The Best, The Hottest, & The Luckiest: A Statistical Tale of Extremes

Sid Redner, Boston University, physics.bu.edu/~redner

Extreme Events: Theory, Observations, Modeling and Simulations

Palma de Mallorca, November 10-14, 2008
The Best
with S. Maslov, P. Chen, H. Xie
J. Informetrics 1, 8-15 (2007)

Observations about scientific citations

Google page rank analysis for “hidden gems”
Phys. Rev. Citation Data  (as of July 03)

353,268 papers, 3,110,839 cites

\( \langle \# \text{ cites} \rangle = 8.81, \langle \text{cite age} \rangle = 6.20 \)

N.B.: Internal citations only; undercount by factor of 3-5.

(for highly cited HEP papers; SPIRES)
Phys. Rev. Citation Data (as of July 03)

353,268 papers, 3,110,839 cites

\langle \# \text{ cites} \rangle = 8.81, \langle \text{cite age} \rangle = 6.20

N.B.: *Internal* citations only; undercount by factor of 3-5.
(for highly cited HEP papers; SPIRES)

11 papers with > 1000 citations
79 papers with > 500 citations
237 papers with > 300 citations
2340 papers with > 100 citations
8073 papers with > 50 citations
Phys. Rev. Citation Data  (as of July 03)

353,268 papers, 3,110,839 cites

$\langle \# \text{ cites} \rangle = 8.81$, $\langle \text{cite age} \rangle = 6.20$

N.B.: *Internal* citations only; undercount by factor of 3-5.
(for highly cited HEP papers; SPIRES)

11 papers with $> 1000$ citations
79 papers with $> 500$ citations
237 papers with $> 300$ citations
2340 papers with $> 100$ citations
8073 papers with $> 50$ citations
245459 papers with $< 10$ citations
84144 papers with 1 citation
23421 papers with 0 citations
## PR papers with >1000 cites

<table>
<thead>
<tr>
<th>article</th>
<th>cites (age)</th>
<th>title</th>
<th>author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR 140</td>
<td>3227 26.7</td>
<td>Self Consistent Equations..</td>
<td>W. Kohn &amp; L. J. Sham</td>
</tr>
<tr>
<td>article</td>
<td>cites (age)</td>
<td>title</td>
<td>author(s)</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>PR 140 A1133 (1965)</td>
<td>4598</td>
<td>26.7</td>
<td>Self Consistent Equations..</td>
</tr>
<tr>
<td>article</td>
<td>cites</td>
<td>(age)</td>
<td>title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1 PR</td>
<td>140</td>
<td>A1133 (1965)</td>
<td>4598</td>
</tr>
<tr>
<td>2 PR</td>
<td>136</td>
<td>B864 (1965)</td>
<td>2460</td>
</tr>
<tr>
<td>3 PRB</td>
<td>23</td>
<td>5048 (1981)</td>
<td>2079</td>
</tr>
<tr>
<td>5 PR</td>
<td>108</td>
<td>1175 (1957)</td>
<td>1364</td>
</tr>
<tr>
<td>6 PRL</td>
<td>19</td>
<td>1264 (1967)</td>
<td>1306</td>
</tr>
<tr>
<td>7 PRB</td>
<td>12</td>
<td>3060 (1975)</td>
<td>1259</td>
</tr>
<tr>
<td>8 PR</td>
<td>124</td>
<td>1866 (1961)</td>
<td>1178</td>
</tr>
<tr>
<td>11 PRB</td>
<td>13</td>
<td>5188 (1976)</td>
<td>1023</td>
</tr>
</tbody>
</table>

Idea: *launch many surfers that surf forever until a steady-state surfer number* $G_i$ *is reached at each site* $i$ *of the network.*

Idea: *launch many surfers that surf forever until a steady-state surfer number* $G_i$ *is reached at each site* $i$ *of the network.*

Equation for number of surfers:

$$G_i = (1 - d) \sum_{nn} \frac{G_j}{k_{out}} + \frac{d}{N}$$

Idea: *launch many surfers that surf forever until a steady-state surfer number* $G_i$ *is reached at each site* $i$ *of the network.*

Equation for number of surfers:

$$G_i = (1 - d) \sum_{j \text{ nn } i} \frac{G_j}{k_{out}^j} + \frac{d}{N}$$

*random walk propagation*  
*manna from heaven*

Idea: *launch many surfers that surf forever until a steady-state surfer number* \( G_i \) *is reached at each site* \( i \) *of the network.*

Equation for number of surfers:

\[
G_i = (1 - d) \sum_{j \text{ nn } i} \frac{G_j}{k_{j \text{ out}}} + \frac{d}{N}
\]

Brin/Page:  \( d=0.15 \) *(bored after 6 clicks)*

We use:  \( d=0.50 \) *(don’t cite beyond 2 generations)*
Correlation between Google & citation counts

