

Is a sub 2 hour marathon in the near future?
Modeling rare events in sports.

Rodney X. Sturdivant, Ph.D., Baylor University and Nick Clark,
Ph.D., West Point

Outline

- ▶ Baseball Rare Events (if needed only)
- ▶ Background
- ▶ Marathon Data
- ▶ Simple Model
- ▶ Self-Exciting Model
- ▶ Further Research

Background

Are we living in a time of records?

- ▶ Include NY Times article screenshot of headline.
- ▶ Brief summary of article's premise.

How can we address this question?

What would randomness look like?

Pictures of Rod and Nick running

Marathon World Record Data

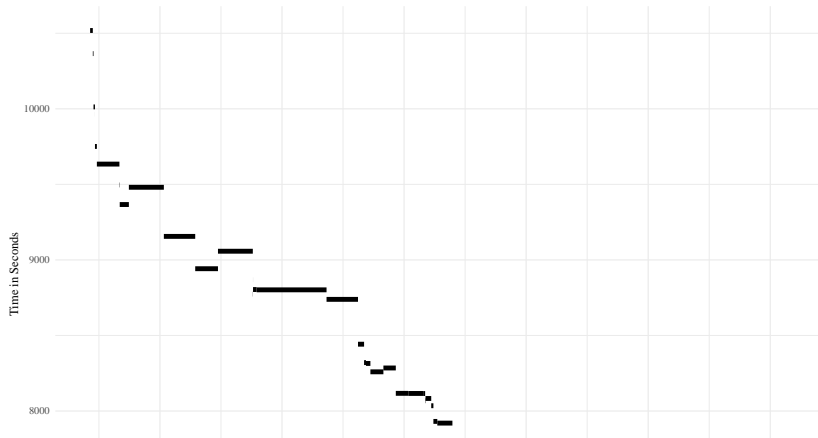
Men's Marathon world records since 1908

NEED TO CLEAN UP - NICER TABLE WITH JUST TIME
NAME NATIONALITY DATE MAYBE INCLUDE A COUPLE OF
PICTURES OF PEOPLE

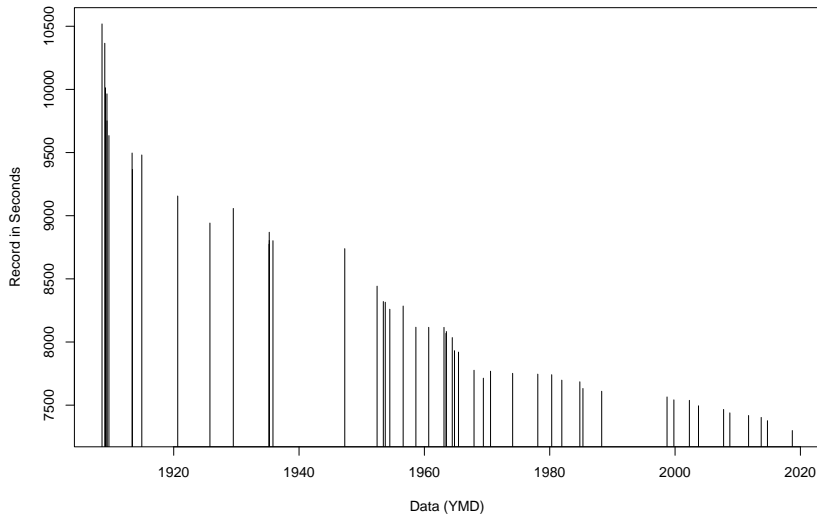
```
## # A tibble: 6 x 11
##   Time   Name   Natio~1 Date   Event~2 Source Notes Time_t
##   <chr> <chr> <chr>   <chr> <chr>   <chr> <chr> <Period
## 1 2:55~ John~ United~ July~ London~ IAAF[~ Time~ 2H 55M
## 2 2:52~ Robe~ United~ Janu~ Yonker~ IAAF[~ Note~ 2H 52M
## 3 2:46~ Jame~ United~ Febr~ New Yo~ IAAF[~ Note~ 2H 46M
## 4 2:46~ Albe~ United~ May ~ New Yo~ IAAF[~ Note~ 2H 46M
## 5 2:42~ Henr~ United~ May ~ Polyte~ IAAF[~ Note~ 2H 42M
## 6 2:40~ Thur~ Sweden Augu~ Stockh~ IAAF[~ Note~ 2H 40M
## # ... with 1 more variable: end <date>, and abbreviated
## # 1: Nationality, 2: 'Event/Place', 3: Time_sec
## # i Use 'colnames()' to see all variable names
```

