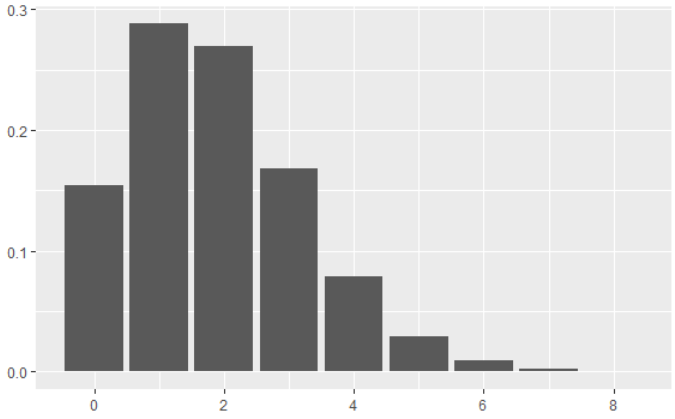
Universiteit Utrecht

Faculteit Betawetenschappen

# Het wiskundig voorspellen van voetbaluitslagen




Bachelorscriptie

Bjorn Buitink, 6272274
*Wiskunde*

*Scriptiebegeleider*
dr. M.C.J. (Martin) Bootsma

18 juni 2021

**Samenvatting**

Er zal met behulp van de Poissonverdeling en Bayesiaanse statistiek een model geschreven worden waarmee het aantal doelpunten in voetbalwedstrijden voorspeld kan worden. Deze voorspellingen zullen worden gebaseerd op het thuisvoordeel en de aanvals- en verdedigingssterkte van beide teams. Aan de hand van deze voorspellingen zetten we in op weddenschappen.

Er worden vervolgens twee inzetstrategieën bekeken: het inzetten op de grootste kanshebber en het inzetten op de weddenschap met de grootste winstverwachting. Er volgt voor beide strategieën dat we uitkomen op een gemiddeld verlies van 6 cent op 1 euro inzet. Wanneer we het inzetten zouden beperken tot alleen inzetten in de tweede seizoenshelft, dan zouden we het verlies kunnen verkleinen naar een gemiddeld verlies van 3 cent per wedstrijd.

# Inhoudsopgave

**1 Introductie** **1**

**2 Kansverdelingen** **2**
2.1 Poissonverdeling . . . 2
2.2 Negatieve binomiale verdeling . . . 4
2.3 Grootste aannemelijkheid . . . 4

**3 Factoren** **6**
3.1 Thuisvoordeel . . . 6
3.1.1 Coronapandemie . . . 6
3.2 Aanvals- en verdedigingssterkte . . . 6
3.3 Overige factoren . . . 6

**4 MCMC** **8**
4.1 Achtergrond . . . 8
4.2 Markov Chain en Monte carlo . . . 8
4.3 Prior, aannemelijkheid en posterior . . . 8
4.4 Het algoritme . . . 9
4.5 Convergentie . . . 9
4.6 Berekening $\lambda_{\text{thuis}}$ en $\lambda_{\text{uit}}$ . . . 9

**5 Het model** **12**
5.1 Data . . . 12
5.2 Stappenplan model . . . 12

**6 Gokkantoren** **13**
6.1 Sportweddenschappen . . . 13
6.2 Gokkantoren en inzetstrategieën . . . 13
6.2.1 Strategie 1: Inzetten op de grootste winkans . . . 13
6.2.2 Strategie 2: Inzetten op de grootste winstverwachting . . . 14
6.3 Resultaten inzetstrategieën . . . 15
6.3.1 Strategie 1: Inzetten op de grootste winkans . . . 15
6.3.2 Strategie 2: Inzetten op de grootste winstverwachting . . . 15
6.4 Kelly criterion . . . 16

**7 Discussie** **18**
7.1 Discussie model . . . 18
7.2 Discussie resultaten . . . 18
7.3 Discussie toekomstig werk . . . 18

**A Appendix A** **19**

**B Appendix B** **21**

**Referenties** **I**

# 1 Introductie

Het lijkt soms iets onmogelijks: het voorspellen van uitslagen bij sportwedstrijden. Niets lijkt zo verraderlijk als de uitslag van een voetbalwedstrijd, of toch niet? In deze scriptie zal er onderzoek worden gedaan naar de mogelijkheden om uitslagen van voetbalwedstrijden te voorspellen met behulp van wiskunde en wedstrijden uit het verleden; met het uiteindelijke doel om het beter te doen dan de gokkantoren.

We gaan op onderzoek of er met behulp van statistiek en kansrekening mogelijkheden zijn om de uitslagen van voetbalwedstrijden te voorspellen. Er zal gekeken worden naar hoe het aantal doelpunten in een wedstrijd wiskundig beschreven kan worden en hoe we de aanvallende en verdedigende capaciteiten van een team kunnen schatten met behulp van Markov Chain Monte Carlo (MCMC).

Er zal een model worden geschreven in R waarbij we kijken naar het aantal thuisdoelpunten, uitdoelpunten en het thuisvoordeel. We gebruiken hierbij de data van vijf verschillende Europese competities en kijken naar de zeven voorgaande seizoenen. Met behulp van de Poissonverdeling kunnen we vervolgens de kans op iedere uitslag bepalen en hiermee de winkansen van ieder team.

Tot slot zullen er twee inzetstrategieen bekeken worden en er zal gekeken welke strategie het meest winstgevend is. Uiteindelijk zal het doel zijn om te kijken of we met de voorspellingen van het model een betere voorspelling kunnen doen dan die van het gokkantoor.

# 2 Kansverdelingen

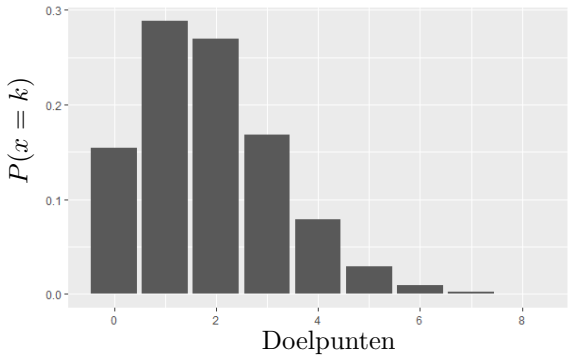
In dit hoofdstuk zal er aandacht besteed worden aan verschillende kansverdelingen die het aantal doelpunten in een voetbalwedstrijd kunnen beschrijven en de methode van de grootste aannemelijkheid.

## 2.1 Poissonverdeling

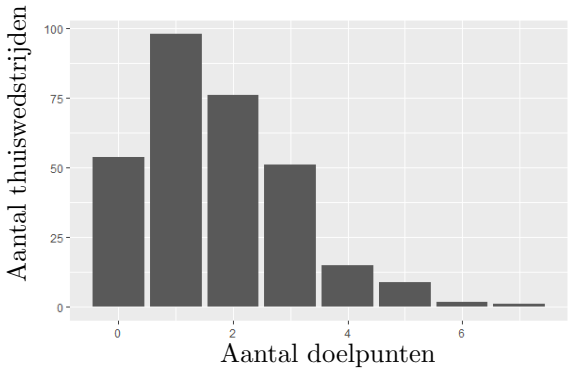
Om de uitslagen van voetbalwedstrijden te kunnen voorspellen, zijn we op zoek naar een kansverdeling die de kans op het aantal doelpunten in een voetbalwedstrijd zo nauwkeurig mogelijk beschrijft. Wanneer we een kansverdeling hebben gevonden, dan kunnen we aan de hand van deze verdeling voorspellingen maken; het vinden van een goede verdeling is dus essentieel. Een van de kansverdelingen die het aantal doelpunten nauwkeurig kan beschrijven is de Poissonverdeling [1]. Om de uitslag van een voetbalwedstrijd te voorspellen kunnen we trekkingen doen uit de Poissonverdelingen van het aantal thuis- en uitdoelpunten om het aantal thuis- en uitdoelpunten te bepalen. We bekijken nu de Poissonverdeling: zij $\lambda$ een vooraf vastgestelde parameter, dan kunnen we de kansfunctie van de Poissonverdeling, met de kans dat er precies $k$ doelpunten vallen, definiëren als

$$f(k;\lambda) = P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} \tag{2.1}$$

We bekijken een Poissonverdeling en een grafiek van het aantal thuisdoelpunten in de Nederlandse competitie, zie Figuur 1 en Figuur 2. Wanneer we deze figuren met elkaar vergelijken dan kunnen we zien dat het gedrag van de Poissonverdeling in overeenstemming is met het aantal thuisdoelpunten in de Nederlandse competitie.




Figuur 1: Poissonverdeling van de doelpunten voor parameter $\lambda = 1.87$




Figuur 2: Grafiek van het aantal thuisdoelpunten per wedstrijd in het seizoen 2017-2018 in de Nederlandse competitie

We definiëren nu twee begrippen die nauw samenhangen met kansverdelingen, dit zijn de verwachtingswaarde en de variantie.

**Definitie 2.1.1.** De verwachtingswaarde van een discrete stochastische variabele wordt gegeven door

$$E[X] = \sum_{k=0}^{\infty} kp(k)$$

**Definitie 2.1.2.** De variantie van een random variabele X wordt gegeven door

$$Var(X) = E[X^2] - (E[X])^2$$

Nu we de deze begrippen gedefinieerd hebben zijn we geïnteresseerd in de waarden van de verwachtingswaarde en de variantie van de Poissonverdeling.

**Claim.** *Voor de Poissonverdeling geldt* $E[X] = \lambda$

*Bewijs.* We gebruiken Definitie 2.1.1 dan volgt voor de Poissonverdeling dat

$$E[X] = \sum_{k=0}^{\infty} kp(k) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}$$

Kies nu $m = k - 1$, dan volgt

$$E[X] = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{m+1}}{m!} = e^{-\lambda} \cdot \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda$$

□

**Claim.** *Voor de Poissonverdeling geldt* $Var[X] = \lambda$

*Bewijs.* We gebruiken Definitie 2.1.1 dan volgt voor de Poissonverdeling dat

$$E[X^2] = \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{m=0}^{\infty} (m+1) \frac{\lambda^{m+1}}{m!} = \lambda \cdot e^{-\lambda} \cdot \sum_{m=0}^{\infty} (m+1) \frac{\lambda^m}{m!} =$$

$$= \lambda \cdot e^{-\lambda} \cdot \Big( \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} + \sum_{m=1}^{\infty} \frac{\lambda^m}{(m-1)!} \Big) = \lambda \cdot e^{-\lambda} \cdot (e^{\lambda} + \lambda e^{\lambda}) = \lambda(1+\lambda) = \lambda + \lambda^2$$

Gebruik nu Definitie 2.1.2, dan volgt dat $Var(X) = E[X^2] - (E[X])^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$ □

Als we dit resultaat toepassen op voetbalwedstrijden, dan zien we dat het verwachte aantal doelpunten gelijk is aan de parameter $\lambda$, tevens geldt dat de variantie gelijk is aan de parameter $\lambda$.
Wanneer we zouden aannemen dat het aantal doelpunten een Poissonverdeling volgt, dan geldt dat de verwachtingswaarde en de variantie van het aantal doelpunten gelijk zijn, maar in de praktijk gaat dit niet altijd op. Karlis en Ntzoufras (2000) onderzochten het verband tussen de verwachtingswaarde en variantie van verschillende teams in verschillende Europese competities. Zij concludeerden dat de verwachtingswaarde en variantie zeker niet altijd hetzelfde zijn [2]. Deze constatering zorgt er voor dat de Poissonverdeling niet altijd gebruikt kan worden. We moeten daarom op zoek naar een andere verdeling die het vallen van doelpunten kan beschrijven én waarvoor geldt dat de verwachtingswaarde en de variantie ongelijk zijn. Deze verdeling zal in de volgende paragraaf worden toegelicht.

