An Objective Assessment of Point Scoring Systems in Formula 1 Motor Racing

Robin K. S. Hankin
Associate Professor
Department of Engineering, Computer and Mathematical Sciences
Auckland University of Technology

Abstract

The points scoring system of Formula 1 motor racing is a long-standing contention among fans and competitors. The inaugural points system was (8,6,4,3,2), that is, 8 points to the winner, 6 to second place, and so on; in 2020 the points system was changed to (25,18,15,12,10,8,6,4,2,1). However, it is difficult to assess statistical nulls using accumulated points. Here, I use Plackett-Luce likelihood to identify a ranking for the competitors, this method being amenable to statistical testing. I go on to assess a number of reasonable points systems objectively: in one well-defined statistical sense a Borda points system is optimal.

Keywords
Reified Bradley-Terry, Likelihood, Formula 1 motor racing, Points systems

Introduction

Formula 1 motor racing is an important and prestigious motor sport (Codling, 2017; Jenkins, 2010). Season ranking is based on a points allocation system wherein competitors are awarded points based on race finishing order; points accumulate additively. The overall competition winner is the competitor who accumulates the most points after the final race. The intent of the points system is to incentivize competitors, stimulate innovation, and to create an exciting sporting spectacle: as such, its study is a practical application of tournament theory (Lazear and Rosen, 1981).

However, in the case of Formula 1 motor racing, the points system is the subject of much controversy, having changed often since the competition's inauguration in 1950 when the points allocation was (8, 6, 4, 3, 2)—eight points to the winner, six for second place, and so on. This system credits only the first five finishers. As of 2020, the current points system of (25, 18, 15, 12, 10, 8, 6, 4, 2, 1) credits the first 10 (we ignore the bonus point awarded for fastest lap and assume a strictly monotonic decrease). Arguably these two systems could introduce different rational behavior under zero-sum assumptions: if, in a race, a driver knows he will place seventh under a low-risk strategy but may place sixth by dint of driving more aggressively, the low-risk strategy might be rational under the first points system (which does not reward the extra ranking), but not under the second, which does.

Still, drivers have strong incentives to maximize their ranking irrespective of any points that may be awarded: sponsors and teams note details of drivers' performance, and a great deal of personal pride may be at stake (Gay-Rees, 2019). It is therefore reasonable to assume that each driver strives to maximize his rank, and this will be done here. If this is so, then changing the points system might change the competitors' rankings (Wood, 2020) but not their behavior: surely a defect of using points to rank the competitors.

Given that racing is a zero-sum game—and that points are monotonically decreasing— each player will try to get as high a rank as possible regardless of the actual points system used. However, there are other consistent interpretations. Bakhrankova (2011), for example, considers the possibility of inter-driver collusion, a phenomenon not pursued here; and Mastromarco and Runkel (2009) suggest that the frequency of rule changes is driven by factors such as driver safety and revenue optimization.

Points systems similar to that of Formula 1 are common in other racing sports; all have the common feature of translating ranks into points which combine additively to generate an overall ranking. Further, we see points systems used in the wider context of competitive situations such as the Eurovision Song Contest and many other such group tournaments. Therefore, the ideas used here for analysis of motorsports furnish a methodology that is directly applicable to a broad range of competitive situations in which points are used to rank competitors.

Bradley-Terry and generalizations for rank statistics

The Bradley-Terry model (Bradley, 1952) assigns non-negative strengths $p_1, p_2, \ldots, p_n$ to each of n competitors in such a way that the probability of i beating $j\neq i$ in pairwise competition is $\frac{p_i}{p_i+p_j}$; it is conventional to normalize so that $\sum p_i = 1$. Further, we use a generalization due to Luce (1959), in which the probability of competitor i winning in a field of $\{1, 2, \ldots, n\}$ is $\frac{p_i}{p_1+\cdots+p_n}$. Noting that there is information in the whole of the finishing order, and not just the first across the line, we can follow Plackett (1975) and consider the runner-up to be the winner among the remaining competitors, and so on down the finishing order. Without loss of generality, if the order of finishing were 1, 2, 3, 4, 5, then a suitable Plackett-Luce likelihood function would be

$$\frac{p_1}{p_1+p_2+p_3+p_4+p_5}\cdot\frac{p_2}{p_2+p_3+p_4+p_5}\cdot\frac{p_3}{p_3+p_4+p_5}\cdot\frac{p_4}{p_4+p_5}$$

