



Over fractionele bewegingen, Noach en olifanten

Fractionele Brownse beweging (fBb) is een stochastisch proces dat in de laatste tientallen jaren gebruikt wordt in uiteenlopende gebieden. Het modelleert bijvoorbeeld beurskoersen, internetverkeer, regenval en bloedstromingen in delen van het menselijk lichaam.

Alle reden om een scriptie te schrijven over hoe dit proces gesimuleerd kan worden.

TON DIEKER

De hydroloog Hurst (1900-1978) werkte aan een project om dammen te bouwen aan de Nijl. Toen hij de waterstanden bestudeerde die de Egyptenaren tussen 622 en 1469 (!) bijgehouden hadden, viel hem op dat er bepaalde patronen in zaten. Een jaar met een bovengemiddeld waterpeil werd vaak gevolgd door weer zo'n bovengemiddeld jaar, wat doorbroken werd door een plotselinge daling. Na deze daling, die het Noach-effect wordt genoemd, volgden enkele ondergemiddelde jaren. De zo ontstane cyclische structuur kan niet verklaard worden met een model gebaseerd op een stochastische wandeling, omdat lange-termijn correlaties een grote rol spelen. Ook de trendwis-

seling heeft een Bijbelse naam gekregen, namelijk het Jozef-effect: zeven overvloedige jaren worden gevolgd door zeven droge jaren.

Dit is het bekendste voorbeeld van gegevens met lange-termijn correlaties die zo heftig zijn dat traditionele modellen geen uitkomst lijken te bieden. Dit fenomeen speelt ook in andere gebieden, zoals in de economie, natuurkunde en biologie. Mijn onderzoek was vooral gemotiveerd door modellen voor communicatienetwerken. In de jaren negentig van de vorige eeuw werden metingen verricht die soortgelijk gedrag ook voor netwerkverkeer suggereerden. De lang-termijn correlaties kunnen bijvoorbeeld veroorzaakt worden

door mensen die films downloaden ('olifanten'), terwijl anderen alleen kleine bestanden versturen ('muizen'). In een model gebaseerd op fractionele Brownse beweging kunnen deze lange-termijn effecten worden opgenomen, dus is een introductie van dit proces op zijn plaats.

Fractionele Brownse beweging

Fractionele Brownse beweging (fBb) is een gecentreerd stochastisch proces in continue tijd met stationaire Gaussische aangroeiingen. Voor zo'n proces worden alle eindig-dimensionale verdelingen vastgelegd door de variantiefunctie. Een fBb B wordt gedefinieerd door een variantiefunctie van de vorm $\text{Var}(B(t)) = t^{2H}$ waarbij H een parameter is tussen 0 en 1, de zogenoemde Hurst parameter. Als deze parameter $\frac{1}{2}$ is, reduceert de fractionele Brownse beweging tot de klassieke Brownse beweging.

Een mogelijke reden voor de populariteit van fBb is een eigenschap die bekend staat onder de naam *self-similarity*. In gewoon Nederlands: hoe veel je ook in- of uitzoomt, het proces ziet er altijd hetzelfde uit (in verdeling). Wiskundig gezien hangt de schaling van de Hurst parameter af. De *self-similarity* houdt ook verband met de benaming 'fractioneel', omdat objecten met deze eigenschap *fractals* worden genoemd. Denk bijvoorbeeld aan sneeuwvlokken, wolken, vertakkingen in onze longen, en het welbekende Drosteplaatje. Intuïtief is het aannemelijk dat deze eigenschap te maken heeft met bepaalde lange-termijn correlaties, die worden aangeduid met de term *long-range dependence*. Voor simulatie heeft zo'n structuur verstrekkende gevolgen.

Simulatie

In veel gevallen is het onmogelijk om expliciete berekeningen te maken in modellen op basis van fBb. Computersimulatie vormt dan een belangrijk