Visualizing the data

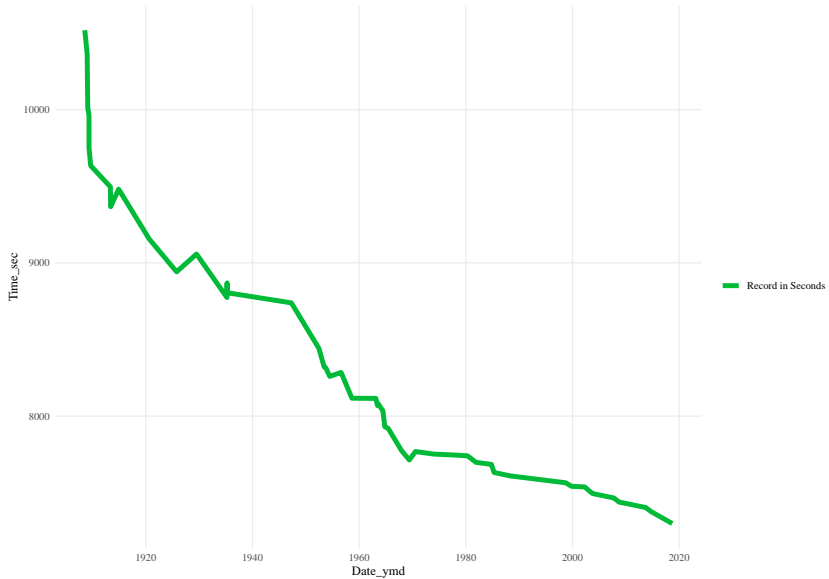
AGAIN NEED CLEANING UP - NEED TIMES IN SOMETHING OTHER THAN SECONDS MAYBE INCLUDE 2 HOUR
HORIZONTAL BAR MAYBE HAVE A SLIDE WHERE ADD PICTURES OF PEOPLE WHO LOWERED RECORD BY A LOT
WHICH PLOT(S) TO USE? MAYBE TWO OF THEM BUT ON ONE SLIDE?



Visualize B



Visualize C



SIMPLE MODEL

POISSON PROCESS (NEED TO SHORTEN, OR PUT ON TWO SLIDES)

A model for a series of discrete events where the average time between events is known, but the exact timing of events is “random” meeting the following criteria:

- ▶ Events are independent of each other. The occurrence of one event does not affect the probability another event will occur.
- ▶ The average rate (events per time period) is constant.
- ▶ Two events cannot occur at the same time.

The time between events (known as the interarrival times) follow an exponential distribution defined as:

$$P(T > t) = e^{-\lambda t}$$

Where T is the random variable of the time until the next event, t is a specific time for the next event, and λ is the rate: the average number of events per unit of time. Note the possible values of T are greater than 0 (positive only).

Reasonableness of Exponential Interarrivals

The exponential distribution has certain attributes, for example:

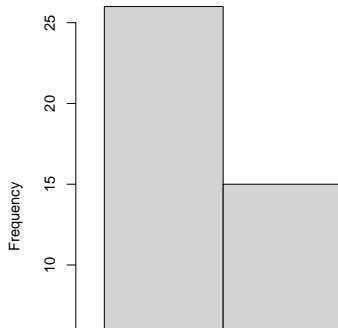
$$E(T) = 1/\lambda \quad SD(T) = 1/\lambda$$

The mean and standard deviation of the years between records:

```
## [1] 2.249427
```

```
## [1] 2.428191
```

Histogram of days_between_mod2



MORE ON THE SIMPLE MODEL

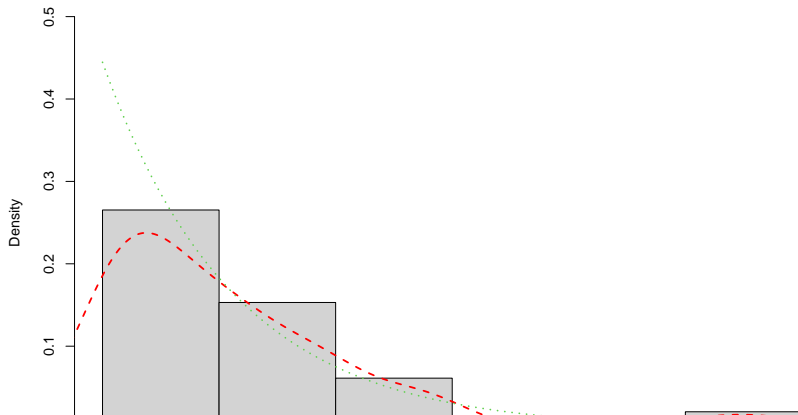
We estimate (MLE) $\lambda = 1/E(T)$

```
##      rate
```

```
## 0.4445577
```