and this would be a forward ranking Plackett-Luce model in the terminology of Mollica and Tardella (2014). A slight generalization allows the incorporation of nonfinishers (DNF etc). If, say, competitors 4 and 5 did not finish, we would have

$$\frac{p_1}{p_1+p_2+p_3+p_4+p_5}\cdot\frac{p_2}{p_2+p_3+p_4+p_5}$$

(observe how this likelihood function, while informative about $p_4 + p_5$, is uninformative about $p_4 | \ p_4 + p_4$). We now use a technique due to Hankin (2010, 2020) and introduce fictional (reified) entities whose nonzero Bradley-Terry strength helps certain competitors or sets of competitors under certain conditions. The canonical example would be the home-ground advantage in association football. If players (teams) 1, 2 with strengths p_1, p_2 compete, and if our observation were a home wins and b away wins for team 1, and c home wins and d away wins for team 2, then a suitable likelihood function would be

$$\left(\frac{p1 + pH}{p1 + p2 + pH}\right)^a\left(\frac{p1 + pH}{p1 + p2 + pH}\right)^b\left(\frac{p1 + pH}{p1 + p2 + pH}\right)^c\left(\frac{p1 + pH}{p1 + p2 + pH}\right)^d$$

where pH is a quantification of the beneficial home ground effect. Similar techniques have been used to account for the first-move advantage in chess, and effective coordination between members of doubles tennis teams; we may use a similar device to account for (e.g.) wet conditions in Formula 1. Here I analyze seasons 2016-2019 using the hyper2 package (Hankin, 2017) which implements the Plackett-Luce likelihood function with additional reified entities Hankin (2020).

One component of Formula 1 motor racing is the starting grid. Placing on the starting grid is determined by time trials usually driven the day before the race itself. Pole position is awarded to the driver with the fastest qualifying time, and this confers a considerable advantage to the sitter. In this analysis we do not consider pole position specifically but attempt to make inferences about the time trials and the race itself in combination (alternatively, we treat grid placing and P-L strengths to be conditionally independent, given race ranking). Similarly, we treat the driver and the team as a single entity about which we wish to make inferences.

Formula 1 dataset

Taking 2017 as an example, Table 1 shows the drivers' ranks. It is straightforward to translate this table into a Plackett-Luce likelihood function using the hyper2 package; for simplicity we will consider only the 11 top-ranked drivers (in the Plackett-Luce likelihood function, the performance of lower-ranked players can be weakly informative about higher-ranked players' strengths. For example, we see that Vettel retired twice—in Singapore and Japan—so any player who placed in those venues will effectively "steal" strength from Vettel, and generally "give" it to Hamilton or Bottas). Although it has many

Table 1*:*
2017 Season Results Table

DRIVER	AUSTRALIA	CHINA	BAHRAIN	RUSSIA	SPAIN	MONACO	…	BRAZIL	ABU DHABI
Hamilton	2	1	2	4	1	7	…	4	2
Vettel	1	2	1	2	2	1	…	1	3
Bottas	3	6	3	1	Ret	4	…	2	1
Räikkönen	4	5	4	3	Ret	2	…	3	4
Ricciardo	Ret	4	5	Ret	3	3	…	6	Ret
…	…	…	…	…	…	…	…	…	…
Hartley	0	0	0	0	0	0	…	0	15
Button	0	0	0	0	0	Ret	…	0	0
Resta	0	0	0	0	0	0	…	0	0

Each row is a driver and each column (after the first) a venue. We see that Hamilton, the first row, came second in Australia, first in China, second in Bahrain, fourth in Russia, and so on (Hartley, Button, and Resta placed last). In the first column, we see the result from Australia in which Hamilton came second, Vettel first, Bottas third, and so on. Here, "Ret" means "retired" and a zero entry means "did not finish".

terms, the overall likelihood expression for the 2017 season is of the general form

$$\frac{p_{\mathrm{Ham}}^{20}\,p_{\mathrm{Mas}}^{16}\,p_{\mathrm{Bot}}^{19}\cdots}{(p_{\mathrm{Per}}+p_{\mathrm{Oco}})(p_{\mathrm{Ver}}+p_{\mathrm{Per}}+p_{\mathrm{Sai}}+p_{\mathrm{Hul}}+p_{\mathrm{Mas}})}$$

$$(p_{\mathrm{Ric}}+p_{\mathrm{Per}}+p_{\mathrm{Mas}})(p_{\mathrm{Sai}}+p_{\mathrm{Mas}})$$

$$(p_{\mathrm{Ver}}+p_{\mathrm{Per}}+p_{\mathrm{Oco}}+p_{\mathrm{Sai}}+p_{\mathrm{Hul}})\cdots$$

