

Het SIR-model verdeelt de populatie (N) in drie aparte groepen: Vatbaar (*Susceptible*, S), besmettelijk (*Infectious*, I) en verwijderd (*Removed*, R). De laatste groep omvat zowel de dodelijke slachtoffers als de genezen patiënten. Het doel van het SIR-model is de evolutie van de aantallen binnen deze groepen te bepalen. Hiervoor worden enkele veronderstellingen gemaakt over de transitie van individuen tussen de groepen in functie van de tijd (t).

De eerste assumptie legt de verplaatsingen vast die mogelijk zijn tussen de compartimenten zoals weergegeven door de pijl in de onderstaande illustratie, i.e. men kan enkel van vatbaar (S) naar besmettelijk (I) en van besmettelijk naar verwijderd (R) evolueren. We veronderstellen dus impliciet dat een genezen persoon niet opnieuw vatbaar kan worden.



De tweede assumptie bepaalt de snelheid waarmee individuen zich verplaatsen tussen de groepen. Deze worden samengevat in drie (gekoppelde) differentiaalvergelijkingen:

$$\frac{dS}{dt} = -\beta I \frac{S}{N}, \quad \frac{dI}{dt} = \beta I \frac{S}{N} - \gamma I, \quad \frac{dR}{dt} = \gamma I,$$

waarbij β en γ twee parameters zijn die mee de evolutie bepalen; β drukt het verwachte aantal besmettelijke contacten van een besmette persoon per dag uit; γ is de proportie van besmette patiënten die genezen of sterven per dag. Initieel zijn deze vrij te kiezen als strikt positieve waarden, maar uiteindelijk zullen deze geschat worden uit de data die beschikbaar is. Een vaak vermelde waarde is de verhouding van deze twee parameters: $R_0 = \beta/\gamma$, ook wel het reproductiegetal van de epidemie genoemd. Ze drukt het verwachte aantal nieuwe besmettingen uit die één besmette persoon veroorzaakt in een populatie van enkel vatbare personen. Hoe groter R_0 , hoe sneller de epidemie zich verspreidt en dit op exponentiele wijze.

Dit stelsel differentiaalvergelijkingen werd voor het allereerst voorgesteld en bestudeerd door *William Ogilvy Kermack* en *Anderson Gray McKendrick* in 1927. Laten we de redenering van deze vergelijkingen toelichten met een discreet voorbeeld en dit vervolgens omzetten naar het continue geval beschreven door de differentiaalvergelijkingen.

We beschouwen een populatie van 500 ($N = 500$) individuen. We veronderstellen dat op dag t er 200 besmettelijk individuen zijn ($I(t) = 200$) en er nog 250 vatbaar zijn ($S(t) = 250$). Bijgevolg zijn er 50 genezen of gestorven individuen ($R(t) = N - S(t) - I(t) = 50$). Om de evolutie naar de volgende dag te beschrijven, veronderstellen we verder dat elk besmettelijk individu één besmettelijk contact heeft ($\beta = 1$) en dat een besmettelijk individu gemiddeld 10 dagen lang de ziekte kan verspreiden ($\gamma = 0.1$). We beschouwen de tijd als discreet en kijken naar de verandering in $S(t)$, $I(t)$ en $R(t)$ de volgende dag.

We hebben 200 besmette personen die elk gemiddeld één besmettelijk contact hebben. Een deel van deze besmettelijke contacten zullen echter zelf al besmet, genezen of gestorven zijn. We moeten dus enkel de contacten met vatbare mensen in rekening brengen. Aangezien er nog maar 50% ($S(t)/N$) van de personen vatbaar zijn, zullen deze besmettelijke contacten leiden tot $1 * 200 * 0.5 = 100$ nieuwe besmettingen. Dit is equivalent met -100 vatbare personen. De verandering van $S(t)$ kunnen we dus samenvatten als:

$$S(t + 1) = S(t) - \beta I(t) \frac{S(t)}{N} = 250 - 1 * 200 * \frac{250}{500} = 150.$$

Voor het aantal besmette personen hebben we reeds de 100 nieuwe besmettingen. Daarvan moeten we nog het aantal genezen of gestorven patiënten aftrekken. We hadden 200 besmettelijke patiënten, die allen gemiddeld 10 dagen lang besmettelijk zijn. Indien we dit uitmiddelen over deze periode, zullen er gemiddeld 20 ($\gamma I(t)$) genezen of sterven per dag. Alles samen levert dit ons een toename van 80 individuen in de besmettelijke groep:

$$I(t + 1) = I(t) + \beta I(t) \frac{S(t)}{N} - \gamma I(t) = 200 + 1 * 200 * \frac{250}{500} - 0.1 * 200 = 280.$$

Het aantal patiënten dat vertrekt uit de besmettelijke categorie, komt in de verwijderde categorie terecht, dus:

$$R(t + 1) = R(t) + \gamma I(t) = 50 + 20 = 70.$$



Op de volgende dag zijn er dus nog 150 vatbare, 280 besmettelijke en 70 genezen of gestorven personen.