The graph shows the correlation between the average Google number and the number of citations. The data points follow a trend line with a slope of 1, indicating a strong positive correlation. The x-axis represents the number of citations, while the y-axis represents the average Google number.
The Hottest

with M. Petersen

Some facts about urban temperatures

Evolution of record temperature events

Role of global warming on records
Philadelphia Temperatures

annual temperature pattern

long-term temperature trends

Aug 7, 1918  106F
Feb 9, 1934  -11F
Philadelphia Temperatures

**annual temperature pattern**

- **Aug 7, 1918**: 106°F
- **Feb 9, 1934**: -11°F

**long-term temperature trends**

IPCC 2001: global warming 1.1°F over past century

Average temperature from 1880 to 2000:
- **1880**: 38°F
- **2000**: 48°F

**annual temperature pattern**:
- **record high**
- **record low**
- **av. high**
- **av. low**

**daily temperature** vs. **day of the year**

- **0°F** to **100°F**
- **0** to **365**

**average temperature** vs. **year**

- **1880** to **2000**
- **32°F** to **70°F**
Philadelphia Temperatures

annual temperature pattern

long-term temperature trends

IPCC 2001: global warming 1.1° F over past century
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption:  

*daily temperatures are iid continuous variables*
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: daily temperatures are iid continuous variables

temperature on a fixed day
Evolution of Temperature Records

Assumption: daily temperatures are iid continuous variables

Temperature on a fixed day

\[ T_0 \]

\[ t_0 \]

Year
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: daily temperatures are iid continuous variables

temperature
on a fixed day

year

$T_0$

t0
Evolution of Temperature Records

assumption: *daily temperatures are iid continuous variables*

temperature on a *fixed day*
Evolution of Temperature Records

assumption: *daily temperatures are iid continuous variables*

temperature on a *fixed day*

$T_0$, $T_1$

$t_0$, $t_1$

year

(Arnold et al., Records, 1998)
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: *daily temperatures are iid continuous variables*

temperature on a *fixed day*
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: *daily temperatures are iid continuous variables*

**temperature on a fixed day**

![Graph showing temperature evolution over years]

$T_0$, $T_1$, $t_0$, $t_1$
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: daily temperatures are iid continuous variables
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: daily temperatures are iid continuous variables

temperature on a fixed day
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: *daily temperatures are iid continuous variables*

temperature on a *fixed day*

![Graph showing evolution of temperature records between years *t₀* and *t₁* with points *T₀* and *T₁*.](image-url)
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: daily temperatures are iid continuous variables

Temperature on a fixed day

\[ T_1 \]
\[ T_0 \]
\[ t_0 \]
\[ t_1 \]

year
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: daily temperatures are iid continuous variables

temperature on a fixed day

\[ T_1 \]
\[ T_0 \]

\[ t_0 \] \[ t_1 \] year
Evolution of Temperature Records

assumption: daily temperatures are iid continuous variables

temperature on a *fixed day*
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: *daily temperatures are iid continuous variables*

temperature on a *fixed day*
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: *daily temperatures are iid continuous variables*

temperature on a *fixed day*

\[ T_3, T_2, T_1, T_0 \]

\[ t_0, t_1, t_2, t_3 \]
Evolution of Temperature Records

assumption: *daily temperatures are iid continuous variables*

temperature on a *fixed day*
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: daily temperatures are iid continuous variables

Goal: compute $T_k, P_k(T)$, $t_k, p_k(t)$
Evolution of Temperature Records

(Arnold et al., Records, 1998)

assumption: *daily temperatures are iid continuous variables*

Goal: compute $T_k, p_k(t)$

distribution independent!
Record Time Statistics

(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)
define $\sigma_i = \begin{cases} 1 & \text{if record in } i^{th} \text{ year} \\ 0 & \text{otherwise} \end{cases}$
define $\sigma_i = \begin{cases} 1 & \text{if record in } i^{th} \text{ year} \\ 0 & \text{otherwise} \end{cases}$

then $\langle \sigma_i \rangle = \frac{1}{i + 1}$
Record Time Statistics

(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)

\[
\sigma_i = \begin{cases} 
1 & \text{if record in } i^{th} \text{ year} \\
0 & \text{otherwise}
\end{cases}
\]

\[
\langle \sigma_i \rangle = \frac{1}{i+1}
\]

\[
\langle \sigma_i \sigma_j \rangle = \langle \sigma_i \rangle \langle \sigma_j \rangle
\]
Record Time Statistics

(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)

Define \( \sigma_i = \begin{cases} 1 & \text{if record in } i^{th} \text{ year} \\ 0 & \text{otherwise} \end{cases} \)

Then \( \langle \sigma_i \rangle = \frac{1}{i + 1} \) and \( \langle \sigma_i \sigma_j \rangle = \langle \sigma_i \rangle \langle \sigma_j \rangle \)

Therefore \( \langle n(t) \rangle = \sum_{i=1}^{t} \langle \sigma_i \rangle \sim \ln t \)

\[ P(n) = \frac{(\ln t)^n}{n!} e^{-\ln t} \]
**Record Time Statistics**

(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)

Define

\[ \sigma_i = \begin{cases} 1 & \text{if record in } i^{\text{th}} \text{ year} \\ 0 & \text{otherwise} \end{cases} \]

