Samenvatting

Op vele gebieden, zoals geologie, biologie of medicijnen worden gegevens verzameld bestaande uit een beeld van een proces van geometrische objecten in de ruimte. Lijnsegmenten kunnen breukvlakken in een rotsformatie voorstellen, waargenomen als rechte lijnen op het rotsoppervlak. Punten kunnen de locaties van bomen in een bos weergeven. Cirkelschijfjes kunnen cellen onder een microscoop zijn. Altijd is ons beeld begrensd, ook al strekt het proces zich ver buiten het beeld uit. Er ontstaat een soort censurering van de gegevens, namelijk dat van sommige objecten alleen een gedeelte waarneembaar is zodat de werkelijke omvang onbekend blijft.

Dit proefschrift gaat over ruimtelijke statistiek, of preciezer: de mathematisch statistische analyse van ruimtelijke data. We observeren een of meerdere beelden, zoals hierboven beschreven, en stellen ons voor dat deze zijn ontstaan als trekkingen uit een kansverdeling. Het doel is om uit onze waarnemingen deze kansverdeling—of belangrijke eigenschappen ervan—zo goed mogelijk te achterhalen. Wellicht verheldert het volgende voorbeeld wat we bedoelen.

Figuur 1.1 in dit proefschrift is een geologische kaart van een gebied van zo'n 160 bij 160 meter van het Canadese schild. In het zwarte deelgebied kunnen we het granieten rotsoppervlak waarnemen. Buiten dit gebied kunnen we het graniet niet zien vanwege begroeiing of water. De witte lijnen in het zwarte deelgebied zijn barsten in de rots. In hoofdstuk 2 geven we een methode om de kansverdeling van de lengte van de barsten te schatten op basis van de kaart.

Men zou kunnen zeggen dat ruimtelijke statistiek moeilijker is dan 'gewone' statistiek. Een typische bijkomende moeilijkheid is ruimtelijke afhankelijkheid: wat men op verschillende locaties in een beeld waarneemt is zelden onafhankelijk van elkaar. Onafhankelijkheid is een standaard aanname in de statistiek, die in de ruimtelijke problemen zelden kan worden gehandhaafd. Andere moeilijkheden komen door censurering en andere rand-effecten. 'Randeffecten' is een verzamelnaam voor wiskundige complicaties die ontstaan als een ruimtelijk proces slechts in een beperkt gebied wordt waargenomen. Als gevolg van al deze moeilijkheden is de ruimtelijke statistiek enigzins een achtergebleven gebied. In mijn proefschrift passen we echter moderne 'mainstream' statistische methoden toe in enige ruimtelijke problemen, waarbij we bovengenoemde moeilijkheden omzeilen. De wijze waarop wij dit doen zijn de 'listen' uit de titel.

In hoofdstuk 1 leggen we ons gereedschap klaar. Nieuw in dit hoofdstuk is een algemene verhandeling over consistentie van de zogeheten meest aannemelijke schatter. Dit resulteert onder andere in Stelling 1.2. Deze stelling geeft voorwaarden waaronder de (niet parametrische) meest aannemelijke schatter in convexe modellen consistent is. In hoofdstuk 2 en 3 kunnen we deze stelling toepassen, hoewel we dit slechts in hoofdstuk 2 daadwerkelijk uitwerken. Ook presenteren we in hoofdstuk 1 een modificatie van de stochastische versie van het bekende EM algorithme. Dit algorithme passen we toe in hoofdstuk 4.

In hoofdstuk 2 beschouwen we 'Laslett's lijnsegmenten probleem'. Het doel is de verdeling van de lengte van lijnsegmenten, waargenomen door een stochastische verzameling, te schatten. Dit probleem heeft een een ruime geschiedenis en werd onder meer bestudeerd in twee eerdere Utrechtse proefschriften, die van Mark van der Laan en Bart Wijers. Wij beschouwen het probleem in grotere algemeenheid en laten de aanname vallen dat het waarnemingsgebied convex is. Als gevolg hiervan kunnen we verschillende fragmenten van één lijnsegment waarnemen. In de praktijk is het onmogelijk te onderscheiden welke fragmenten bij elkaar horen, en zodoende zijn ze op een onontwarbare wijze afhankelijk. Onze 'list' is een eenvoudiger, maar vergelijkbaar probleem te beschouwen zonder deze afhankelijkheid. Voor dit probleem kunnen we de meest aannemelijke schatter afleiden. We tonen vervolgens aan dat de schatter ook goede eigenschappen heeft in het oorspronkelijke probleem.

In hoofdstuk 3 schatten we de kansverdeling van de lengte van een 'typische koorde' van een stochastische verzameling. Een koorde is het langste lijnsegment door een gegeven punt van de verzameling in een gegeven richting, dat geheel binnen de verzameling past. De moeilijkheid is dat we de kansverdeling van de stochastische verzamling in termen van de koordelengte verdeling zouden moeten kennen om de meest aannemelijke schatter uit te kunnen rekenen. Dit is helaas zelden het geval. Wederom beschouwen we een eenvoudiger variant van het probleem, leiden de meest aannemelijke schatter af voor *dat* probleem en laten zien dat de schatter ook goed is in de oorspronkelijke situatie.

In hoofdstuk 4 bestuderen we het 'bommen' model. We nemen de vereniging waar van enige toevallig rondgestrooide cirkelschijfjes. Men zou zich kunnen voorstellen dat deze schijfjes bom-kraters zijn, hetgeen de naam van dit model verklaart. Op basis van onze waarnemingen willen we het verwachte aantal schijfjes schatten, waarvan het centrum zich in een verzameling van gegeven oppervlakte bevindt. Voor de kenner: We schatten de intensiteit van het Poisson process dat het waargenomen Boolese model genereerde. Het probleem is dat sommige schijfjes geheel bedekt kunnen worden door andere en zodoende volledig onzichtbaar zijn. Hier bestaat onze list uit een algorithme dat ons in staat stelt een steekproef te nemen uit de voorwaardelijke verdeling van de niet-geobserveerde schijfjes, gegeven wat we wel observeren. Met behulp van dergelijke steekproeven kunnen we de meest aannemelijke schatter schatten, zo gezegd. Het algorithme is in grotere algemeenheid toe te passen om puntprocessen te genereren waarvan de zogeheten Papangelou intensiteit van nul is wegbegrensd.