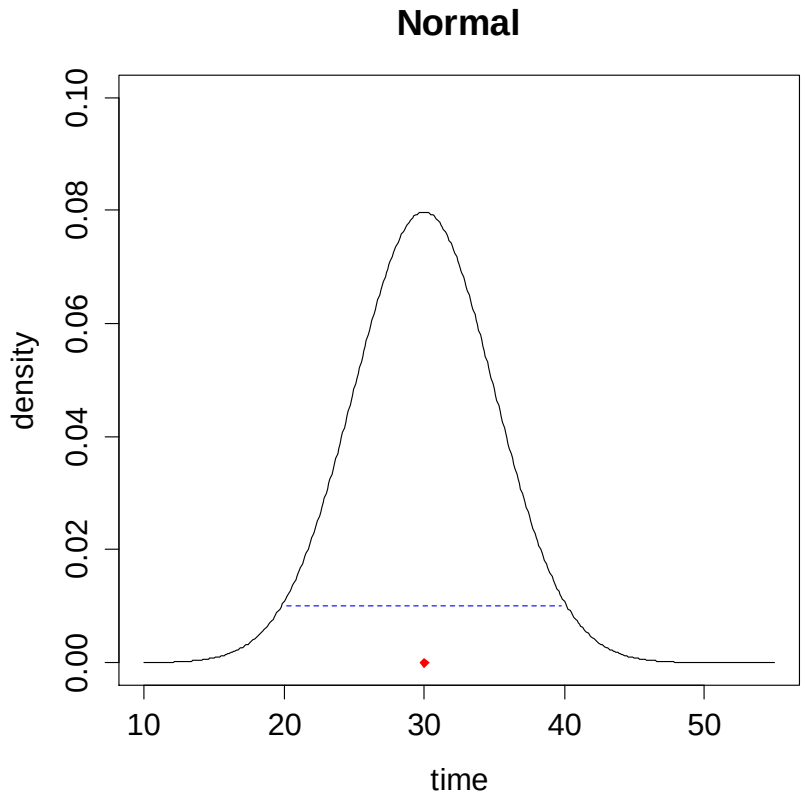
Method

- Time to run is a continous variable time >= 0
- Distribution determined by its density function f(time)
- Expected value
- Confidence interval
- Likelihood



Method

- Time to run is a continous variable time >= 0
- Distribution determined by its density function f(time)
- Expected value
- Confidence intervals
- Likelihood



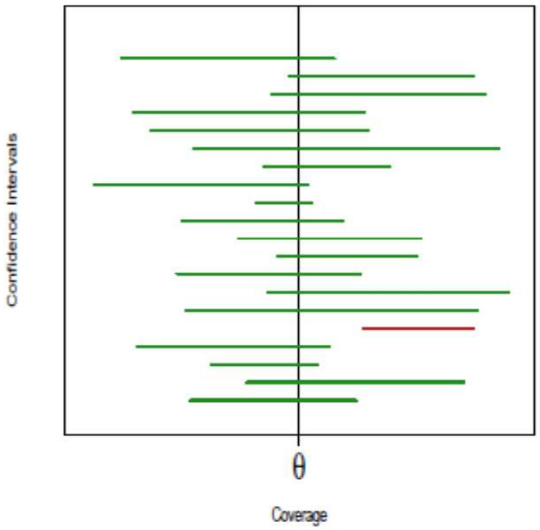
Different interpretations of the confidence interval

Frequentist:

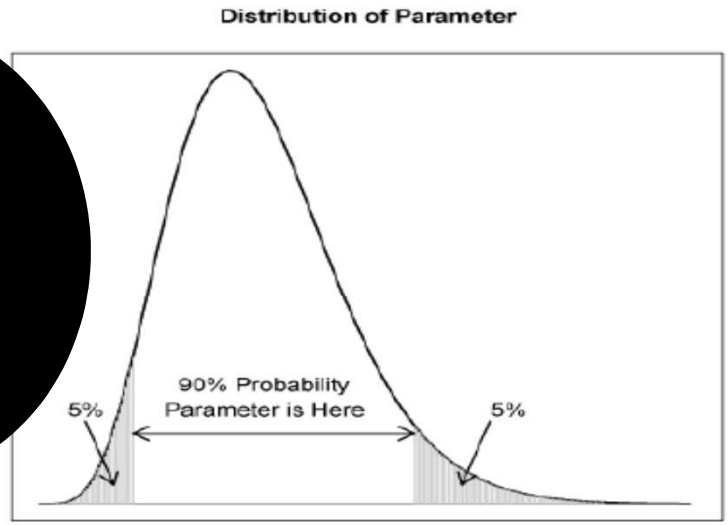
In repeated sampling 90% of the derived intervals will cover the true parameter value

Bayesian:

With these data, the parameter value is inside the interval with 90% probability



Which one of these interpretations do you prefer?

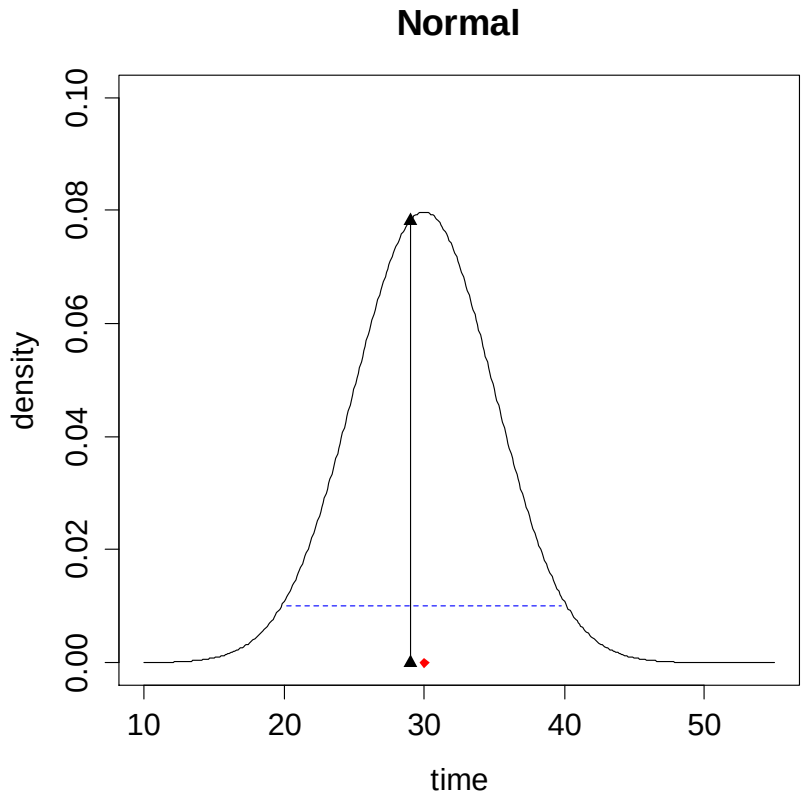


Stolen from Casella
http://www.stat.ufl.edu/archived/casella/Talks/BayesRefresher.pdf

Method

Here are the measuring instruments I used to describe predictions and how good they were

- Time to run is a continous variable time >= 0
- Distribution determined by its density function f(time)
- Expected value
- Confidence intervals
- Likelihood