Then

\[ \langle \sigma_i \rangle = \frac{1}{i + 1} \quad \langle \sigma_i \sigma_j \rangle = \langle \sigma_i \rangle \langle \sigma_j \rangle \]

Therefore

\[ \langle n(t) \rangle = \sum_{i=1}^{t} \langle \sigma_i \rangle \sim \ln t \quad P(n) = \frac{(\ln t)^n}{n!} e^{-\ln t} \]

Probability that current record is broken in \(i^{\text{th}}\) year:
Record Time Statistics

(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)

\[ \sigma_i = \begin{cases} 1 & \text{if record in } i^{th} \text{ year} \\ 0 & \text{otherwise} \end{cases} \]

Define \( \sigma_i = \begin{cases} 1 & \text{if record in } i^{th} \text{ year} \\ 0 & \text{otherwise} \end{cases} \)

Then \( \langle \sigma_i \rangle = \frac{1}{i + 1} \)  \( \langle \sigma_i \sigma_j \rangle = \langle \sigma_i \rangle \langle \sigma_j \rangle \)

Therefore \( \langle n(t) \rangle = \sum_{i=1}^{t} \langle \sigma_i \rangle \sim \ln t \)  \( P(n) = \frac{(\ln t)^n}{n!} e^{-\ln t} \)

Probability that current record is broken in ith year:

\[ = \frac{1}{i(i + 1)} \rightarrow \text{time until next record} = \infty \]
Records If Global Warming Is Occurring

Assume temperature distribution as a soluble example only

\[ p(T; t) = \begin{cases} e^{-(T-\nu t)} & T > \nu t \\ 0 & T < \nu t \end{cases} \]

Exceedance probability

\[ p_{\geq}(T_k; t_k + j) = \int_{T_k}^{\infty} e^{-[T-\nu(t_k+j)]} \, dT = e^{-(T_k-\nu t_k)} e^{j\nu} \equiv X e^{j\nu} \]

Prob of record at year \( n \)

\[ q_n(T_k) \equiv p_{\geq}(T_k) p_{\leq}(T_k)^{n-1} \]

\[ \rightarrow e^{n\nu} X \prod_{j=1}^{n-1} (1 - e^{j\nu} X) \]

When 0 record must occur

(Over)estimate of record time

\[ e^{j\nu} X = 1 \rightarrow (t_{k+1} - t_k)\nu = T_k - \nu t_k \]

\[ \rightarrow t_k \sim \frac{k}{\nu} \] records ultimately occur at constant rate

(Ballerini & Resnick, 85; Borokov, 99)
Frequency of Record High Temperatures

![Graph showing the frequency of record high temperatures over time. The x-axis represents time on a logarithmic scale, and the y-axis represents probability on a logarithmic scale. There are three curves labeled with probability values: v=0.003, v=0.006, and v=0.012.](image)
Frequency of Record Low Temperatures

The graph shows the frequency of record low temperatures over time. The x-axis represents time, and the y-axis represents probability. The graph includes lines for different values of v (0.003, 0.006, 0.012) with corresponding markers and labels for each line.
Record Frequencies: Multiple Stations

Potsdam
years after 1893

CET
years after 1878

Brussels
years after 1833

Milan
years after 1763

Prague
years after 1775

△ highs

▽ lows
The Luckiest
with C. Sire

Statistics of team win/loss records

Consecutive-game winning/losing streaks
Fun Facts About Team Win/Loss Records

Best all-time record:
1906 Chicago Cubs: 116/36 .763

Worst all-time record:
1916 Philadelphia Athletics: 36/117 .235

Historically most successful 1901-1960:
NY Highlanders/Yankees 1903-1960: 5105/8809 .5795

Historically least successful 1901-1960:
Philadelphia Phillies 1901-1960: 3828/8786 .4357
Boston Braves 1905-1952: 2627/5923 .4435
Bradley-Terry Competition Model

Zermelo (1929)
Bradley & Terry (1952)
Bradley-Terry Competition Model

- Each team $i$ has strength $x_i$, $i=1,2,\ldots,n$ (fixed in each season, can change between seasons)

Zermelo (1929)
Bradley & Terry (1952)
Bradley-Terry Competition Model

- each team \( i \) has strength \( x_i \), \( i=1,2,\ldots,n \)
  (fixed in each season, can change between seasons)

- probability that team \( i \) beats \( j \):
  actually, assume only:

\[
 p_{ij} = \frac{x_i}{x_i + x_j}
\]

\[
 p_{ij} = \frac{f(x_i)}{f(x_i) + f(x_j)}
\]

numerical studies: Anthology of Statistics in Sports, Albert et al. (2005)
Bradley-Terry Competition Model

- Each team $i$ has strength $x_i$, $i=1,2,...,n$ (fixed in each season, can change between seasons)