Finding the maximum likelihood estimate for the players' strengths is straightforward numerically. The hyper2 package includes a suite of numerical optimization routines, and because they have access to derivatives, convergence is rapid. A graphical diagram of the strengths is given in Figure 1. We see that in 2016 the driver with the largest estimated strength was Rosberg at about 30%, and in years 2017-2019 was Hamilton at about 29%, 37%, and 42% respectively. As an illustration of the value of likelihood methods (as opposed to points-based methods), a likelihood ratio test [samep.test(), supplied with the hyper2 package] rejects the null that Hamilton and Vettel have the same strength in 2018 (H_0: $p_{\mathrm{Ham}} = p_{\mathrm{Vet}}$) with a likelihood ratio of $e^{2.76} \simeq 15.8$, corresponding by Wilks's theorem to an asymptotic p-value of about 0.02. We may also use the reified entity concept to test the null hypothesis that Hamilton's strength was unchanged from 2016, where Rosberg had the highest estimated strength, to 2017-2019, where Hamilton did; we fail to reject this null.

Likelihood scoring vs points scoring

Applying the current points system, for example, to the 2017 results table we would rank the drivers as follows:

Hamilton > Vettel > Bottas > Räikkönen > Ricciardo > Verstappen >Peréz > Ocon > Sainz > Hülkenberg > Massa

(this happens to be identical to the ranking after the extra point was awarded for fastest lap). However, if we were to adopt a Zipfian points system of $(1, \frac{1}{2}, \frac{1}{3}, \ldots)$ we would have

Hamilton > Vettel > Bottas > Ricciardo > Verstappen > Räikkönen > Peréz > Ocon > Massa> Sainz > Hülkenberg

[note that "Zipf's law" refers to the probability mass function (Zipf 1949); here I am using it simply as a monotonically decreasing sequence]. Thus, these two systems agree on the first three places but fourth is awarded to Räikkönen under the current F1 system and Ricciardo under Zipf. Compare the likelihood ranking:

Hamilton > Vettel > Bottas > Räikkönen > Ocon > Ricciardo > Verstappen > Peréz > Massa > Sainz > Hülkenberg

So, for 2017 at least, we see that the current points system agrees with likelihood ranking for the top four places, while a Zipfian system agrees to three. We can plot one ranking against the other, shown in Figure 2. Note that the historically correct points awarded to the drivers differs from that calculated here. That is for two reasons: firstly, "fastest lap" points are not included here, and also the truncation of the order table to the first 11 drivers can

