

# DEMOGRAFIE

## 1. Het belang van demografie

Demografie vervult een niet weg te denken rol in onze huidige samenleving. Diverse instanties, zoals in Nederland het Centraal Bureau voor de Statistiek (CBS) leveren jaarlijks tellingen van de bevolking en andere gegevens die te maken hebben met de toename of afname van de bevolking, zoals het aantal overledenen en geboren. Deze tellingen zijn op zichzelf interessant genoeg; het is de taak van de demograaf om zo nauwkeurig mogelijk voorspellingen te doen over hoe de bevolking van een stad, land of van de gehele wereld zich in de toekomst zal ontwikkelen.

Met hoeveel mensen moeten wij de aarde delen? Hoelang duurt het voordat de mensheid de gehele hoeveelheid aardgas en -olie heeft opgesoupeerd? Hoe belangrijk is het dat wij gaan uitzien naar andere vormen van energie?

De PTT is in oktober 1995 overgegaan tot 10-nummerige telefoonnummers in geheel Nederland, onder andere omdat de verwachte bevolkingsgroei en de daarmee samenhangende telefoonbehoefte dat nodig maakte.

Een hot item is de vergrijzing van de maatschappij; de babies van de baby-boom die na de oorlog heeft plaatsgevonden beginnen zo langzamerhand ouder te worden en de aanwas van kinderen is afgenomen sindsdien. Dit heeft grote gevolgen voor de samenstelling van de bevolking. Kunnen we daarover voorspellingen doen?

Dit zijn slechts enkele voorbeelden van vragen die tot het terrein van de demografie behoren. In deze Wiskunde Werkt opdracht zullen wij ons bezig houden met bevolkingsgroei, bevolkingssamenstelling en geboortegolven; we zullen aan de hand van gegevens van het CBS de grootte en samenstelling van de Nederlandse bevolking en de invloed van geboortebeperving of -verruiming en medische vooruitgang proberen te voorspellen.

## 2. Bevolkingsgroei

Bevolkingsgroei is in principe geometrisch. Om dat in te zien volstaat het een voorbeeld te bekijken met een eenvoudige fictieve populatie: een paar konijnen. Dit paar konijnen produceert aan het eind van het eerste jaar van hun leven een eerste paar konijnen en aan

het eind van het tweede jaar nog een tweede paar konijnen. Het oorspronkelijke paar blijft daarna wel in leven, maar produceert verder geen nageslacht meer. Al hun nakomelingen planten zich voort volgens dezelfde regels en blijven nadat ze hun tweede paar voortgebracht hebben ook in leven. Het aantal paren konijnen aan het eind van  $n$  jaren noemen we  $u_n$ . Om te beginnen is  $u_0 = 1$  en  $u_1 = 2$ .

### Opdracht 2.1

- Leid een recurrente betrekking en/of een formule af voor  $u_n$ .
- Laat zien dat voor grote waarden van  $n$  de bevolking na  $n + 1$  jaar zich tot de bevolking na  $n$  jaar verhoudt als min of meer constant voor grote waarden van  $n$ .
- Hoe snel groeit deze bevolking in procenten per jaar?

In het vorige voorbeeld hebben we kunnen zien dat een bevolking geometrisch groeit. Als een bevolking van 1 miljard mensen groeit met een snelheid van 1.5% per jaar, dan kunnen we het aantal mensen na 10 jaar uitrekenen; dat moet gelijk zijn aan  $(1.015)^{10} \approx 1.16$  miljard mensen. Als we de groeisnelheid per jaar  $r\%$  is en het aantal mensen nu gelijk is aan  $N_0$ , dan is het aantal mensen na  $t$  jaren gelijk aan

$$(2.1) \quad N_t = N_0 \left(1 + \frac{r}{100}\right)^t .$$

Een begrip dat demografen vaak hanteren omdat het gemakkelijk werkt is het zogenaamde verdubbelingsgetal, d.w.z. het aantal jaar dat een bevolking nodig heeft om te verdubbelen. Dit idee is niet nieuw; het wordt ook gebruikt in de financiële wereld en toegepast in nucleaire fysica (halveringstijd).

**Opdracht 2.2** Als benadering voor het verdubbelingsgetal wordt vaak genomen  $70/r$ , waar  $r$  weer de groeisnelheid is per jaar in procenten.

- Wat is het verdubbelingsgetal voor een groeisnelheid van 1.5%, zowel exact als volgens deze formule?
- Waar komt de formule vandaan?
- Geef een verklaring waarom de formule minder goed werkt voor de konijnenpopulatie.

Ontwikkelingslanden hebben vaak een hogere bevolkingsgroei, dus een lager verdubbelingsgetal.

**Opdracht 2.3** In 1970 had de Verenigde Staten een bevolking van 205 miljoen mensen en een groeisnelheid van 1.2% per jaar; buurland Mexico had op dat moment een bevolking van 50 miljoen, maar een groeisnelheid van 3.5% per jaar.