## 2.2 Negatieve binomiale verdeling

We zagen in de vorige paragraaf dat de Poissonverdeling een kansverdeling is die gebruikt kan worden om doelpunten te beschrijven. Naast de Poissonverdeling bestaat er nog een andere verdeling die gebruikt kan worden, namelijk de negatieve binomiale verdeling [3].

Zij $r$ het aantal successen, $k$ het aantal mislukkingen en $p$ de kans op succes, dan wordt de kansfunctie van de negatieve binomiale verdeling gegeven door

$$f(k;r,p) = P(X=k) = \binom{k+r-1}{r-1}(1-p)^k p^r$$

Waar de Poissonverdeling dezelfde verwachtingswaarde en variantie heeft, zijn deze bij de negatieve binomiale verdeling verschillend. Voor de verwachtingswaarde en de variantie van de negatieve binomiale verdeling geven we een claim zonder bewijs.

**Claim.** *Voor de negatieve binomiale verdeling geldt* $E[X] = \frac{pr}{1-p}$

**Claim.** *Voor de negatieve binomiale verdeling geldt* $Var[X] = \frac{pr}{(1-p)^2}$

De negatieve binomiale verdeling kan dus ook gebruikt worden voor het beschrijven van doelpunten. In het vervolg zullen we alleen gebruik maken van de Poissonverdeling omdat dit de meest gebruikelijke verdeling is.

## 2.3 Grootste aannemelijkheid

In deze paragraaf bekijken we de methode van de grootste aannemelijkheid (maximum likelihood method). Dit principe zal ook terugkomen in Hoofdstuk 4. Voordat we naar de methode van de grootste aannemelijkheid kunnen kijken, definiëren we eerst de aannemelijkheidsfunctie.

**Definitie 2.3.1.** De aannemelijkheidsfunctie $L(\theta)$ is de functie

$$L(\theta) = P(X_1 = x_1, \ldots, X_n = x_n) = p_\theta(x_1)\cdots p_\theta(x_n)$$

De methode van de grootste aannemelijkheid kiest, gegeven een dataset, de parameter zodat de data het meest aannemelijk is.

**Definitie 2.3.2.** De meest-aannemelijke schatter (maximum-likelihood-estimator) van $\theta$ is de waarde $t = h(x_1, x_2, ..., x_n)$ waarvoor de aannemelijkheidsfunctie $L(\theta)$ maximaal is. De variabele $T = h(X_1, X_2, ..., X_n)$ heet de meest aannemelijke schatter van $\theta$ [4].

De meest aannemelijke schatter is interessant omdat bij deze schatter de aannemelijkheid het hoogst is (wat de naam al doet vermoeden). We bekijken nu de meest aannemelijke schatter voor de Poissonverdeling, zie de volgende claim.

**Claim.** *De meest aannemelijke schatter voor de Poissonverdeling is* $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} k_i$

*Bewijs.* $L(\lambda) = \dfrac{\lambda^{k_1}}{k_1!}e^{-\lambda}\cdot\dfrac{\lambda^{k_2}}{k_2!}e^{-\lambda}\cdot\ldots\cdot\dfrac{\lambda^{k_n}}{k_n!}e^{-\lambda} = \dfrac{\lambda^{k_1+k_2+\cdots+k_n}}{k_1!\cdot k_2!\cdot\ldots\cdot k_n!}e^{-n\lambda} = \dfrac{\lambda^{\sum_{i=1}^n k_i}}{k_1!\cdot k_2!\cdot\ldots\cdot k_n!}e^{-n\lambda}$

Neem nu aan beide kanten het logaritme

$$\log(L(\lambda)) = \log(\lambda^{\sum_{i=1}^n k_i}e^{-n\lambda}) - \log(k_1!\cdot k_2!\cdot\ldots\cdot k_n!) =$$

$$= \log(\lambda)\sum_{i=1}^{n} k_i - n\lambda\log(e) - (\log(k_1! + \log(k_2!)\cdot\ldots\cdot\log(k_n!) = \log(\lambda)\sum_{i=1}^{n} k_i - n\lambda - \sum_{i=1}^{n}\log(k_i!)$$

We leiden af naar $\lambda$ om vervolgens de loglikelihood te bekijken

$$\frac{\partial}{\partial\lambda}l(\lambda) = \frac{\partial}{\partial\lambda}(\log(\lambda)\sum_{i=1}^{n} k_i) - \frac{\partial}{\partial\lambda}(n\lambda) - \frac{\partial}{\partial\lambda}(\sum_{i=1}^{n}\log(k_i!)) \implies l'(\lambda) = \frac{1}{\lambda}\sum_{i=1}^{n} k_i - n$$

Los op voor $l'(\lambda) = 0$, dan volgt

$$\frac{1}{\lambda}\sum_{i=1}^{n} k_i - n = 0 \implies \lambda = \frac{1}{n}\sum_{i=1}^{n} k_i$$

□

We zien dat de meest aannemelijke schatter gelijk is aan $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} k_i$. Als we de betekenis van deze schatter vertalen naar voetbalwedstrijden, dan zien we dat het het aantal verwachte doelpunten in een voetbalwedstrijd gelijk is aan het competitiegemiddelde van alle doelpunten.

# 3 Factoren

In dit hoofdstuk zullen we bepalen welke factoren we mee laten wegen in het model en hoe we deze factoren meenemen.

## 3.1 Thuisvoordeel

Er zijn verschillende onderzoeken gedaan naar het effect van thuisvoordeel binnen sportwedstrijden [5]. Thuisvoordeel is het positieve voordeel dat een team kan hebben doordat het in zijn eigen stadion speelt en dus beïnvloed kan worden door het aanwezige publiek. Bij voetbalwedstrijden is het effect van thuisvoordeel ook aanwezig [6]; een goed voorbeeld hiervan is de Engelse club Chelsea. Chelsea was onder leiding van trainer José Mourinho in de seizoenen 2004/2005 tot en met 2007/2008 thuis ongeslagen: van de 60 thuiswedstrijden werden er 46 gewonnen en 14 gelijk gespeeld [7].

### 3.1.1 Coronapandemie

Sinds de uitbraak van het virus SARS-CoV-2, worden sinds mei 2020 vrijwel alle voetbalwedstrijden in Europa zonder publiek gespeeld. Er moet daarom rekening gehouden worden met het feit dat het thuisvoordeel voor de thuisspelende ploeg kleiner zou kunnen worden, omdat er geen publiek aanwezig is in het stadion. Het ontbreken van thuisvoordeel zou invloed kunnen hebben op de parameters in het model en daarmee op de voorspelde resultaten. De invloed op het thuisvoordeel zal verder behandeld worden in de discussie, zie Hoofdstuk 7.

## 3.2 Aanvals- en verdedigingssterkte

Binnen een voetbalcompetitie zijn er vaak 18-20 verschillende voetbalteams actief. Een team bestaat ongeveer uit 30 spelers en de prestaties van het team zijn afhankelijk van welke (typen) spelers ze tot hun beschikking hebben. Zo zijn er teams die veel sterke aanvallende spelers hebben, maar ook teams met veel sterke verdedigende spelers. De aantal doelpunten die er gescoord wordt en de aantal doelpunten die geïncasseerd wordt, hangt dus af van de aanvallende en de verdedigende capaciteiten van beide teams.

Je zou kunnen stellen dat het aantal thuisdoelpunten dat gescoord wordt, afhangt van de aanvalssterkte van het thuisspelende team en van de verdedigingssterkte van het uitspelende team [8] [1]. Idem geldt voor het aantal uitdoelpunten dat dit afhangt van aanvalssterkte van het uitspelende team en van de verdedigingssterkte van het thuisspelende team. De aanvals- en verdedigingssterkte zullen de basis zijn voor het berekenen van de verwachte aantal doelpunten en daarmee de basis voor het model. Meer informatie over de berekening en benadering van aanvals- en verdedigingssterkte is te lezen in Hoofdstuk 4.

## 3.3 Overige factoren

Er zijn natuurlijk veel meer factoren naast de aanvalssterkte, verdedigingssterkte en thuisvoordeel meewegen in het resultaat van een voetbalwedstrijd. We benoemen een aantal van deze factoren en leggen uit hoe deze van invloed kunnen zijn op de resultaten. De onderstaande factoren zullen zelf niet meegenomen worden in het model.

**Vorm** De term 'vorm' is een bekend begrip bij sportwedstrijden. Vorm heeft betrekking op de laatste 5 wedstrijden waarbij gekeken wordt naar de prestaties van een team in deze laatste 5 wedstrijden. Als een team in de laatste 5 wedstrijden goede prestaties heeft geleverd ten opzichte van de verwachte prestaties, dan hebben zij een goede vorm (ze doen het dus beter dan verwacht). Andersom geldt ook dat als een team in de laatste 5 wedstrijden slechte prestaties heeft geleverd ten opzichte van de verwachte prestaties, dat ze dan een slechte vorm hebben. Voor toekomstige wedstrijden zal een positieve vorm een positieve invloed hebben op het aantal doelpunten, terwijl een negatieve vorm een negatieve invloed zal hebben op het aantal doelpunten.

**Kaarten en corners** Tijdens een wedstrijd kunnen er gele en rode kaarten worden uitgedeeld. Wanneer een team vaak een agressieve tactiek hanteert en dus relatief vaak een rode kaart krijgt, is de kans groot dat een team met 10 man verder moet spelen. Het spelen met een speler minder vergroot de kans dat de tegenstander een doelpunt scoort. Het aantal corners die een team voor en/of tegen heeft, kan ook van invloed zijn op het aantal doelpunten [9]. Een corner zorgt er namelijk voor dat een team direct bij het strafschopgebied van de tegenstander in de buurt is; dit kan de kans vergroten op een doelpunt.

Naast al deze factoren kan er ook gekeken worden naar schorsingen en blessures van spelers. De afwezigheid van een topscorer zou bijvoorbeeld kunnen leiden tot een lagere aanvalssterkte, terwijl het ontbreken van een centrale verdediger kan leiden tot een lagere verdedigingssterkte.

# 4 MCMC

In dit hoofdstuk zal aandacht besteed worden aan Markov Chain Monte Carlo (MCMC) waarmee de parameters van de aanvals- en verdedigingssterkte bepaald zullen worden.

## 4.1 Achtergrond

Voor het gebruik van MCMC zullen we de Bayesiaanse statistiek gebruiken. De Bayesiaanse statistiek heeft de eigenschap dat de kansen voortdurend aangepast kunnen worden indien er nieuwe informatie beschikbaar komt. De methode waarmee we de aanvals- en verdedigingssterkte gaan bepalen is met MCMC. Zoals de naam al zegt, is MCMC een combinatie tussen de Markov Chain en de Monte Carlo methode. In verschillende wetenschappelijke publicaties wordt ook van deze techniek gebruik gemaakt om achter de aanvals- en verdedigingssterkte te komen [10]. Voor definities en notatie zullen we in dit hoofdstuk grotendeels de lijn volgen van het boek Pattern Recognition and Machine Learning van Christopher Bishop [11].