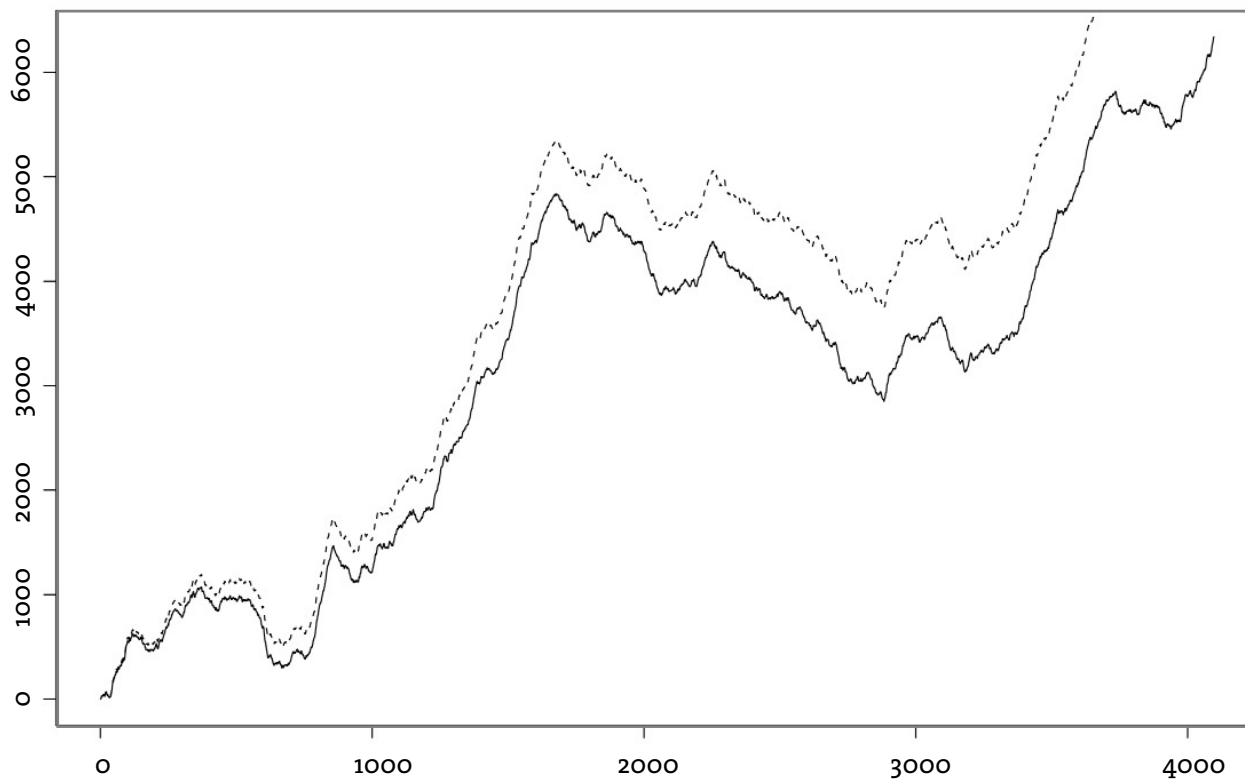
instrument om inzichten te verschaffen. Hoewel er verschillende simulatiemethoden bestaan voor fBb, kunnen simulatiemethoden die dit proces benaderen interessant zijn vanwege hun snelheid. Voor een gegeven benadering is het dan van belang te weten in welke opzichten de verkregen realisatie op fBb lijkt. Er bleek een drietal methoden veelbelovend: een methode gebaseerd op handig conditioneren (hierna: Random Midpoint Displacement (RMD) methode), een spectrale methode, en een wavelet methode.

Theoretische vergelijking

Hoewel theoretisch onderzoek naar de benaderingen slechts beperkt mogelijk is, levert dit toch interessante inzichten op.

Voor de RMD methode beschrijf ik in mijn scriptie een algoritme om de covariantiematrix van het benaderende proces te berekenen. Bij vergelijking met de gewenste covariantiematrix (die van fBb) valt op dat de benadering vooral veel moeite heeft om de lange-termijn correlaties correct te simuleren. Toch is dit te overzien: de benadering wordt een stuk beter als bijvoorbeeld twee keer zo veel punten gesimuleerd worden en alleen het middelste deel daadwerkelijk gebruikt wordt.

De spectrale methode, die in mijn scriptie het grondigst bestudeerd wordt, heeft een aantal mooie theoretische eigenschappen. Zo presteert de methode beter als het aantal te simuleren punten groter wordt. Bovendien is de methode nauw verwant met een tweetal andere methodes uit de literatuur, waaronder een exacte methode. Dit maakt het mogelijk om de fout in de benadering onder de loep te nemen. Ter illustratie: in Figuur 1 is zowel een 'exacte' als een 'spectrale' realisatie gegeven, waarbij dezelfde willekeurige getallen zijn gebruikt. Het verschil tussen de twee lijnen wordt steeds groter, hetgeen typisch is voor de spectrale methode. Op dit voorbeeld komen we later terug.



Figuur 1. Een vergelijking tussen een exacte realisatie (zwarte lijn) en een spectrale realisatie (onderbroken lijn).

De wavelet-methode is theoretisch lastig te bestuderen, maar het is bijvoorbeeld wel mogelijk om de covariantiematrix te simuleren en te vergelijken met de covariantiestructuur van fBb. De resultaten zijn teleurstellend, maar er kan geen simpele verklaring gegeven worden waarom deze methode zo slecht presteert.