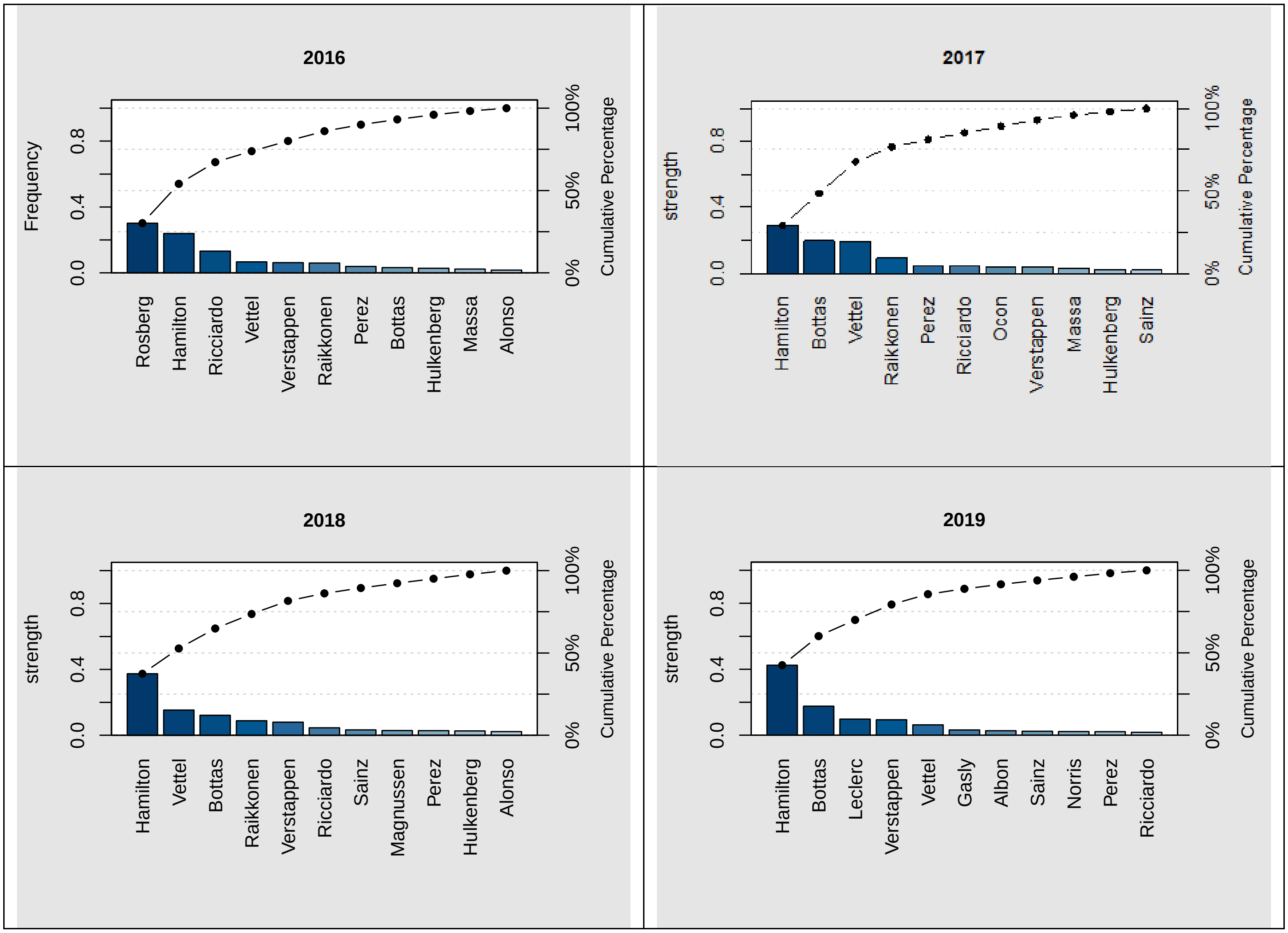

Figure 1:
Maximum likelihood estimates of the strengths of the top-ranked 11 drivers in Formula 1 motor racing, seasons 2016-2019

increase the rank of a driver if non-first-11 drivers are placed.

We thus see a comparison between two ordering systems. Taking 2017 as an example, drivers Hamilton and Vettel are respectively first and second according to both ranking procedures; but Räikkönen and Bottas are third and fourth, and fourth and third, according to points and likelihood respectively. We define the degree of agreement between the two ranking systems as the maximal value of r such that places 1, 2,…, r all match. Thus, from Figure 2, the degree of agreement between likelihood ranking and points ranking for the years 2016-2019 would be 11, 1, 2, 2 respectively.

However, the points system used is essentially arbitrary. We could use, for example, a Zipfian points system to rank the drivers: award one point to the winner, half a point to second place, one third of a point to third, and so on; see Figure 3 in which we see generally poorer agreement between points-based ranks and likelihood-based ranks, with a degree of agreement of 0,1,2,2 for the years 2016-2019 respectively. This might be an indication that using a Zipfian points allocation is objectively worse than the current points system. There are a number of plausible points systems that might be used:

- The current Formula 1 system (25, 18, 15, 12, 10, 8, 6, 4, 2, 1)
- The inaugural Formula 1 system (8, 6, 4, 3, 2)
- Zipfian
- Borda (n, n−1, n−2,..., 3, 2, 1, 0)