## 4.2 Markov Chain en Monte carlo

Om MCMC te kunnen begrijpen, geven we een korte uitleg over de begrippen Markov Chain en de Monte Carlo methode. Een Markov Chain is een keten die afhangt van een reeks random variabelen $z^1, z^2 \ldots, z^m$ zodat de onderstaande eigenschap geldt voor alle $m \in \{1, 2, \ldots, m-1\}$

$$p(z^{m+1} \mid z^1, z^2, \ldots, z^m) = p(z^{m+1} \mid z^m)$$

De kans om naar een andere staat te verplaatsen is dus alleen afhankelijk van de huidige staat en hangt niet af van de voorgaande staten. Een belangrijk aspect voor MCMC is dat de Markov Chain een stationaire verdeling kent; dit betekent dat op den duur de Markov Chain naar een vaste verdeling zal convergeren. Voor een stationaire verdeling geldt dat wanneer de Markov Chain overgangskansen $T(z', z)$ heeft, dat de verdeling $p'(z)$ stationair is wanneer geldt dat

$$p'(z) = \sum_{z'} T(z', z) p'(z')$$

De Monte Carlo methode is een techniek waarbij een proces vele malen herhaald wordt, maar telkens met verschillende waarden. Het combineren van beide concepten leidt vervolgens tot de Markov Chain Monte Carlo.

## 4.3 Prior, aannemelijkheid en posterior

Nu we voldoende informatie hebben met betrekking tot MCMC, kunnen we ons richten op de werking van MCMC. De basis van MCMC is gebaseerd op drie verschillende begrippen, namelijk de posteriorverdeling, de prior en de aannemelijkheid. We zullen alle drie deze begrippen toelichten.

We definiëren de parameters als $w$ en de data als $D$. We bekijken eerst de prior: de prior $P(w)$ is de verwachting die we van de parameter hebben, deze informatie hebben we voor elke parameter $w$. Voor de prior kunnen we onderscheid maken tussen informative priors en uninformative priors. Informative priors zijn priors die specifieke informatie geven over de prior, terwijl uninformative priors meer algemene informatie over de prior geven. Vervolgens bekijken we de aannemelijkheid (likelihood): de aannemelijk $P(D|w)$ geeft aan hoe aannemelijk de data is gegeven de parameters. Voor elke parameter $w$ kunnen we de aannemelijkheid $P(D|w)$ berekenen. De posteriorverdeling is hetgeen waar we naar op zoek zijn, deze wordt gegeven door $P(w|D)$. Wanneer we de stelling van Bayes gebruiken, dan volgt dat de de posteriorverdeling op de volgende manier uit kunnen drukken

$$P(w|D) = \frac{P(D|w) \cdot P(w)}{P(D)}$$

Hierbij is $P(D)$ een normaliserings factor, die op de volgende manier berekend kan worden

$$P(D) = \int P(D|w) P(w) dw$$

Er geldt dat de posterior evenredig is met het product van de aannemelijkheid en de prior, ofwel posterior $\propto$ aannemelijkheid $\cdot$ prior. Zoals we eerder al noemden heeft de Markov Chain een stationaire verdeling; wanneer we het MCMC algoritme uitvoeren zal het algoritme convergeren naar de posteriorverdeling als de stationaire verdeling.

## 4.4 Het algoritme

Het meest gebruikte algoritme voor MCMC is het Metropolis–Hastings algoritme. De werking van het algoritme zal hieronder, zowel in formule als in woorden, uitgelegd worden [12] [13].

**Algorithm 1:** Metropolis-Hastings

Kies een $X$ en bereken $P_X$ = prior $\cdot$ aannemelijkheid
Doe een voorstel om te verplaatsen naar $Y$ waarbij $Y$ uit de proposal distribution $q(Y|X)$ komt
Bereken nu voor $Y$, $P_Y$ = prior $\cdot$ aannemelijkheid
Bereken $\alpha = \dfrac{P_Y \cdot q(X|Y)}{P_X \cdot q(Y|X)}$
**if** $P_Y > P_X$ **then**
  *Accepteer $Y$ en herhaal voorgaande stappen met $Y$ als nieuwe waarde*
**else**
  *Accepteer $P_Y$ met kans $\alpha$ en herhaal voorgaande stappen met $Y$ als nieuwe waarde*
  *Weiger $P_Y$ met kans $1-\alpha$ herhaal voorgaande stappen met $X$ als nieuwe waarde*

In woorden doet het algoritme het volgende: we pakken een parameter $X$ en we berekenen voor die parameter $X$ de waarde $P_X$ = prior $\cdot$ aannemelijkheid. Vervolgens doen we een voorstel om te verplaatsen naar een andere waarde van de parameter, namelijk $Y$. Deze $Y$ komt uit een proposal distribution $q(Y|X)$ (bijvoorbeeld de normale verdeling). Nu berekenen we voor $Y$ de waarde $P_Y$ = prior $\cdot$ aannemelijkheid. We vergelijken nu $P_X$ en $P_Y$, wanneer $P_Y > P_X$ dan accepteren we de nieuwe waarde $Y$ en wordt dat ons nieuwe punt en herhalen we alle voorgaande stappen opnieuw. Wanneer $P_Y < P_X$ dan accepteren we $Y$ met kans $\alpha$ en nemen we $Y$ als nieuwe waarde en herhalen we alle voorgaande stappen opnieuw. We weigeren $Y$ met kans $1-\alpha$ en nemen dan $X$ als nieuwe waarde en herhalen we alle voorgaande stappen.

## 4.5 Convergentie

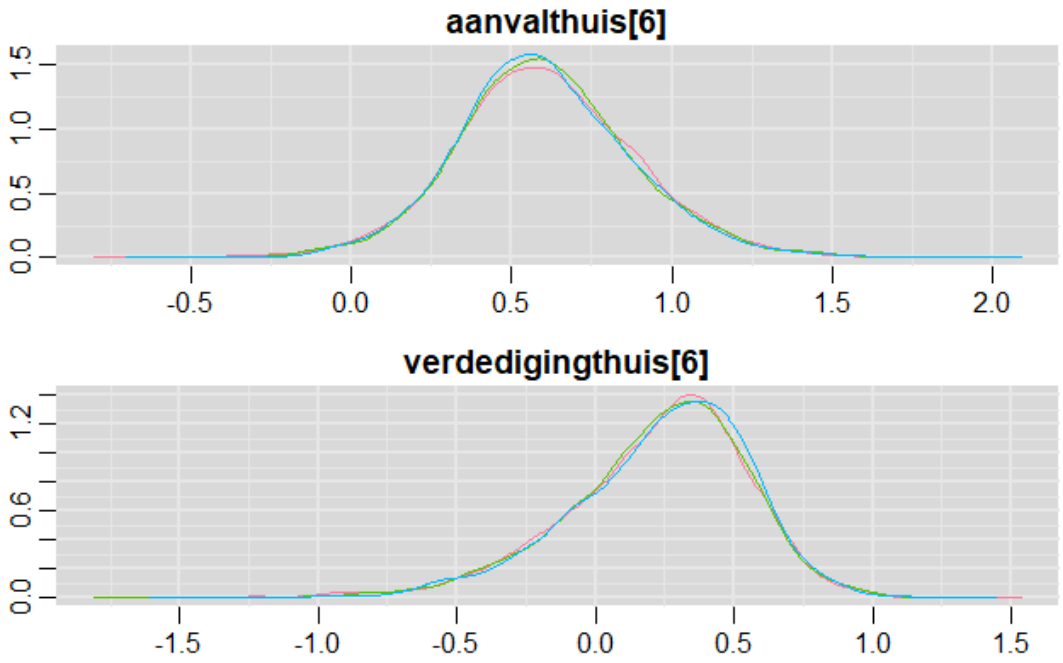
We zijn in het bijzonder geïnteresseerd in de convergentie van het algoritme, want het algoritme zal convergeren naar de meest aannemelijke waarde van onze parameters. Als voorbeeld nemen we de wedstrijd Borussia Dortmund - VfB Stuttgart in de Duitse competitie en bekijken we de aanvals- en verdedigingssterktes van beide teams, zie hiervoor Figuren 3, 4, 5 en 6. In deze figuren zijn de trace van het algoritme te zien en de dichtheid voor de parameters. De verschillende kleuren die hierbij te zien zijn, staan voor de verschillende chains in de Markov chain.

## 4.6 Berekening $\lambda_{\text{thuis}}$ en $\lambda_{\text{uit}}$

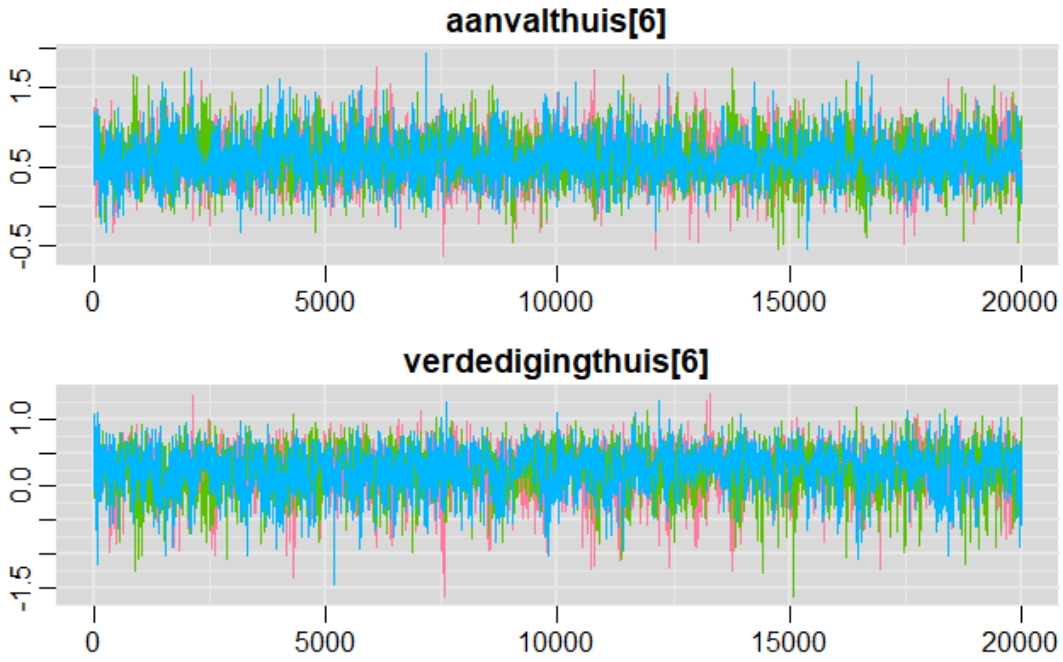
Nu we met behulp van MCMC de waardes van aanvals- en verdedigingssterktes van beide teams hebben berekend, kunnen we de waarde van de parameters $\lambda_{\text{thuis}}$ en $\lambda_{\text{uit}}$ voor de Poissonverdeling berekenen. Als team $i$ tegen team $j$ speelt, dan berekenen we de parameters berekenen op de volgende manier

$$\lambda_{\text{thuis}} = e^{\text{aanvalthuis[i]} + \text{verdedigingsuit[j]}}$$
$$\lambda_{\text{uit}} = e^{\text{aanvaluit[j]} + \text{verdedigingthuis[i]}}$$

Het thuisvoordeel is hierbij al meegenomen in de waarde van aanvalthuis en verdedigingthuis.