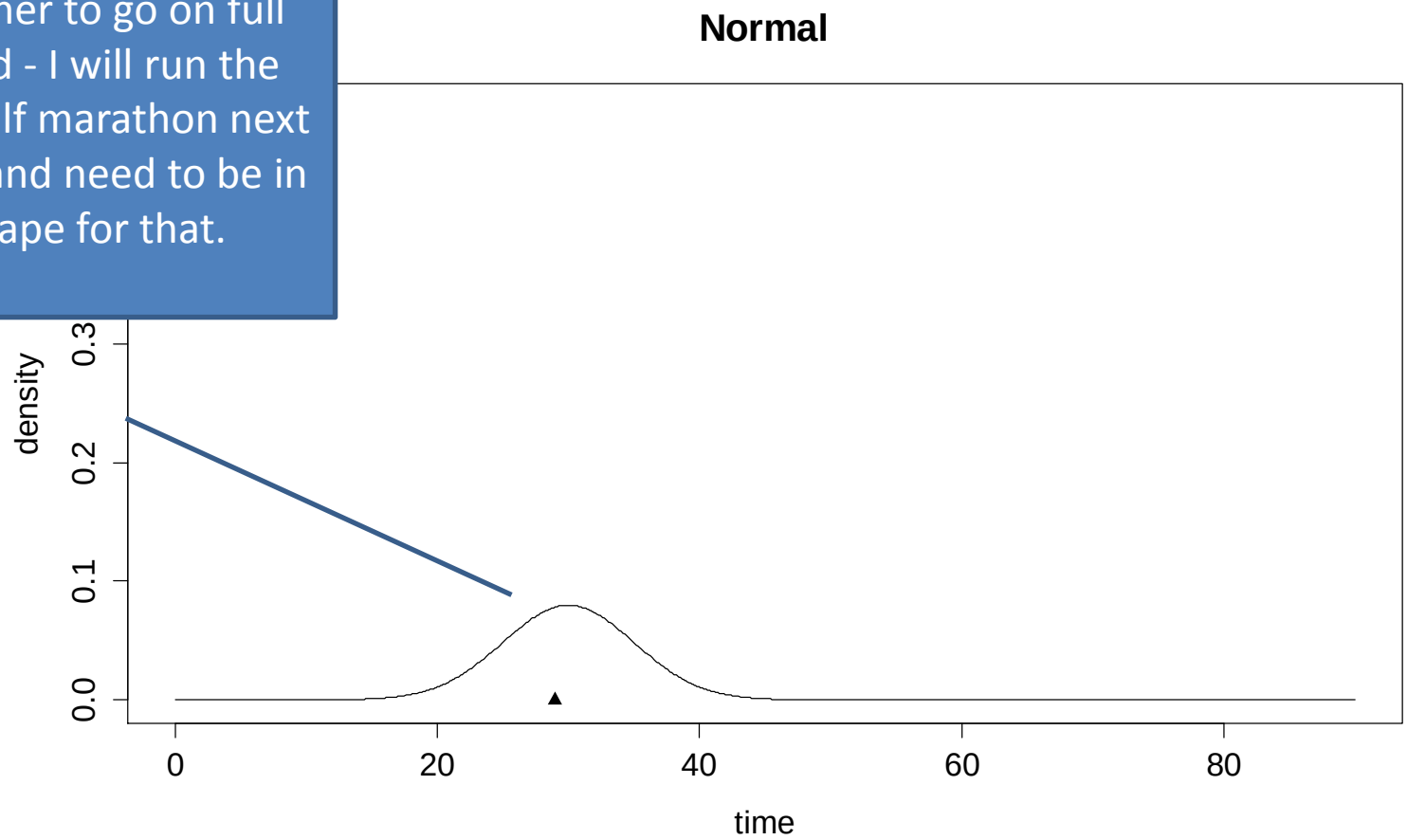
Results

These are the rules set up to evaluate the winners in the different categories

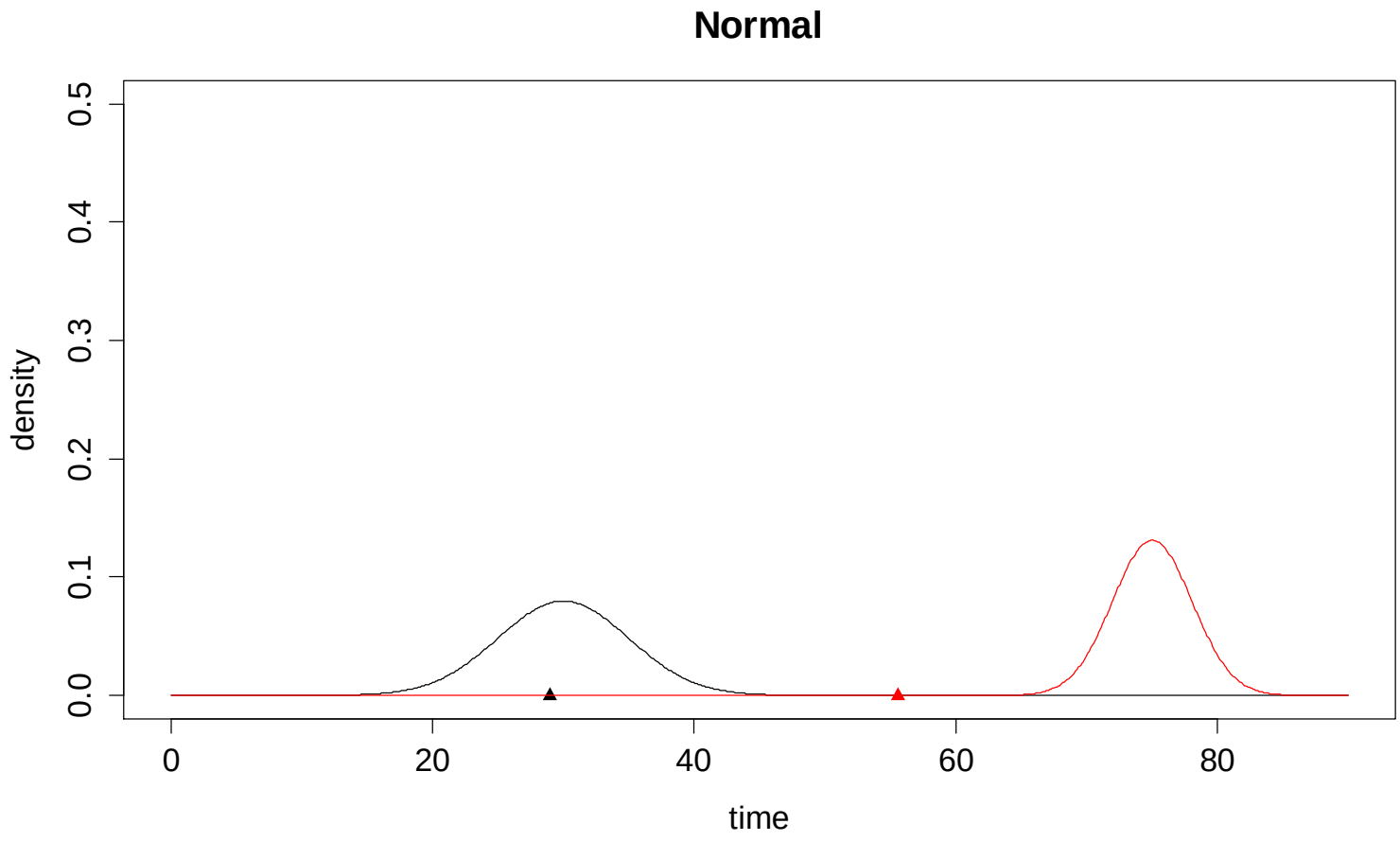
- Accurate – if inside 95th confidence interval
- Most accurate prediction – highest likelihood
- Most precise prediction – smallest 95th confidence interval
- Most safe prediction – widest 95th confidence interval
- Most pessimistic and optimistic predictions – largest positive and negative difference to the expected value
- Most unexpected failure – my own judgment

The Normals

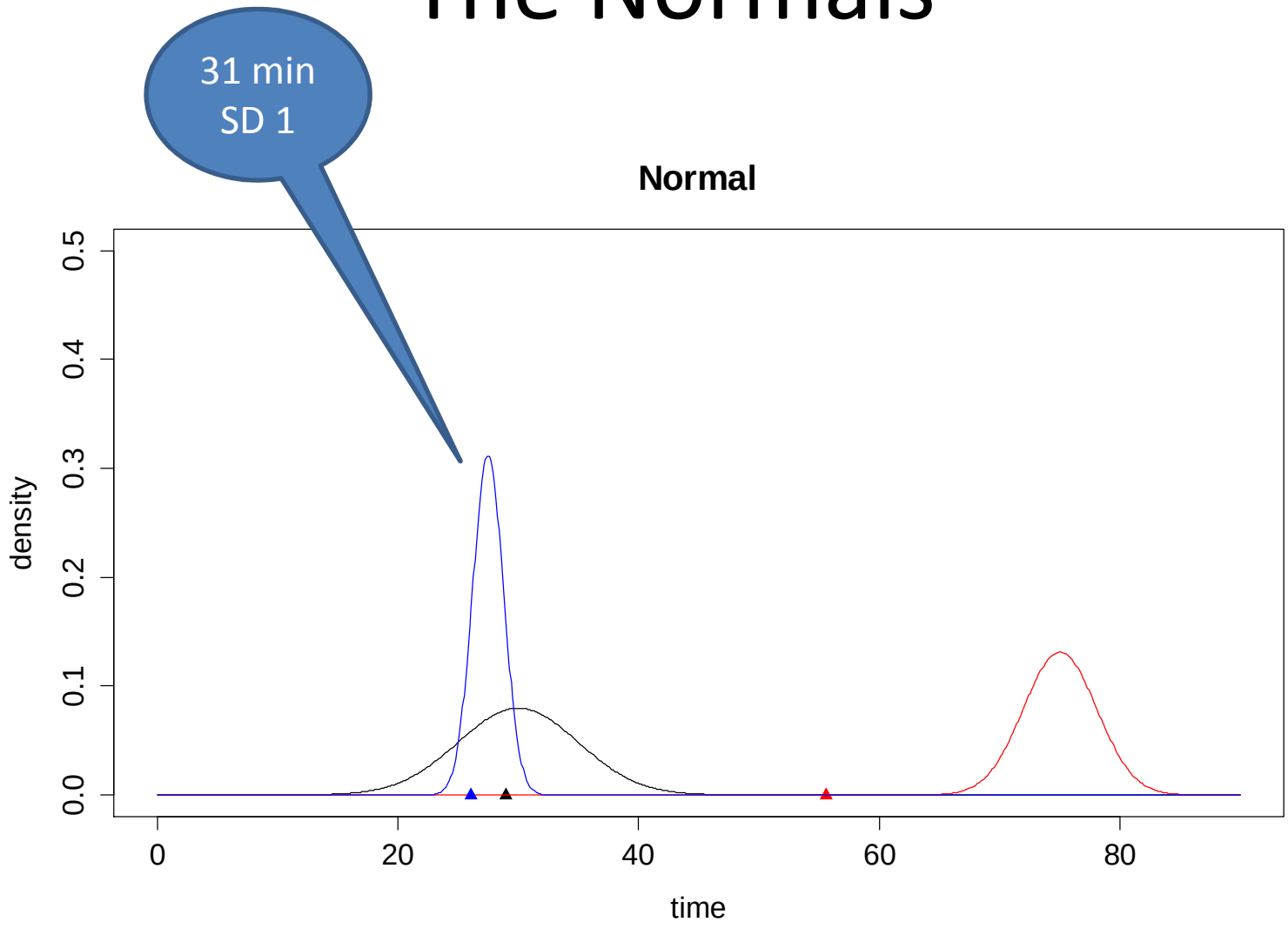
Between 42 and 46 min with 95% confidence (large uncertainty since I haven't decided whether to go on full speed - I will run the Gbg half marathon next week and need to be in shape for that.