- Maak een schatting van het jaartal waarin Mexico meer inwoners zal tellen dan de Verenigde Staten, als beide landen op dezelfde snelheid blijven groeien (we negeren hier ook even het feit dat er nog steeds veel mensen vanuit Mexico de VS binnenkomen).

### 3. Bevolkingssamenstelling

Stel je een stad voor met 100.000 inwoners. We kunnen de bevolking van de stad onderverdelen in leeftijdsgroepen van 5 jaar en voor elk van de groepen tellen hoeveel personen erin vallen. Een mogelijke uitslag zou gegeven kunnen worden door de volgende grafiek.

Figuur 3.1. De bevolking van een stad

Het is duidelijk dat zowel de groei als de samenstelling van de bevolking beïnvloed wordt door onzekere factoren als geboortes en sterftes. Het is voor de gemeente van groot belang om in ieder geval een idee te hebben van hoe beide aspecten van de bevolking zich in de komende jaren zullen ontwikkelen. Beleidsplannen en begrotingen zullen daarop afgestemd moeten worden. Hier zullen we ons concentreren op de samenstelling van de bevolking op grond van leeftijd. Laten we eens kijken wat er in onze stad over 5 jaar aan de hand is. De mensen die nu in de leeftijdsgroep 20–24 zitten, zullen, als ze na 5 jaar nog in leven zijn, doorschuiven naar de leeftijdsgroep 25–29. Het aantal mensen in de leeftijdsgroep 25–29 over 5 jaar zal dus in ieder geval kleiner of gelijk zijn aan het aantal personen nu in de leeftijdsgroep 20–24, tenminste als we aannemen dat er geen mensen van buitenaf naar de stad verhuizen. Als we de leeftijdsgroep 20–24 over 5 jaar willen bekijken, zijn we geïnteresseerd in de groep die nu een leeftijd van 15–19 heeft. Dus aan de ene kant neemt de bevolking af door sterfte; aan de andere kant neemt de bevolking natuurlijk toe door

geboortes. Over 5 jaar zullen in de leeftijdsgroep 0–4 alle mensen opgenomen zijn die in de tussengelegen 5 jaar geboren zijn en inmiddels niet overleden. Voor alle duidelijkheid zij erop gewezen dat bijvoorbeeld de leeftijdsgroep 0–4 alle kinderen bevat in hun eerste tot en met hun 5de levensjaar op het moment van de telling. Zij hebben op dat moment geen van allen nog hun 5de verjaardag gevierd.

Laten we de bevolking weer onderverdeeld zien in 20 leeftijdsgroepen van 5 jaren. Het doel is om de samenstelling van een stationaire bevolking, zo die er is, te achterhalen. Eerst nemen we aan dat de bevolking gesloten is voor migratie. Er komen dus van buiten af geen mensen bij en er gaan geen mensen weg. De bevolkingstoename of -afname wordt dan volledig bepaald door geboortes en sterftes. Verder nemen we aan dat er over de komende 5 jaar  $B$  mensen geboren worden die over 5 jaren van nu nog in leven zijn.

Wat kunnen we met deze aannames bereiken? Nummer de leeftijdsgroepen van 0 t/m 19, dus groep nr.  $i$  bevat personen in hun  $(5i)$ de tot en met hun  $(5i+4)$ de levensjaar. Laat  $N_i$  het aantal personen in leeftijdsgroep  $i$  zijn op dit moment. We definiëren  $p_i$  als de kans dat een willekeurig gekozen persoon uit die groep over 5 jaar nog in leven is. Het aantal personen in de leeftijdsgroep 0–4 op dit moment is dus  $N_0$ . Uit bovenstaande aannames volgt dat het verwachte aantal personen in de leeftijdsgroep 5–9 over 5 jaar gelijk is aan  $N_0p_0$ . Het aantal personen nu in de leeftijdsgroep 5–9 is gelijk aan  $N_1$ . Als de bevolking stationair is moeten die twee uitdrukkingen gelijk zijn aan elkaar. Hetzelfde moet gelden voor alle andere leeftijdsgroepen. Een uitzondering vormt de eerste leeftijdsgroep. Over 5 jaar bevat deze naar verwachting  $B$  personen. Hieruit volgt dat  $B = N_0$ . Schematisch weergegeven:

leeftijdsgroep	aantal nu	overlevingskans	verwachte aantal over 5 jaar
0 – 4	$N_0$	$p_0$	$B$
5 – 9	$N_1$	$p_1$	$N_0p_0$
10 – 14	$N_2$	$p_2$	$N_1p_1$
$(5i) - (5i + 4)$	$N_i$	$p_i$	$N_{i-1}p_{i-1}$
90 – 94	$N_{18}$	$p_{18}$	$N_{17}p_{17}$
$\geq 95$	$N_{19}$	$p_{19}$	$N_{18}p_{18}$

Tabel 3.2 Schema bevolkingsopbouw

In bovenstaand schema worden alle mensen ouder dan 95 bij elkaar genomen. Omdat het aantal mensen boven de 100 jaar verwaarloosbaar klein is, doen we net alsof de groep van 95 en ouder 5 jaar later in zijn geheel verdwenen is. De kans  $p_{19}$  dat personen in de leeftijdsgroep  $\geq 95$  nu, over 5 jaar nog in leven zijn stellen we dus gelijk aan 0. Als  $B$  en alle overlevingskansen gegeven zijn kan de stationaire bevolking dus berekend worden.