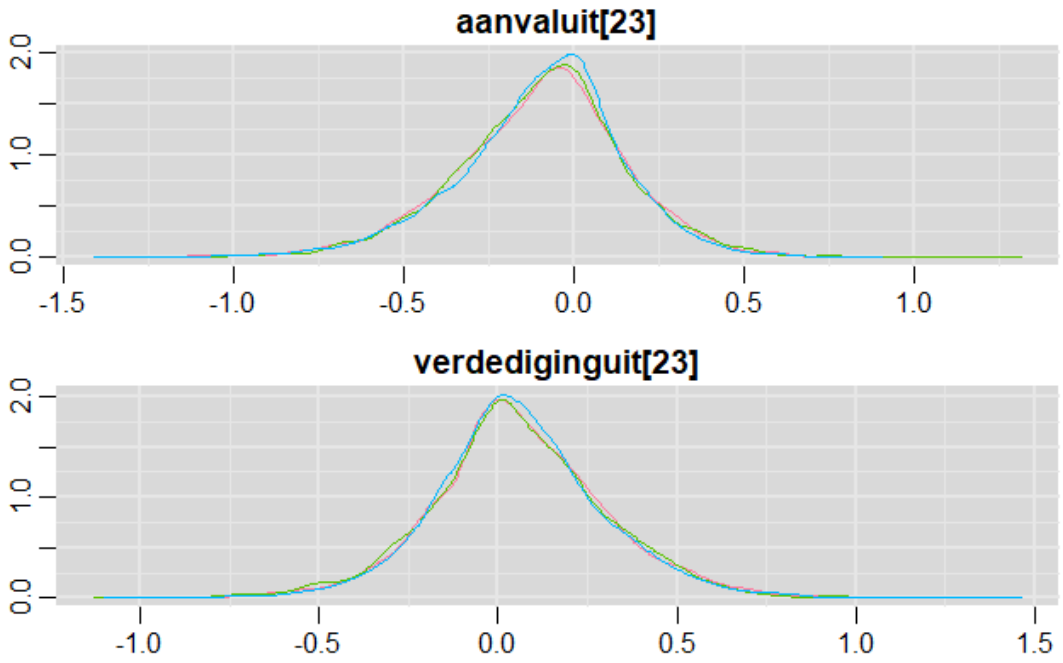



Figuur 3: De distributie van de aanvals- en verdedigingssterkte van het thuisspelende team Borussia Dortmund

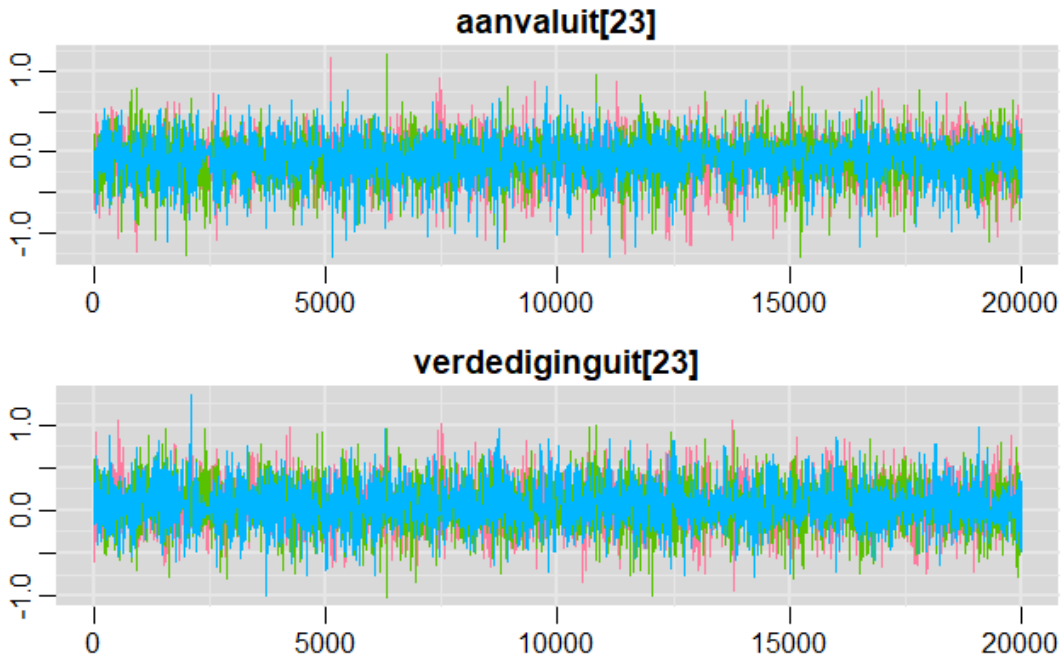



Figuur 4: De trace van de aanvals- en verdedigingssterkte van het thuisspelende team Borussia Dortmund

.




Figuur 5: De distributie van de aanvals- en verdedigingssterkte van het uitspelende team VfB Stuttgart




Figuur 6: De trace van de aanvals- en verdedigingssterkte van het uispelende team VfB Stuttgart

.

# 5 Het model

In de voorgaande hoofdstukken hebben we allerlei aspecten van het model behandeld, zoals de factoren die meegenomen worden en de bepaling van parameters; in dit hoofdstuk bekijken we hoe het model er globaal uit komt te zien.

Omdat er veel gebruik gemaakt wordt van statistiek en data-analyse is ervoor gekozen om het model te schrijven in R. Als naslagwerk voor het programmeren is gebruik gemaakt van het boek 'Machine learning essentials: Practical guide in R' [14]. De code zelf kan gevonden worden in Appendix B.

## 5.1 Data

De data voor het model zal worden verkregen via de website `https://www.football-data.co.uk/`. Hierbij worden de data van 5 grote Europese competities gebruikt (Engeland, Spanje, Nederland, Duitsland en Frankrijk). We gebruiken hiervoor 7 opeenvolgende seizoenen, vanaf seizoen 2013/2014 tot en met seizoen 2019/2020. Omdat we de data van 7 seizoenen gebruiken, kunnen we onze voorspellingen op voldoende data baseren. De beschikbare informatie in de data varieert van doelpunten tot kaarten en van quoteringen van verschillende gokkantoren tot corners. Naast de 7 seizoenen die we gebruiken om het model te trainen, zal de data van het seizoen 2020/2021 gebruikt worden om het model te testen.

## 5.2 Stappenplan model

We nemen nu stap voor stap het model door. Allereerst wordt de data geladen en deze wordt zodanig aangepast zodat alleen de relevante data bewaard wordt (de data omtrent gele kaarten, corners en overige quoteringen van gokkantoren worden verwijderd). We scheiden vervolgens de trainingsdata (de seizoenen 2013/2014 t/m 2019/2020) van het seizoen dat we willen voorspellen (2020/2021). Hierna wordt het MCMC algoritme met 20000 iteraties uitgevoerd om de aanvals- en verdedigingssterktes van beide teams te berekenen. De data die het MCMC algoritme gebruikt zijn alle trainingsdata + de uitslagen van de voorgaande speelronden in het huidige seizoen. Wanneer we speelronde 7 willen voorspellen, gebruiken we dus de trainingsdata + speelronde 1 t/m 6. Met MCMC worden aan de hand van het aantal thuisdoelpunten, uitdoelpunten en de thuisspelende teams de waardes voor de aanvalthuis, aanvaluit, verdedigingthuis en verdediginguit verkregen. Het stuk in de code over MCMC is grotendeels overgenomen uit [10] en [15] en aangepast naar ons model.

Na het berekenen van de aanvals- en verdedigingssterktes, worden volgens de formules in Hoofdstuk 4.6 de $\lambda$ voor verwachte thuis- en uitdoelpunten berekend, ($\lambda_{\text{thuis}}$ en $\lambda_{\text{uit}}$).

$$\lambda_{\text{thuis}} = e^{\text{aanvalthuis[i]} + \text{verdediginguit[j]}}$$
$$\lambda_{\text{uit}} = e^{\text{aanvaluit[j]} + \text{verdedigingthuis[i]}}$$

Wanneer we de waardes van $\lambda$ hebben gevonden, kunnen we met de Poissonverdeling de kans op $k = 0, 1, 2, 3, \ldots$ doelpunten bepalen. Wanneer we dit voor beide teams uitvoeren, kunnen we vervolgens de kans op thuiswinst, gelijkspel en uitwinst bepalen.

Nu we weten hoe we de kansen op thuiswinst, gelijkspel en uitwinst kunnen bepalen, kunnen we ons nu richten op het inzetten op weddenschappen. Verdere informatie over dit stuk van het model is te lezen in Hoofdstuk 6.

# 6 Gokkantoren

In dit hoofdstuk zal er aandacht besteed worden aan de werking van sportweddenschappen, de implementatie van gokkantoren in het model en er worden verschillende inzetstrategieën bekeken.

## 6.1 Sportweddenschappen

Vroeger kon er bij de plaatselijke boekwinkel of benzinepomp een papiertje met voorspellingen ingevuld worden, maar tegenwoordig gebeurt dit veelal online. Het online voorspellen biedt veel meer mogelijkheden wat betreft het inzetten, er kan ingezet worden op uitslagen, doelpunten, kaarten, corners en zelfs op doelpuntenmakers; wij zullen ons voornamelijk richten op het voorspellen van de uitslagen. Voordat we dit kunnen doen, zullen we begrippen omtrent sportweddenschappen duidelijk maken aan de hand van een voorbeeld. We bekijken de wedstrijd Borussia Dortmund - VfB Stuttgart in de Duitse competitie (2020/2021).

Een voetbalwedstrijd in competitieverband heeft altijd drie mogelijke uitslagen; deze worden aangeduid op de volgende manier

$$1 = \text{Thuisteam wint}, \quad X = \text{Gelijkspel}, \quad 2 = \text{Uitteam wint}$$

Voor de voetbalwedstrijd Borussia Dortmund - VfB Stuttgart betekent een 1 dat Borussia Dortmund wint, een $X$ dat er gelijkgespeeld wordt en een 2 dat VfB Stuttgart wint. Aan de gebeurtenissen $1, X, 2$ zitten quoteringen (ook wel odds genoemd) verbonden, dit zijn de getallen die in Figuur 7 onder de gebeurtenissen $1, X, 2$ te zien zijn. Deze quoteringen geven aan hoeveel er verdiend kan worden met het inzetten op een bepaalde weddenschap. We noemen de quotering $q$ en de inzet $i$, dan geldt voor het geld dat het gokkantoor jou uitkeert gelijk is aan $i \cdot q$.

| 1 | X | 2 |
|---|---|---|
| 1.5 | 4.5 | 6 |

Figuur 7: Quoteringen voor de wedstrijd Borussia Dortmund - VfB Stuttgart

We gebruiken nu een voorbeeld om dit te verduidelijken. Stel je zet $i = 1.5$ euro in op $X$, we zien in Figuur 7 dat $X$ een quotering heeft van $q = 4.5$. Wanneer vervolgens de wedstrijd eindigt in een gelijkspel, dan keert het gokkantoor jou $i \cdot q = 1.5 \cdot 4.5 = 6.75$ uit.

## 6.2 Gokkantoren en inzetstrategieën

In Hoofdstuk 5 zagen we dat er ook data beschikbaar zijn van verschillende gokkantoren. We zullen ons beperken tot één gokkantoor: bet365. Het specificeren van één type gokkantoor brengt zowel voordelen als nadelen met zich mee. Voordelen zijn dat het overzichtelijker en duidelijker is om met één gokkantoor te werken dan met meerdere gokkantoren tegelijk. Hierdoor kunnen we uiteindelijk beter concluderen hoe de prestaties van het model zijn ten opzichte van het gokkantoor. Het nadeel is dat er elke wedstrijd bij hetzelfde gokkantoor ingezet moet worden. Zo kan het bijvoorbeeld zijn dat je een gokkantoor met aantrekkelijkere quoteringen misloopt (en hiermee dus ook de kans op hogere winst).

