

HET SIMULEREN VAN FORMULE 1 RACESTRATEGIEËN

Dat het kiezen van de juiste racestrategie in de Formule 1 een cruciale rol kan spelen, werd duidelijk tijdens het wereldkampioenschap in 2010. Met wel vier coureurs in de strijd om de titel was het seizoen ongekend spannend. Fernando Alonso had de beste papieren, maar liep door een strategische fout van zijn team de wereldtitel mis. Ferrari dacht namelijk dat een vroege pitstop de sleutel tot de winst zou zijn, maar achteraf bleek dat Alonso beter langer door had kunnen rijden. Sebastian Vettel, die wel langer doorreed, won uiteindelijk de race en staat nu te boek als de jongste wereldkampioen ooit in de Formule 1.



CLAUDIA SULSTERS

Om een fout in de racestrategie en de daarbij horende gevolgen zoveel mogelijk te voorkomen, gebruiken Formule 1-teams simulatiesoftware om de race vooraf te simuleren met verschillende variabelen. Voor deze simulatiemodellen worden grote hoeveelheden data gebruikt die gegenereerd worden door de sensoren aan boord van de Formule 1-auto's. Dit artikel* beschrijft een simulatiemodel op basis van *open source data*, dat kan worden gebruikt om de optimale racestrategie voor een Formule 1-team te bepalen. Het simulatiemodel beschrijft onder andere de invloed van het brandstofverbruik en de bandenslijtage op de rondetijden. Daarnaast imiteert het de meeste gebeurtenissen op het circuit, zoals de chaos na de start, pitstops, inhaalacties, en het uitvallen van coureurs door een crash of een technisch probleem.

Ontwerp van het simulatiemodel

Elke simulatie begint met het bepalen van de coureurs die de race niet zullen finishen. Zij raken bijvoorbeeld betrokken bij een crash of krijgen een technisch probleem. Voor elke coureur kiezen we willekeurig een racestrategie. Vervolgens verzamelen alle coureurs zich op de *starting grid*, de lichten gaan uit en de race gaat van start. Vlak na de start is het vaak chaos: alle coureurs proberen één of meer plaatsen te winnen en crashes komen vaak voor op dit moment in de race. De volgende stap is het simuleren van een rondetijd voor elke coureur en elke ronde van de race. Deze rondetijd wordt gesimuleerd met behulp van de kwalificatietijd en parameters die de invloed van het brandstofverbruik en de bandenslijtage op de rondetijd beschrijven. Inhaalacties worden toegevoegd door te kij-

ken naar het verschil in (cumulatieve) rondetijd tussen opeenvolgende coureurs. Net als bij een echte Formule 1-race is een inhaalactie niet altijd succesvol. Dit hangt namelijk af van het verschil in rondetijd, waar de inhaalactie wordt ingezet (in een bocht of op een recht stuk) en hoe agressief de andere coureur verdedigt. Aan het einde van elke ronde simuleren we eventuele pitstops. Het simuleren van rondetijden herhaalt zich totdat de zwart-wit geblokte vlag wordt gezwaaid en we weten welke coureur als eerste over de streep is gekomen.

Did Not Finish

Het komt bijna nooit voor dat alle coureurs de finish halen. Het uitvallen van coureurs wordt gesimuleerd met

behulp van een *Did Not Finish*-kans (DNF-kans). Deze kans wordt per coureur bepaald aan de hand van data over het wel of niet finishen in vorige races van dat seizoen. We gebruiken hiervoor de Bayesiaanse methode. Het aantal keer dat een coureur is uitgevallen in n races kan gemodelleerd worden als een binomiale verdeling met parameters n en p , met p de DNF-kans. Voor elke race zijn er namelijk twee mogelijke uitkomsten: wel of niet finishen. Het doel van de Bayesiaanse methode is om de priorverdeling van de DNF-kans te updaten aan de hand van nieuwe data. We kiezen de bètaverdeling als prior omdat deze verdeling twee belangrijke eigenschappen heeft: ten eerste is de bètaverdeling gedefinieerd op het interval $(0,1)$, net als onze DNF-kans, en ten tweede is de bètaverdeling de zogeheten *conjugate prior* van de *Bernoulli likelihood*. Dit laatste betekent dat de posterior-

verdeling ook een bèta-verdeling is. De parameters van de priorverdeling schatten we op basis van alle data van alle coureurs. Voor een individuele coureur updaten we deze prior vervolgens met data over het niet finishen van de betreffende coureur. Zo heeft Daniel Ricciardo, die relatief weinig uitviel, bijvoorbeeld een DNF-kans van 0,12 terwijl Daniil Kvyat, een coureur die veel vaker uitviel, een DNF-kans van 0,24 heeft.