- Probability that team $i$ beats $j$: actually, assume only:

$$ p_{ij} = \frac{x_i}{x_i + x_j} $$

$$ p_{ij} = \frac{f(x_i)}{f(x_i) + f(x_j)} $$


- Uniform strength distribution:

$$ x \in [\epsilon, 1] \text{ with } 0 \leq \epsilon \leq 1 $$

Fundamental parameter: $\epsilon$

Common assumption: Log-normal distribution
Win Fraction Versus Rank

MLB data for 1905, 10, 15,..., 1960

data from www.shrpsports.com
Average Win Fraction Versus Rank

MLB 1901-60

![Graph showing the average win fraction versus rank for MLB 1901-60.](image-url)
Average Win Fraction Versus Rank

MLB 1901-60

The graph shows the average win fraction $W(r)$ versus the rank $r$. The data points follow a clear upward trend, indicating an increase in win fraction as the rank increases. The win fraction at the lowest rank is approximately 0.32, and it increases to around 0.67 at the highest rank. The trend suggests a strong correlation between rank and win fraction.
Average Win Fraction Versus Rank

MLB 1901-60, 1961-2005

W(r)

0.32
0.36
0.63
0.67

1901-60
1961-2005
Average Win Fraction Versus Rank

Simulations of BT model for $10^6$ seasons

$W(r)$

$\epsilon = 0.278$

$\epsilon = 0.435$

balanced schedule

1961-2005

1901-60

0.36

0.32
Theory for the Average Win Fraction $W$
Theory for the Average Win Fraction $W$

for a team of strength $x$:

$$W(x) = \frac{1}{N} \sum_{j=1}^{N} \frac{x}{x + x_j}$$
Theory for the Average Win Fraction $W$

for a team of strength $x$:

$$W(x) = \frac{1}{N} \sum_{j=1}^{N} \frac{x}{x + x_j}$$

$$\rightarrow \frac{1}{1 - \epsilon} \int_{\epsilon}^{1} \frac{x}{x + y} \, dy$$

- infinite season length
- infinite number of teams
- balanced schedule
- uniform strengths on $[\epsilon, 1]$
Theory for the Average Win Fraction $W$

for a team of strength $x$:

$$W(x) = \frac{1}{N} \sum_{j=1}^{N} \frac{x}{x + x_j}$$

$$\rightarrow \frac{1}{1 - \epsilon} \int_{\epsilon}^{1} \frac{x}{x + y} \, dy$$

$$= \frac{x}{1 - \epsilon} \ln \left[ \frac{x + 1}{x + \epsilon} \right]$$

- infinite season length
- infinite number of teams
- balanced schedule
- uniform strengths on $[\epsilon, 1]$
Theory for the Average Win Fraction $W$

for a team of strength $x$:

$$ W(x) = \frac{1}{N} \sum_{j=1}^{N} \frac{x}{x + x_j} \rightarrow \frac{1}{1 - \epsilon} \int_{\epsilon}^{1} \frac{x}{x + y} \, dy $$

$$ = \frac{x}{1 - \epsilon} \ln \left[ \frac{x + 1}{x + \epsilon} \right] $$

strength $x \rightarrow$ rank $r$: $x = \epsilon + (1 - \epsilon)r \rightarrow$

$$ W(r) = \left[ \epsilon + (1 - \epsilon)r \right] \ln \left[ \frac{1 + \epsilon + (1 - \epsilon)r}{2\epsilon + (1 - \epsilon)r} \right] $$

- infinite season length
- infinite number of teams
- balanced schedule
- uniform strengths on $[\epsilon, 1]$
Average Win Fraction Versus Rank

\[ W(r) = \varepsilon = 0.278 \]

\[ \varepsilon = 0.435 \]

1961-2005

1901-60

0.32

0.36

0.67

0.63
Average Win Fraction Versus Rank

BT model analytics

\[
W(r) = \varepsilon \cdot r
\]

\(\varepsilon = 0.278\) for 1901-60
\(\varepsilon = 0.435\) for 1961-2005

Win fraction increases with rank, with different slopes for different time periods.
Role of Season Length on Win Fraction

![Graph](image-url)
Role of Season Length on Win Fraction

$W(r)$ vs $r$ for game data, BT model, and 300 game season.
Role of Season Length on Win Fraction

![Graph showing the relationship between season length and win fraction](image)

- **Game data**
- **BT model**
- **300 game season**
- **500 game season**
Role of Season Length on Win Fraction

![Graph showing the relationship between win fraction and season length with data points and model predictions for different season lengths.](image-url)
Role of Season Length on Win Fraction

$W(r)$

- Game data
- BT model
- 300 game season
- 500 game season
- 1000 game season
- 10000 game season
- Theory
Fun Facts About Winning and Losing Streaks
**Fun Facts About Winning and Losing Streaks**