### 6.2.1 Strategie 1: Inzetten op de grootste winkans

De eerste strategie die we bekijken is strategie 1: het inzetten op de grootste winkans. Als we volgens strategie 1 inzetten, dan zetten we in op het team dat de grootste kans heeft om de wedstrijd te winnen. Hiermee hebben we dus ook de grootste kans om de weddenschap te winnen.
We bekijken de wedstrijd Borussia Dortmund - VfB Stuttgart, zie Figuur 7. We vinden met het simuleren van deze wedstrijd in ons model dat de volgende winkansen gelden: Borussia Dortmund wint 55.55% van de wedstrijden, in 22.13% van de wedstrijden wordt het een gelijkspel en VfB Stuttgart wint in 22.32% van de wedstrijden. We zullen inzetten volgens strategie 1, en we zullen dus kiezen voor het team met de grootste winkans. In dit geval zullen we dus inzetten op Borussia Dortmund.

### 6.2.2 Strategie 2: Inzetten op de grootste winstverwachting

De tweede strategie die we bekijken is strategie 2: het inzetten op de grootste winstverwachting. Als we volgens strategie 2 inzetten, dan zetten we in op het team dat de grootste winstverwachting heeft.
We bekijken wederom de wedstrijd Borussia Dortmund - VfB Stuttgart met de berekende winkansen 55.55% / 22.13% / 22.32%. We zijn op zoek naar de weddenschap met de grootste winstverwachting (en dus de weddenschap die het meest zal opleveren). Om dit te kunnen berekenen, definiëren we drie begrippen: de verwachte winst, het verwachte verlies en de verwachte opbrengst. Deze zijn gelijk aan

$$\text{Verwachte winst} = (\text{Kans op winst}) \cdot (\text{Quotering} - 1) \cdot \text{Inzet}$$
$$\text{Verwacht verlies} = (1 - \text{Kans op winst}) \cdot \text{Inzet}$$
$$\text{Verwachte opbrengst} = \text{Verwachte winst} - \text{Verwacht verlies}$$

We kunnen aan de hand van de bovenstaande vergelijkingen de verwachte opbrengst als het volgt beschrijven

$$\text{Verwachte opbrengst} = (\text{Kans op winst}) \cdot (\text{Quotering} - 1) \cdot \text{Inzet} - (1 - \text{Kans op winst}) \cdot \text{Inzet} \tag{6.1}$$

Nu we de begrippen gedefinieerd hebben, kunnen we de verwachte opbrengsten berekenen met behulp van vergelijking 6.1 voor de wedstrijd Borussia Dortmund - VfB Stuttgart, zie Figuur 7. Voor het versimpelen van het rekenwerk zullen we kiezen voor een inzet van 1 euro. We verkrijgen de volgende verwachte opbrengsten

$$\text{Verwachte opbrengst } 1 = -0.16683$$

$$\text{Verwachte opbrengst X} = -0.00426965$$

$$\text{Verwachte opbrengst } 2 = 0.3385964$$

We zien dat de verwachte opbrengst voor zowel weddenschap 1 als weddenschap X negatief is; het is dus niet aantrekkelijk om hier op in te zetten. De verwachte opbrengst voor weddenschap 2 is het grootst, wat betekent dat we op weddenschap 2 in zullen zetten.
Belangrijk is om op te merken dat we bij een negatieve verwachte opbrengst niet zullen inzetten, want dit zal op de lange termijn niet voor winst zorgen. Wanneer we dus voor voor elke weddenschap een negatieve winstverwachting hebben, dan zetten we niet in.

Uiteindelijk was de uitslag van de wedstrijd Borussia Dortmund - VfB Stuttgart 1-5, dat betekent dat de underdog gewonnen heeft. Als we nu kijken naar de inzetstrategieën, dan zien we dat we met strategie 1 een verlies hebben van 1 euro en dat we bij strategie 2 een opbrengst hebben van 6 euro.

## 6.3 Resultaten inzetstrategieën

In deze paragraaf bespreken we de relevante en interessante uitkomsten van de verschillende inzetstrategieën die we verkregen hebben met simulaties van het model. We simuleren elke speelronde van iedere competitie en berekenen hierbij hoeveel er verdiend wordt per speelronde. De bedragen die iedere speelronde verdiend worden zijn te zien in de volledige resultaten in Appendix A.We zullen elke speelronde voor elke wedstrijd 1 euro inzetten. In de landen Duitsland en Nederland zal er 306 euro ingezet worden ingezet (34 speelrondes met 9 wedstrijden per speelronde) voor de landen Spanje, Frankrijk en Engeland is dit 380 euro (38 speelrondes met 10 wedstrijden per speelronde).
Hieronder is een beknopt overzicht gegeven van de resultaten, waarbij we kijken naar de opbrengst per wedstrijd, de opbrengst per wedstrijd in de eerste seizoenshelft en de opbrengst per wedstrijd in de tweede seizoenshelft.

### 6.3.1 Strategie 1: Inzetten op de grootste winkans

Wanneer we naar de resultaten in Tabel 1 kijken, zien we dat er alleen in Engeland winst wordt gemaakt over het gehele seizoen, namelijk 16 cent per wedstrijd. Als we kijken naar de andere landen, dan zien we dat we in Nederland, Spanje en Frankrijk het verlies kunnen beperken tot maximaal 10 cent per wedstrijd. We kunnen ook constateren dat er in tweede seizoenshelft (5 cent verlies per wedstrijd) gemiddeld minder verlies geleden wordt dan in de eerste seizoenshelft (7 cent verlies per wedstrijd).

| | **Duitsland** | **Nederland** | **Spanje** | **Frankrijk** | **Engeland** | **Gemiddeld** |
|---|---|---|---|---|---|---|
| **Opbrengst per wedstrijd** | -19 cent | -8 cent | -10 cent | -7 cent | 16 cent | -6 cent |
| **1e helft per wedstrijd** | -23 cent | -13 cent | -14 cent | -2 cent | 16 cent | -7 cent |
| **2e helft per wedstrijd** | -15 cent | -3 cent | -6 cent | -12 cent | 16 cent | -5 cent |

Tabel 1: Resultaten strategie 1

### 6.3.2 Strategie 2: Inzetten op de grootste winstverwachting

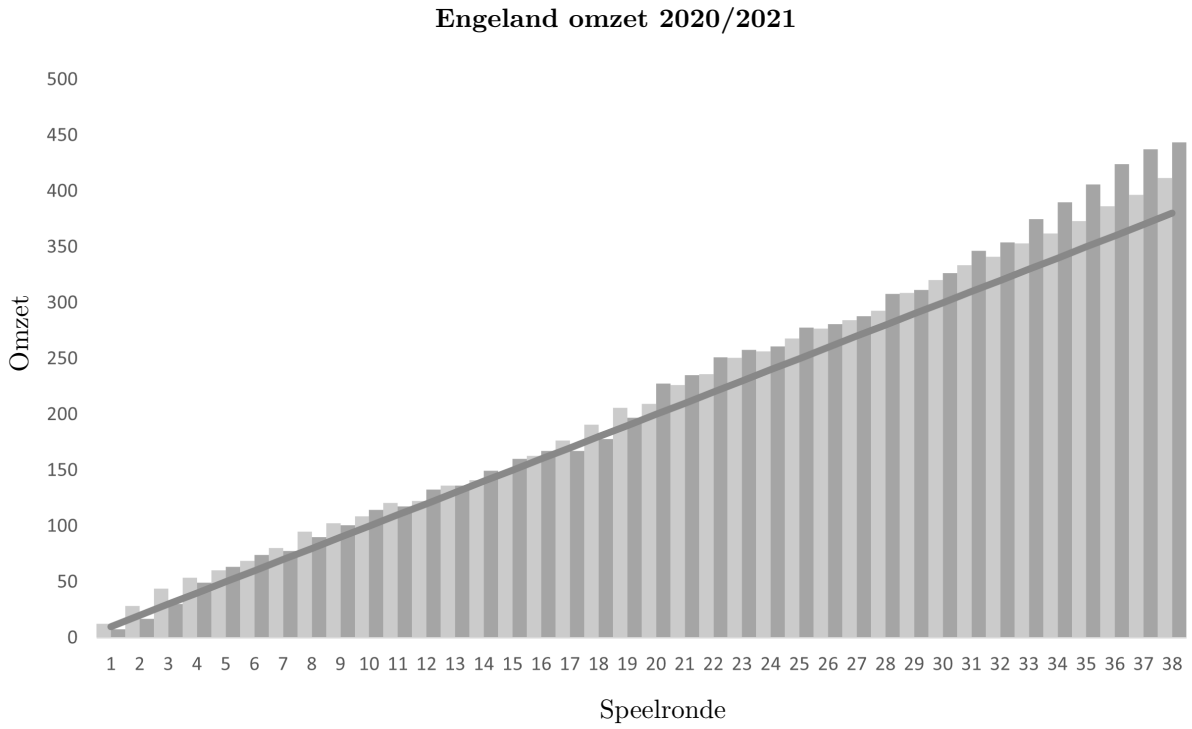
Ten opzichte van de resultaten van strategie 1 zijn de seizoensbrede resultaten gelijkwaardig. We kunnen zien dat seizoensbreed het gemiddelde verlies voor strategie 1 en strategie 2 hetzelfde is (6 cent). Seizoensbreed is het wederom alleen de Engelse competitie waar we geld kunnen verdienen. Wanneer we de twee seizoenshelften bekijken, zien we dat hier een duidelijk verschil ontstaat. Het verlies in de eerste seizoenshelft is erg groot (11 cent), terwijl we in de tweede seizoenshelft quitte spelen. In de tweede seizoenshelft kunnen we in drie competities winst boeken, maar in de overige twee competities vergroten we het verlies.

Voor beide seizoenen kunnen we constateren dat de tweede seizoenshelft winstgevender is dan de eerste seizoenshelft.Waarom dit zo is, is mogelijk te verklaren doordat er meer actuele data in het model wordt opgenomen. De prestaties van voorgaande seizoenen hoeven namelijk geen garantie te zijn voor de prestaties in de toekomst. Hoe meer recentelijke data we hebben, des te beter kunnen we een beeld schetsen van de situatie in de competitie.

| | **Duitsland** | **Nederland** | **Spanje** | **Frankrijk** | **Engeland** | **Gemiddeld** |
|---|---|---|---|---|---|---|
| **Opbrengst per wedstrijd** | -8 cent | -21 cent | -13 cent | -3 cent | 17 cent | -6 cent |
| **1e helft per wedstrijd** | -21 cent | -18 cent | -9 cent | -11 cent | 4 cent | -11 cent |
| **2e helft per wedstrijd** | 5 cent | -24 cent | -16 cent | 6 cent | 30 cent | 0 cent |

Tabel 2: Resultaten strategie 2

Een interessante visualisatie van de omzet in de Engelse competitie is te zien in Figuur 8. Hierbij staan de lichtgrijze staven voor strategie 1, de donkergrijze staven voor strategie 2 en de grijze lijn voor de inzet. Alles boven de grijze lijn levert winst op, alles onder de grijze lijn levert verlies.
Er is te zien dat in de eerste seizoenshelft de opbrengsten van zowel strategie 1 als strategie 2 variëren rond de inzet. Wanneer we kijken naar de tweede seizoenshelft, zien we dat de opbrengsten stijgen en allemaal boven de grijze lijn komen liggen.