Chaos bij de start

De start van een F1-race is vaak de beste kans voor een coureur om posities te winnen in het veld. Het simulatiemodel gebruikt de empirische verdeling van het aantal posities dat een coureur bij de start in vorige races heeft gewonnen of verloren om de chaos bij de start te simuleren. In deze empirische verdelingen wordt echter maar aan een beperkt aantal posities kansmassa toegekend omdat ze zijn gebaseerd op een kleine steekproef. Om deze reden zijn de empirische verdelingen niet geschikt om een aantal gewonnen of verloren posities uit te simuleren voor het simulatiemodel. De oplossing is om een mix van normaalverdelingen te gebruiken, waarbij de verwachting van elk van deze normaalverdelingen een positie met kansmassa is. Op deze manier ontstaat een *smooth* empirische verdeling waarmee we wel het aantal gewonnen of verloren posities kunnen simuleren.

Rondetijden

Nu bekend is welke coureurs ongeschonden de eerste bocht zijn doorgekomen, simuleren we een rondetijd voor elke coureur met behulp van een regressiemodel. Dit model beschrijft de invloed van het brandstofverbruik en de bandenslijtage op de rondetijd. Het is namelijk bekend dat de rondetijden dalen naarmate de race vordert, omdat Formule 1-auto's lichter worden door het brandstofverbruik. Daarnaast heeft de slijtage van de banden invloed op de rondetijd. Wanneer een coureur net is gewisseld naar nieuwe banden nemen de rondetijden iets toe, omdat de banden eerst moeten worden opgewarmd. Als de banden de optimale temperatuur hebben bereikt, kan de coureur een paar snelle

rondes rijden. Wanneer de coureur echter te lang doorrijdt op de banden nemen de rondetijden weer toe als gevolg van de bandenslijtage. We nemen dus aan dat er een lineair verband bestaat tussen de rondetijd en het brandstofverbruik en een kwadratisch (convex) verband tussen de rondetijd en de bandenslijtage. Een rondetijd bestaat dan uit een basis rondetijd, een correctie voor het brandstofverbruik, een correctie voor de bandenslijtage en willekeurige fluctuaties:

(rondetijd) = (basisrondetijd) + (brandstofcorrectie) + (bandenslijtage) + (willekeurige fluctuaties).

Voor elke coureur schatten we vervolgens het effect van het brandstofverbruik en de bandenslijtage voor elke type band op de rondetijd met behulp van een regressiemodel. De parameters van het regressiemodel worden geschat met behulp van de kleinste kwadratenschatter. Als basisrondetijd gebruiken we de kwalificatietijd van de coureur. De geschatte parameters kunnen vervolgens worden gebruikt om nieuwe rondetijden te simuleren.

Inhalen

Tot nu toe hebben we interacties tussen coureurs, zoals inhaalacties, genegeerd. Inhaalacties zijn aan het simulatiemodel toegevoegd met behulp van het 'inhaalmodel'. Stel je twee coureurs voor die vlak achter elkaar rijden: als de achterste coureur sneller is dan de voorste coureur kan hij proberen in te halen. Of deze inhaalactie succesvol is hangt af van onder andere de relatieve snelheid tussen de coureurs, de afstand tussen de coureurs, waar op het circuit de inhaalactie wordt ingezet en hoe agressief de voorste coureur zijn positie verdedigt. Voor alle opeenvolgende paren van coureurs op het circuit berekenen we het verschil in cumulatieve rondetijd als $\delta_{jk} = C_k - C_j$, waarbij coureur j voor coureur k ligt. Als dit verschil negatief is en kleiner dan een bepaalde 'inhaal drempel', dan betekent dit dat coureur k snel genoeg is om coureur j in te halen. We simuleren aan de hand van de 'inhaal kans' of de inhaalactie succesvol is of niet. Als de inhaalactie succesvol is, wisselen de coureurs van positie. Als de inhaalactie niet succesvol is, dan heeft de coureur de volgende ronde mogelijk opnieuw een kans om in te halen.