Statistische vergelijking

Nu we een eerste indruk hebben van de drie methoden om fBb te simuleren, kan de analyse gecompleteerd worden met behulp van statistische toetsen. De moeilijkheid is dat niet duidelijk is wat we eigenlijk zouden willen toetsen. De statistische literatuur over gegevens met long-range dependence richt zich vooral op het schatten van de Hurst parameter. De schattingen kunnen dan vergeleken worden met de gewenste Hurst parameter.

Omdat de Hurst parameter een fBb uniek karakteriseert, speelt deze een rol in elke eigenschap van fBb waarop een schattingsprocedure gebaseerd kan worden. Er zijn daarom erg veel manieren om deze parameter te schatten. In mijn scriptie heb ik negen verschillende schattingsmethoden onder de loep genomen, maar dat is slechts een selectie.

De RMD methode komt vrij goed door deze analyse heen, maar de wavelet methode blijkt dramatisch slecht. De spectrale methode is redelijk te noemen, maar net als bij de RMD methode blijken de lange-termijn correlaties niet goed uit de verf te komen. Ook in Figuur 1 is dit te zien: het verschil tussen de twee lijnen is lokaal vrijwel constant, maar globaal missen we een grote 'golf' in de spectrale simulatie. Deze golf zou precies voor de lange-termijn correlaties moeten zorgen.

Snelheidsvergelijking

Een van de meest belangrijke aspecten van een (benaderende) methode voor het genereren van fBb is zijn snelheid, met name als realistische systemen gesimuleerd dienen te worden. Toch hangt het erg af van de situatie welke methode het beste gebruikt kan worden. Hoewel de lange-termijn correlaties niet geheel naar wens zijn, kan het dan toch interessant zijn om een snelle methode te kiezen in plaats van een exacte methode.

Van de drie methoden is de RMD methode het snelst, maar er moet nog een aantal parameters gekozen worden die invloed hebben op de snelheid. De spectrale methode is goede tweede, en de wavelet method is uiterst traag.

Conclusies

Het simuleren van fBb is vrij lastig; alle bestudeerde benaderende methoden hebben moeite met de lange-termijn correlaties. De RMD methode presteert het beste van de benaderende methoden, maar toch is het in veel gevallen niet de juiste keuze. De snelste exacte methode, gebaseerd op een eigenwaarde decompositie van een circulant matrix, kan zelfs op snelheid erg goed concurreren met de RMD methode. De RMD methode is lineair in de grootte van de realisatie N , terwijl deze exacte methode orde $N \log N$ is. Voor ongeveer 15.000 punten is de exacte methode ongeveer tien keer zo langzaam, hetgeen in veel gevallen een redelijke prijs is om exacte realisaties te kunnen genereren.

De scriptie en gerelateerde bestanden kunnen gedownload worden van www.cwi.nl/~ton/fbm.

TON DIEKER is als promovendus verbonden aan het CWI en aan de de Universiteit van Twente. Zijn scriptie schreef hij als student aan de Vrije Universiteit. E-mail: ton.dieker@cwi.nl.



Opheffing Sectie Statistische Programmatuur

Op de laatste Statistische Dag heeft de sectie SSP op een goed bezochte bijeenkomst (met Brian Ripley als discussant) aangegeven dat statistische programmatuur inmiddels zo verweven is in het werk en de belangstelling van de andere secties van de VVS-OR, dat een speciale sectie SSP niet langer meer nodig is. Het Algemeen Bestuur van de VVS-OR heeft daarop in haar vergadering van mei 2005 de sectie SSP op verzoek, en ervol, opgeheven.

Leden van de sectie SSP kunnen zich aanmelden bij een van de andere secties van de VVS-OR. Dit kan worden opgegeven bij het administratiekantoor.

Statistiek en OR in het nieuws

Na het succes van de lezingenmiddag over de zaak Lucy de B. heeft het bestuur van de VVS-OR besloten tot het instellen van een werkgroep die statistiek en OR in de actualiteit gaat volgen. De activiteiten van deze werkgroep kunnen, naast het signaleren van actuele gebeurtenissen op het werkterrein van de vereniging, bestaan uit het organiseren van lezingenmiddagen, het op gang brengen van discussie, bijvoorbeeld via het internet, of het voorbereiden van een verenigingsstandpunt. Een belangrijk deel van de discussie zal plaatsvinden via het internet. Het bestuur nodigt leden die aan deze werkgroep willen bijdragen uit zich te melden bij de voorzitter.