Figuur 8: Visualisatie van de omzet in Engeland volgens strategie 1 en strategie 2.

Een complete discussie over resultaten en uitgebreidere analyse kan gevonden worden in Hoofdstuk 7.

## 6.4 Kelly criterion

Er zijn verschillende speelstijlen voor het inzetten van geld op weddenschappen. Er kan gekozen worden om al je geld op één bepaalde weddenschap in te zetten, of juist je geld te verdelen door de inzet te spreiden over meerdere weddenschappen. Om de optimale balans te vinden tussen inzet en de kans op een gebeurtenis, kunnen we gebruik maken van het Kelly criterion.

**Definitie 6.4.1.** Het Kelly criterion geeft de fractie $f^*$ van je bankrol wat je in kunt zetten. Deze fractie wordt gedefinieerd als

$$f^* = \frac{bp - q}{b} \qquad [16]$$

waarbij $p$ de winkans, $q = 1 - p$ en $b =$ (quotering - 1) .

**Stelling 6.4.2.** *Het inzetten volgens het Kelly criterion zorgt voor de maximale winst op de lange termijn*

*Bewijs.* We leveren een kort bewijs. We starten met 1 euro en zetten vervolgens $f$ centen in op een weddenschap. Laat $p$ de kans op winst zijn en daarmee $1-p$ de kans op verlies. Bij winst is het nieuwe bedrag $1+fb$, en bij verlies $1-f$. Wanneer we op $n$ weddenschappen inzetten, dan verwachten we dat we $pn$ keer winst en $(1-p)n$ keer verlies hebben. In formule vorm levert dit de opbrengst

$$\text{Opbrengst} = (1+fb)^{pn}(1-f)^{(1-p)n}$$

Neem nu het logaritme aan beide kanten, we krijgen dan

$$\log(\text{Opbrengst}) = \log((1+fb)^{pn}(1-f)^{(1-p)n})$$

Wanneer we dit herschrijven krijgen we

$$\log(\text{Opbrengst}) = \log((1+fb)^{pn}(1-f)^{(1-p)n})) = \log((1+fb)^{pn}) + \log((1-f)^{(1-p)n}) =$$

$$= pn \cdot \log(1+fb) + (1-p)n \cdot \log(1-f)$$

Nu volgt dat

$$\frac{\log(\text{Opbrengst})}{n} = p \cdot \log(1+fb) + q \cdot \log(1-f)$$

We willen nu de winst maximaliseren, dus we leiden af

$$\frac{d}{df}\frac{\log(\text{Opbrengst})}{n} = \frac{d}{df} \cdot p \cdot \log(1+fb) + \frac{d}{df} \cdot q \cdot \log(1-f) \implies 0 = \frac{bp}{1+fb} - \frac{q}{1-f}$$

Wanneer we dit herschrijven levert dit

$$\frac{bp}{1+fb} = \frac{q}{1-f} \implies bp(1-f) = q(1+fb) \implies bp - bpf = q + qfb \implies bp - q = fb(q+p) \implies f = \frac{bp-q}{b(q+p)}$$

Hieruit volgt dat voor $f = \dfrac{bp-q}{b}$ de maximale winst op de lange termijn behaald wordt.

□

We kunnen dus concluderen dat inzetten volgens het Kelly criterion ons op de lange termijn de meeste winst oplevert.

# 7 Discussie

In dit hoofdstuk zullen er verschillende onderwerpen ter discussie worden gesteld en er zal worden geanalyseerd hoe deze aangepast en/of verholpen kunnen worden. Er wordt ook gekeken welke stukken eventueel in de toekomst uitgewerkt kunnen worden.

## 7.1 Discussie model

De factoren waar het model nu op gebaseerd is zijn het thuisvoordeel en de aanvals- en verdedigingssterkte van beide teams. We zagen eerder al in Hoofdstuk 3 dat er meerdere factoren zijn die invloed kunnen hebben op de resultaten. Er waren plannen om de factor 'vorm' toe te voegen aan het model, maar door gebrek aan tijd is deze factor niet meegenomen in het model. Wanneer we kijken naar toekomstig werk, dan zal een van de eerste toevoegingen de factor 'vorm' zijn.

Tevens is er benoemd dat de negatieve binomiale verdeling een goed alternatief is voor de Poissonverdeling, wanneer er geldt dat de verwachtingswaarde en variantie niet gelijk zijn. Echter is de implementatie van de negatieve binomiale verdeling niet toegepast. De implementatie van deze verdeling is voor in de toekomst een goede toevoeging om te kijken of dit eventueel tot andere resultaten leidt.

Zoals eerder verteld is in het afgelopen seizoen bij de meeste wedstrijden geen publiek aanwezig geweest. Het ontbreken van het publiek kan geleid hebben tot een lager thuisvoordeel en daarmee tot andere resultaten. Echter, uit recent onderzoek is gebleken dat het wegvallen van publiek invloed heeft op het aantal kaarten en op het schoten op doel, maar niet direct op het thuisvoordeel zelf [17]. Hieruit volgt dat we hiermee geen rekening hoeven te houden.

## 7.2 Discussie resultaten

Wanneer we kijken naar de resultaten die we hebben verkregen, zien we dat over het algemeen beide strategieën evengoed werken. Wanneer we ons nu specificeren op de tweede seizoenshelft, zien we dat strategie 2 beduidend beter werkt dan strategie 1. De logische oorzaak hiervoor zou moeten zijn dat er extra data van het huidige seizoen beschikbaar is, maar het toevoegen van een extra 876 wedstrijden op een bestaande dataset van 12264 wedstrijden zou niet zo'n enorm verschil moeten maken. Wanneer we meerdere seizoenen zouden bekijken, zouden we kunnen controleren of deze bevinding van de grotere winst in de tweede seizoenshelft een toevalstreffer was, of dat dit in het algemeen echt geldt.

In Hoofdstuk 6 zagen we dat we de winst kunnen maximaliseren door in te zetten volgens het Kelly criterion. Uiteindelijk is het inzetten met deze methode niet meegenomen in het model. De nadruk zou moeten liggen op de resultaten van de strategieën en welke strategie het meeste oplevert en niet zo zeer op het maximaliseren van deze winst. Uiteindelijk zal het inzetten op basis van het Kelly criterion altijd de inzet moeten verhogen en dit is niet bepalend voor de keuze van de strategie. Als toekomstig werk is het goed om te kijken naar het daadwerkelijk inzetten met behulp van het Kelly criterion om te kijken hoeveel meer winst er gemaakt kan worden.

## 7.3 Discussie toekomstig werk

Zoals eerder genoemd willen we, wanneer het model in de toekomst uitgebreid zou worden, kijken naar de toevoeging van de factor 'vorm'. Tevens willen we kijken hoe de resultaten zouden verschillen wanneer er gebruik gemaakt wordt van de negatieve binomiale verdeling. De resultaten zijn gebaseerd op de prestaties van 1 seizoen. Om statisch significantere resultaten te kunnen verkrijgen zou het model getest moeten worden op meerdere seizoenen. Voor toekomstig werk is het dus goed om te kijken naar meerdere seizoenen, waarbij we kijken wat de gemiddelde resultaten van het model zijn. Mede door te kijken naar meerdere seizoenen vergroten we de nauwkeurigheid van de resultaten. Hierbij kan er ook gekeken worden of er inderdaad een verschil is tussen het inzetten in de eerste en tweede seizoenshelft, zoals we eerder concludeerden.

[]graphicx[]color
framed
alltt upquote

# A Appendix A

Tabel 3: Resultaten Strategie 1: Inzetten op de winnaar

| | **Duitsland** | **Nederland** | **Spanje** | **Frankrijk** | **Engeland** |
|---|---|---|---|---|---|
| **Speelronde 1** | 11.02 | 7.57 | 7.82 | 6.08 | 12.58 |
| **Speelronde 2** | 5.42 | 10.61 | 10.3 | 10.6 | 15.9 |
| **Speelronde 3** | 7.69 | 5.03 | 5.82 | 13.68 | 15.29 |
| **Speelronde 4** | 6.01 | 8.25 | 8.58 | 7.76 | 10.04 |
| **Speelronde 5** | 3.92 | 9.08 | 11.1 | 5 | 6.4 |
| **Speelronde 6** | 7.53 | 3.97 | 15.45 | 10.05 | 8.8 |
| **Speelronde 7** | 5.86 | 9.56 | 17.4 | 9.78 | 11.49 |
| **Speelronde 8** | 11.1 | 11.04 | 10.56 | 8.56 | 14.53 |
| **Speelronde 9** | 2.61 | 4 | 1.85 | 11 | 7.4 |
| **Speelronde 10** | 5.83 | 6.06 | 5.7 | 14.2 | 6.46 |
| **Speelronde 11** | 4.05 | 26.12 | 0 | 9.78 | 11.67 |
| **Speelronde 12** | 6.74 | 4.23 | 4.6 | 12.23 | 2 |
| **Speelronde 13** | 9.15 | 9.11 | 13.57 | 8.52 | 13.58 |
| **Speelronde 14** | 7.65 | 4.65 | 9.26 | 12.7 | 5.05 |
| **Speelronde 15** | 5.3 | 6.2 | 3.53 | 11.44 | 6.88 |
| **Speelronde 16** | 4.87 | 2.5 | 14.78 | 6.85 | 14.82 |
| **Speelronde 17** | 13.01 | 5.08 | 8.4 | 10.15 | 13.52 |
| **Speelronde 18** | 6.55 | 15.13 | 12.1 | 7.03 | 14.21 |
| **Speelronde 19** | 9.57 | 3.8 | 3.21 | 12.73 | 15.45 |
| **Speelronde 20** | 0 | 14.53 | 10.32 | 4.34 | 3.32 |
| **Speelronde 21** | 16.03 | 7.63 | 10.36 | 15.36 | 16.85 |
| **Speelronde 22** | 3.79 | 17.94 | 7.73 | 8.96 | 9.95 |
| **Speelronde 23** | 11.26 | 8.1 | 8.88 | 7.78 | 14.33 |
| **Speelronde 24** | 8.53 | 5.99 | 8.5 | 6.41 | 6.04 |
| **Speelronde 25** | 9.72 | 3.5 | 9.75 | 8.08 | 11.27 |
| **Speelronde 26** | 2.37 | 3.3 | 8.71 | 19.98 | 9.02 |
| **Speelronde 27** | 5.54 | 7.75 | 6.16 | 6.11 | 7.48 |
| **Speelronde 28** | 7.1 | 4.14 | 8.19 | 10.5 | 8.36 |
| **Speelronde 29** | 11.22 | 12.91 | 13.1 | 8.12 | 16.04 |
| **Speelronde 30** | 13.14 | 10.53 | 10.1 | 8.29 | 11.59 |
| **Speelronde 31** | 11.46 | 11.1 | 7.43 | 9.61 | 13.15 |
| **Speelronde 32** | 6.12 | 8.74 | 10.94 | 10.74 | 7.63 |
| **Speelronde 33** | 1.44 | 4.6 | 5.71 | 12.01 | 12.13 |
| **Speelronde 34** | 6.7 | 9.09 | 16.27 | 5.36 | 8.94 |
| **Speelronde 35** | - | - | 10.85 | 10.92 | 10.99 |
| **Speelronde 36** | - | - | 9.9 | 6.7 | 13.27 |
| **Speelronde 37** | - | - | 8.41 | 9.04 | 10.37 |
| **Speelronde 38** | - | - | 5.99 | 1.3 | 14.95 |
| **Totaal** | 248.3 | 281.84 | 341.33 | 351.67 | 441.75 |
| **1e seizoenshelft** | 117.76 | 133.06 | 164.03 | 186.4 | 206.67 |
| **2e seizoenshelft** | 130.54 | 148.78 | 177.3 | 165.27 | 205.08 |