**Opdracht 3.1** Neem  $B = 20.000$ , en stel dat

$$p = (p_0, p_1, \dots, p_{18}) = (0.999, 0.999, 0.999, 0.998, 0.997, 0.994, 0.991, 0.988, 0.980, 0.968, 0.948, 0.912, 0.880, 0.815, 0.727, 0.593, 0.356, 0.171, 0.007) .$$

– Bepaal de stationaire bevolking en maak een plaatje of tabel ervan.

## 4. Geboortegolven

De bevolking van Nederland, ingedeeld in leeftijdsgroepen van 15 jaar, zag er in 1960 ongeveer als volgt uit.

leeftijdsgroep	aantal in 1960 (x 1000)	15 – jaars overlevingskans	gem. aantal kinderen/vrouw over 15 jaar
0 – 14	3410	0.996	0.649
15 – 29	2526	0.994	1.292
30 – 44	2151	0.973	0.193
45 – 59	1839	0.899	--
60 – 74	1137	0.654	--
75 – 89	369	0.153	--
≥ 90	11	--	--

Tabel 4.1. De bevolking van Nederland in 1960

Met behulp van de technieken van Paragraaf 3 is het niet moeilijk een voorspelling te doen van de samenstelling van de bevolking van Nederland in 1975, tenminste wat betreft de leeftijdsgroepen ouder dan 15 jaar. Rest ons het aantal mensen in de leeftijdsgroep 0–14 jaar in 1975 te voorspellen. Iemand die in 1975 in die leeftijdsgroep valt moet tussen 1960 en 1975 geboren zijn en vervolgens tot 1975 hebben overleefd. Hoeveel mensen er tussen 1960 en 1975 geboren zijn hangt ten duidelijkste af van het aantal potentiële moeders in 1960 en hun gedrag ten aanzien van kinderen krijgen. In Tabel 4.1 is gegeven het gemiddelde aantal *overlevende* kinderen dat een vrouw in 1960 uit een bepaalde leeftijdsgroep tussen 1960 en 1975 zal krijgen. Dit wordt wel de *vruchtbaarheidsfactor* genoemd van de leeftijdsgroep. Het aantal tot 1975 overlevende kinderen dat tussen 1960 en 1975 naar verwachting geboren zal worden uit vrouwen die in 1960 tussen de 0 en 15 jaar oud zijn is dus gelijk aan  $3410/2$  duizenden (het aantal vrouwen is ongeveer gelijk aan de helft van het totaal aantal mensen) vermenigvuldigd met 0.649. Door dezelfde berekening uit te voeren voor de vrouwen uit leeftijdsgroepen 15–29 en 30–44 jaar en alles op te tellen krijgen we het verwachte aantal kinderen in de leeftijdsgroep 0–14 jaar in 1975.

### Opdracht 4.1

– Maak een voorspelling van de bevolkingssamenstelling voor 1975 en extrapoleer deze (aangenomen dat het geboorte- en sterftegedrag hetzelfde blijven) tot aan het jaar 2050.

Sinds 1975 is het geboorte-gedrag echter ingrijpend veranderd. Het is algemeen bekend dat onder gunstiger economische omstandigheden en vooruitzichten minder kinderen geboren worden. Vanaf 1975 is het gemiddeld aantal overlevende kinderen per vrouw gedaald tot 0.586 voor vrouwen in de leeftijdsgroep 0–14 jaar (in 1975), 1.192 (15–29 jaar) en 0.196 (30–44 jaar). Alleen het aantal kinderen per vrouw in de leeftijdsgroep 30–44 jaar steeg, als gevolg van de tendens dat vrouwen steeds later kinderen te krijgen.

### Opdracht 4.2

- Wat zijn de gevolgen van deze verandering in geboortegedrag?
- Maak een nieuwe voorspelling van de bevolkingssamenstelling tot aan 2050.
- Bespreek de verschillen met de voorspelling van Opdracht 4.1.

De zogenaamde baby-boom ontstond na de oorlog, voornamelijk in de jaren 1945-1960, toen er relatief veel kinderen geboren werden. In de jaren van de recessie en de Tweede Wereldoorlog (1930–1945) werden er juist heel weinig kinderen geboren. In de telling van 1960 waren de mensen van de baby-boom vertegenwoordigd in de leeftijdsgroep 0–14 jaar. De leeftijdsgroep 15–29 jaar is in deze telling juist ondervertegenwoordigd. Tegenwoordig wordt er veel gesproken over de vergrijzing van de Nederlandse samenleving, omdat de baby-boomers even na de eeuwwisseling de pensioengerechtigde leeftijd zal bereiken. Een term die door demografen in dit verband gehanteerd wordt is de *demografische druk*, dat wil zeggen de verhouding van het aantal personen van 0–19 jaar en van 65 jaar en ouder tot het aantal personen in de zogenaamde productieve leeftijdsgroep van 20–64 jaar.