**longest win streaks**

26  1916 New York Giants    (1 tie)
21  1935 Chicago Cubs
20  2002 Oakland Athletics
19  1906 Chicago White Sox    (1 tie)
19  1947 New York Yankees
18  1904 New York Giants
18  1953 New York Yankees
17  1907 New York Giants
17  1912 Washington Senators
17  1916 New York Giants
17  1931 Philadelphia Athletics
16  1909 Pittsburgh Pirates
16  1912 New York Giants
16  1926 New York Yankees
16  1951 New York Giants
16  1977 Kansas City Royals
15  1903 Pittsburgh Pirates
15  1906 New York Highlanders
15  1913 Philadelphia Athletics
15  1924 Brooklyn Dodgers
15  1936 Chicago Cubs
15  1936 New York Giants
15  1946 Boston Red Sox
15  1960 New York Yankees
15  1991 Minnesota Twins
15  2000 Atlanta Braves
15  2001 Seattle Mariners
### Fun Facts About Winning and Losing Streaks

#### Longest Win Streaks

<table>
<thead>
<tr>
<th>Length</th>
<th>Year</th>
<th>Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>1916</td>
<td>New York Giants (1 tie)</td>
</tr>
<tr>
<td>21</td>
<td>1935</td>
<td>Chicago Cubs</td>
</tr>
<tr>
<td>20</td>
<td>2002</td>
<td>Oakland Athletics</td>
</tr>
<tr>
<td>19</td>
<td>1906</td>
<td>Chicago White Sox (1 tie)</td>
</tr>
<tr>
<td>19</td>
<td>1947</td>
<td>New York Yankees</td>
</tr>
<tr>
<td>18</td>
<td>1904</td>
<td>New York Giants</td>
</tr>
<tr>
<td>18</td>
<td>1953</td>
<td>New York Yankees</td>
</tr>
<tr>
<td>17</td>
<td>1907</td>
<td>New York Giants</td>
</tr>
<tr>
<td>17</td>
<td>1912</td>
<td>Washington Senators</td>
</tr>
<tr>
<td>17</td>
<td>1916</td>
<td>New York Giants</td>
</tr>
<tr>
<td>17</td>
<td>1931</td>
<td>Philadelphia Athletics</td>
</tr>
<tr>
<td>16</td>
<td>1909</td>
<td>Pittsburgh Pirates</td>
</tr>
<tr>
<td>16</td>
<td>1912</td>
<td>New York Giants</td>
</tr>
<tr>
<td>16</td>
<td>1926</td>
<td>New York Yankees</td>
</tr>
<tr>
<td>16</td>
<td>1951</td>
<td>New York Giants</td>
</tr>
<tr>
<td>16</td>
<td>1977</td>
<td>Kansas City Royals</td>
</tr>
<tr>
<td>16</td>
<td>1903</td>
<td>Pittsburgh Pirates</td>
</tr>
<tr>
<td>15</td>
<td>1906</td>
<td>New York Highlanders</td>
</tr>
<tr>
<td>15</td>
<td>1913</td>
<td>Philadelphia Athletics</td>
</tr>
<tr>
<td>15</td>
<td>1924</td>
<td>Brooklyn Dodgers</td>
</tr>
<tr>
<td>15</td>
<td>1936</td>
<td>Chicago Cubs</td>
</tr>
<tr>
<td>15</td>
<td>1936</td>
<td>New York Giants</td>
</tr>
<tr>
<td>15</td>
<td>1946</td>
<td>Boston Red Sox</td>
</tr>
<tr>
<td>15</td>
<td>1960</td>
<td>New York Yankees</td>
</tr>
<tr>
<td>15</td>
<td>1991</td>
<td>Minnesota Twins</td>
</tr>
<tr>
<td>15</td>
<td>2000</td>
<td>Atlanta Braves</td>
</tr>
<tr>
<td>15</td>
<td>2001</td>
<td>Seattle Mariners</td>
</tr>
</tbody>
</table>

#### Longest Losing Streaks

<table>
<thead>
<tr>
<th>Length</th>
<th>Year</th>
<th>Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>1961</td>
<td>Philadelphia Phillies</td>
</tr>
<tr>
<td>21</td>
<td>1988</td>
<td>Baltimore Orioles</td>
</tr>
<tr>
<td>20</td>
<td>1906</td>
<td>Boston Americans</td>
</tr>
<tr>
<td>20</td>
<td>1906</td>
<td>Philadelphia As</td>
</tr>
<tr>
<td>19</td>
<td>1906</td>
<td>Boston Beaneaters</td>
</tr>
<tr>
<td>19</td>
<td>1914</td>
<td>Cincinnati Reds</td>
</tr>
<tr>
<td>19</td>
<td>1975</td>
<td>Detroit Tigers</td>
</tr>
<tr>
<td>19</td>
<td>2005</td>
<td>Kansas City Royals</td>
</tr>
<tr>
<td>18</td>
<td>1920</td>
<td>Philadelphia As</td>
</tr>
<tr>
<td>18</td>
<td>1948</td>
<td>Washington Senators</td>
</tr>
<tr>
<td>18</td>
<td>1959</td>
<td>Washington Senators</td>
</tr>
<tr>
<td>17</td>
<td>1926</td>
<td>Boston Red Sox</td>
</tr>
<tr>
<td>17</td>
<td>1962</td>
<td>NY Mets (first year)</td>
</tr>
<tr>
<td>17</td>
<td>1977</td>
<td>Atlanta Braves</td>
</tr>
<tr>
<td>16</td>
<td>1911</td>
<td>Boston Braves</td>
</tr>
<tr>
<td>16</td>
<td>1907</td>
<td>Boston Doves</td>
</tr>
<tr>
<td>16</td>
<td>1907</td>
<td>Boston Americans</td>
</tr>
<tr>
<td>15</td>
<td>1909</td>
<td>St. Louis Browns</td>
</tr>
<tr>
<td>15</td>
<td>1911</td>
<td>Boston Rustlers</td>
</tr>
<tr>
<td>15</td>
<td>1927</td>
<td>Boston Braves</td>
</tr>
<tr>
<td>15</td>
<td>1927</td>
<td>Boston Red Sox</td>
</tr>
<tr>
<td>15</td>
<td>1935</td>
<td>Boston Braves</td>
</tr>
<tr>
<td>15</td>
<td>1937</td>
<td>Philadelphia As</td>
</tr>
<tr>
<td>15</td>
<td>2002</td>
<td>Tampa Bay</td>
</tr>
<tr>
<td>15</td>
<td>1972</td>
<td>Texas Rangers (first year)</td>
</tr>
</tbody>
</table>