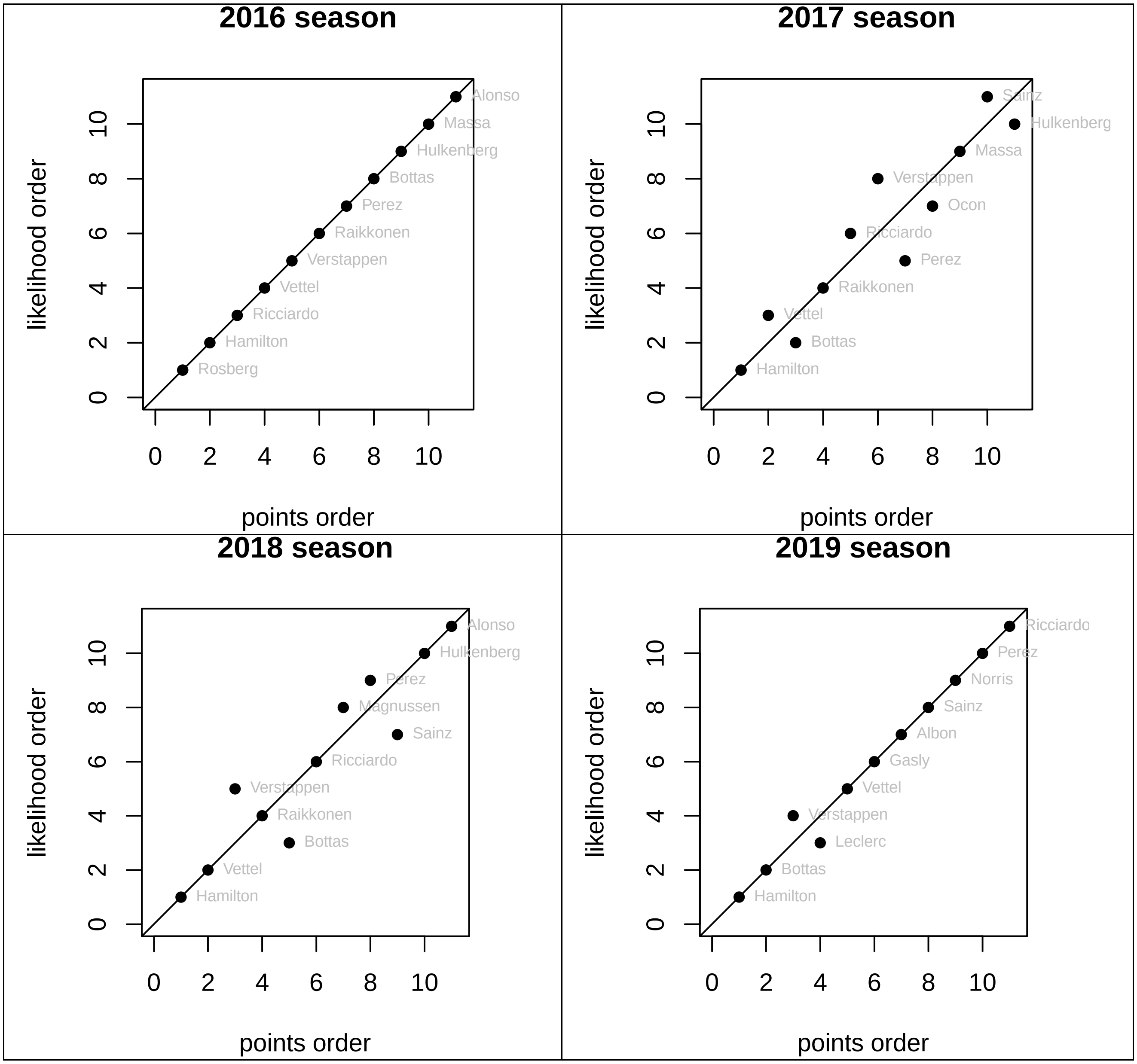

Figure 2:
Ordering of the top-ranked 11 drivers in Formula 1, seasons 2016-2019. Horizontal axis gives official (points-based) order, and the vertical axis gives the likelihood order. Thus, taking 2017 as an example, the points-based ordering would be Hamilton first, then Bottas, then Verstappen; while the likelihood ordering is (reading vertically) Hamilton, Verstappen, Bottas.

- Halving system:
- A “winner takes all” system (1, 0, 0, 0,…)

We note that some of these may be generalized. We might consider a more general Borda-like points system (r, r – 1, r – 2,…, 3, 2, 1, 0) for fixed integer r with 0 < r < n (Emerson, 2007); the halving system can be generalized to a geometric distribution; and the “winner takes all” system can be replaced by giving equal points to the top r competitors (1, 1,…,1, 0, 0,…) where there are r 1’s and (n-r) 0’s, for some integer r<n.