### Opdracht 4.3

- Waarom is het begrip demografische druk van belang?
- Maak, onder het scenario van Opdracht 4.2, een overzicht van de ontwikkeling van de demografische druk in Nederland.
- Geef een verklaring voor eventuele schommelingen in de gevonden waarden.

Onder het laatste scenario van lagere geboorte-aantallen hebben we in Opdracht 4.2 een voorspelling gedaan tot aan het jaar 2050 van de bevolking van Nederland. We hebben het al gehad over de invloed van de baby-boom, die een overschot tot gevolg had van mensen in de leeftijdsgroep 0–14 in 1960, 15–29 in 1975 enz. Dit overschot gaat als een soort golf door de tijd heen. Een tweede golf, die een gedempte invloed heeft maar toch zeker aanwezig is, is de golf van kinderen van de baby-boomers, mensen in de leeftijdsgroep 0–14 jaar in 1990, 15–29 in 2005 enz. Een nog weer meer gedempte golf is die veroorzaakt door de kleinkinderen van de baby-boomers. De invloed van die golven kan men duidelijker naar voren brengen met behulp van de volgende techniek.

Het is bekend (in de volgende paragraaf zullen we hiervan een wiskundig bewijs proberen te geven) dat onder gelijk blijvende geboorte- en sterftegetallen, de invloed van die geboortegolven wegebt en dat de relatieve bevolkingssamenstelling zich zal stabiliseren. Dat wil zeggen, dat als je maar lang genoeg zou door gaan met het maken van voorspellingen, de verhouding van het aantal mensen in een bepaalde leeftijdsgroep tot het totale aantal mensen van de bevolking op dat moment zal convergeren naar een constante. Dit wordt wel de stabiele of *stationaire (bevolkings)verdeling* genoemd. Als je deze constante fracties, eventueel uitgedrukt in percentages, weet of dicht genoeg kan benaderen, kan je terugrekenen wat in bijvoorbeeld 1960 (onder die stationaire verdeling) het aantal mensen had moeten zijn in die bepaalde leeftijdsgroep, gegeven het totaal aantal mensen in 1960. Doorrekenend kan men hetzelfde doen voor elke leeftijdsgroep in 1975 en ook in 1990, 2005 enz. Afwijkingen van deze berekende aantallen onder de stationaire verdeling met

de waargenomen aantallen (in 1960) en voorspelde aantallen (in 1975 en later) geven een duidelijk inzicht in welke leeftijdsgroepen in bepaalde jaren een overschot hebben.

#### Opdracht 4.4

– Maak zo'n tabel met afwijkingen van de stabiele waarden voor de jaren 1960–2050. (Voor het bepalen van de stabiele percentages moet misschien nog langer doorgegaan worden met voorspellen dan 2050.)

## 5. Stationaire bevolkingsverdeling

We hebben in de vorige paragraaf al een voorbeeld gezien van een bevolking die door de tijd heen dezelfde overlijdingskansen en vruchtbaarheidsfactoren had. Het bleek dat de leeftijdsproporties, de proporties van die populatie, ingenomen door de verschillende leeftijdsgroepen, naar een constante vector convergeerde, naarmate de tijd voortging. Deze verdeling van relatieve aantallen in de verschillende leeftijdsgroepen is de stationaire leeftijds- of bevolkingsverdeling behorende bij deze overlevingskansen en vruchtbaarheidsfactoren. In 1931 merkte de wiskundige demograaf Lotka dit verschijnsel voor het eerst op. We zullen hier trachten voor deze bewering een wiskundig bewijs te geven.

Beschouw daartoe een populatie, bestaande uit  $l + 1$  leeftijdsgroepen, genummerd  $0, \dots, l$ . We bekijken de populatie op tijdstippen  $t = 0, 1, 2, \dots$ , zodanig dat elke persoon uit leeftijdsgroep  $i$  op tijdstip  $t$  bij overleving terecht komt in leeftijdsgroep  $i + 1$  op tijdstip  $t + 1$ , voor  $i = 0, 1, \dots, l - 1$ . De leeftijdsgroepen zijn dus alle (op de laatste na) één tijdseenheid breed. Na leeftijdsgroep  $l$  komt ieder lid van de bevolking te overlijden. De kans op overleving op tijdstip  $t$  van iemand van leeftijdsgroep  $i$  is gelijk aan  $p_i$  voor  $i = 0, 1, \dots, l - 1$ , *onafhankelijk van het gekozen tijdstip  $t$* . We nemen aan dat geen der  $p_i$ 's gelijk is aan 0. Verder nemen we weer aan dat de populatie op elk tijdstip  $t$  gesloten is voor migratie.