*Note: Some streaks include additional notes or clarifications.*
## Fun Facts About Winning and Losing Streaks

### Longest Win Streaks

<table>
<thead>
<tr>
<th>Length</th>
<th>Year</th>
<th>Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>1916</td>
<td>New York Giants (1 tie)</td>
</tr>
<tr>
<td>21</td>
<td>1935</td>
<td>Chicago Cubs</td>
</tr>
<tr>
<td>20</td>
<td>2002</td>
<td>Oakland Athletics</td>
</tr>
<tr>
<td>19</td>
<td>1906</td>
<td>Chicago White Sox (1 tie)</td>
</tr>
<tr>
<td>19</td>
<td>1947</td>
<td>New York Yankees</td>
</tr>
<tr>
<td>18</td>
<td>1904</td>
<td>New York Giants</td>
</tr>
<tr>
<td>18</td>
<td>1953</td>
<td>New York Yankees</td>
</tr>
<tr>
<td>17</td>
<td>1907</td>
<td>New York Giants</td>
</tr>
<tr>
<td>17</td>
<td>1912</td>
<td>Washington Senators</td>
</tr>
<tr>
<td>17</td>
<td>1916</td>
<td>New York Giants</td>
</tr>
<tr>
<td>17</td>
<td>1931</td>
<td>Philadelphia Athletics</td>
</tr>
<tr>
<td>16</td>
<td>1909</td>
<td>Pittsburgh Pirates</td>
</tr>
<tr>
<td>16</td>
<td>1912</td>
<td>New York Giants</td>
</tr>
<tr>
<td>16</td>
<td>1926</td>
<td>New York Yankees</td>
</tr>
<tr>
<td>16</td>
<td>1951</td>
<td>New York Giants</td>
</tr>
<tr>
<td>16</td>
<td>1977</td>
<td>Kansas City Royals</td>
</tr>
<tr>
<td>15</td>
<td>1903</td>
<td>Pittsburgh Pirates</td>
</tr>
<tr>
<td>15</td>
<td>1906</td>
<td>New York Highlanders</td>
</tr>
<tr>
<td>15</td>
<td>1913</td>
<td>Philadelphia Athletics</td>
</tr>
<tr>
<td>15</td>
<td>1924</td>
<td>Brooklyn Dodgers</td>
</tr>
<tr>
<td>15</td>
<td>1936</td>
<td>Chicago Cubs</td>
</tr>
<tr>
<td>15</td>
<td>1936</td>
<td>New York Giants</td>
</tr>
<tr>
<td>15</td>
<td>1946</td>
<td>Boston Red Sox</td>
</tr>
<tr>
<td>15</td>
<td>1960</td>
<td>New York Yankees</td>
</tr>
<tr>
<td>15</td>
<td>1991</td>
<td>Minnesota Twins</td>
</tr>
<tr>
<td>15</td>
<td>2000</td>
<td>Atlanta Braves</td>
</tr>
<tr>
<td>15</td>
<td>2001</td>
<td>Seattle Mariners</td>
</tr>
</tbody>
</table>