Tabel 4: Resultaten Strategie 2: Inzetten op de grootste winstverwachting

| | **Duitsland** | **Nederland** | **Spanje** | **Frankrijk** | **Engeland** |
|---|---|---|---|---|---|
| **Speelronde 1** | 8.32 | 8.5 | 2.1 | 6.4 | 7.43 |
| **Speelronde 2** | 24.3 | 2.25 | 7.25 | 8.4 | 9.5 |
| **Speelronde 3** | 0 | 2.15 | 10.25 | 14.5 | 13.1 |
| **Speelronde 4** | 0 | 15.75 | 15.45 | 0 | 19.45 |
| **Speelronde 5** | 0 | 3.6 | 21.25 | 3.3 | 14 |
| **Speelronde 6** | 8.75 | 0 | 23 | 5.5 | 10.5 |
| **Speelronde 7** | 4.5 | 6 | 13.05 | 10.35 | 3.6 |
| **Speelronde 8** | 3.4 | 3 | 3 | 7.4 | 12.4 |
| **Speelronde 9** | 13 | 7 | 12.2 | 13.25 | 10.75 |
| **Speelronde 10** | 6.7 | 0 | 14.52 | 13.1 | 13.75 |
| **Speelronde 11** | 8.3 | 32.62 | 22.2 | 4.5 | 3.25 |
| **Speelronde 12** | 2.3 | 2.7 | 2.5 | 15.1 | 15 |
| **Speelronde 13** | 11.65 | 8.95 | 4.4 | 16.9 | 3.3 |
| **Speelronde 14** | 1.65 | 11.7 | 5.65 | 12.8 | 13.65 |
| **Speelronde 15** | 11.95 | 7.85 | 0 | 11.15 | 10.65 |
| **Speelronde 16** | 2.15 | 6.3 | 6.15 | 6.65 | 6.85 |
| **Speelronde 17** | 13.28 | 6.65 | 1.83 | 0 | 0 |
| **Speelronde 18** | 17.47 | 8.55 | 7.3 | 9.7 | 10.8 |
| **Speelronde 19** | 0 | 15.05 | 0 | 10.37 | 18.8 |
| **Speelronde 20** | 17.05 | 0 | 13.9 | 22 | 30.7 |
| **Speelronde 21** | 0 | 8.2 | 2.6 | 3.4 | 7.8 |
| **Speelronde 22** | 12.68 | 8.15 | 1.8 | 20.3 | 15.75 |
| **Speelronde 23** | 4.5 | 15.15 | 13.3 | 0 | 6.7 |
| **Speelronde 24** | 7.4 | 6.55 | 0 | 12.1 | 3.2 |
| **Speelronde 25** | 15.55 | 3.5 | 11.65 | 7.63 | 16.7 |
| **Speelronde 26** | 7.2 | 3.3 | 11.25 | 18.8 | 3.4 |
| **Speelronde 27** | 2.37 | 3.1 | 7.9 | 3.1 | 6.95 |
| **Speelronde 28** | 10.3 | 0 | 1.95 | 13.1 | 20 |
| **Speelronde 29** | 10.5 | 12.2 | 8.6 | 17 | 3.5 |
| **Speelronde 30** | 5.55 | 5.95 | 8.1 | 15.75 | 15 |
| **Speelronde 31** | 15.3 | 16.5 | 2.05 | 0 | 19.8 |
| **Speelronde 32** | 5.75 | 7.73 | 7.03 | 0 | 7.83 |
| **Speelronde 33** | 17.75 | 0 | 19.2 | 8.35 | 20.9 |
| **Speelronde 34** | 11 | 2.45 | 18.15 | 7.2 | 15 |
| **Speelronde 35** | - | - | 11.15 | 7.75 | 15.8 |
| **Speelronde 36** | - | - | 4.95 | 8.25 | 18.5 |
| **Speelronde 37** | - | - | 16.24 | 3.75 | 13 |
| **Speelronde 38** | - | - | 0 | 32.25 | 6.5 |
| **Totaal** | 280.62 | 241.4 | 331.92 | 370.01 | 443.81 |
| **1e seizoenshelft** | 120.25 | 125.02 | 172.1 | 169.37 | 196.78 |
| **2e seizoenshelft** | 160.37 | 116.38 | 159.82 | 200.73 | 247.03 |