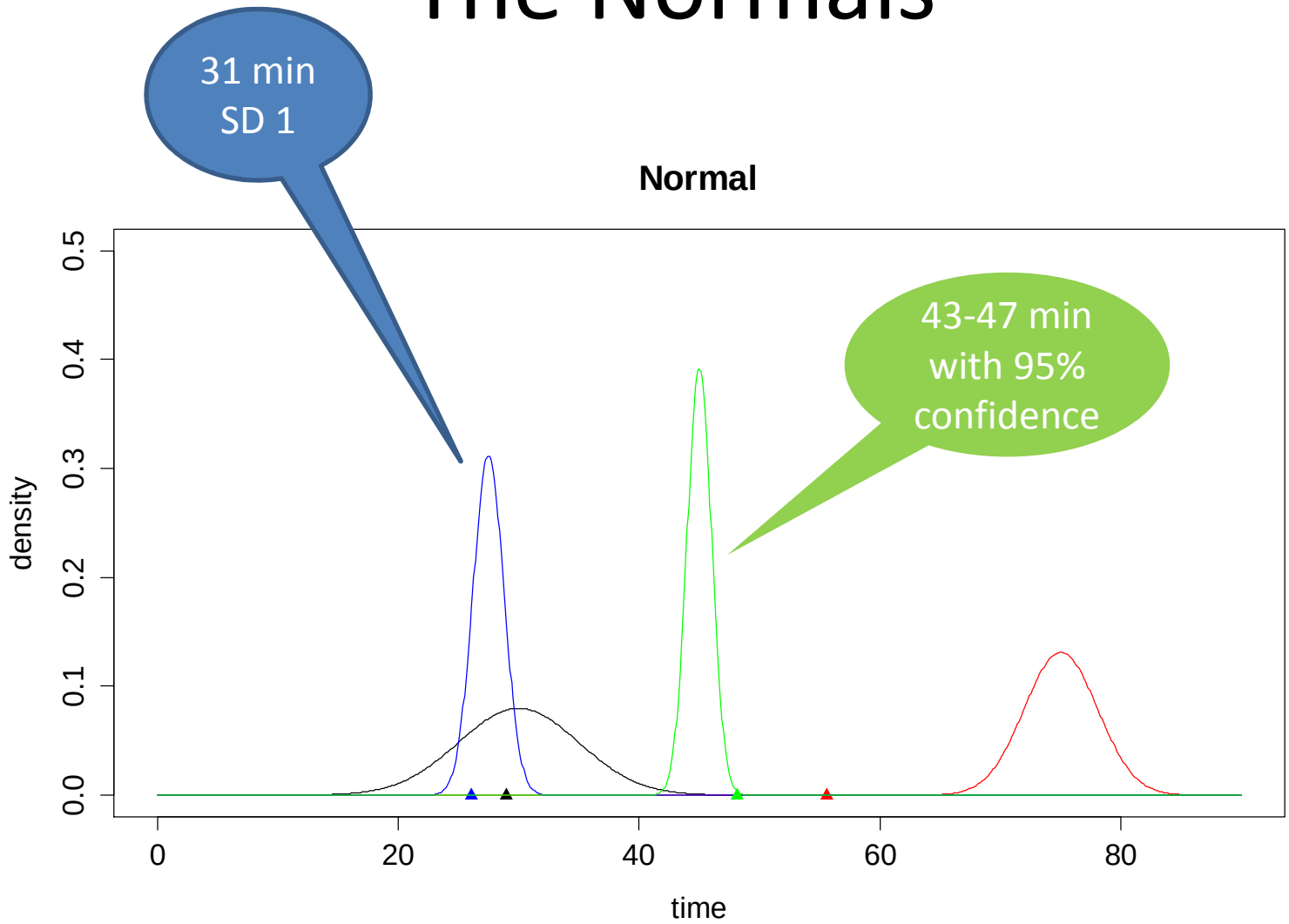
The Normals



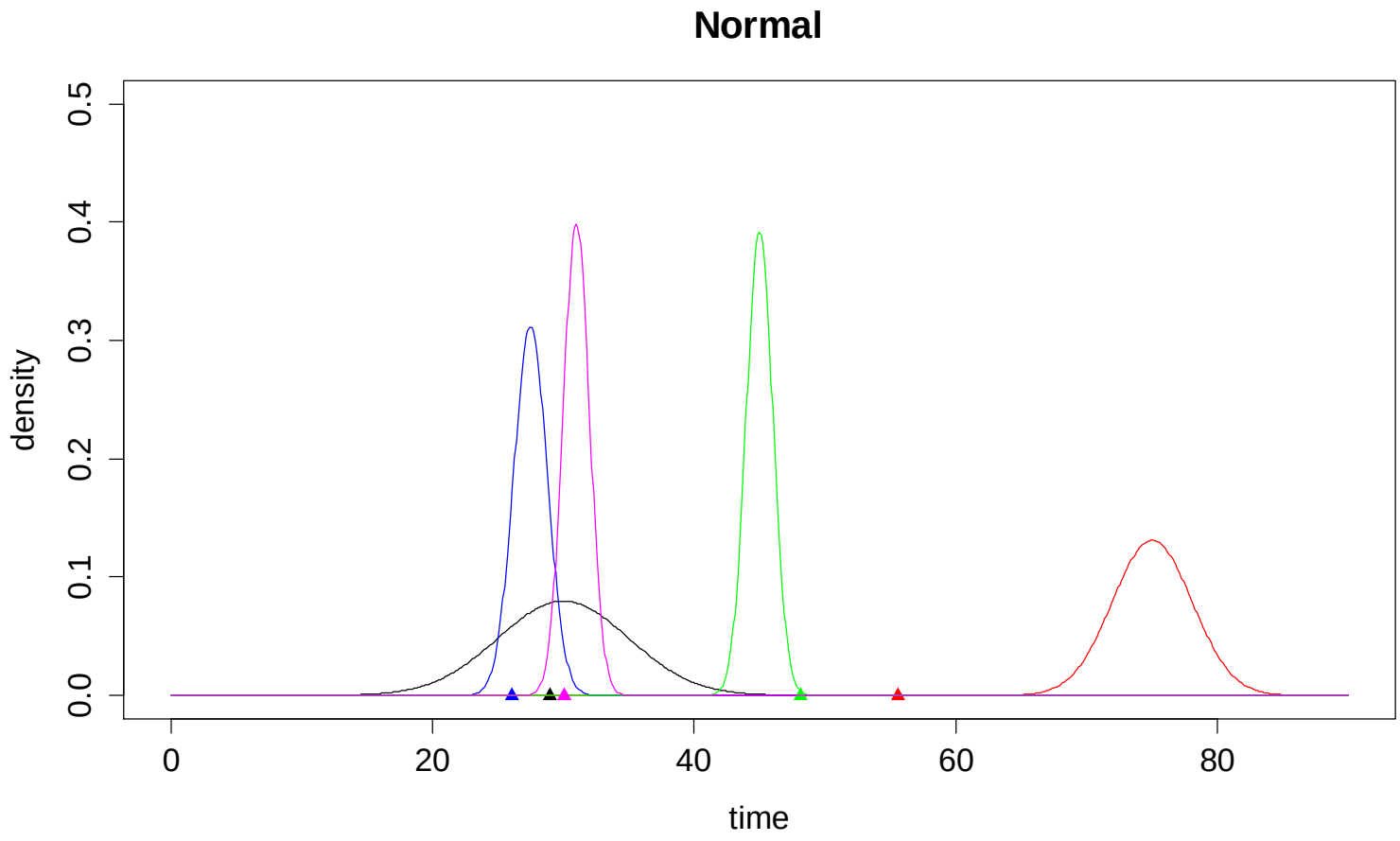
The Normals



The Normals

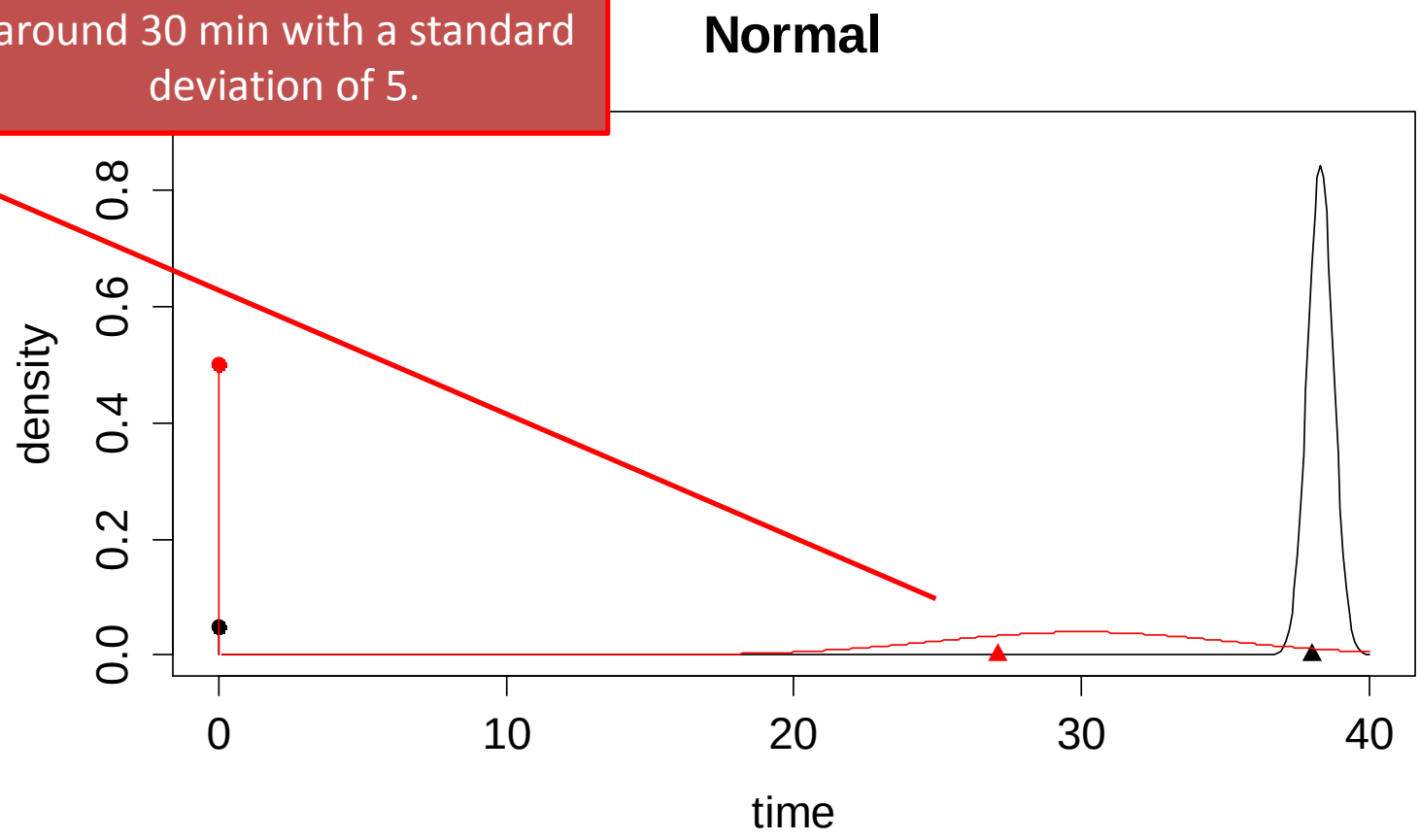

31 min
SD 1
Normal
43-47 min
with 95%
confidence
density
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
time