Nu zou je de terechte opmerking kunnen maken dat individuen die besmet geraken doorheen de dag, zelf ook nog andere mensen kunnen besmetten tijdens dezelfde dag. Dit scenario lijkt niet opgenomen in de discrete vergelijkingen. We moeten de veranderingen uitdrukken over een 'kleine' tijdstap (Δt), voldoende klein dat dit soort situatie zich niet kan voordoen. We normaliseren het aantal besmettelijke contacten over de tijdsplan Δt , zodat

$$S(t + \Delta t) = S(t) - \frac{\beta I(t)S(t)}{N} \Delta t$$

Door de vergelijkingen een beetje te transformeren, bekommen we:

$$\frac{S(t + \Delta t) - S(t)}{\Delta t} = - \frac{\beta I(t)S(t)}{N}$$

Indien we hier de limiet nemen voor $\Delta t \rightarrow 0$, krijgen we in het linkerlid de afgeleide van $S(t)$ en verschijnt de eerste differentiaalvergelijking van het SIR-model. Net zoals de snelheid waarmee je beweegt de afgeleide is van de plaats naar de tijd, kan je de afgeleide van $S(t)$ interpreteren als de ogenblikkelijke snelheid waarmee de vatbare aantallen veranderen. Analoog krijg je de differentiaalvergelijkingen voor $I(t)$ en $R(t)$.

Met een goed begrip van wat de SIR-vergelijkingen precies modelleren, kunnen we ons focussen op het oplossen ervan. Bij differentiaalvergelijkingen is de onbekende niet een waarde of veranderlijke, maar een functie. We zoeken dus drie functies, $S(t)$, $I(t)$ en $R(t)$ wiens afgeleiden (de ogenblikkelijke veranderingen) voldoen aan de vooropgestelde vergelijkingen.

Het voordeel van de abstractie naar differentiaalvergelijkingen is dat we kunnen terugvallen op de wiskundige tools die ontworpen zijn gedurende de 400 jaar sinds de introductie van de Calculus door Newton en Leibniz. De ontwikkeling van analytische en numerieke oplossingen is voortgevloeid uit de talloze toepassingen binnen de wetenschappen en uit de interesse voor de wiskunde zelf. Het is

bovendien een tak van de wiskunde die nog lang niet af is; er reteren dus nog talloze uitdagende en interessante problemen voor de toekomstige wiskundigen!

Hoewel er (ingewikkelde) analytische oplossingen bestaan voor het SIR-model, is het ook een mooie toepassing voor het gebruik van numerieke methoden. Hierbij beperkt zich men tot een benadering van de functies en dit enkel in een select aantal waarden (ook wel een **grid** genoemd). De werking van deze methoden valt buiten het doel van dit artikel, maar de ervaren lezer kan zelf snel aan de slag met de *Runge-Kutta* methoden, zoals de voor- of achterwaartse Euler methoden.

We beschouwen een simulatie met een populatie van één miljoen, waarvan in het begin 100 besmettelijk zijn en geen enkele genezen of gestorven. We stellen $\gamma = 0.1$ (10 dagen besmettelijk) en laten β variëren. In Figuren 1 & 2 zien we het verloop van de epidemie. Indien $\beta = 1$ (*gemiddeld 1 contact per dag*) zien we na 8 dagen een piek bereiken van rond de 600.000 besmettelijke patiënten. Terwijl voor $\beta = 0.25$ (*gemiddeld 1 contact per 4 dagen*) de piek pas na 45 dagen valt en er slechts rond de 200.000 besmettingen zijn; bovendien blijft in dit geval ongeveer 10% van de populatie gevrijwaard van een besmetting. Dit toont aan dat we met kleine aanpassingen die effect hebben op het aantal besmettelijke contacten, we de piek van besmettingen hevig kunnen dempen. Vandaar dat de huidige maatregelen zo streng zijn. Ze zorgen er niet enkel voor dat de gezondheidszorg overeind blijft, maar dat ook een deel van de bevolking beschermd wordt van zodra de epidemie uitdooft.