It is straightforward to calculate the degree of agreement for the observed rank table for years 2016-2019, shown in

Table 2:

Points (P-) and likelihood (L-) based rankings for the drivers' championship, Formula 1 seasons 2016-2019. Bold font indicates agreement.

	2016		2017		2018		2019	
Rank	P	L	P	L	P	L	P	L
1	Rosberg	Rosberg	Hamilton	Hamilton	Hamilton	Hamilton	Hamilton	Hamilton
2	Hamilton	Hamilton	Vettel	Bottas	Vettel	Vettel	Bottas	Bottas
3	Ricciardo	Ricciardo	Bottas	Vettel	Verstappen	Bottas	Verstappen	Leclerc
4	Vettel	Vettel	Raikkonen	Raikkonen	Raikkonen	Raikkonen	Leclerc	Verstappen
5	Verstappen	Verstappen	Ricciardo	Perez	Bottas	Verstappen	Vettel	Vettel

---

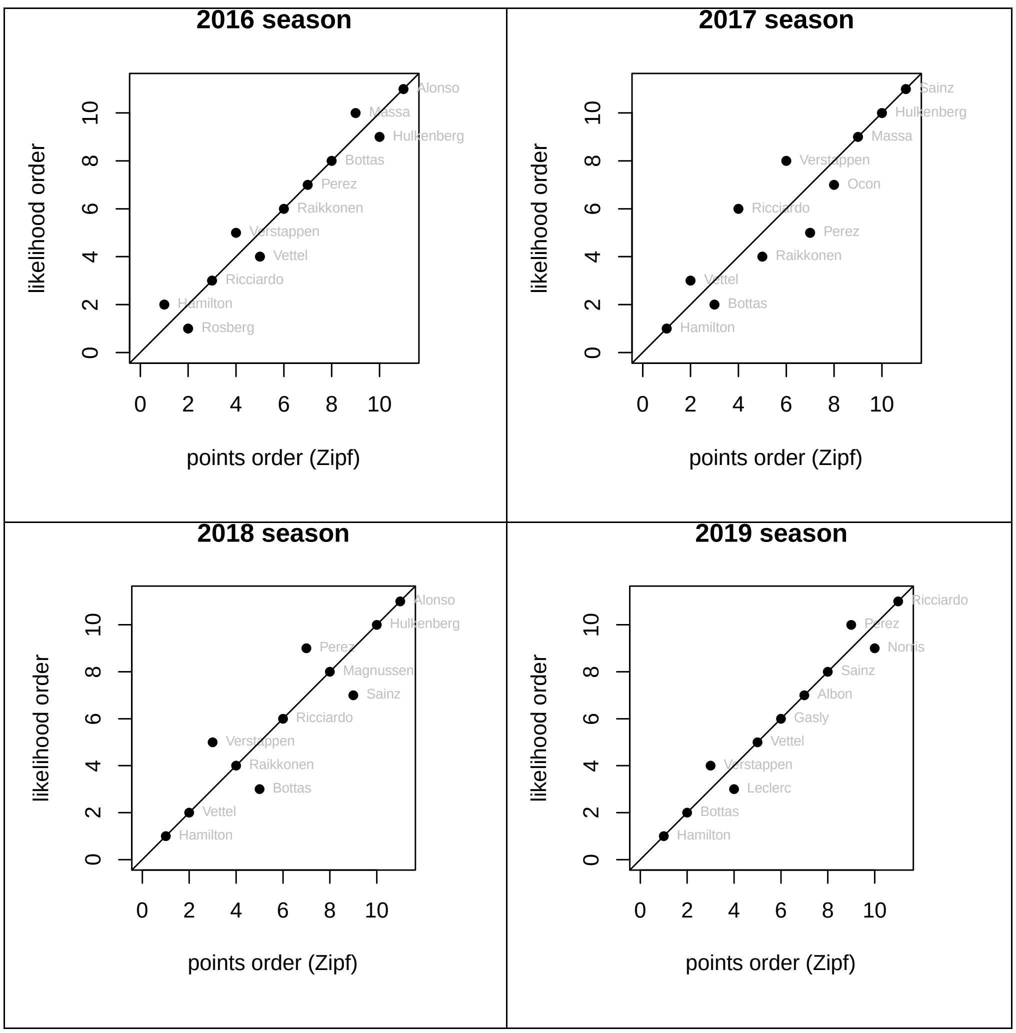

Figure 3:

Points ranking calculated by a Zipfian system. Note the generally poorer agreement between the two ranking systems.