The Normals



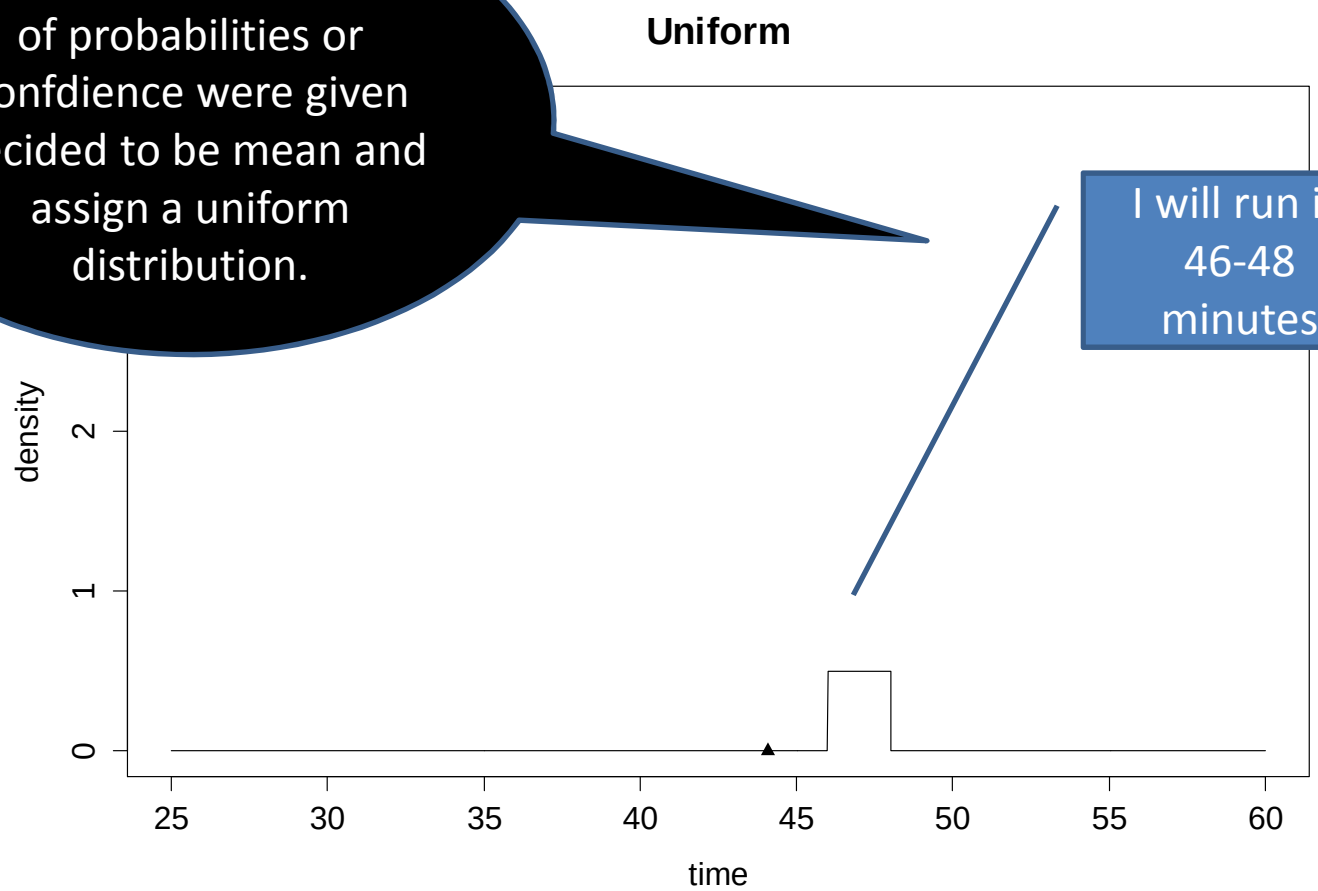
The Normal mixtures

50% chance of finishing. If I finish, I should have made it in around 30 min with a standard deviation of 5.



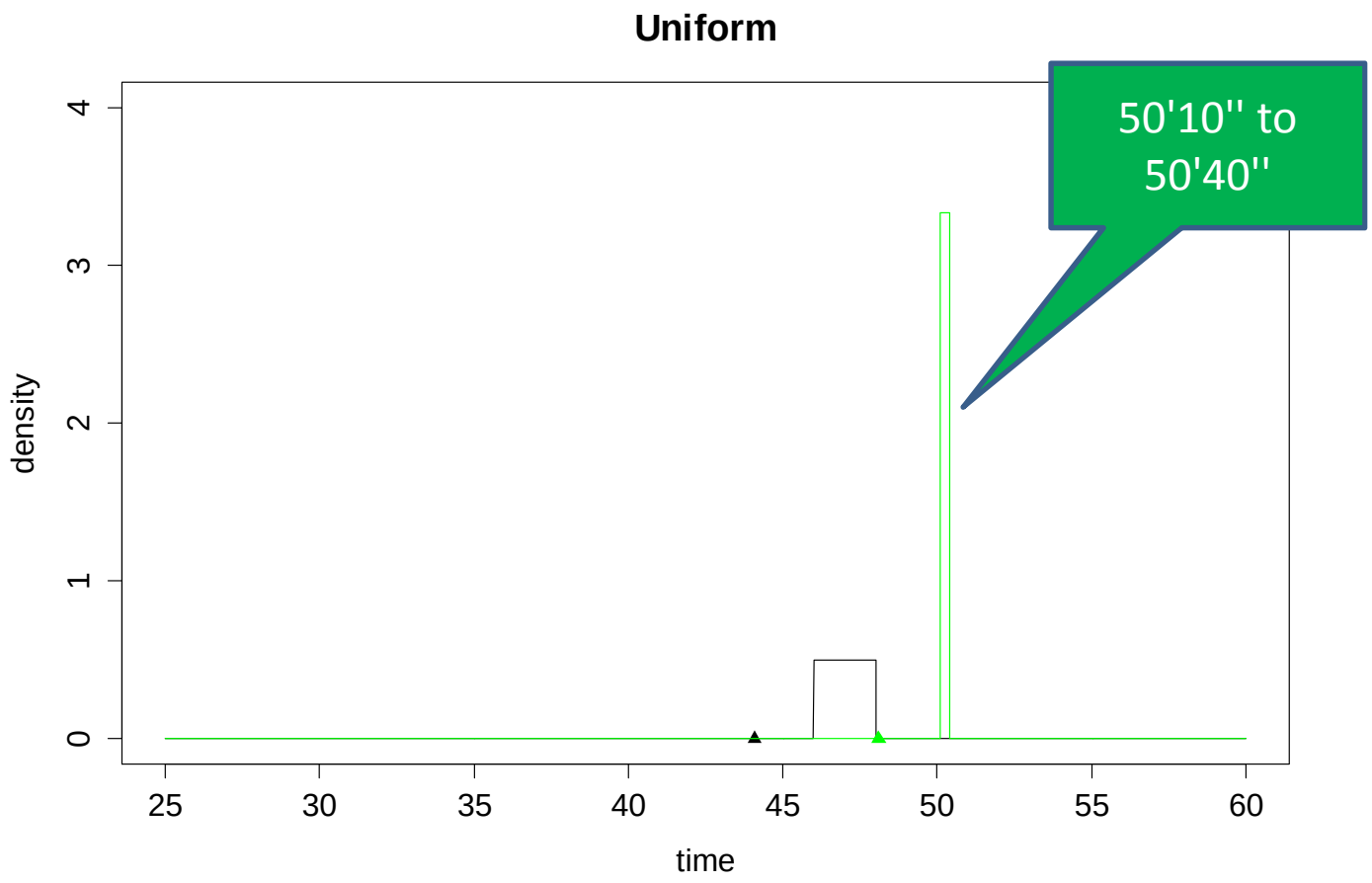
The Uniformists (explicit or implicit)

Whenever no statement of probabilities or confdience were given decided to be mean and assign a uniform distribution.