# B Appendix B

De code van het model in R staat hieronder beschreven. De onderstaande code gaat over de voorspellingen in Duitsland, de code voor de andere 4 landen gaat analoog, maar is weggelaten om de ruimte te beperken.

```
Negeer de onderstaande regel, deze is voor de visualisatie van de code
({r setup, include=FALSE} knitr::opts_chunk$set(warning = FALSE, message = FALSE))

#Vanaf hier begint het model
#We bekijken het model voor Duitsland

#Importeer de data
setwd("C:\\Users\\bjorn\\Scriptie\\Data\\Duitsland")
Duitsland <- ldply(list.files(), read.csv, header=TRUE)

#Verwijder overbodige data
Duitsland = Duitsland[,-c(8:22,26:129)]

#Bekijk hoeveel verschillende teams er de afgelopen 8 jaar mee hebben gedaan
teams = unique(factor(Duitsland$HomeTeam))

#Verwijder het seizoen 2020/2021
Duitsland <- as.data.frame(Duitsland)
Duitsland.tabel <- Duitsland[c(2143:2448),]

data <- with(Duitsland,list(
  N=length(Div),
  nclub = nlevels(factor(HomeTeam)),
  thuisteam=as.numeric(factor(HomeTeam)),
  uitteam=as.numeric(factor(AwayTeam)),
  doelpuntenthuis=FTHG,
  doelpuntenuit=FTAG))

#Scheid de trainseizoenen van het seizoen dat we willen voorspellen
data <- as.data.frame(data)
Duitsland.train = data[-c(2143:2448),]
Duitsland.test = data[c(2143:2448),]

#Voeg het aantal teams toe
Duitsland.train$N <- c(26)
Duitsland.train$nclub<- c(26)
Duitsland.test$N <- c(26)
Duitsland.test$nclub<- c(26)

#In de onderstaande regels zorgt [1:90,] dat we de data tot en met speelronde 10
#meenemen bij de voorspelling van speelronde 11.

Duitsland.train <- rbind(Duitsland.train, Duitsland.test[1:90,])
Duitsland.train <- as.list(Duitsland.train)
Duitsland.train$N <- c(26)
Duitsland.train$nclub<- c(26)

winst = numeric()
model <- function() {
```

```
  for (i in 1:N) {
    doelpuntenthuis[i] ~ dpois(HS[i])
    doelpuntenuit[i]  ~ dpois(OS[i])
    log(HS[i]) <-  aanvalthuis[thuisteam[i]] + verdediginguit[uitteam[i]]
    log(OS[i]) <-  verdedigingthuis[thuisteam[i]] + aanvaluit[uitteam[i]]
  }
  for (i in 1:nclub) {
    aanvalthuis[i] <-  algemeen[i]+rest[i] + thuisvoordeel
    aanvaluit[i] <-  algemeen[i]+rest[i]
    verdedigingthuis[i] <-  algemeen[i]-rest[i] + thuisvoordeel
    verdediginguit[i] <-  algemeen[i]-rest[i]
    algemeen[i] ~ dnorm(0,pow(sigmatop,-2))
    rest[i]     ~ dnorm(0,pow(sigmarest,-2))
  }
  sigmarest ~ dunif(0,1000)
  sigmatop ~ dunif(0,1000)
  thuisvoordeel ~ dnorm(0,0.0001)
}


parameters <- c("aanvalthuis","verdedigingthuis","aanvaluit", "verdediginguit")
inits <- function(){
  list(algemeen=rnorm(Duitsland.train$nclub),
       rest=rnorm(Duitsland.train$nclub),
       sigmarest=runif(1),
       sigmatop=runif(0,0.5),
       thuisvoordeel=rnorm(1)
  )
}
jagsDuitsland<- jags(Duitsland.train, model=model, inits=inits,
                parameters=paramaters,progress.bar="gui",
                n.iter=15000)

#We simuleren nog een keer tot convergentie
jagsDuitsland <- autojags(jagsDuitsland,n.iter=20000)
jagsDuitsland.mcmc <- as.mcmc(jagsDuitsland)

#Sla data op
jags.smat <- jagsDuitsland$BUGSoutput$summary


#Met deze tools kunnen de density en de trace van een variabele geplot worden
#de invoer is gelijk aan aanvalthuis[i]/verdedigingthuis[i] etc. waarbij de i
#staat voor het team waarvan je de plot wilt opvragen

denplot(jagsDuitsland.mcmc, parms = c("aanvaluit[23]", "verdediginguit[23]"))
traplot(jagsDuitsland.mcmc, parms = c("aanvaluit[23]", "verdediginguit[23]"))
```

```
#In de onderstaande regels zorgt [91:99,] er voor dat we de 11 speelronde gaan
#voorspellen.

Duitsland.tabel <- Duitsland.tabel[91:99,]

#Bereken het aantal thuisdoelpunten voor met behulp van de verkregen tabel
dpt_thuis_voor<- numeric()
for (i in Duitsland.test[91:99,]$thuisteam) {
  thuisdpt <-  jags.smat[i,1]
  dpt_thuis_voor<- c(dpt_thuis_voor,thuisdpt)
}

#Bereken het aantal thuisdoelpunten tegen met behulp van de verkregen tabel
dpt_thuis_tegen <- numeric()
for (i in Duitsland.test[91:99,]$uitteam) {
  uitdpt <-  jags.smat[79+i,1]
  dpt_thuis_tegen <- c(dpt_thuis_tegen,uitdpt)
}

#Bereken het totale aantal thuisdoelpunten
samengesteldthuis <- exp(c(dpt_thuis_voor + dpt_thuis_tegen))

#Voeg het aantal thuisdoelpunten toe als kolom
Duitsland.tabel$thuisdoelpunten <- c(samengesteldthuis)

#Bereken het aantal uitdoelpunten voor met behulp van de verkregen tabel
dpt_uit_voor<- numeric()
for (i in Duitsland.test[91:99,]$thuisteam) {
  thuisdpt <-  jags.smat[53+i,1]
  dpt_uit_voor<- c(dpt_uit_voor,thuisdpt)
}

#Bereken het aantal uitdoelpunten tegen met behulp van de verkregen tabel
dpt_uit_tegen <- numeric()
for (i in Duitsland.test[91:99,]$uitteam) {
  uitdpt <-  jags.smat[26+i,1]
  dpt_uit_tegen <- c(dpt_uit_tegen,uitdpt)
}

#Bereken het totale aantal uitdoelpunten
samengestelduit <- exp(c(dpt_uit_voor + dpt_uit_tegen))

#Voeg het aantal uitdoelpunten toe als kolom

Duitsland.tabel$uitdoelpunten <- c(samengestelduit)

#Bereken nu met behulp van de Poissonverdeling de kansen op alle doelpunten
Duitsland.tabel <- as.data.frame(Duitsland.tabel)
x <- 0:8
thuis = character()
gelijk = character()
uit = character()
for(i in 1:nrow(Duitsland.tabel)){
```

```
  poissonthuis = dpois(x,Duitsland.tabel[i,11])
  poissonuit = dpois(x,Duitsland.tabel[i,12])

  tabel <- poissonthuis %o% poissonuit
  rownames(tabel) = 0:8
  colnames(tabel) = 0:8

  #De diagonaal van de tabel is gelijk aan de kans op een gelijkspel
  X = sum(diag(tabel))

  #De benedendriehoek van de tabel is gelijk aan de kans op een winst voor thuis
  Y = sum(tabel[lower.tri(tabel)])

  #De bovendriehoek van de tabel is gelijk aan de kans op een winst voor uit
  Z = sum(tabel[upper.tri(tabel)])

  thuis <- c(thuis, Y)
  gelijk <- c(gelijk, X)
  uit <- c(uit,Z)
}

#Voeg nu de kolommen toe met de kans op thuiswinst, gelijkspel en uitwinst
Duitsland.tabel$thuiswinst <- as.numeric(thuis)
Duitsland.tabel$gelijk <- as.numeric(gelijk)
Duitsland.tabel$uitwinst <- as.numeric(uit)


##########################################################################

#We bekijken nu het inzetten via strategie 1: het inzetten op de winnaar.

#Bekijk welk team de grootste kans heeft om te winnen
inzet_strategie1 = character()
resultaat_strategie1 = character()
for(i in 1:nrow(Duitsland.tabel)){
  if (Duitsland.tabel[i,13] >Duitsland.tabel[i,15]) {
    resultaat_strategie1<- "H"
  } else if (Duitsland.tabel[i,13]<Duitsland.tabel[i,15]) {
    resultaat_strategie1<- "A"
  } else
    resultaat_strategie1<- "D"
  inzet_strategie1 <- c(inzet_strategie1,resultaat_strategie1)
}

#Voeg nu een kolom toe met de voorspelling
Duitsland.tabel$voorspelling_strategie1 <- inzet_strategie1

#Controleer of de voorspelling overeenkomt met het werkelijke resultaat
Duitsland.tabel$klopt_voorspelling1 <- Duitsland.tabel$voorspelling_strategie1
== Duitsland.tabel$FTR
```

```
#We koppelen nu de quotering die hoort bij de wedstrijd aan de voorspelling
inzetten_strategie1 = character()
resultaatinzetten_strategie1 = character()
for(i in 1:nrow(Duitsland.tabel)){
  if (Duitsland.tabel[i,16] == "H") {
    inzetten_strategie1 <- Duitsland.tabel[i,8]
  } else if (Duitsland.tabel[i,16] == "A") {
    inzetten_strategie1 <- Duitsland.tabel[i,10]
  } else
    inzetten_strategie1 <- Duitsland.tabel[i,9]
  resultaatinzetten_strategie1 <- c(resultaatinzetten_strategie1,
                                    inzetten_strategie1)
}

#Voeg nu de kolom met quoteringen toe
Duitsland.tabel$quotering_voorspelling1 <- resultaatinzetten_strategie1

#We kijken nu hoeveel geld de voorspellingen hebben opgeleverd

uitkering_strategie1 = character()
winst_strategie1 = character()
for(i in 1:nrow(Duitsland.tabel)){
  if (Duitsland.tabel[i,17] == "TRUE"){
    uitkering_strategie1  <- Duitsland.tabel[i,18]
  } else{
    uitkering_strategie1  <- 0
  }
  winst_strategie1 <- c(winst_strategie1,uitkering_strategie1)
}
Duitsland.tabel$Opbrengst_strategie1 <- as.numeric(winst_strategie1)
Totale_winst_strategie1 = sum(Duitsland.tabel$Opbrengst_strategie1)


################################################################################

#We bekijken nu strategie 2, het inzetten op de hoogste verwachte opbrengst

#Bereken de verwachte opbrengst voor het inzetten op thuiswinst
opbrengstthuis <- numeric()
for(i in 1:nrow(Duitsland.tabel)){
  Winstthuis <- Duitsland.tabel[i,13]*(Duitsland.tabel[i,8] - 1)
  Verliesthuis <- 1- Duitsland.tabel[i,13]
  Opbrengstthuis <- Winstthuis - Verliesthuis
  opbrengstthuis <- c(opbrengstthuis,Opbrengstthuis)
}

#Bereken de verwachte opbrengst voor het inzetten op gelijkspel
opbrengstgelijk <- numeric()
for(i in 1:nrow(Duitsland.tabel)){
  Winstgelijk <- Duitsland.tabel[i,14]*(Duitsland.tabel[i,9] - 1)
  Verliesgelijk <- 1- Duitsland.tabel[i,14]
  Opbrengstgelijk <- Winstgelijk - Verliesgelijk
  opbrengstgelijk <- c(opbrengstgelijk,Opbrengstgelijk)
}
```

```
#Bereken de verwachte opbrengst voor het inzetten op uitwinst
opbrengstuit <- numeric()
for(i in 1:nrow(Duitsland.tabel)){
  Winstuit <- Duitsland.tabel[i,15]*(Duitsland.tabel[i,10] - 1)
  Verliesuit <- 1- Duitsland.tabel[i,15]
  Opbrengstuit <- Winstuit - Verliesuit
  opbrengstuit <- c(opbrengstuit,Opbrengstuit)
}
#Voeg nu kolommen met de verwachte opbrengsten toe
Duitsland.tabel$Verwacht_thuis <- c(opbrengstthuis)
Duitsland.tabel$Verwacht_gelijk <- c(opbrengstgelijk)
Duitsland.tabel$Verwacht_uit <- c(opbrengstuit)

#Bepaal nu aan de hand van de verwachte opbrengst op welke weddenschap er in
#gezet moet worden, en koppel de voorspelling aan de quotering

inzet_strategie2 = character()
resultaat_strategie2 = character()
for(i in 1:nrow(Duitsland.tabel)){
  if (Duitsland.tabel[i,20] > Duitsland.tabel[i,22] &&
      Duitsland.tabel[i,20] > Duitsland.tabel[i,21] &&
      Duitsland.tabel[i,20] > 0) {
    resultaat_strategie2 <- Duitsland.tabel[i,8]
  } else if (Duitsland.tabel[i,22] > Duitsland.tabel[i,20] &&
             Duitsland.tabel[i,22] > Duitsland.tabel[i,21] &&
             Duitsland.tabel[i,22] > 0) {
    resultaat_strategie2 <- Duitsland.tabel[i,10]
  } else if (Duitsland.tabel[i,21] > Duitsland.tabel[i,20] &&
             Duitsland.tabel[i,21] > Duitsland.tabel[i,22] &&
             Duitsland.tabel[i,21] > 0) {
    resultaat_strategie2 <- Duitsland.tabel[i,9]
  } else
    resultaat_strategie2 <- 0
  inzet_strategie2 <- c(inzet_strategie2,resultaat_strategie2)
}

#Voeg nu de kolom met de inzetstrategie toe
Duitsland.tabel$Inzet_strategie2 <- inzet_strategie2

#We maken de voorspelling aan de hand van de eerder berekende verwachte opbrengst
inzetten_strategie2 = character()
voorspelling_strategie2 = character()
for(i in 1:nrow(Duitsland.tabel)){
  if (Duitsland.tabel[i,20] > Duitsland.tabel[i,22] &&
      Duitsland.tabel[i,20] > Duitsland.tabel[i,21]  ) {
    voorspelling_strategie2 <- "H"
  } else if (Duitsland.tabel[i,22] > Duitsland.tabel[i,20] &&
             Duitsland.tabel[i,22] > Duitsland.tabel[i,21] ) {
    voorspelling_strategie2 <- "A"
  } else
    voorspelling_strategie2 <- "D"
  inzetten_strategie2 <- c(inzetten_strategie2,voorspelling_strategie2)
}
```

```
#Voeg nu de kolom met de voorspelling toe
Duitsland.tabel$Voorspelling_strategie2 <- inzetten_strategie2

#Controleer of de voorspelling overeenkomt met het werkelijke resultaat
Duitsland.tabel$klopt_voorspelling2 <- Duitsland.tabel$Voorspelling_strategie2
== Duitsland.tabel$FTR


#We kijken nu hoeveel geld de voorspellingen hebben opgeleverd
uitkering_strategie2 = character()
winst_strategie2 = character()
for(i in 1:nrow(Duitsland.tabel)){
  if (Duitsland.tabel[i,25] == "TRUE"){
    uitkering_strategie2 <- Duitsland.tabel[i,23]
  } else{
    uitkering_strategie2 <- 0
  }
  winst_strategie2 <- c(winst_strategie2,uitkering_strategie2)
}
Duitsland.tabel$Opbrengst_strategie2 <- as.numeric(winst_strategie2)
Totale_winst_strategie2 <- sum(Duitsland.tabel$Opbrengst_strategie2)

Negeer onderstaande errormelding

## Error:  <text>:2:8:  unexpected symbol
## 1:
## 2:  Negeer de
##           ^
```

# Referenties


[1] M. J. Maher. „Modelling association football scores". In: *Statistica Neerlandica* 36.3 (sep 1982), p. 109–118.

[2] Dimitris Karlis en Ioannis Ntzoufras. „On modelling soccer data". In: *Student* 3.4 (2000), p. 229–244.

[3] Tugbay Inan. „Using poisson model for goal prediction in European football". In: (2020).

[4] H.P. Lopuhaä F.M. Dekking C. Kraaikamp. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer Texts in Statistics. Springer, 2005. ISBN: 9781852338961,1852338962.

[5] Mark J. Dixon en Stuart G. Coles. „Modelling Association Football Scores and Inefficiencies in the Football Betting Market". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 46.2 (1997), p. 265–280. ISSN: 00359254, 14679876.

[6] Richard Pollard. „Home advantage in football: A current review of an unsolved puzzle". In: *The open sports sciences journal* 1.1 (2008).

[7] Wikipedia. *Jose Mourinho*. URL: `https://nl.wikipedia.org/wiki/Jose_Mourinho`.

[8] Dimitris Karlis en Ioannis Ntzoufras. „Analysis of sports data by using bivariate Poisson models". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 52.3 (2003), p. 381–393.

[9] Thomas Reilly, Jan Cabri en Duarte Araújo. *Science and Football V: The Proceedings of the Fifth World Congress on Sports Science and Football*. Routledge, 2005.

[10] Gianluca Baio en Marta Blangiardo. „Bayesian hierarchical model for the prediction of football results". In: *Journal of Applied Statistics* 37.2 (2010), p. 253–264.

[11] Christopher M Bishop. „Pattern recognition and machine learning". In: (2006).

[12] Ioannis Ntzoufras. *Bayesian modeling using WinBUGS*. Deel 698. John Wiley & Sons, 2011.

[13] Jianming Ma en Kara M Kockelman. „Bayesian multivariate Poisson regression for models of injury count, by severity". In: *Transportation Research Record* 1950.1 (2006), p. 24–34.

[14] Alboukadel Kassambara. *Machine learning essentials: Practical guide in R*. Sthda, 2018.

[15] Putting a football model into JAGS. *Wingfeet*. URL: `https://www.r-bloggers.com/2012/10/putting-a-football-model-into-jags/`.

[16] J. L. Kelly. „A new interpretation of information rate". In: *The Bell System Technical Journal* 35.4 (1956), p. 917–926.

[17] Fabian Wunderlich e.a. „How does spectator presence affect football? Home advantage remains in European top-class football matches played without spectators during the COVID-19 pandemic". In: *Plos one* 16.3 (2021), e0248590.