### Longest Losing Streaks

<table>
<thead>
<tr>
<th>Length</th>
<th>Year</th>
<th>Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>1961</td>
<td>Philadelphia Phillies</td>
</tr>
<tr>
<td>21</td>
<td>1988</td>
<td>Baltimore Orioles</td>
</tr>
<tr>
<td>20</td>
<td>1906</td>
<td>Boston Americans</td>
</tr>
<tr>
<td>20</td>
<td>1906</td>
<td>Philadelphia As</td>
</tr>
<tr>
<td>19</td>
<td>1906</td>
<td>Boston Beaneaters</td>
</tr>
<tr>
<td>19</td>
<td>1914</td>
<td>Cincinnati Reds</td>
</tr>
<tr>
<td>19</td>
<td>1975</td>
<td>Detroit Tigers</td>
</tr>
<tr>
<td>19</td>
<td>2005</td>
<td>Kansas City Royals</td>
</tr>
<tr>
<td>18</td>
<td>1920</td>
<td>Philadelphia As</td>
</tr>
<tr>
<td>18</td>
<td>1948</td>
<td>Washington Senators</td>
</tr>
<tr>
<td>18</td>
<td>1959</td>
<td>Washington Senators</td>
</tr>
<tr>
<td>17</td>
<td>1926</td>
<td>Boston Red Sox</td>
</tr>
<tr>
<td>17</td>
<td>1962</td>
<td>NY Mets (first year)</td>
</tr>
<tr>
<td>17</td>
<td>1977</td>
<td>Atlanta Braves</td>
</tr>
<tr>
<td>16</td>
<td>1911</td>
<td>Boston Braves</td>
</tr>
<tr>
<td>16</td>
<td>1907</td>
<td>Boston Doves</td>
</tr>
<tr>
<td>16</td>
<td>1907</td>
<td>Boston Americans (2 ties)</td>
</tr>
<tr>
<td>16</td>
<td>1944</td>
<td>Brooklyn Dodgers (1 made-up game)</td>
</tr>
<tr>
<td>15</td>
<td>1909</td>
<td>St. Louis Browns</td>
</tr>
<tr>
<td>15</td>
<td>1911</td>
<td>Boston Rustlers</td>
</tr>
<tr>
<td>15</td>
<td>1927</td>
<td>Boston Braves</td>
</tr>
<tr>
<td>15</td>
<td>1927</td>
<td>Boston Red Sox</td>
</tr>
<tr>
<td>15</td>
<td>1935</td>
<td>Boston Braves</td>
</tr>
<tr>
<td>15</td>
<td>1937</td>
<td>Philadelphia As</td>
</tr>
<tr>
<td>15</td>
<td>2002</td>
<td>Tampa Bay</td>
</tr>
<tr>
<td>15</td>
<td>1972</td>
<td>Texas Rangers (first year)</td>
</tr>
</tbody>
</table>

- **27 ≥15-game win/loss streaks between 1901-30**
- **13 between 1931-60**
- **15 between 1961-present**
Winning and Losing Streak Distributions

MLB data

$P_n$ vs. $n$ for 1901-60 and 1961-2005.
Winning and Losing Streak Distributions

MLB data

\[ P_n \]

\[ 1961-2005 \]

\[ 1901-60 \]

\[ 2^{-n} \]
Winning and Losing Streak Distributions

BT model simulations

\[ P_n \]

\[ 2^{-n} \]

1961-2005

1901-60

\[ \varepsilon = 0.435 \quad \varepsilon = 0.278 \]
Streak Length Distribution in BT Model

\[ P_n(x) = \prod_{j=1}^{n} \frac{x}{x + x_j} \cdot \frac{x_0}{x + x_0} \cdot \frac{x_{n+1}}{x + x_{n+1}} \]
Streak Length Distribution in BT Model

\[ P_n(x) = \prod_{j=1}^{n} \frac{x}{x + x_j} \frac{x_0}{x + x_0} \frac{x_{n+1}}{x + x_{n+1}} \]

\[ \langle P_n(x) \rangle \{x_j\} = x^n \left\langle \frac{1}{x + y} \right\rangle^n \left\langle \frac{y}{x + y} \right\rangle^2 \]

\[ \left\langle \frac{1}{x + y} \right\rangle = \frac{1}{1 - \epsilon} \int_\epsilon^1 \frac{dy}{x + y} = \frac{1}{1 - \epsilon} \ln \left( \frac{x + 1}{x + \epsilon} \right) \]

\[ \left\langle \frac{y}{x + y} \right\rangle = 1 - \frac{x}{1 - \epsilon} \ln \left( \frac{x + 1}{x + \epsilon} \right) \]
Streak Length Distribution in BT Model

\[ P_n(x) = \prod_{j=1}^{n} \frac{x}{x+x_j} \frac{x_0}{x+x_0} \frac{x_{n+1}}{x+x_{n+1}} \]

\[
\langle P_n(x) \rangle \{x_j\} = x^n \left\langle \frac{1}{x+y} \right\rangle^n \left\langle \frac{y}{x+y} \right\rangle^2 \left\langle \frac{1}{x+y} \right\rangle = \frac{1}{1-\epsilon} \int_{\epsilon}^{1} \frac{dy}{x+y} = \frac{1}{1-\epsilon} \ln \left( \frac{x+1}{x+\epsilon} \right)
\]

\[
\left\langle \frac{y}{x+y} \right\rangle = 1 - \frac{x}{1-\epsilon} \ln \left( \frac{x+1}{x+\epsilon} \right)
\]

\[
\langle P_n \rangle = \frac{1}{1-\epsilon} \int_{\epsilon}^{1} \left[ 1 - \frac{x}{1-\epsilon} \ln \left( \frac{x+1}{x+\epsilon} \right) \right]^2 e^{ng(x)} \, dx
\]