Het vergelijken van pitstopstrategieën

We kunnen het simulatiemodel gebruiken om de optimale pitstopstrategie voor Mercedes te bepalen tijdens de Grand Prix van Japan in 2016. Mercedes reed destijds met het rijdersduo Lewis Hamilton en Nico Rosberg. Volgens de reglementen van Pirelli, de bandenleverancier van de Formule 1, waren er drie soorten banden beschikbaar: de zachte band, de medium band en de harde band. Daarnaast was de top 10 volgens de reglementen verplicht om de race op de zacht band te starten en moesten alle coureurs verplicht tenminste één pitstop te maken en de harde band gebruiken. Deze restricties zorgden ervoor dat niet alle combinaties van bandensoorten mogelijk waren. Ten slotte verwachtte Pirelli dat een twee-stop strategie optimaal zou zijn, dus nemen we aan dat elke coureur twee keer een pitstop maakt.

De verzameling van mogelijke bandenstrategieën voor Lewis Hamilton en Nico Rosberg bestaat dus uit:

- Zacht – medium – hard
- Zacht – hard – medium
- Zacht – hard – hard

Daarnaast gebruiken we de gemiddelde duur van een pitstop van Mercedes tijdens de Grand Prix van Japan in 2015 als pitstopduur.

Op basis van data weten we wat de werkelijke pitstopstrategie was van Lewis Hamilton en Nico Rosberg tijdens de Grand Prix van Japan in 2016. Lewis Hamilton wisselde zijn banden in rondje 13 van de zachte band naar de harde band en in ronde 33 opnieuw naar de harde band. Nico Rosberg gebruikte dezelfde bandenstrategie, maar zijn pitstops vonden plaats in ronde 12 en ronde 29. We

gebruiken het simulatiemodel om drie alternatieve strategieën te simuleren en te vergelijken met de huidige strategie:

- Strategie 1: dezelfde bandenstrategie, maar de pitstops vinden eerder plaats, namelijk in ronde 10 en ronde 28.
- Strategie 2: bandenstrategie zacht – hard – zacht met pitstops in ronde 12 en ronde 29.
- Strategie 3: bandenstrategie zacht – hard – medium met pitstops in ronde 12 en ronde 29.

We simuleren elke combinatie van strategieën 1000 keer en berekenen het gemiddelde van de som van de posities van Nico Rosberg en Lewis Hamilton. De combinatie van strategieën die resulteert in het laagste gemiddelde is de optimale strategie voor Mercedes. Volgens tabel 1 is dat de werkelijke strategie voor Lewis Hamilton in combinatie met strategie 2 voor Nico Rosberg. Met behulp van statistische testen kunnen we bepalen dat deze combinatie van strategieën significant beter is dan de werkelijke strategie. De conclusie is dus dat Mercedes de juiste strategie had gekozen voor Lewis Hamilton, maar voor Nico Rosberg beter strategie 2 had kunnen kiezen.

* Dit artikel is gebaseerd op het artikel 'Simulating Formula One Racestrategies', winnaar van de Business Analytics Best Paper Award 2018. Het volledige artikel is te vinden op: https://beta.vu.nl/nl/Images/werkstuk-sulsters_tcm235-877826.pdf.

CLAUDIA SULSTERS is recentelijk afgestudeerd van de masteropleiding Business Analytics aan de Vrije Universiteit en werkt momenteel als data scientist bij Data Science Lab. in Amsterdam.
Email: claudia.maria@live.nl
LinkedIn: <https://www.linkedin.com/in/claudiasulsters/>

STRATEGIE		GEMIDDELDE VAN DE SOM VAN DE POSITIES	P-VALUE T.O.V WERKELIJKE STRATEGIE
Lewis Hamilton	Nico Rosberg		
Werkelijk	Werkelijk	5,02	-
Werkelijk	Strategie 2	4,59	$1,003 \times 10^{-9}$
Werkelijke	Strategie 3	4,91	$2,445 \times 10^{-5}$

Tabel 1. Het vergelijken van racestrategieën