Model fit

Histogram, density curve and exponential model



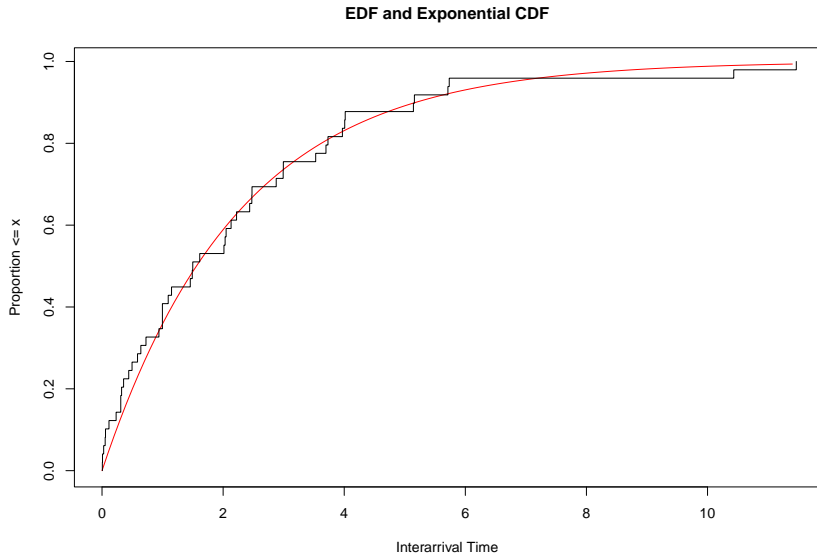
Fit of simple model

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  days_between_mod2
## D = 0.078053, p-value = 0.9264
## alternative hypothesis: two-sided

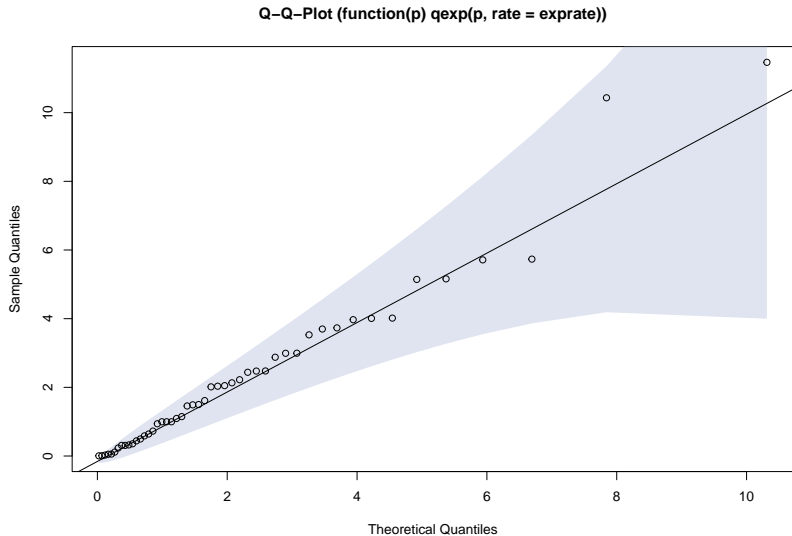
##
##  Cramer-von Mises test of goodness-of-fit
##  Braun's adjustment using 7 groups
##  Null hypothesis: exponential distribution
##  with parameter rate = 0.444557679401457
##  Parameters assumed to have been estimated from data
##
## data:  days_between_mod2
## omega2max = 0.29762, p-value = 0.6416

##
##  Anderson-Darling test of goodness-of-fit
```

Fit of simple model B



Are records then random?



What are the poorly fit points?

LOOK BACK AT THE ORIGINAL DATA HERE... LONGEST TIMES BETWEEN EVENTS (I THINK)... ONE IS WW2 PRETTY SURE... THE OTHER NEED TO LOOK AGAIN - MAYBE AN UNUSUALLY LARGE LOWERING OF THE RECORD OR SOMETHING?

References

Data source: Wikipedia (https://en.wikipedia.org/wiki/Marathon_world_record_progression)
scraped August 12, 2022

Poisson process: <https://towardsdatascience.com/the-poisson-distribution-and-poisson-process-explained-4e2cb17d459>