Verder is er een aantal leeftijdsgroepen waarvoor de personen vruchtbaar zijn. De index-verzameling van die leeftijdsgroepen duiden we aan met  $\mathcal{F}$ . Voor een vruchtbare leeftijdsgroep  $j \in \mathcal{F}$  noemen we  $f_j$  de vruchtbaarheidsfactor, het gemiddelde aantal overlevende personen dat een persoon uit leeftijdsgroep  $j$  tussen twee opeenvolgende tijdstippen voortbrengt. Deze vruchtbaarheidsfactoren veranderen ook niet in de tijd, net zo min als de overlevingskansen. Merk op dat de definitie van vruchtbaarheidsfactor hier iets verschilt van die in de vorige paragraaf. We nemen aan dat er minstens twee vruchtbare leeftijdsgroepen zijn, d.w.z. twee indices  $i$  en  $j$  waarvoor  $f_i$  en  $f_j$  beiden groter dan nul zijn.

Tot slot nog een definitie: met  $N_i(t)$  bedoelen we het aantal personen in leeftijdsgroep  $i$  op tijdstip  $t$  van een bevolking  $N$ . De relatie tussen de bevolking op tijdstip  $t$  en op

tijdstip  $t + 1$  is gegeven door

$$(5.1) \quad \begin{aligned} N_{i+1}(t+1) &= N_i(t) p_i, & i = 0, \dots, l-1, \\ N_0(t+1) &= \sum_{j \in \mathcal{F}} N_j(t) f_j. \end{aligned}$$

Met  $\nu(t) = (\nu_0(t), \nu_1(t), \dots, \nu_l(t))$  zullen we de verhouding bedoelen van het aantal personen in elke leeftijdsgroep tot het totaal aantal personen in de bevolking, m.a.w.

$$(5.2) \quad \nu_i(t) = \frac{N_i(t)}{\sum_{j=0}^l N_j(t)}, \quad i = 0, 1, \dots, l.$$

We willen laten zien dat voor  $i = 0, 1, \dots, l$ ,  $\nu_i(t) \rightarrow \nu_i$  als  $t \rightarrow \infty$ . Voor alle duidelijkheid: een stabiele leeftijdsverdeling hoeft nog niet te betekenen dat  $N_i(t)$  niet met  $t$  verandert; de bevolking als geheel kan best groeien terwijl de leeftijdsopbouw gelijk blijft. De stabiele leeftijdsverdeling van deze bevolking zullen we aanduiden met  $\nu = (\nu_0, \nu_1, \dots, \nu_l)$ . Over het algemeen zullen we als we over een vector met grootheden voor verschillende leeftijdsgroepen praten, eenvoudig de leeftijdsindex weglaten. Zo staat  $N(t)$  voor de vector  $N(t) = (N_0(t), N_1(t), \dots, N_l(t))$  en is  $p = (p_0, p_1, \dots, p_l)$ .

Hoe gaan we nu bewijzen dat de proporties van het aantal personen in de leeftijdsgroepen naar een constante vector  $(\nu_0, \dots, \nu_l)$  convergeert? We zullen dit bereiken door ons twee populaties  $M$  en  $N$  voor te stellen, met op tijdstip  $t = 0$  verschillende samenstellingen, maar wier ontwikkeling wel door dezelfde vruchtbaarheids- en overlevingsparameters beschreven wordt. We zullen laten zien dat voor deze populaties het verschil in samenstelling kleiner wordt, naarmate de tijd vordert. Voor één van beide populaties zullen we dan op het laatst de stabiele populatie invullen, waarmee ons bewijs dan voltooid is.

Om eerst een beetje gevoel voor de zaak te krijgen, beschouwen we een proefpopulatie  $N$ , bestaande uit 3 leeftijdsgroepen, dus  $l = 2$ . Op tijdstip  $t = 0$  geldt  $N_0(0) = 3$ ,  $N_1(0) = 2$  en  $N_2(0) = 1$ . Voor een tweede populatie  $M$  stellen we  $M_0(0) = 6$ ,  $M_1(0) = 8$  en  $M_2(0) = 2$ . Voor beide populaties is  $p_0 = p_1 = \frac{1}{2}$ ,  $f_1 = 1$ ,  $f_2 = 2$ .

### Opdracht 5.1

- Bepaal voor tijdstippen  $t = 1, \dots, 6$  de verwachte ontwikkeling van beide populaties.
- Bepaal tevens voor tijdstippen  $t = 0, \dots, 6$  de vector  $(\frac{N_0(t)}{M_0(t)}, \frac{N_1(t)}{M_1(t)}, \frac{N_2(t)}{M_2(t)})$  en de vectoren  $\nu(t)$  en  $\mu(t)$ .