Figure 1

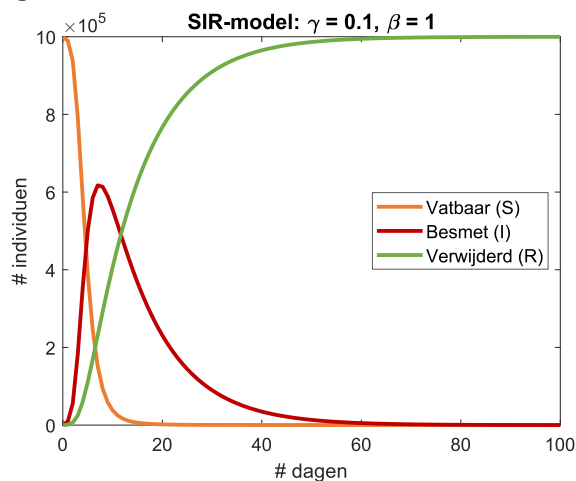
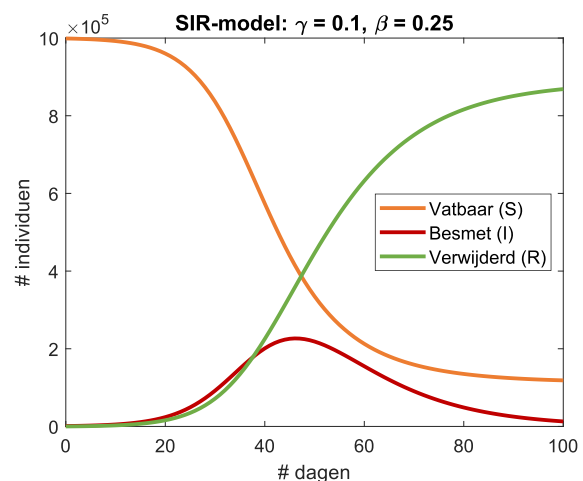


Figure 2



We zien dus dat de parameters van het SIR-model, β en γ , een sterke invloed hebben op het voorspelde verloop van de epidemie. Dus rest er nog de vraag: hoe bepalen we β en γ ? Het korte antwoord: **Data Science & Statistiek!** Op basis van enkele observaties tracht men de parameters te schatten. Indien we enkel data hebben over de aantallen en niet de individuele personen, moeten we gebruik maken van 'curve fitting'. We selecteren de parameters die ervoor zorgen dat de bijhorende functies het 'beste' de data benaderen. Het is uiteraard mogelijk en ook te verwachten, dat geen enkele keuze van β en γ oplossingen zal genereren die de data perfect doorkruisen. Dit omdat een model zelden een perfecte weergave is van de realiteit. Er bestaan verschillende criteria om te bepalen welke parameters optimaal zijn, maar dit zou ons opnieuw te ver leiden in de details van de wiskundige statistiek.

Het verhaal eindigt niet met het SIR-model. Integendeel, het is nog maar het begin.

- We zouden bijvoorbeeld een expliciet onderscheid kunnen maken tussen gestorven en genezen. Verder kan men een betere weergave krijgen van de belasting op de gezondheidszorg door meer compartimenten toe te voegen, zoals hospitalisaties en intensieve zorgen. De transities tussen deze groepen vergen dan weer bijkomende parameters om de verandering te modeleren en voegen telkens vergelijkingen toe aan het stelsel op analoge wijze als het originele SIR-model. Deze parameters moeten geschat worden, wat betekent dat je voldoende data nodig hebt om te vermijden dat je de ruis modelleert i.p.v. het signaal (een typisch *overfitting* probleem). Zo heb je steeds een afwisseling tussen wat de data toelaat en de verfijning van je model.
- Door externe factoren, zoals de maatregelen van de overheid, zou je verwachten dat het verloop van de epidemie verandert. In dit geval zijn de parameters van je SIR-model zelf ook afhankelijk van de tijd. In principe zou je hiermee dus het effect van bijvoorbeeld overheidsmaatregelen kunnen kwantificeren in functie van de schommelingen in de parameters.
- Ten slotte is de kwaliteit van de data ook niet onmiskenbaar. Omdat niet alle landen of deelstaten op dezelfde manier de besmettingen tellen, zijn deze data niet eenvoudig cumuleerbaar. Ook dit brengt een uitdaging met zich mee om ervoor te zorgen dat er kwaliteitsvolle data ter beschikking staat opdat de statistische analyse ervan zinvol is. Het is een oud gezegde in de statistiek, maar nog steeds van toepassing: *"garbage in, garbage out!"*.

Uiteindelijk is het modeleren van epidemie is een knap voorbeeld van het gebruik van wiskunde in data science. Het specifieke SIR-model biedt niet enkel inzicht in wiskundige abstractie, het is ook een combinatie van klassieke analyse, numerieke methoden en statistiek, die samen bijdragen tot één bepaalde toepassing. Het helpt om de verspreiding van een epidemie te verklaren, alsook het verdere verloop te voorspellen op basis van voorgaande observaties. Het is verder een domein waarin de vraag naar wiskundig onderbouwde resultaten groot zal blijven. Heb je verder nog interesse in de technische details, bekijk dan zeker eens het onderzoek van [prof. Barbé](#).

Gilles Inghelbrecht

Mei 2020