\[
g(x) = \ln x + \ln \left[ \frac{1}{1-\epsilon} \ln \left( \frac{x+1}{x+\epsilon} \right) \right]
\]

monotonic
Streak Length Distribution in BT Model

\[ P_n(x) = \prod_{j=1}^{n} \frac{x}{x + x_j} \frac{x_0}{x + x_0} \frac{x_{n+1}}{x + x_{n+1}} \]

\[ \langle P_n(x) \rangle \{x_j\} = x^n \left( \frac{1}{x + y} \right)^n \left( \frac{y}{x + y} \right)^2 \left\langle \frac{1}{x + y} \right\rangle = \frac{1}{1 - \epsilon} \int_{\epsilon}^{1} \frac{dy}{x + y} = \frac{1}{1 - \epsilon} \ln \left( \frac{x + 1}{x + \epsilon} \right) \]

\[ \left\langle \frac{y}{x + y} \right\rangle = 1 - \frac{x}{1 - \epsilon} \ln \left( \frac{x + 1}{x + \epsilon} \right) \]

\[ \langle P_n \rangle = \frac{1}{1 - \epsilon} \int_{\epsilon}^{1} \left[ 1 - \frac{x}{1 - \epsilon} \ln \left( \frac{x + 1}{x + \epsilon} \right) \right]^2 e^{ng(x)} \, dx \]

\[ g(x) = \ln x + \ln \left[ \frac{1}{1 - \epsilon} \ln \left( \frac{x + 1}{x + \epsilon} \right) \right] \]

\[ \sim \frac{1}{ng'(1)} e^{ng(1)} \quad \text{monotonic} \]
Asymptotic Behavior

\[ \langle P_n \rangle \sim \frac{1}{ng'(1)} e^{ng(1)} \]

\[ g(x) = \ln x + \ln \left[ \frac{1}{1 - \epsilon} \ln \left( \frac{x + 1}{x + \epsilon} \right) \right] \]

\[ g(1) = -\ln(1 - \epsilon) + \ln \ln \left( \frac{2}{1 + \epsilon} \right) \]
Asymptotic Behavior

\[ \langle P_n \rangle \sim \frac{1}{ng'(1)} e^{ng(1)} \]

\[ g(x) = \ln x + \ln \left[ \frac{1}{1 - \epsilon} \ln \left( \frac{x + 1}{x + \epsilon} \right) \right] \]

\[ g(1) = -\ln(1 - \epsilon) + \ln \ln \left( \frac{2}{1 + \epsilon} \right) \]

fastest decay when \( \epsilon = 1 \): \( \langle P_n \rangle \sim 2^{-n} \)

slowest decay when \( \epsilon = 0 \): \( \langle P_n \rangle \sim (\ln 2)^n \approx (0.693)^n \)
Winning and Losing Streak Distributions
Winning and Losing Streak Distributions numerically integrate

\[
\langle P_n \rangle = \frac{1}{1-\epsilon} \int_\epsilon^1 \left[ 1 - \frac{x}{1-\epsilon} \ln \left( \frac{x+1}{x+\epsilon} \right) \right]^2 e^{ng(x)} \, dx
\]
Winning and Losing Streak Distributions

numerically integrate

$$\langle P_n \rangle = \frac{1}{1 - \epsilon} \int_\epsilon^1 \left[ 1 - \frac{x}{1 - \epsilon} \ln \left( \frac{x + 1}{x + \epsilon} \right) \right]^2 e^{ng(x)} \, dx$$

same as: $10^5$ randomizations of all 2166 win/loss histories
Summary/Outlook
Summary/Outlook

Citations:
Page rank analysis: *helps uncover hidden “gems”*
Future: *contextual analysis, specialization*
Summary/Outlook

Citations:
Page rank analysis: *helps uncover hidden “gems”*
Future: *contextual analysis, specialization*

Climate:
Global warming seems to affect temperature records
Open: *seasonality, high/low asymmetry, changing variability*
Inter-day Temperature Correlations

\[ C_\alpha(t) = \frac{1}{365} \sum_{i=1}^{365} \frac{\langle T_i T_{i+t} \rangle - \langle T_i \rangle \langle T_{i+t} \rangle}{\langle T_i^2 \rangle - \langle T_i \rangle^2} \quad \alpha = \text{low, middle, high} \]
Seasonal Variance

daily variances (10-day averages)
day of the year

low
average
high
Seasonal Variance & Record Numbers

- Seasonal Variance
- Daily variances (10-day averages)
- Number of record highs (20-day average)
Summary/Outlook

Citations:
- Page rank analysis: *helps uncover hidden “gems”*
- Future: *contextual analysis, specialization*

Climate:
- Global warming seems to affect temperature records
- Open: *seasonality, high/low asymmetry, changing variability*
Summary/Outlook

Citations:
- Page rank analysis: *helps uncover hidden “gems”*
- Future: *contextual analysis, specialization*

Climate:
- Global warming seems to affect temperature records
- Open: *seasonality, high/low asymmetry, changing variability*

Competition:
- Statistical model → *win/loss records, streak distributions*
  *keener competition; harder to maintain advantage*