Als het goed is, blijkt dat de fluctuaties in de gevonden quotienten  $\frac{N_i(t)}{M_i(t)}$  kleiner worden. Om dit exact te maken beschouwen we nu twee algemene populaties  $N$  en  $M$  en definiëren we

$$(5.3) \quad C(t) = \max_{i=0, \dots, l} \frac{N_i(t)}{M_i(t)}, \quad c(t) = \min_{i=0, \dots, l} \frac{N_i(t)}{M_i(t)}, \quad d(t) = C(t) - c(t).$$

### Opdracht 5.2

- Stel, analoog aan (5.1), een stelsel vergelijkingen op voor  $M(t+1)$  in termen van  $M(t)$ ,

$p$  en  $f$ .

– Gebruik dit stelsel samen met (5.1) om te laten zien dat

$$(5.4) \quad c(t) \leq c(t+1) \leq C(t+1) \leq C(t) ,$$

en dus dat  $d(t+1) \leq d(t)$ , voor  $t = 0, 1, 2, \dots$

We concluderen uit (5.4) dat de verhoudingen tussen aantallen personen in  $M$  en  $N$  in identieke leeftijdsgroepen op elk volgend tijdstip in ieder geval niet extremer wordt. Maar we willen meer: we willen dat die verhoudingen krimpen, dus dat  $c(t)$  en  $C(t)$  beiden naar dezelfde constante, zeg  $C$ , convergeren. We willen dus graag dat naast (5.4) ook

$$(5.5) \quad d(t) \rightarrow 0 , \quad \text{als } t \rightarrow \infty .$$

Om dit te bereiken moeten we de afschattingen van de vorige opdracht wat verfijnen.

Daartoe kijken we iets nauwkeuriger naar de verhouding van het aantal nieuwgeborenen op tijdstip  $t+1$ . We zien dat geldt

$$(5.6) \quad \frac{N_0(t+1)}{M_0(t+1)} = \frac{\sum_{j \in \mathcal{F}} N_j(t) f_j}{\sum_{j \in \mathcal{F}} M_j(t) f_j} = \frac{\sum_{j \in \mathcal{F}} \frac{N_j(t)}{M_j(t)} M_j(t) f_j}{\sum_{j \in \mathcal{F}} M_j(t) f_j} .$$

We schrijven nu  $r$  voor de cardinaliteit van  $\mathcal{F}$ , zetten de  $r$  termen  $\frac{N_j(t)}{M_j(t)}$  in oplopende volgorde en sommeren de termen van  $\sum_{j \in \mathcal{F}} \frac{N_j(t)}{M_j(t)} M_j(t) f_j$  in die volgorde. Zo zien we dat de teller van de laatste uitdrukking van (5.6) kan worden begrensd door

$$(5.7) \quad \begin{aligned} \sum_{j \in \mathcal{F}} \frac{N_j(t)}{M_j(t)} M_j(t) f_j &= \sum_{i=1}^r \frac{N_i(t)}{M_i(t)} M_i(t) f_i \\ &= \frac{N_1(t)}{M_1(t)} M_1(t) f_1 + \frac{N_r(t)}{M_r(t)} M_r(t) f_r + \sum_{i=2}^{r-1} \frac{N_i(t)}{M_i(t)} M_i(t) f_i \\ &\leq C(t) \sum_{j=1}^r M_j(t) f_j - [C(t) - c(t)] M_1(t) f_1 . \end{aligned}$$

Combineren van (5.6) en (5.7) geeft

$$(5.8) \quad \frac{N_0(t+1)}{M_0(t+1)} \leq C(t) - [C(t) - c(t)] \frac{M_1(t) f_1}{\sum_{j \in \mathcal{F}} M_j(t) f_j} .$$

Op dit punt zullen we voor de bevolking  $M$  een heel speciale kiezen, namelijk een die op  $t=0$  en dus ook voor alle  $t$  daarna de stabiele verdeling heeft. Voor deze bevolking geldt

$$(5.9) \quad \frac{M_i(t)}{\sum_{j=0}^l M_j(t)} = \frac{M_i}{\sum_{j=0}^l M_j} = \mu_i , \quad t = 0, 1, \dots .$$

### Opdracht 5.3

– Ga na dat uit (5.9) en een argumentatie, analoog aan (5.6)–(5.8) volgt dat

$$c(t) + \rho d(t) \leq \frac{N_0(t+1)}{M_0(t+1)} \leq C(t) - \rho d(t),$$

met

$$\rho = \min_{i \in \mathcal{F}} \frac{\mu_i f_i}{\sum_{j \in \mathcal{F}} \mu_j f_j}.$$

– Laat zien dat onder onze voorwaarden  $\rho > 0$ , onafhankelijk van  $t$ .

– Voltooi het bewijs door te laten zien dat  $C(t)$  en  $c(t)$  beiden naar dezelfde constante convergeren.

– Ga het hele bewijs nog eens na en controleer of er ergens door 0 gedeeld wordt.

Hint: stel dat  $d(t) > \varepsilon$  voor alle  $t$ , voor een bepaalde  $\varepsilon > 0$ .

## 6. Schatten van parameters uit meer dan één census

Een census is een momentopname van een bevolking. Stel dat de bevolking is onderverdeeld in leeftijdsgroepen van gelijke lengte, zeg  $L$  jaren. De leeftijdsgroep met nummer 0 loopt van 0 tot en met  $L - 1$ , nummer 1 van  $L$  tot en met  $2L - 1$  en de laatste, nummer  $l$  van  $l \cdot L$  en hoger. De census van 1995 bijvoorbeeld bevat (met  $L = 5$ ) voor elk van die leeftijdsgroepen het aantal personen in die leeftijdsgroep. Stel je nou voor dat we meerdere censi hebben, genomen op tijdstippen  $t = 0, 1, \dots, m$ , waarbij een tijdseenheid precies overeen komt met  $L$  jaar. De breedte van de leeftijdsgroepen komt dus precies overeen met de periode tussen twee opeenvolgende tijdstippen. Het aantal waargenomen personen op tijdstip  $t$  in leeftijdsgroep  $i$  noemen we  $n_i(t)$ .

Om een stochastisch model te formuleren beschouwen we de vectoren  $n(0), \dots, n(m)$ , waar  $n(t) = (n_1(t), \dots, n_l(t))^T$  als realisaties van stochastische vectoren  $N(0), \dots, N(m)$ , waar  $N(t) = (N_1(t), \dots, N_l(t))^T$ . De stochastische vector  $N(t)$  beschrijft dus de leeftijdsopbouw van de bevolking op tijdstip  $t$ . Met deze notatie kijken we eerst wat er van tijdstip 0 naar tijdstip 1 gebeurt. Bekijk de leeftijdsgroep  $i - 1$  op tijdstip 0, bestaande uit  $n_{i-1}(0)$  personen. Ieder van deze personen heeft, onafhankelijk van alle andere personen in de gehele populatie, dezelfde kans  $p_{i-1}$  om op het volgende tijdstip nog in leven te zijn. Verder nemen we aan dat de populatie op elk tijdstip gesloten is voor migratie. Voorwaardelijk, gegeven  $N(0) = n(0)$  zijn  $N_1(1), \dots, N_l(1)$  onafhankelijk.

### Opdracht 6.1

– Wat is de voorwaardelijke verdeling van  $N_i(1)$ , gegeven dat  $N(0) = n(0)$ ?

Het voorgaande is niet specifiek voor de periode tussen de tijdstippen 0 en 1. Als de bevolking op tijdstippen 0 en 1 nu bekend is, bekijk dan de leeftijdsgroep  $i - 1$  op tijdstip  $t = 1$ , bestaande uit  $n_{i-1}(1)$  personen. Net zoals voorheen heeft ieder van deze personen, onafhankelijk van alle andere personen, dezelfde kans  $p_{i-1}$  om op het volgende tijdstip nog in leven te zijn. We nemen dus weer aan dat deze overlevingskansen niet in de tijd veranderen.

### Opdracht 6.2

- Wat is de voorwaardelijke verdeling van  $(N_1(2), \dots, N_l(2))$ , gegeven dat  $N(0) = n(0)$  en  $N(1) = n(1)$ ?
- Hangt deze af van  $n(0)$ ?

En dit gaat natuurlijk precies zo door van tijdstip 2 naar tijdstip 3 enzovoorts tot en met tijdstip  $m$ . We vinden zo steeds de voorwaardelijke verdeling van  $N(t)$ , gegeven  $N(0) = n(0), N(1) = n(1), \dots, N(t-1) = n(t-1)$  en zien dat deze niet afhangt van  $n(0), \dots, n(t-2)$ , maar alleen van  $n(t-1)$ .

Samenvattend: volgens ons kansmodel is voor  $t = 1, \dots, m$  de voorwaardelijke verdeling van  $N(t)$ , gegeven  $N(t-1) = n(t-1), N(t-2) = n(t-2), \dots, N(0) = n(0)$  als volgt bepaald: de elementen  $N_1(t), \dots, N_l(t)$  zijn onderling onafhankelijk en  $N_i(t)$  heeft de verdeling die aan de hand van de vorige twee opdrachten bepaald is. We merken op dat voor dit schattingsprobleem  $N_0(t)$  voor  $t = 0, \dots, m$  niet stochastisch verondersteld wordt.

We weten dus de voorwaardelijke verdeling van  $N(t)$ , gegeven  $N$  op tijdstippen  $t - 1$  en eerder. Maar we willen graag

$$(6.1) \quad P(N(1) = n(1), N(2) = n(2), \dots, N(m) = n(m) \mid N(0) = n(0))$$

weten. Dit is namelijk de aannemelijkheidsfunctie, of beter gezegd de voorwaardelijke aannemelijkheidsfunctie, gegeven dat  $N(0) = n(0)$ . Door deze uitdrukking te maximaliseren naar  $p_0, \dots, p_{l-1}$  vinden we de meest aannemelijke schatters  $\hat{p}_i$  van  $p_i$ . Om (6.1) te vinden, gebruiken we het volgende principe, namelijk dat voor gebeurtenissen  $A_0, A_1, \dots, A_m$  met  $P(A_t) \neq 0, t = 0, 1, \dots, m$ ,

$$(6.2) \quad P(A_m \cap A_{m-1} \cap \dots \cap A_1 \mid A_0) = P(A_m \mid A_{m-1}, \dots, A_0) \cdot P(A_{m-1} \mid A_{m-2}, \dots, A_0) \cdots P(A_1 \mid A_0) .$$

### Opdracht 6.3

- Bewijs dit.

### Opdracht 6.4

- Pas (6.2) toe met  $A_t = \{N(t) = n(t)\}$ , voor  $t = 0, 1, \dots, m$ .

– Maximaliseer de logaritme van (6.1) naar  $p$  en laat zien dat de meest aannemelijke schatters  $\hat{p}_i$  van  $p_i$  voldoen aan

$$(6.3) \quad \sum_{t=1}^m \left\{ \frac{n_{i+1}(t)}{\hat{p}_i} - \frac{n_i(t-1) - n_{i+1}(t)}{1 - \hat{p}_i} \right\} = 0, \quad i = 0, 1, \dots, l-1.$$

– Leid hieruit een expliciete uitdrukking af voor  $\hat{p}_i$ .

**Opdracht 6.5** In de bibliotheek van Economie (hoofdgebouw, 4de verdieping), staan jaargangen van het CBS-jaarboek (ET.00271). Neem de meest recente jaargang en 5 jaargangen daarvoor, steeds met tussenpozen van 5 jaar. Zoek in elk boek de tabel “bevolking naar burgerlijke staat geslacht en leeftijd” op (alleen de vrouwen).

– Schat op grond van deze 6 censi met de methoden van deze paragraaf de  $p_i$ 's. In de censi van voor 1969 worden sommige bevolkingsgroepen (bijvoorbeeld die van 0–14) niet verder onderverdeeld in groepen van 5 jaar. Deze census is dus niet in zijn geheel bruikbaar voor het schatten van alle  $p_i$ 's. Laat voor het schatten van de betreffende  $p_i$ 's deze census buiten beschouwing.

– Zoek in het meest recente CBS-jaarboek ook de bijbehorende vruchtbaarheidsfactoren op en maak vervolgens, onder de aanname van gelijk blijvende overlevingskansen en vruchtbaarheidsfactoren, een projectie van de bevolking tot aan het jaar 2050. We nemen hierbij aan dat de vruchtbare leeftijdsgroepen 15 – 19, 20 – 24, ..., 45 – 49 zijn.

Onder de door ons gestelde voorwaarden in deze Wiskunde Werkt opdracht is het in principe onmogelijk om waarden van  $\hat{p}_i$  te krijgen die buiten het interval  $[0, 1]$  liggen. Bij het uitvoeren van Opdracht 6.5 is het echter in het geheel niet ondenkbaar dat een waarde van  $\hat{p}_i$  buiten  $[0, 1]$  voorkomt.

### Opdracht 6.6

- Laat zien dat onder de gestelde voorwaarden  $\hat{p}_i$  altijd in  $[0, 1]$  zit.
- Leg uit waarom een waarde van  $\hat{p}_i$  buiten  $[0, 1]$  toch kan voorkomen.
- Tot welke problemen zou dit kunnen leiden?
- Welke oplossing suggereer je?

Tot slot gaan we wiskundig, op grond van de geschatte  $p_i$ 's de stabiele leeftijdsverdeling bepalen. Laten we een bevolking op tijdstip  $t$  bekijken en aannemen dat stabiliteit op dat moment al aanwezig is. Dan geldt dus voor  $i = 0, 1, \dots, l-1$ ,

$$(6.4) \quad \frac{n_i(t)}{n_{i+1}(t)} = \frac{n_i(t+1)}{n_{i+1}(t+1)}.$$

### Opdracht 6.7

– Bepaal de stabiele leeftijdsverdeling  $\nu = (\nu_0, \nu_1, \dots, \nu_l)$  voor de  $\hat{p}_i$ 's en de  $f_k$ 's gevonden bij Opdracht 6.5 of Opdracht 6.6.

Hint: Een mogelijke methode is de volgende. Leid een recurrente betrekking af voor de quotiënten  $n_{i+1}/n_i$ . Definieer daartoe  $\frac{n_{j+1}(t)}{n_j(t)} = w_j$  en gebruik (6.4) om de quotiënten uit te drukken in  $w_0$ . Laat vervolgens zien dat  $w_0$  aan de volgende de polynoomvergelijking voldoet:

$$(6.5) \quad \sum_{j \in \mathcal{F}} \left( \frac{w_0}{p_0} \right)^{j+1} q_j f_j = 1 ,$$

met  $q_j = p_0 p_2 \cdots p_{j-1}$ . Voor het oplossen van (6.5) kan bijvoorbeeld Maple worden gebruikt.

NB. De hint hoeft niet persé gevolgd te worden.