Table 2. It is clear that there is no points system that is the best for all four years. However, observing that both the likelihood ranking and the points ranking are random variables in this paradigm suggests a method whereby we can objectively assess a given points system. Using sampling techniques, we can repeatedly generate an order table in silico, using estimated driver strengths from the observed tables. For each of, say, 1000 such synthetic tables, calculate drivers' maximum likelihood Plackett strengths, and also their points awarded according to any given points system. We then compare rankings generated by the Plackett strengths and the points awarded and note the degree of agreement between the two, as measured by the number of rankings correctly predicted. This furnishes an objective assessment of the points system used.

Numerical results

We now assess the six points systems using the methodology outlined above, using 1000 in silico trials. For each of the six points systems, each of the 1000 trials results in a single non-negative integer: the degree of agreement between the Plackett-Luce ranking and the rankings according to the points system considered. There are three measures that might be used to assess the distribution of degree of agreement: (1), the mean degree of agreement d; (2), the probability of correctly predicting the winner, Prob($d \geq 1$); and (3), the probability of correctly predicting the complete order statistic Prob($D = n$). These summaries are shown for each of the six points systems in Figures 4, 5 and 6 respectively.

We see that the Borda points system (n, n-1, n-2,…, 3, 2, 1, 0) gives the highest mean number of places, and also the highest probability of correctly predicting the winner. However, the probability of predicting the complete order statistic is maximized using the winner takes all system (1, 0, 0,…, 0).

To summarize, the winner takes all system is the optimum points system in the following sense: simulated race results have a higher probability of matching the likelihood-based complete order statistic when using a winner takes all points system than when using any other points system.

Conclusions

Many competitive situations involve ranking the participants and one way of doing this is to assign points for the winner, second place, third place, etc. The overall ranking for a sequence of observations is decided on the basis of accumulated points. Because changing the points system can change participants' overall ranking but does not affect their behavior, accumulated points should not be used to make inferences about participants' skills. Maximum likelihood estimation of Plackett-Luce strengths furnishes a ranking system that does not suffer from the arbitrariness of a points system. Further, this allows one to conduct statistical tests on a range of interesting nulls in the context of an established suite of software.

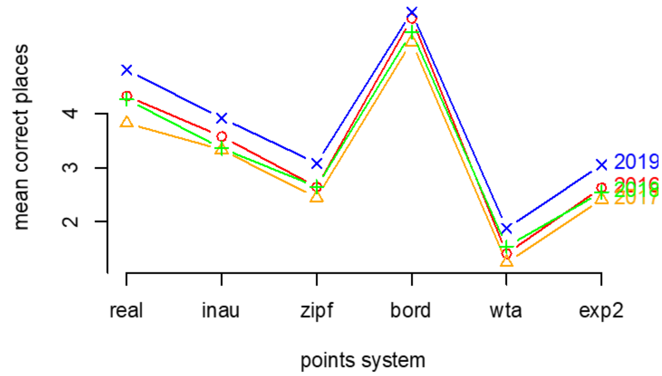

Figure 4:
Simulated Formula 1 races, seasons 2016-2019: mean number of agreeing places for each of six points systems when compared with a Plackett-Luce strength ordering

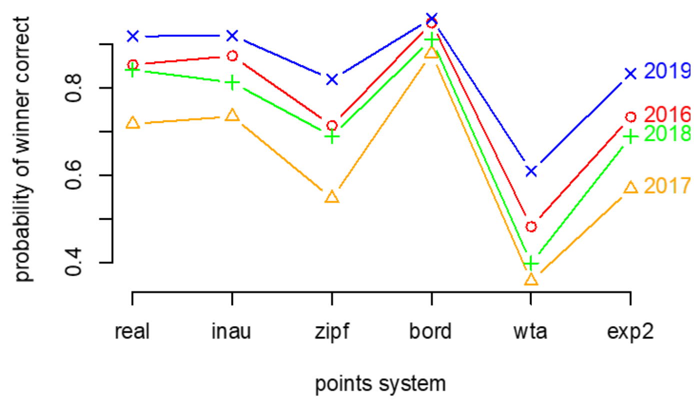

Figure 5:
Simulated Formula 1 races, seasons 2016-2019: probability of identifying the Plackett-Luce winner for each of six points systems