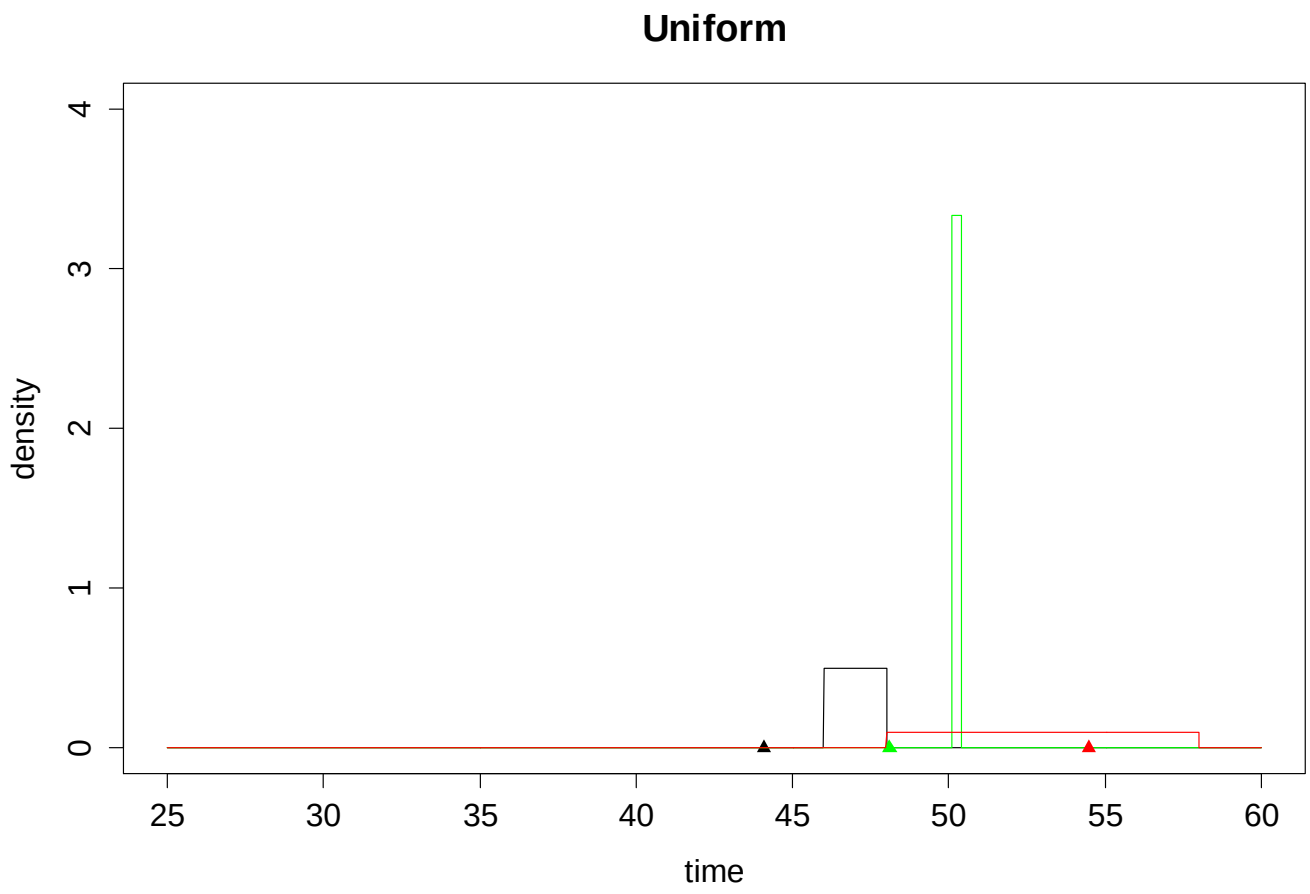


I will run in 46-48 minutes

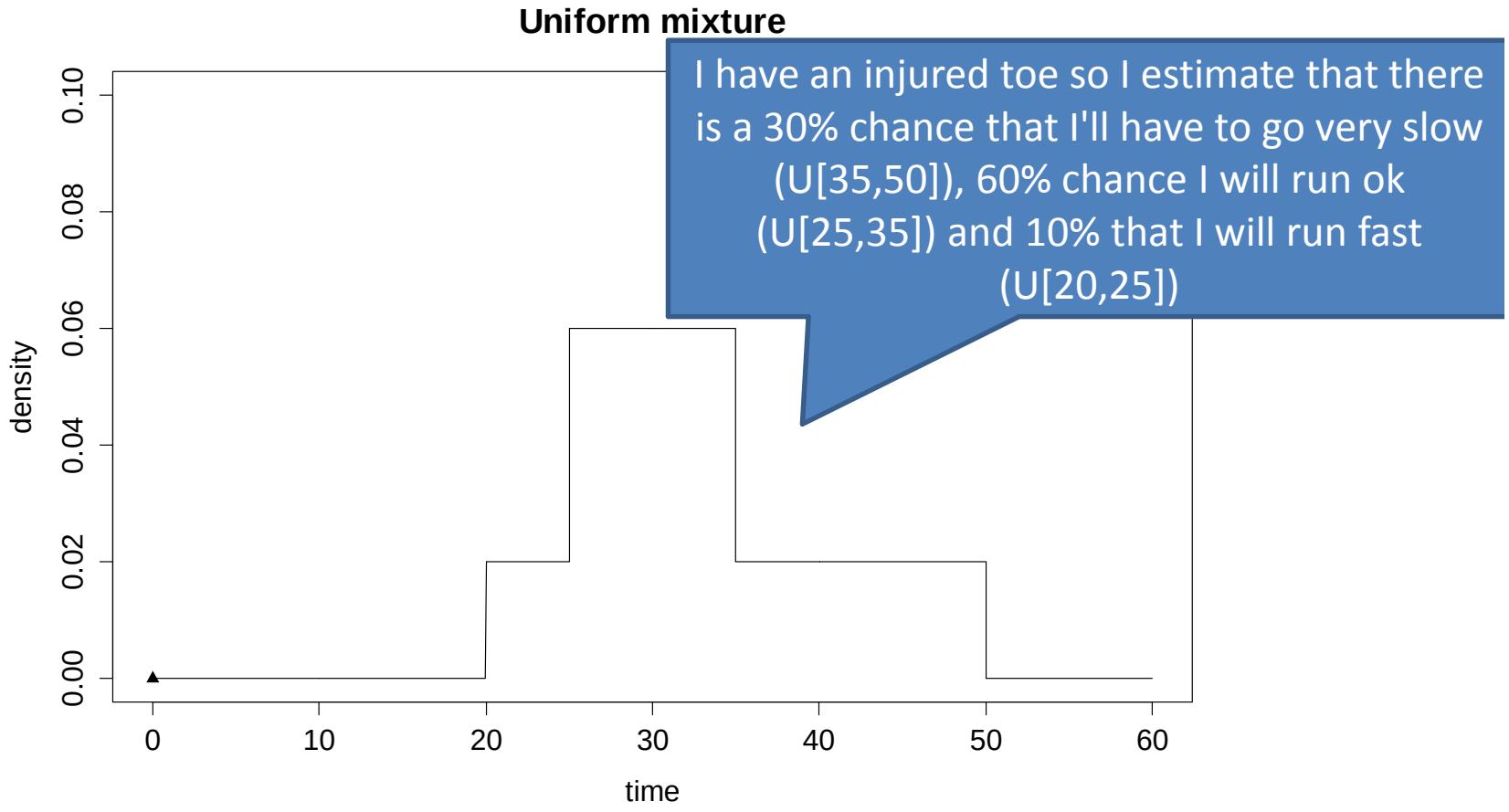
The Uniformists (explicit or implicit)



The Uniformists (explicit or understated)

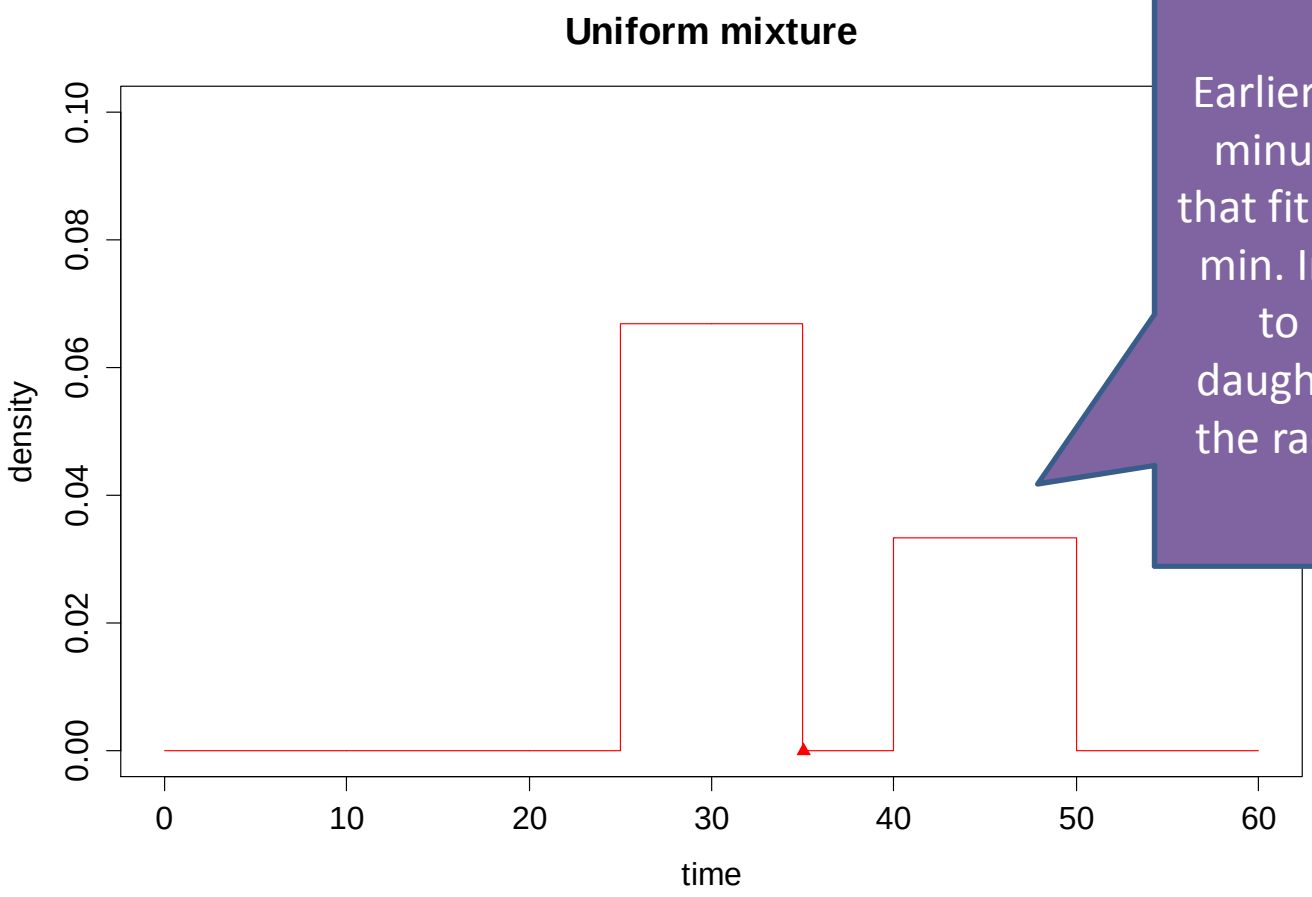


The Uniform mixtures



I have an injured toe so I estimate that there is a 30% chance that I'll have to go very slow (U[35,50]), 60% chance I will run ok (U[25,35]) and 10% that I will run fast (U[20,25])

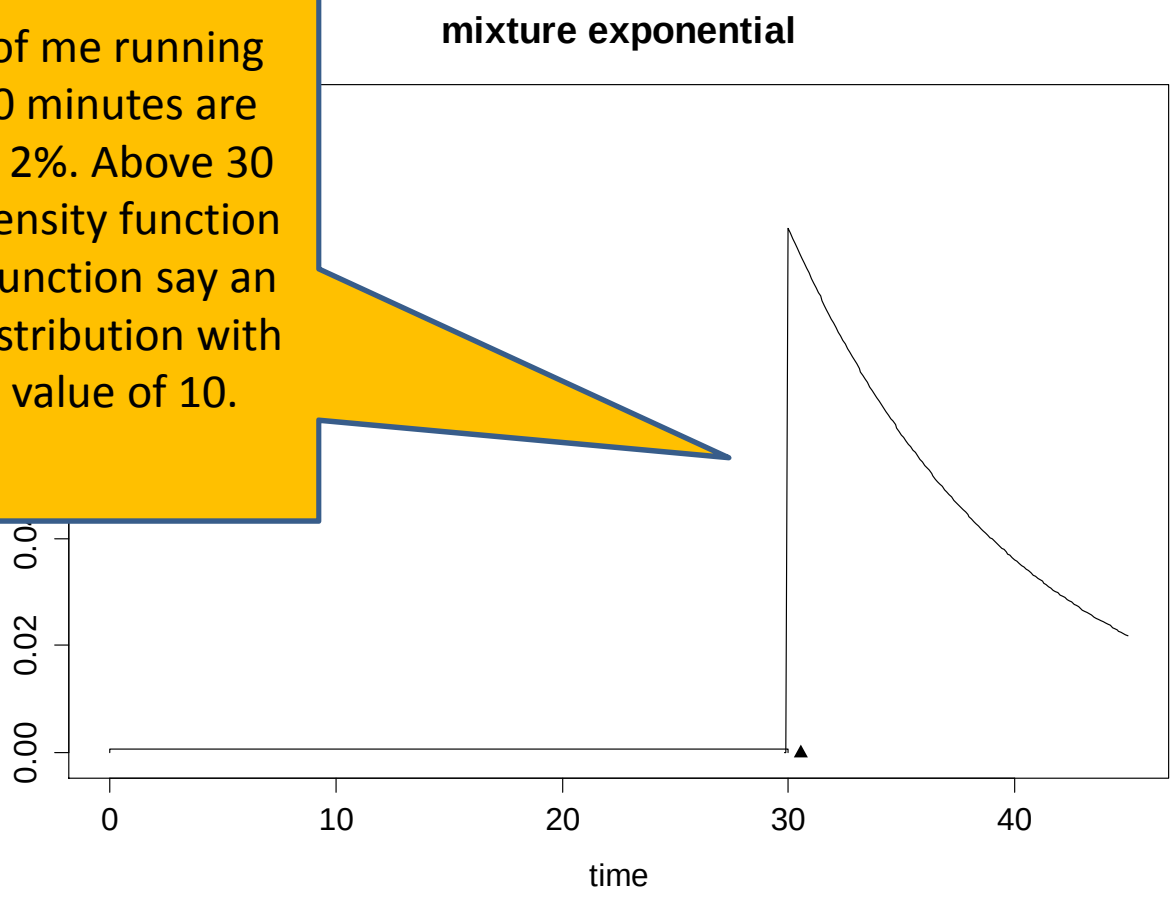
The Uniform mixtures



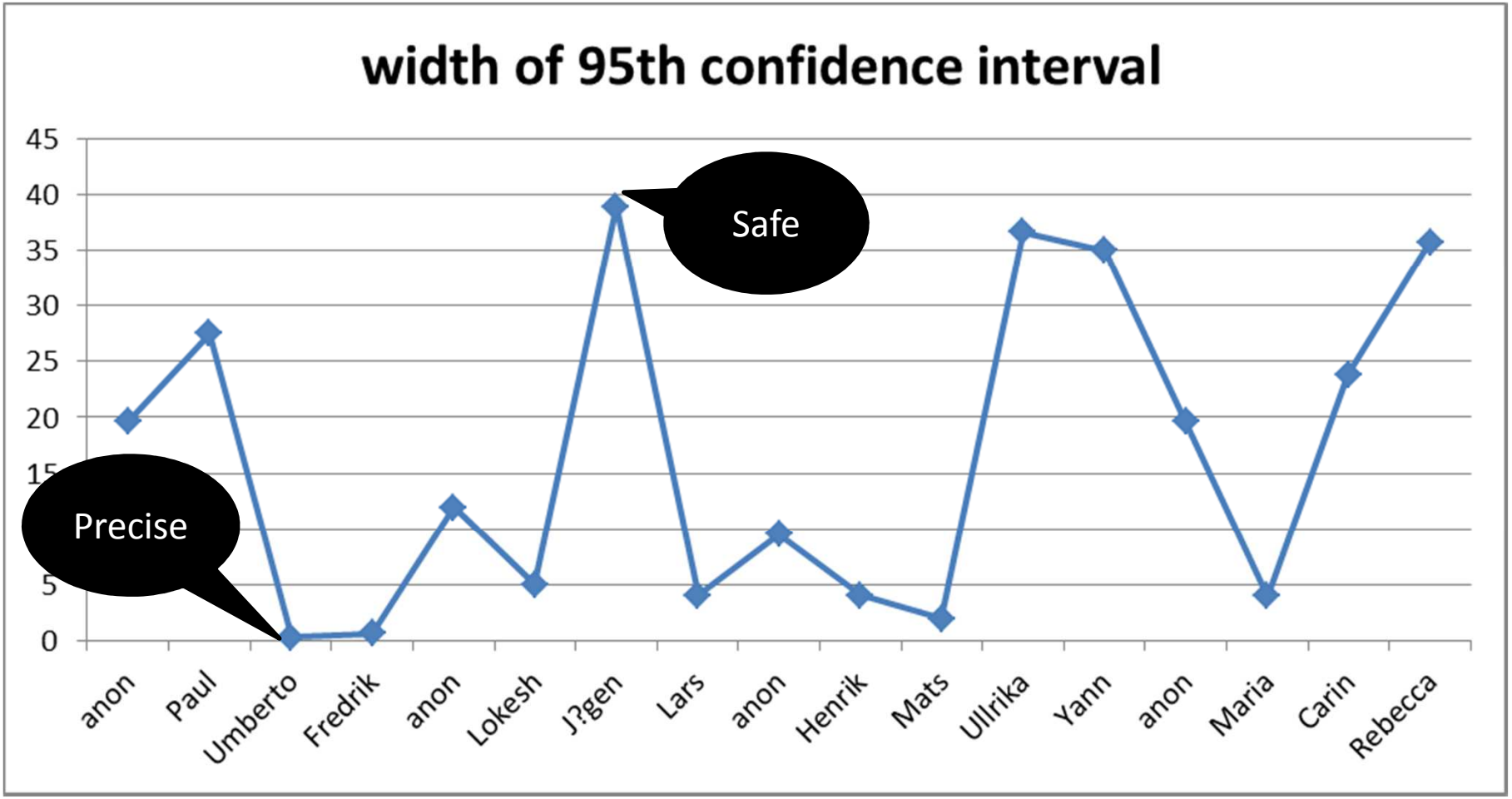
Earlier I used to run ca 25-30 minutes, but now I am not that fit so it may take ca 30-35 min. In addition, I may need to bring my 8 year old daughter. In that case we do the race in about 40-50 min.

A process example

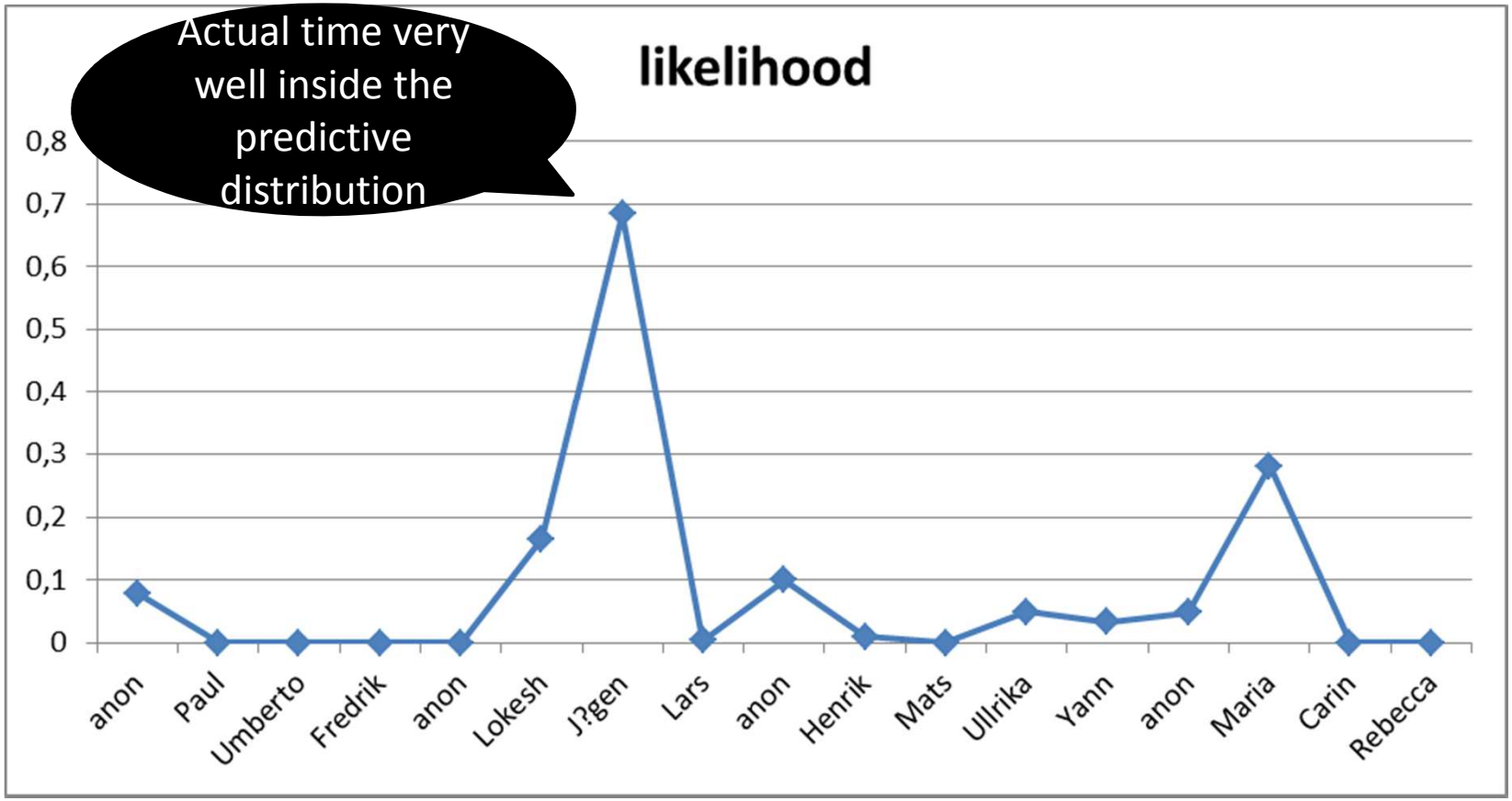
The chances of me running faster than 30 minutes are very small say 2%. Above 30 minutes the density function is a decaying function say an Exponential distribution with an expected value of 10.



Precise or safe

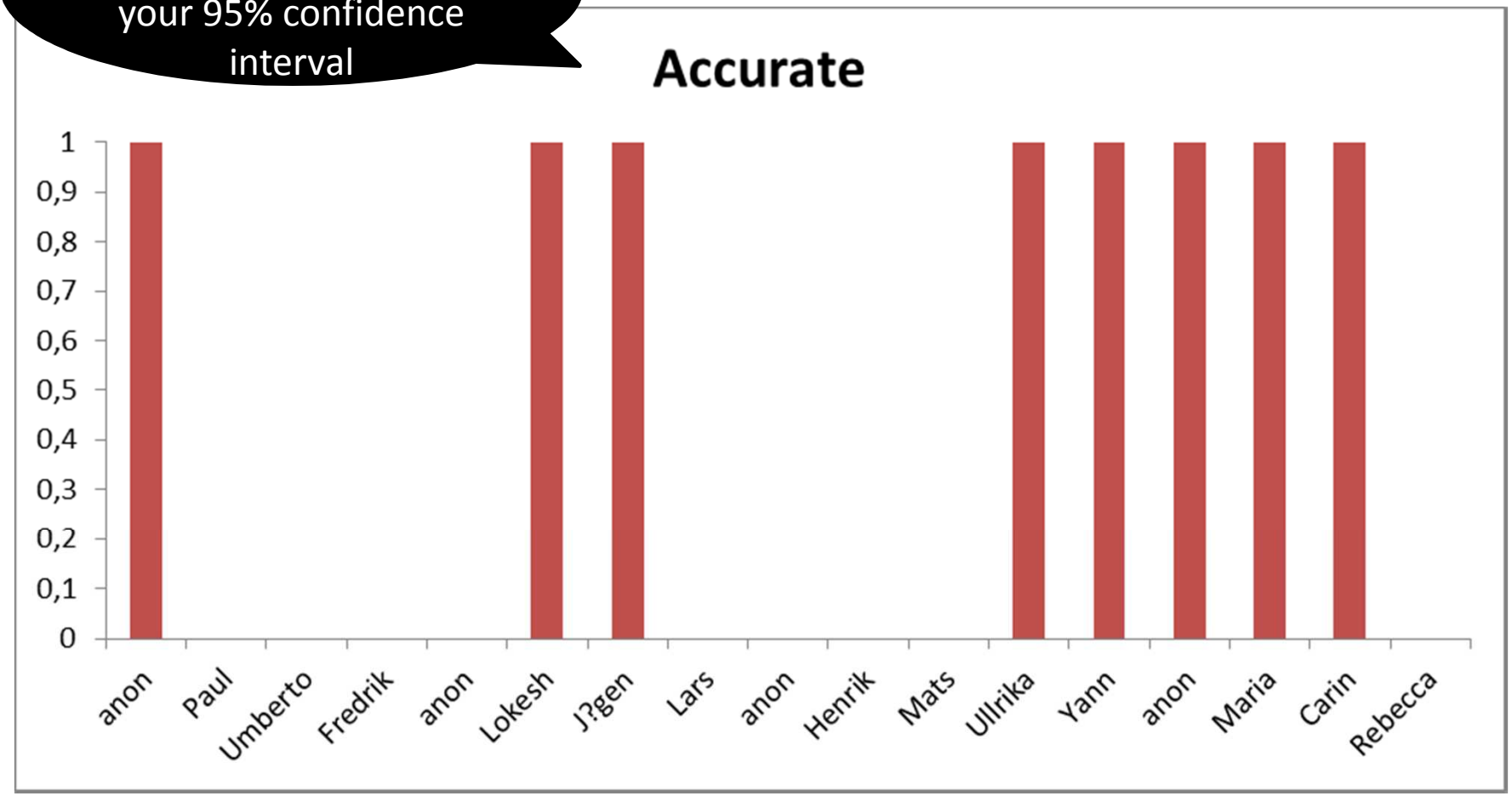


The highest likelihood award

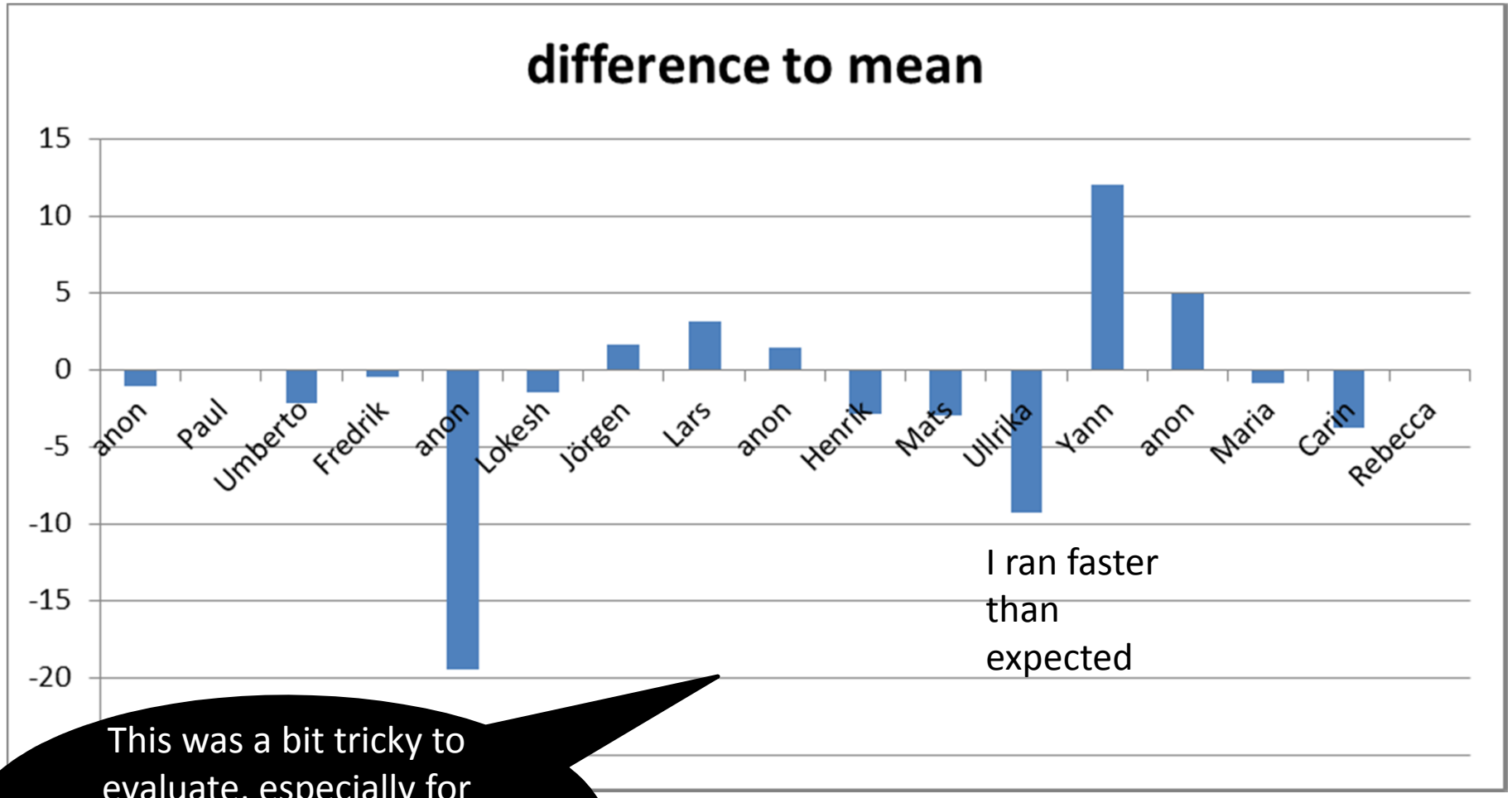


Who was accurate?

You are accurate as long as you are inside your 95% confidence interval



Mr and Mrs bias

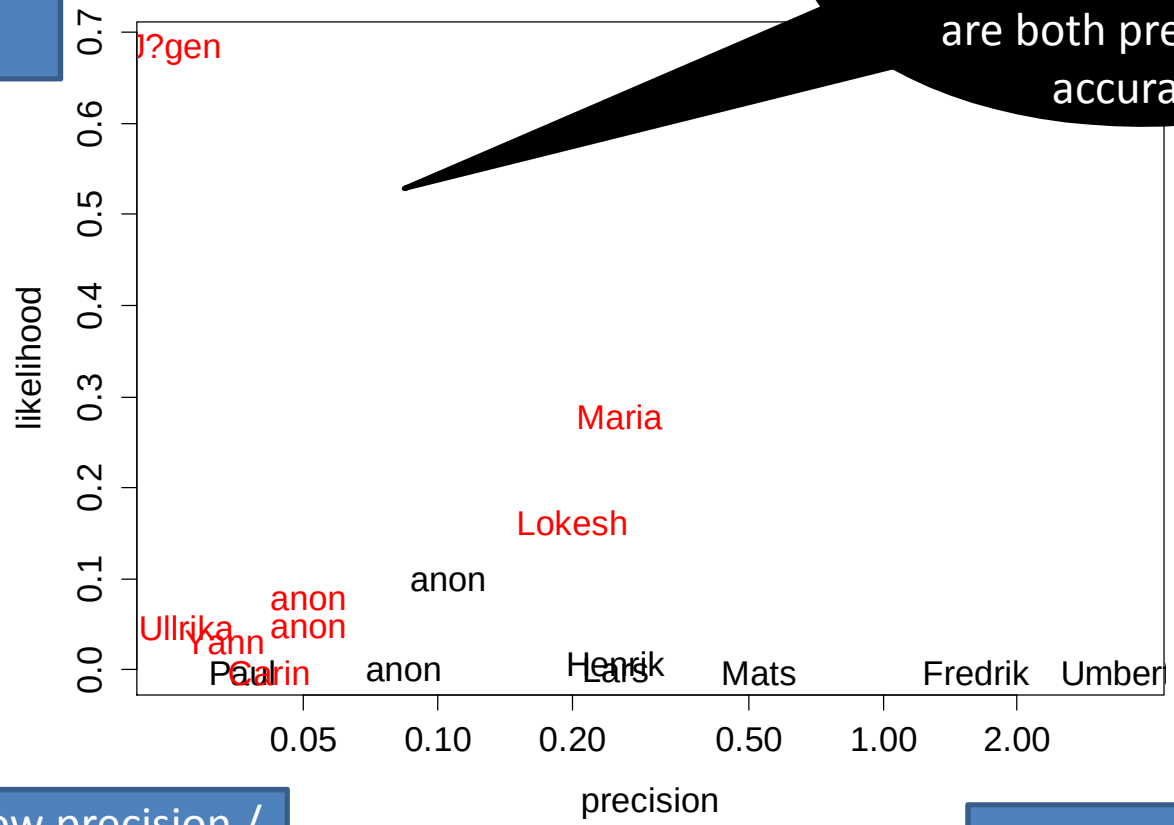


This was a bit tricky to evaluate, especially for the ones giving mixture distributions where no running was an option

The precision award and the least risk taker

Is there a trade-off between being precise and accurate?
Not necessarily, some are both precise and accurate.

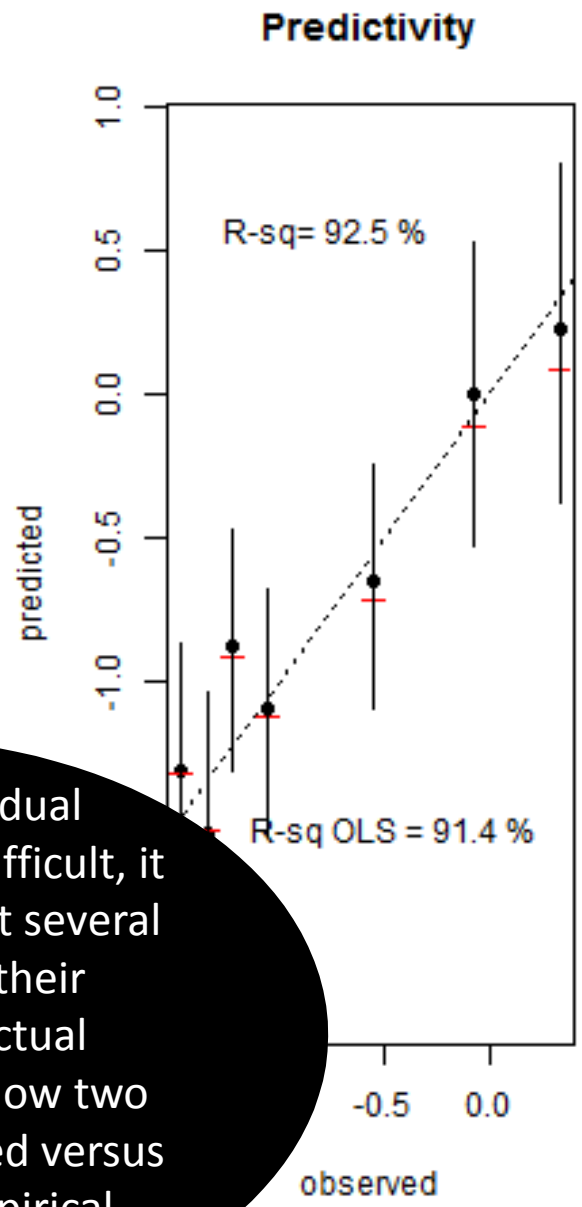
High likelihood / accurate prediction

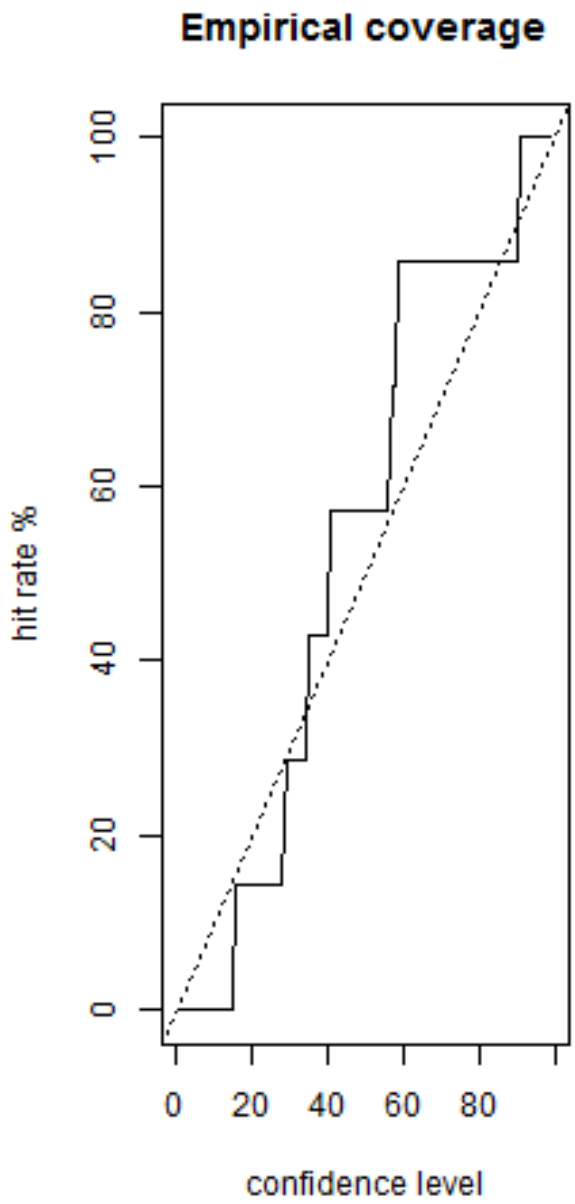


Low precision / "Better safe than sorry"

High precision

Sample -based

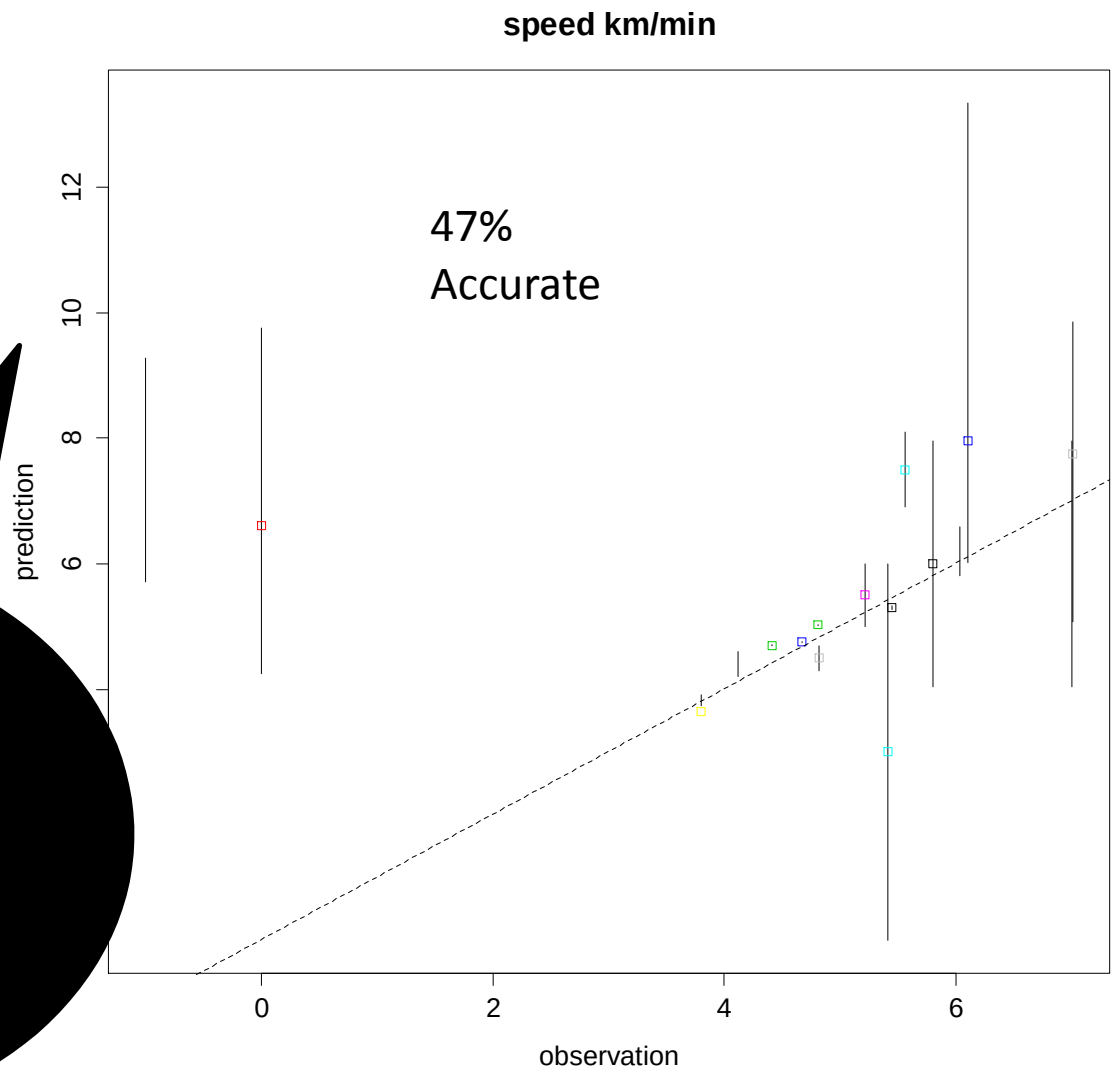




Evaluating individual predictions is a bit difficult, it is common to look at several predictions and their corresponding actual outcomes. Here I show two useful plots: observed versus predicted and empirical coverage

Observed versus predicted

Sample -based



This is how the observed versus predicted plot looks for the 17 respondents. The one to the left ran but did not provide any observation

Some final awards to hand out

Failures, errors and black swans

- The uniformists - Why use a uniform distribution (or an interval) and risk being outside?

Most incomplete prediction award

- Anon –Estimated the time with her mobile phone – different precision in the measurements. Measurement error is unknown, but manageable.

Most Unknown observation error

- Rebecca – Did not finalize the race. Is this an event to consider. All I know is that time for Rebecca > 0.

Most partially observable award

- Paul – Made his prediction but missed to sign up for the race. Did not run. Most unexpected event.

The black swan award