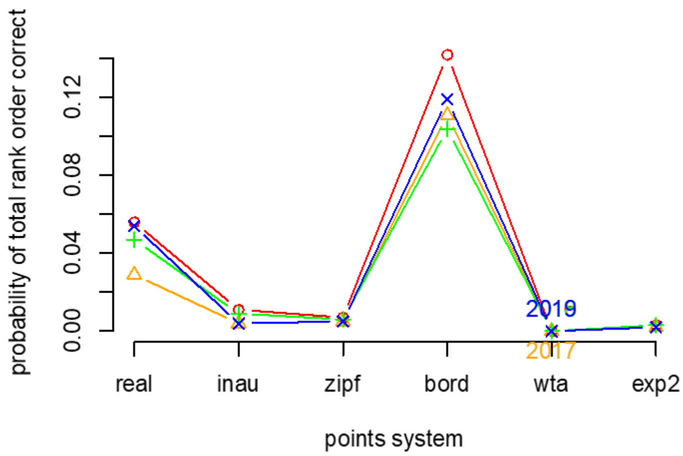

Figure 6:

Simulated Formula 1 races, seasons 2016-2019: probability of complete agreement between Plackett-Luce ranks and likelihood calculated by each of six points systems

The points system used in Formula 1 motor racing is a source of lively debate from many perspectives, with changes being controversial. By treating total points scored as a random variable, it is possible to compare different point allocation schemes against objective Plackett-Luce ranks. Of the six points systems considered here, a Borda system or a winner-takes-all-system appear to be closest to objective Plackett-Luce, depending on the exact definition of "closest". The analysis could help to better understand the impact of different points systems on the overall ranking of drivers and indeed teams. This could be particularly useful for team managers, who may be looking to optimize their team strategy based on the points system in use. Additionally, the analysis could help the motor racing community to evaluate the fairness of different points systems. This could be particularly important for race organizers or governing bodies, who may be seeking to design a points system that is as fair and objective as possible. The analysis could be used to inform debates or discussions around potential changes to the Formula 1 points system. Practitioners might be able to make more informed decisions on strategy; and to use the techniques presented here to guide decision-making.

References

Bakhrankova, K., (2011). Points scoring systems in Formula 1: A game theoretic perspective, in: Klatte, D. (Ed.), *Operations Research Proceedings*, Springer-Verlag. pp. 211–216.

Bradley, R.A., (1952). Incomplete block rank analysis: On the appropriateness of the model for a method of paired comparisons. *Biometrics 10*, 375–390.

Codling, S., (2017). *Speed read F1*. *Motorbooks*. ISBN 978-0-7603-5562-6.

Emerson, P., (2007). *Designing an All-Inclusive Democracy—Consensual Voting Procedures for use in Parliaments, Councils and Committees*. Springer-Verlag. ISBN 978-3-540-33163-6.

Gay-Rees, J., (2019). *Formula 1: drive to survive*. Netflix documentary.

Hankin, R.K.S., (2010). A generalization of the Dirichlet distribution. *Journal of Statistical Software 33*, 1–18.

Hankin, R.K.S., (2017). Partial rank data with the hyper2 package: likelihood functions for generalized Bradley-Terry models. *The R Journal 9*, 429–439.

Hankin, R.K.S., (2020). A generalization of the Bradley-Terry model for draws in chess with an application to collusion. *Journal of Economic Behavior and Organization 180*, 325–333.

Jenkins, M., (2010). Technological discontinuities and competitive advantage: a historical perspective on Formula 1 motor racing 1950-2006. *Journal of Management Studies 47*, 884–910.

Lazear, E.P., Rosen, S., (1981). Rank-order tournaments as optimum labour contracts. *The Journal of Political Economy 89*, 841–864.

Luce, R., (1959). *Individual choice behaviour: A theoretical analysis*. Wiley, New York.

Mastromarco, C., Runkel, M., (2009). Rule changes and competitive balance in Formula One motor racing. *Applied Economics 41*, 3003–3014.

Mollica, C., Tardella, L., (2014). Epitope profiling via mixture modeling of ranked data. *Statistics in Medicine 33*, 3738–3758.

Plackett, R.L., (1975). The analysis of permutations. *Applied Statistics 24*, 193–202.

Wood, E., (2020). When points systems played a part in history. *Formula Scout* (published online).

Zipf, G., (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley.