

‘n Ondersoek na die Eindige Steekproefgedrag
van Inferensiemetodes
in Ekstreemwaarde-teorie

Dewald van Deventer

Werkstuk ingelewer ter gedeeltelike vervulling
van die vereistes vir die
graad Magister
The crest of the University of Stellenbosch, featuring a shield with various symbols, topped with a crown and a banner.
in die Handelswetenskappe
aan die
Universiteit van Stellenbosch

Promotor: Prof. Tertius de Wet

Februarie 2005

Opsomming

Ekstreme waardes is rare, ontpiepte waarnemings wat slegs by uitsondering realiseer. Wanneer sulke gebeurtenisse – byvoorbeeld aardbewings, fratsgolwe en beleggingsmark ineenstortings - egter wel plaasvind, gaan dit dikwels met enorme lewensverlies en/of finansiële skade gepaard. Om hierdie rede is die akkurate modellering van ekstreme waardes van kritieke belang. Ekstreemwaarde-teorie behels die ontwikkeling van statistiese modelle en tegnieke om sulke waarnemings te beskryf en te modelleer.

In hierdie werkstuk word aspekte van ekstreemwaarde-teorie bespreek. Dié teorie bestaan uit twee breë benaderings: `n Klassieke maksimummetode, gebaseer op die eienskappe van die maksimum van `n steekproef en `n meer moderne drempelteorie, gebaseer op die eienskappe van oorskrydings van `n gekose drempelwaarde. Hierdie werkstuk rus die praktisyn toe met die nodige teoretiese en praktiese toerusting ten opsigte van beide benaderings, om sodoende ekstreemwaarde-analises met vertroue uit te voer.

Ekstreemwaarde-teorie – vir beide benaderings - is op asimptotiese argumente gebaseer. Die limietresultaat vir steekproefmaksima geld slegs by benadering vir eindige steekproewe, terwyl die limietverdeling vir drempeloorskrydings slegs by benadering geld as die drempelwaarde eindig is. Ons ondersoek in hierdie werkstuk die gehalte van ekstreemwaarde-gebaseerde inferensies ten opsigte van die onbekende onderliggende verdeling wanneer die steekproefgrootte of drempelwaarde eindig is. Aangesien die beraming van ekstreme stertkwantiele van die onderliggende verdeling, sowel as die berekening van vertrouensintervalle vir hierdie kwantiele tipies die vernaamste doelwit van `n ekstreemanalise is, evalueer ons die akkuraatheid van bogenoemde inferensies aan die hand van hierdie beramings. Dit geskied met behulp van `n simulasiestudie, uitgevoer in die sagteware pakket S-Plus.

Summary

Extremes are unusual or rare events. However, when such events – for example earthquakes, tidal waves and market crashes - do take place, they typically cause enormous losses, both in terms of human lives and monetary value. For this reason, it is of critical importance to accurately model extremal events. Extreme value theory entails the development of statistical models and techniques in order to describe and model such rare observations.

In this document we discuss aspects of extreme value theory. This theory consists of two approaches: The classical maxima method, based on the properties of the maximum of a sample and the more popular threshold theory, based upon the properties of exceedances of a specified threshold value. This document provides the practitioner with the theoretical and practical tools for both these approaches. This will enable him/her to perform extreme value analyses with confidence.

Extreme value theory – for both approaches - is based upon asymptotic arguments. For finite samples, the limiting result for the sample maximum holds approximately only. Similarly, for finite choices of the threshold, the limiting distribution for exceedances of that threshold holds only approximately. In this document we investigate the quality of extreme value based inferences with regard to the unknown underlying distribution when the sample size or threshold is finite. Estimation of extreme tail quantiles of the underlying distribution, as well as the calculation of confidence intervals, are typically the most important objectives of an extreme analysis. For that reason, we evaluate the accuracy of extreme based inferences in terms of these estimates. This investigation was carried out using a simulation study, performed with the software package S-Plus.

Verklaring

Hiermee verklaar ek, die ondergetekende, dat die inhoud van hierdie werkstuk my eie, oorspronklike werk is, wat ek nie voorheen in enige formaat – hetsy deels of in geheel - by enige universiteit of ander akademiese instelling vir 'n graad aangebied het nie.

Geteken:

Datum:

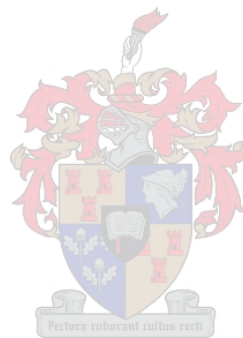


Bedankings

Hierdie werkstuk sou nie voltooi kon word sonder die bydraes van die volgende persone nie; daarom bedank ek baie graag:

- My Hemelse Vader, wat my deurgaans met Sy groot genade deurdrenk het. Ek sê baie dankie vir al die energie, insig en deurbrake waarmee U my deurentyd gesien het. Soms moes ek langer wag vir dié insigte; mag dit my daaraan herinner om altyd nederig te bly en alle eer altyd aan U te gee.
- My wonderlike ouers en my meisie Heidi, wat my eindeloos ondersteun en nuwe moed ingepaat het. Dankie Pa vir al die oproepe om na my vordering te verneem, dankie Ma vir al die ure wat Ma vir my op die knieë deurgebring het – hulle is sekerlik nou behoorlik deurgebid! `n Baie groot dankie aan Heidi, wat ten spyte van die verlies van haar vader in November 2004, aangehou het om my te onderskraag. En baie dankie aan my broer Jaco, wat so getrou gehelp het met my data vestiging; ek waardeer dit, Boet!
- My promotor, professor Tertius de Wet, wat oneindig baie ure afgestaan het om my van bystand, insig en kennis te voorsien. U het telkens die regte oomblik gekies om te prys, om te vermaan en om aan te moedig. Vir hierdie leiding – op persoonlike sowel as akademiese vlak – sê ek aan u baie dankie. Dit het my gehelp om as mens geweldig te groei en ek stel u bydrae daartoe hoog op prys. Dit was voorwaar `n plesier en `n voorreg om met u saam te werk.
- Professor Stuart Coles, vir `n uiters leesbare publikasie wat my in staat gestel het die veld van ekstreemwaarde-teorie baie beter te verstaan. Professor Coles was ook so gaaf om e-posse– waarmee ek hom meer as een maal lastig geval het - van `n volslae vreemdeling te beantwoord; dit boonop gewoonlik binne `n dag of twee! Aan hierdie *gentleman* wat ek graag eendag sal wil ontmoet, sê ek baie dankie.
- Carl du Toit, my gewaardeerde kollega en vriend, wat altyd daar was om te luister en te bemoedig. Baie dankie vir al die verpligtinge wat jy van my skouers gelig het toe die tesis-sperdatum begin dreig het. Ek waardeer dit baie, ou maat!

- Tom Berning, wat altyd bereid was om `n geselsie aan te knoop en my dikwels goeie raad gegee het. Baie dankie ook vir die laaste nippertjie proeflees van hierdie werkstuk, dit word waardeer!
- Ten slotte `n baie groot dankie aan die Departement Statistiek en Aktuariële Wetenskap aan die Universiteit van Stellenbosch. Die uiters kundige personeel van hierdie departement het my van die nodige kennis en insig voorsien om hierdie werkstuk met vertroue aan te pak. Baie dankie ook vir die finansiële steun in die vorm van `n deelydse lektoraat, wat my voortgesette studie moontlik gemaak het.



Inhoudsopgawe

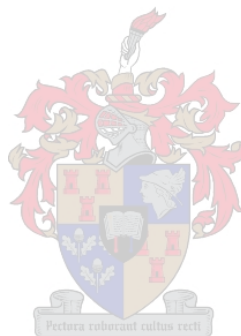
Hoofstuk 1: Die Ekstreemwaarde Probleem	1
Hoofstuk 2: Die Asimptotiese Gedrag van Maksima	6
2.1 Inleiding	6
2.2 Definisies en Notasie	8
2.3 Die Fisher-Tippett Stelling	11
2.4 Maksimum Domeine van Aantrekking	27
2.4.1 Die Frechet-geval	27
2.4.2 Die Weibull-geval	30
2.4.3 Die Gumbel-geval	32
2.5 Parameter-inferensie	34
2.5.1 Algemene Oorwegings	34
2.5.2 Maksimum Anneemlikheid Beraming	35
2.5.3 Inferensie met behulp van die Profiel Anneemlikheid funksie	39
2.5.4 Evaluering van die Gepaste Model	43
2.6 Voorbeeld	45
2.7 Gevolgtrekking	51
Hoofstuk 3: Drempelmodelle	53
3.1 Inleiding	53
3.2 Definisies en Notasie	55
3.3 Die Pickands-Balkema-De Haan Stelling	56



3.4 Die Modelling van Absolute Drempeloorskrydings	66
3.4.1 Die Keuse van 'n Drempel	67
3.4.2 Maksimum Anneemlikheid Beraming	69
3.4.3 Kwantielberaming	70
3.4.4 Evaluering van die Gepaste Model	72
3.5 'n Limietverdeling vir Relatiewe Drempeloorskrydings	74
3.6 Die Modelling van Relatiewe Drempeloorskrydings	75
3.6.1 Die Bepaling van 'n Drempel	76
3.6.2 Maksimum Anneemlikheid Beraming	76
3.6.3 Kwantielberaming	77
3.6.4 Evaluering van die Gepaste Model	79
3.7 Voorbeeld	79
3.8 Gevolgtrekking	85
Hoofstuk 4: Eindige Steekproefgedrag van Maksima-inferensie	86
4.1 Inleiding	86
4.2 Simulasie-ontwerp	87
4.3 Bespreking van Resultate	90
4.3.1 Die Frechet-domein	92
4.3.2 Die Gumbel-domein	100
4.3.3 Die Weibull-domein	105
4.4 Gevolgtrekking	107
Aanhangsel 4A	109
4A.1 Die Frechet-domein	110
4A.2 Die Gumbel-domein	119



4A.3 Die Weibull-domein	128
Hoofstuk 5: Die Gedrag van Drempelinferensie by Eindige Drempelwaardes	
5.1 Inleiding	133
5.2 Simulasie-ontwerp	135
5.3 Bespreking van Resultate	138
5.3.1 Die Frechet-domein	139
5.3.2 Die Gumbel-domein	148
5.3.3 Die Weibull-domein	152
5.4 Gevolgtrekking	152
Aanhangsel 5A	155
5A.1 Die Frechet-domein	156
5A.2 Die Gumbel-domein	165
5A.3 Die Weibull-domein	174
Hoofstuk 6: Samevatting	175
Verwysings	179



Hoofstuk 1: Die ekstreemwaarde probleem

As 'n statistiese dissipline is ekstreemwaarde-teorie onkonvensioneel in die opsig dat dit tegnieke en modelle ontwikkel ten einde rare, ongewone voorkomste – eerder as meer alledaagse gebeure – te beskryf. Die teorie het sy ontstaan gehad met die kernresultaat van Fisher en Tippett in 1928 (Fisher en Tippett, 1928). Vir etlike jare daarna sou slegs gedempte belangstelling onder statistici in hierdie studieveld heers. Die boek van Gumbel (1958) het daartoe bygedra dat ekstreemwaarde-metodologie sedert die vyftigerjare meer wydverspreid aangewend is om natuurlike voorkomste te modelleer. Hierdie aanvanklike toepassings het grotendeels in die siviele ingenieurswese plaasgevind: Ingenieurs moes op grond van beperkte historiese data strukture sterk genoeg bou sodat dit verwagte toekomstige natuurkragte sou kon trotseer. Oor die afgelope 50 jaar het die ekstreemteorie ontluik tot een van die belangrikste statistiese dissiplines in die toegepaste wetenskappe. Ekstreemwaarde-teorie word tans in 'n wye verskeidenheid van toepassingsvelde aangewend, onder andere in finansiële risikobestuur, korttermyn versekering, die voorspelling van verkeer in telekommunikasie en steeds in die ingenieurswese.

Ekstreemanalises word gekenmerk deur die doelwit om die gedrag van 'n stogastiese proses by uitermate hoë vlakke te beskryf. Dit behels tipies die beraming van ekstreme stertkwantiele – dus dié met 'n uiters lae oorskrydingswaarskynlikheid. Dikwels word van hierdie analises verwag om die waarskynlikheid van ekstreme gebeure wat nog nooit vantevore plaasgevind het nie, te beraam. 'n Belangrike voorbeeld hiervan is die laagliggende kusstaat Nederland. Hierdie land – waarvan groot dele laer as die seevlak geleë is – het deur die eeue gereeld rampspoedige vloede (die mees onlangse oorstromings het in 2002 plaasgevind) ondervind. Die Februarie-vloed van 1953 was veral verwoestend: 1836 mense is gedood en 201 000 beeste het versuip. Daarby is 49 000 huise en plase, sowel as 200 000 hektaar se grond oorstrom. 'n Opskrif

in *De Yssel- en Lekstreek*, 6 Februarie 1953, het gelui: “Springtij en orkaan veroorzaken nationale ramp. Nederland in grote watersnood.” As `n klein kusstaat waar grond drooggelê word (sulke areas staan as polders bekend) ten einde die bewoonbare landoppervlakte te vermeerder, maak Nederland staat op stuitwalle genaamd dyke om die land teen oorstromings vanaf die omringende see te beskerm. Hier is dit noodsaaklik dat siviele ingenieurs die dyke hoog genoeg sal bou sodat die polders veilig sal wees teen alle seevlakke wat in die volgende (sê) 100 jaar verwag word. Tipies is daar egter nie 100 jaar se historiese plaaslike seevlakdata beskikbaar nie, maar slegs vir `n baie korter tydperk. Selfs waar `n langer tydperk van historiese data beskikbaar is, mag die aard van die data sodanig wees dat ekstreme vlakke wat nog nie in die verlede voorgekom het nie, in die toekoms mag realiseer.

Ekstreemwaarde-teorie het meer onlangs ook in finansiële risikobestuur `n natuurlike toepassingsveld gevind. Die toename in die volume en kompleksiteit van finansiële instrumente wat verhandel word, het daartoe gelei dat finansiële risikobestuur `n uiters belangrike kwessie in finansiële instellings en korporasies geword het. Onlangse finansiële katastrofes soos die ondergang van Metallgesellschaft, Barings Bank en Enron beklemtoon die belangrikheid van gesonde risikobestuur. Regulasies, soos die Basel II akkoord vir banke, sowel as Solvency 2 vir lewens- en korttermyn-versekeraars, verplig deesdae finansiële instellings om streng risikobestuur toe te pas en aan sekere tegniese vereistes te voldoen. Dit behels veral dat hierdie instansies voldoende kapitaalreserwes moet opbou ten einde hulle risiko-posisie in beleggingsmarkte te verskans. Een van die mees bekende regulatoriese maatstawwe van hierdie risiko-blootstelling staan bekend as Waarde op Risiko, oftewel “Value at Risk” (VAR). Vir `n gegewe tydshorison t en vertrouensvlak p word VAR gedefinieer as daardie verlies in markwaarde oor die tydshorison t wat met waarskynlikheid $1-p$ oorskry word. Anders gestel: Die waarskynlikheid dat ons meer as VAR sal verloor oor die tydshorison t is minder as $1-p$. Ten einde `n bank se reserwe vir die dekking van markrisiko te bepaal, neem regulators $t=10$ dae en $p=0.99$. Hier word dus

gefokus op die vlak van verliese wat oor 'n periode van 10 dae met 'n waarskynlikheid van slegs 0.01 (oftewel een persent van die tyd) oorskry word. Statisties gesproke is die berekende VAR die eerste persentiel van die waarskynlikheidsverdeling van die verandering in die waarde van die portefeulje oor 'n 10-dag periode. Hierdie bedrag staan bekend as die Kapitaal op Risiko oftewel "Capital at Risk" (CAR). Vir die berekening van interne risikomaatstawwe kan selfs meer ekstreme stertkwantiele van die portefeulje-verliesverdeling beskou word deur p nog kleiner te neem, byvoorbeeld $p = 0.999$.

In die meer konvensionele statistiese analises waar die beraming van middelwaardes van belang is, is daar tipies volop steekproef waarnemings in die omgewing van hierdie middelwaardes beskikbaar. In hierdie ondersoeke kom die beraming van middelwaardes dus neer op interpolasie tussen die omliggende datapunte. In 'n ekstreemstudie modelleer ons egter dikwels die voorkoms van gebeure buite die variasiewydte van die beskikbare steekproef. Dit beteken dat ons in 'n ekstreemanalise dikwels moet ekstrapoleer vanaf waargenome waardes na voorkomste wat nog nie vantevore plaasgevind het nie. Ekstreemwaarde-metodologie gebruik asimptotiese teorie en verskaf klasse van limietverdelings wat hierdie soort ekstrapolasie moontlik maak. In hierdie werkstuk ondersoek ons die doeltreffendheid van sodanige asimptotiese teorie in die geval van eindige steekproewe.

Die teorie van ekstreemwaardes bestaan uit die klassieke maksima-metode, sowel as die meer moderne drempelbenadering. In die tweede hoofstuk van hierdie werkstuk word die maksimateorie in detail uiteengesit. Die hoofresultaat is die Fisher-Tippett stelling, waarvolgens die algemene vorm van 'n unieke klas van limietverdelings – indien sekere voorwaardes bevredig word sodat die limiet wel bestaan - vir die steekproef maksimum herlei is. Hierdie klas van verdelings staan as die veralgemeende ekstreemwaarde-verdeling bekend. Afgesien van die teoretiese bespreking, word aandag ook aan die praktiese modellering van ekstreme waardes met behulp van die blokmaksima benadering, gegee. Ons

sluit die hoofstuk af met `n praktiese voorbeeld hiervan, naamlik die Port Pirie datastel, wat bestaan uit die jaarlikse maksimum seevlak soos waargeneem by Port Pirie, net noord van Adelaide in Suid-Australië, gedurende die periode 1923-1987.

In hoofstuk drie word aspekte van die drempelgebaseerde ekstreemwaarde-teorie bespreek. Drempelteorie behels dat alle waarnemings wat `n sekere drempel oorskry, as ekstreme waardes beskou word. Hier is ons hoofresultaat die Pickands-Balkema-De Haan stelling, waarvolgens die voorwaardelike verdeling van die drempeloorskrydings by hoë drempels “naby” is aan die veralgemeende Pareto-verdeling. Hierdie stelling is van toepassing op absolute oorskrydings, maar die teoretiese eienskappe van relatiewe oorskrydings word ook belig. Soos in die geval van die tweede hoofstuk, word die praktiese aspekte van drempelmodellering in detail behandel. Die hoofstuk word weer eens met `n praktiese voorbeeld hiervan afgesluit, naamlik die bestudering van die daaglikse reënval in die suidweste van Engeland oor die tydperk 1914-1962.

Ekstreemwaarde-metodologie maak gebruik van asimptotiese teorie wat slegs by benadering geld in die geval van eindige steekproewe (maksimateorie) en eindige drempelwaardes (drempelmetode). In hoofstuk vier word die volgende vraag ondersoek: Hoe geldig is maksima-gebaseerde inferensie ten opsigte van die onbekende onderliggende verdeling by eindige steekproefgroottes? Ons maak van `n simulasiestudie gebruik om hierdie vraag te evalueer. Die beraming van stertkwantiele van die onderliggende verdeling en vertrouensintervalle vir hierdie kwantiele is een van die vernaamste doelwitte van `n ekstreemwaarde-analise. Om hierdie rede oorweeg ons die geldigheid van bogenoemde inferensie in terme van die akkuraatheid van hierdie beramings. In hoofstuk vyf ondersoek ons die geldigheid van drempelgebaseerde inferensie rakende die onderliggende verdeling by eindige waardes van die drempel. Soos in hoofstuk vier word daar van `n simulasiestudie gebruik gemaak om hierdie kwessie aan te spreek. Weer eens word puntberamings vir stertkwantiele van die onderliggende verdeling,

sowel as beraamde vertrouensintervalle vir hierdie kwantiele verkry ten einde die geldigheid van dié inferensie te evalueer.

Ekstreemwaarde-teorie maak van asimptotiese argumente gebruik om vanaf voorheen waargenome voorkomste te ekstrapoleer na vlakke wat nog nie vantevore aangeteken is nie. Hierdie ekstrapolasie na die onbekende – en dit gebaseer op asimptotiese teorie wat slegs by benadering geld – is `n moontlike punt van kritiek teen die ekstreemteorie. Daarteenoor geld die argument dat toepassings ekstrapolasie verg en dat ekstreemwaarde-teorie tans die mees wetenskaplike benadering is tot hierdie probleem.

There is always going to be an element of doubt, as one is extrapolating into areas one doesn't know about. But what extreme value theory (EVT) is doing is making the best use of whatever data you have about extreme phenomena.

- Richard Smith (soos aangehaal in Chavez-Demoulin en Roehrl, 2004)

The key message is that EVT cannot do magic – but it can do a whole lot better than empirical curve-fitting and guesswork. My answer to the sceptics is that if people aren't given well-founded methods like EVT, they'll just use dubious ones instead. – Jonathan Tawn (soos aangehaal in Chavez-Demoulin en Roehrl, 2004)

Hoofstuk 2: Die Asimptotiese Gedrag van Maksima

2.1 Inleiding

In die klassieke teorie stel ons dikwels belang in die eienskappe van middelwaardes soos die gemiddelde en mediaan. In die studie van ekstreme waardes ondersoek ons ontpiepte gebeurtenisse, daardie waarnemings wat – anders as die waarnemings in die omgewing van die bogenoemde middelwaardes – selde voorkom. Hierdie sogenaamde ekstreme waarnemings word dikwels bestudeer omdat hulle voorkoms potensiële risiko's inhou. Ekstreme weerstoestande soos uiters hoë windsnelhede of baie lae temperature kan gevaar vir menselewens inhou, terwyl 'n enorme groot versekeringseis 'n herversekeraar kan laat ondergaan. Die akkurate modellering van hierdie ekstreme waarnemings is dus van groot belang.

As 'n eerste stap in hierdie modellering word maksima gedefinieer as ekstreme waardes. Dit is bekend dat $P(M_n \leq x) = F^n(x)$, waar M_n die maksimum oor n waarnemings met 'n onderliggende verdeling F voorstel. Dit kan bewys word dat M_n byna seker konvergeer na x_F , die regter eindpunt van die verdeling F . Hierdie resultate is egter van mindere praktiese belang. Gevolglik word eerder ondersoek ingestel na die limietverdeling van M_n , dus of daar konstantes $a_n > 0$ en b_n gevind kan sodat M_n , genormeer met hierdie konstantes, na 'n nie-ontaarde limietverdeling konvergeer as $n \rightarrow \infty$. In hierdie hoofstuk word aandag gegee aan sodanige limietresultate en word aangetoon hoe dit vir inferensie doeleindes aangewend kan word.

In paragraaf 2.2 word 'n aantal noodsaaklike definisies en notasie uitgelig. Die hoofresultaat van hierdie hoofstuk is die Fisher-Tippett stelling, gegee en bewys in paragraaf 2.3. Hierdie stelling bewys dat indien genormeerde maksima na 'n limietverdeling konvergeer, dan is hierdie limietverdeling die veralgemeende ekstreemwaarde-verdeling G_γ . Die subklasse van G_γ , naamlik die Frechet ($\gamma > 0$), Gumbel ($\gamma = 0$) en Weibull ($\gamma < 0$) klasse word gegee, terwyl G_γ verder veralgemeen word tot die vorm $G_{\gamma,\mu,\sigma}$. Hier is $G_{\gamma,\mu,\sigma}$ van dieselfde klas as G_γ , met μ 'n skuif- en σ 'n skaalparameter. Voorwaardes op die onderliggende verdeling F word

bespreek, terwyl gepaste keuses van $a_n > 0$ en b_n aangebied word vir 'n aantal bekende kontinue verdelings, waaronder die normaalverdeling. In paragraaf 2.4 word die maksimum domein van aantrekking van elk van die subklasse van G_γ gekarakteriseer. Hierdie karakterisering geskied in terme van die stertkwantiefunksie U sowel as die stert-verdelingsfunksie \bar{F} , terwyl die Von Mises stelling telkens 'n voldoende voorwaarde op die onderliggende verdeling F gee om tot 'n bepaalde domein van G_γ te behoort. Voorbeelde word ook verskaf van bekende verdelings binne elk van die genoemde domeine van aantrekking.

Paragraaf 2.5 word aan parameter-inferensie gewy. By praktiese modellering word die bestaande steekproef in blokke van gelyke grootte opgedeel en die ekstreemwaarde-verdeling $G_{\gamma,\mu,\sigma}$ word op die blokmaksima gepas. Ten eerste word 'n aantal algemene oorwegings by die keuse van 'n optimale blokopdeling in paragraaf 2.5.1 toegelig. Die keuse van die blokgrötte en aantal blokke lei naamlik tot 'n afruiling tussen sydigheid en variansie. Die metode van maksimum aanneemlikheid (MA) om beramers vir die parameters γ, μ en σ van die veralgemeende eksteemwaarde-verdeling $G_{\gamma,\mu,\sigma}$ te vind, word in paragraaf 2.5.2 ondersoek. Die asimptotiese normaliteit van die MA beramers word gebruik om benaderde vertrouensintervalle vir die parameters van die veralgemeende eksteemwaarde-verdeling te herlei. Die profiel aanneemlikheid funksie, gebaseer op die χ^2 -verdeling word ook vir hierdie doel aangewend. Beramers vir ekstreme kwantiele word herlei. In paragraaf 2.5.4 word grafiese tegnieke om die pasgehalte van die veralgemeende ekstreemwaarde-model op die maksima te beoordeel, bespreek.

Die hoofstuk word afgesluit met 'n praktiese voorbeeld. Die datastel onder beskouing is die maksimum jaarlikse seevlak soos waargeneem by Port Pirie, net noord van Adelaide in Suid-Australië, gedurende die periode 1923 tot 1987.

2.2 Definiesies en notasie

Die volgende definiesies sal gebruik word:

Definisie 2.2.1

Die funksie $F_X(x) = P(X \leq x)$ word gedefinieer as die verdelingsfunksie van die stogastiese veranderlike X . Waar daar geen moontlikheid van verwarring bestaan nie, sal vir gerief slegs $F(x)$ geskryf word.

Δ

Definisie 2.2.2

Die funksie $f_X(x) = \frac{d}{dx} F_X(x)$ word gedefinieer as die digtheidsfunksie van die stogastiese veranderlike X . Waar daar geen moontlikheid van verwarring bestaan nie, sal vir gerief slegs $f(x)$ geskryf word.



Δ

Definisie 2.2.3

Die funksie $Q(p) = \inf\{x : F(x) \geq p\}$ word gedefinieer as die kwantiefunksie, die inverse van die verdelingsfunksie F .

Δ

Definisie 2.2.4

Die stert-kwantiefunksie U word as volg gedefinieer:

$$U(y) = Q\left(1 - \frac{1}{y}\right) = F^{-1}\left(1 - \frac{1}{y}\right), \quad y \geq 1.$$

Δ

Definisie 2.2.5

Die regter eindpunt van die stogastiese veranderlike X met verdelingsfunksie $F_X(x)$ word gedefinieer as

$$x_F = \sup_x \{x \in \mathfrak{R} : F(x) < 1\} \leq \infty.$$

Δ

Definisie 2.2.6

Veronderstel X_1, X_2, \dots, X_n is onafhanklike, identies-verdeelde stogastiese veranderlikes, verdeel soos 'n stogastiese veranderlike X met verdelingsfunksie $F(x, \underline{\theta})$ en digtheidsfunksie $f(x, \underline{\theta})$, waar $\underline{\theta}$ 'n onbekende parametervektor verteenwoordig. Die aanneemlikheidsfunksie van X_1, X_2, \dots, X_n word gedefinieer as

$$\begin{aligned} L(\underline{x}, \underline{\theta}) &= f_{X_1, X_2, \dots, X_n}(x_1, \dots, x_n, \underline{\theta}) = f_X(x_1, \underline{\theta}) f_X(x_2, \underline{\theta}) \dots f_X(x_n, \underline{\theta}) \\ &= \prod_{i=1}^n f_X(x_i, \underline{\theta}). \end{aligned}$$

In praktiese toepassings is die logaritme van die aanneemlikheidsfunksie dikwels van groter belang en die volgende notasie word vir hierdie funksie gereserveer:

$$l(\underline{x}, \underline{\theta}) = \log L(\underline{x}, \underline{\theta}).$$



Δ

Definisie 2.2.7

Veronderstel X_1, X_2, \dots, X_n is onafhanklike, identies-verdeelde stogastiese veranderlikes, verdeel soos 'n stogastiese veranderlike X met verdelingsfunksie $F(x, \underline{\theta})$ en digtheidsfunksie $f(x, \underline{\theta})$, waar $\underline{\theta}$ 'n onbekende parametervektor verteenwoordig. Die maksimum aanneemlikheids beramer $\hat{\underline{\theta}}$ van $\underline{\theta}$ word gedefinieer as daardie waarde van $\underline{\theta}$ wat die aanneemlikheidsfunksie $L(\underline{x}, \underline{\theta})$ maksimeer. Omdat die logaritmiëse funksie monotoon stygend is, kan die maksimum aanneemlikheids beramer $\hat{\underline{\theta}}$ van $\underline{\theta}$ ook gedefinieer word as daardie waarde van $\underline{\theta}$ wat die log-aanneemlikheidsfunksie $l(\underline{x}, \underline{\theta})$ maksimeer.

Δ

Definisie 2.2.8

Laat $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ die waargenome waardes voorstel van 'n geordende steekproef van onafhanklike waarnemings uit 'n populasie met verdelingsfunksie F . Die empiriese verdelingsfunksie word gedefinieer as

$$\tilde{F}(x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{i}{n}, & x_{(i)} \leq x < x_{(i+1)} \\ 1, & x \geq x_{(n)} \end{cases}$$

vir $i = 1, 2, \dots, n-1$.

Dan bestaan die waarskynlikheidstipping uit die punte

$$\{(\hat{F}(x_{(i)}), \frac{i}{n}), i = 1, 2, \dots, n\}$$

en die kwantielstipping uit die punte

$$\{(\hat{F}^{-1}(\frac{i}{n}), x_{(i)}), i = 1, 2, \dots, n\},$$

waar \hat{F} op 'n beraamde verdelingsfunksie – soos gepas op die waarnemings x_1, x_2, \dots, x_n - dui.



Δ

Definisie 2.2.9

Laat ℓ 'n positiewe, reële en meetbare funksie wees. Dan is ℓ reëlmatig-variërend as en slegs as daar 'n reële konstante ρ bestaan sodanig dat $\lim_{x \rightarrow \infty} \frac{\ell(xt)}{\ell(x)} = t^\rho$ vir alle $t > 0$. Ons skryf $\ell \in \mathfrak{I}_\rho$ en ons noem ρ die indeks van reëlmatige variëring. In die geval $\rho = 0$, noem ons die funksie ℓ stadig-variërend. Vir 'n stadig-variërende funksie ℓ geld dus: $\lim_{x \rightarrow \infty} \frac{\ell(xt)}{\ell(x)} = 1$. Voortaan word die simbool ℓ vir hierdie klas van funksies gereserveer.

Δ

Definisie 2.1.10

Laat $\{X_n\}$ 'n ry stogastiese veranderlikes met verdelingsfunksies $\{F_n(x)\}$ en X 'n stogastiese veranderlike met verdelingsfunksie $F(x)$ wees. Die ry $\{X_n\}$ konvergeer met waarskynlikheid een – of byna seker – na X indien

$$P[\lim_{n \rightarrow \infty} X_n = X] = 1,$$

wat ook aangedui word as $X_n \xrightarrow{b.s.} X$. Die ry $\{X_n\}$ konvergeer in verdeling na X indien

$\lim_{n \rightarrow \infty} F_n(x) = F(x)$ vir elke kontinue punt x van $F(x)$ en dit word aangedui deur

$$X_n \xrightarrow{D} X.$$

△

2.3 Die Fisher-Tippett stelling

Laat X_1, X_2, \dots, X_n 'n steekproef van grootte n voorstel van 'n stogastiese veranderlike X met 'n verdelingsfunksie F . X_1, X_2, \dots, X_n is onafhanklik en identies verdeel soos X .



In tradisionele statistiek word daar dikwels in middelwaardes (soos die gemiddelde en die mediaan) en in die besonder in die asimptotiese eienskappe daarvan belang gestel. Die ruggraat van sulke ondersoeke is die bekende sentrale limietstelling, waarbinne die normaalverdeling as limietverdeling aangebied word. Volgens die sentrale limietstelling konvergeer die verdeling van

$$\frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n X_i - E(X)}{\sqrt{\text{Var}(X)}}$$

onder geskikte voorwaardes na die Normaal(0,1) verdeling indien die steekproefgrootte $n \rightarrow \infty$. Die sentrale limietstelling beskou dus die som $S_n = X_1 + X_2 + \dots + X_n$ en poog om konstantes $a_n > 0$ en b_n te vind sodat $Y_n = a_n^{-1}(S_n - b_n)$ in verdeling na 'n nie-ontaarde verdeling konvergeer. In die geval

van die sentrale limietstelling is die normaalverdeling ons nie-ontaarde limietverdeling.

In ekstreemwaarde-teorie daarenteen, val die klem op die studie van ekstreme waardes, eerder as op die bestudering van meer tipiese waarnemings wat in die sentrum van verdelings aangetref word. Hier word ekstreme waardes gedefinieer as die maksima van 'n reeks stogastiese veranderlikes.

As 'n eerste stap stel ekstreemwaarde-teorie belang in die verdeling van

$$M_n = \max(X_1, X_2, \dots, X_n).$$

Ons weet die volgende:

$$P(M_n \leq x) = F^n(x).$$

Ongelukkig is hierdie (eksakte) resultaat van beperkte onmiddellike nut in die praktyk aangesien die verdeling F tipies onbekend is. Ons word dus genoop om die asimptotiese eienskappe van M_n te ondersoek.

Ekstreme waardes vind plaas in die omgewing van die boonste stert (indien ons in maksima belangstel) of in die onderste stert (indien minima van belang is) van die onderliggende verdeling F . Dus kan verwag word dat die asimptotiese gedrag van M_n verwant is aan die verdelingsfunksie F in die omgewing van die regter eindpunt in die regter stert.

Dit volg dat vir $x < x_F$,

$$P(M_n \leq x) = F^n(x) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

In die geval $x \geq x_F$ en $x_F < \infty$,

$$P(M_n \leq x) = F^n(x) = 1 \quad \text{as } n \rightarrow \infty.$$

Dus, $M_n \xrightarrow{P} x_F$ as $n \rightarrow \infty$ waar $x_F < \infty$.

Aangesien die ry $\{M_n\}$ nie-dalend is in n , konvergeer dit byna seker:

$$M_n \xrightarrow{b.s.} x_F, \quad n \rightarrow \infty.$$

Hierdie feit bied egter weinig bruikbare insig in die asimptotiese verdeling van maksima.

Die volgende vraag duik dan op: Bestaan gepaste konstantes $a_n > 0$ en b_n sodat indien M_n genormeer word daarmee, dit na 'n nie-ontaarde limietverdeling konvergeer? Dus, bestaan daar 'n analoog aan die sentrale limietstelling vir die genormeerde maksimum van 'n steekproef?

Die antwoord hierop, naamlik dat so 'n nie-ontaarde limietverdeling vir genormeerde maksima wel bestaan, is in 1928 deur Fisher en Tippett (Fisher en Tippett, 1928) gegee en deur Gnedenko (1943) verder verfyn.

Ons sê nou dat F in die maksimum domein van aantrekking van die verdeling G lê, indien daar konstantes $a_n > 0$ en b_n bestaan sodat die genormeerde maksima

$\frac{M_n - b_n}{a_n}$ met onderliggende verdeling F na 'n verdeling G konvergeer as $n \rightarrow \infty$.

Dus,

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \rightarrow G(x).$$

Ons skryf hiervoor $F \in MDA(G)$.

Let op dat deurgaans aanvaar sal word dat die onderliggende populasie-verdeling F kontinu is.

Stelling 2.3.1: Die Fisher-Tippett stelling

Indien $F \in MDA(G)$ moet G noodwendig van die vorm G_γ wees, met

$$G_\gamma(x) = \exp[-(1 + \gamma x)^{-1/\gamma}], \text{ vir } 1 + \gamma x > 0. \quad (2.1)$$

Hier neem ons vir $\gamma = 0$,

$$G_0(x) = \exp[e^{-x}] \quad (2.2)$$

Δ

Die volgende resultaat, die sogenaamde Helly-Bray stelling, vorm die basis vir die bewys van die Fisher-Tippett stelling. Sien Billingsley (1995) vir 'n bewys van die Helly-Bray stelling.

Stelling 2.3.2 : Die Helly-Bray stelling

Veronderstel die ry stogastiese veranderlikes $\{Y_n\}$ het verdelingsfunksies $\{F_n\}$ en die stogastiese veranderlike Y het 'n verdelingsfunksie F . Dan sal $Y_n \xrightarrow{D} Y$ as en slegs as vir alle reële, begrensde en kontinue funksies z , $E(z(Y_n)) \rightarrow E(z(Y))$ as $n \rightarrow \infty$.

Δ

Voordat die Fisher-Tippett stelling bewys word, bewys ons eers 'n aantal hulpresultate.

Lemma 2.3.3

Gestel 'n funksie h bestaan sodat

$$\lim_{x \rightarrow \infty} \left\{ \frac{U(xu) - U(x)}{a(x)} \right\} = h(u), \quad (2.3)$$

vir 'n positiewe funksie a en alle $u > 0$. Dan bestaan

$$g(u) \equiv \lim_{x \rightarrow \infty} \frac{a(xu)}{a(x)}$$

en g voldoen aan

$$g(uv) = g(u)g(v), \text{ vir } u, v > 0.$$

Bewys

Veronderstel

$$\lim_{x \rightarrow \infty} \left\{ \frac{U(xu) - U(x)}{a(x)} \right\} \equiv h(u)$$

bestaan. Laat $u, v > 0$, dan

$$\frac{U(xuv) - U(x)}{a(x)} = \frac{U(xuv) - U(xu) + U(xu) - U(x)}{a(x)} = \frac{U(xuv) - U(xu)}{a(xu)} \frac{a(xu)}{a(x)} + \frac{U(xu) - U(x)}{a(x)} \quad (2.4)$$

Aangesien die limiete van $\frac{U(xuv) - U(x)}{a(x)}$, $\frac{U(xuv) - U(xu)}{a(xu)}$ en $\frac{U(xu) - U(x)}{a(x)}$ almal

bestaan as $x \rightarrow \infty$, volg dat die verhouding $\frac{a(xu)}{a(x)}$ moet konvergeer as $x \rightarrow \infty$. Laat

$$g(u) = \lim_{x \rightarrow \infty} \frac{a(xu)}{a(x)}.$$

Nou, vir alle $u, v > 0$ geld dat

$$\frac{a(xuv)}{a(x)} = \frac{a(xuv)}{a(xv)} \frac{a(xv)}{a(x)}$$

sodat deur $x \rightarrow \infty$, volg dat $g(uv) = g(u)g(v)$.

Δ

Let op dat in lemma 2.3.3 word bewys dat

$$g(u) \equiv \lim_{x \rightarrow \infty} \frac{a(xu)}{a(x)}$$

dus 'n variant van die Cauchy funksionaalvergelyking (sien Paneah, 2003), naamlik

$$f(a+b) = f(a) + f(b),$$

bevredig. Die oplossing van so 'n Cauchy vergelyking is altyd van die vorm $f(a) = \lambda a$, $\lambda \in \mathfrak{R}$. Deur hierin 'n log-transformasie te maak, kan maklik aangetoon word dat enige positiewe, meetbare oplossing van die vergelyking $g(uv) = g(u)g(v)$; $u, v > 0$ noodwendig van die vorm $g(u) = u^\gamma$ is, vir 'n reële γ .

Lemma 2.3.4

Indien

$$\lim_{x \rightarrow \infty} \left\{ \frac{U(xu) - U(x)}{a(x)} \right\} \equiv h(u)$$

vir 'n positiewe funksie a en positiewe u bestaan, dan is h noodwendig van die vorm

$$h(u) \equiv h_\gamma(u) = \frac{u^\gamma - 1}{\gamma}.$$

(C_γ)



Hier neem ons

$$h_0(u) = \log u. \tag{C_0}$$

Bewys

Herskryf vergelyking (2.3) as volg:

$$\lim_{x \rightarrow \infty} \left\{ \frac{U(xu) - U(x)}{a(x)} \right\} \equiv h_\gamma(u).$$

Dan volg uit (2.4) dat

$$h_\gamma(uv) = h_\gamma(v)u^\gamma + h_\gamma(u).$$

As $\gamma = 0$ dan geld

$$h_0(uv) = h_0(v) + h_0(u)$$

sodat

$$h(u) = \log(u).$$

As $\gamma \neq 0$ dan vir $u, v > 1$ geld

$$h_\gamma(uv) = h_\gamma(vu)$$

oftewel

$$h_\gamma(v)u^\gamma + h_\gamma(u) = h_\gamma(u)v^\gamma + h_\gamma(v)$$

sodat

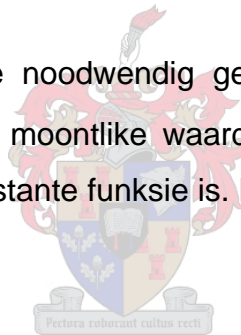
$$\frac{h_\gamma(u)}{u^\gamma - 1} = \frac{h_\gamma(v)}{v^\gamma - 1}.$$

Laat

$$i(v) = \frac{h_\gamma(v)}{v^\gamma - 1}.$$

Dit volg dat $i(v) = i(u)$ vir u, v nie noodwendig gelyk nie. Die funksie i lewer dus dieselfde funksiewaarde by elke moontlike waarde in die definisieversameling. Dit impliseer dat die funksie i 'n konstante funksie is. Hieruit volg weer dat

$$h_\gamma(u) = d(u^\gamma - 1).$$



Die konstante d kan vervang word deur die konstante $c = \gamma d$ oftewel $d = \frac{c}{\gamma}$ ten einde die geval $\gamma = 0$ (die limiet as $\gamma \rightarrow 0$) te inkorporeer. Dit sou impliseer dat

$$h_\gamma(u) = \frac{c(u^\gamma - 1)}{\gamma}.$$

Aangesien die stertkwantielfunksie U monotoon nie-dalend is, impliseer dit dat $h_\gamma(u)$ ook monotoon nie-dalend is. Indien $\gamma > 0$, dan is $u^\gamma - 1$ positief sodat d nie-negatief moet wees. As $\gamma < 0$ dan is $u^\gamma - 1$ negatief en moet d gevolglik ook negatief wees. Dit volg dat $c = \gamma d$ altyd positief sal wees sodat c in die positiewe funksie a geabsorbeer kan word. Gevolglik het ons

$$h_\gamma(u) = \frac{u^\gamma - 1}{\gamma}.$$

Δ

Hierdie hulpresultate stel ons nou in staat om ons hoofresultaat, die Fisher-Tippett stelling, te bewys.

Bewys van die Fisher-Tippett stelling


Stel $Y_n = \frac{M_n - b_n}{a_n}$.

Dan weet ons uit die Helly-Bray stelling dat $Y_n \xrightarrow{D} Y_\gamma$ met $Y_\gamma \sim G_\gamma$ as en slegs as

$$E\left\{z\left(\frac{M_n - b_n}{a_n}\right)\right\} \rightarrow E(z(Y)) = \int_{-\infty}^{\infty} z(y) dG_\gamma(y)$$

indien $n \rightarrow \infty$, vir enige begrensde, kontinue funksie z .

Beskou nou, met $f(x) = \frac{dF(x)}{dx}$,

$$\begin{aligned} E\left\{z\left(\frac{M_n - b_n}{a_n}\right)\right\} &= \int_{-\infty}^{\infty} z\left(\frac{x - b_n}{a_n}\right) dF^n(x) \\ &= \int_{-\infty}^{\infty} z\left(\frac{x - b_n}{a_n}\right) nF^{n-1}(x) f(x) dx \end{aligned}$$


Maak hierin die transformasie

$$F(x) = 1 - \frac{v}{n}$$

om te kry, met

$$U(x) = F^{-1}\left(1 - \frac{1}{x}\right),$$

$$E\left\{z\left(\frac{M_n - b_n}{a_n}\right)\right\} = n \int_n^0 z\left(\frac{U\left(\frac{n}{v}\right) - b_n}{a_n}\right) \left[1 - \frac{v}{n}\right]^{n-1} \left(-\frac{1}{n}\right) dv$$

$$= \int_0^n z\left(\frac{U\left(\frac{n}{v}\right) - b_n}{a_n}\right) \left[1 - \frac{v}{n}\right]^{n-1} dv$$

Aangesien $(1 - \frac{v}{n})^{n-1} \rightarrow e^{-v}$, sal die linkerkant konvergeer indien $\frac{U(\frac{n}{v}) - b_n}{a_n}$ konvergeer vir alle $v > 0$. Kies ons nou $b_n = U(n)$, verkry ons konvergensie van laasgenoemde indien 'n positiewe funksie a bestaan sodat vir $v > 0$ geld

$$\lim_{x \rightarrow \infty} \frac{U(xu) - U(x)}{a(x)} = h(u),$$

vir een of ander funksie h , nie identies nul nie. Nou, volgens lemma 2.3.5 kan ons h neem van die vorm

$$h(u) \equiv h_\gamma(u) = \frac{u^\gamma - 1}{\gamma},$$

vir γ reëel en $u > 1$.

Maak ons die keuse $a_n = a(n)$, kry ons dus

$$\lim_{n \rightarrow \infty} E\left\{z\left(\frac{M_n - b_n}{a_n}\right)\right\} = \int_0^\infty z\left(h_\gamma\left(\frac{1}{v}\right)\right) e^{-v} dv. \quad (2.5)$$

Vir $\gamma \neq 0$, maak die transformasie $u = h_\gamma\left(\frac{1}{v}\right)$ in vergelyking (2.5). Dan is

$$u = h_\gamma\left(\frac{1}{v}\right) = \frac{\left(\left(\frac{1}{v}\right)^\gamma - 1\right)}{\gamma} \quad \text{sodat}$$

$$v = (u\gamma + 1)^{\frac{1}{\gamma}}$$

en

$$dv = -(u\gamma + 1)^{\frac{1}{\gamma}-1} du$$

Dit volg dat:

$$\begin{aligned} \lim_{n \rightarrow \infty} E\left[z\left(\frac{M_n - b_n}{a_n}\right)\right] &= - \int_{a = \lim_{v \rightarrow 0} \frac{(\frac{1}{v})^\gamma - 1}{\gamma}}^{b = \lim_{v \rightarrow \infty} \frac{(\frac{1}{v})^\gamma - 1}{\gamma}} z(u) e^{-(u\gamma+1)^{\frac{1}{\gamma}}} (u\gamma + 1)^{\frac{1}{\gamma}-1} du \\ &= \int_b^a z(u) e^{-(u\gamma+1)^{\frac{1}{\gamma}}} (u\gamma + 1)^{\frac{1}{\gamma}-1} du \end{aligned}$$

Let op dat die term

$$e^{-(u\gamma+1)^{\frac{1}{\gamma}}}(u\gamma+1)^{-\frac{1}{\gamma}-1} = \frac{dG_\gamma(u)}{du},$$

waar

$$G_\gamma(u) = e^{-(u\gamma+1)^{\frac{1}{\gamma}}}.$$

Dus volg dat

$$\lim_{n \rightarrow \infty} E\left[z\left(\frac{M_n - b_n}{a_n}\right)\right] = \int_b^a z(u) dG_\gamma(u) \quad (2.6)$$

Die grense a en b in (2.6) word maklik as volg verkry:

Indien $\gamma > 0$,

$$a = +\infty \text{ en } b = -\frac{1}{\gamma}.$$

Indien $\gamma < 0$,

$$a = -\frac{1}{\gamma} \text{ en } b = -\infty.$$

Vir $\gamma = 0$ is

$$h_0(u) = \log u$$

en uitdrukking (2.5) word

$$\lim_{n \rightarrow \infty} E\left\{z\left(\frac{M_n - b_n}{a_n}\right)\right\} = \int_0^\infty z\left(\log\left(\frac{1}{v}\right)\right) e^{-v} dv \quad (2.7)$$

Maak in (2.7) die transformasie

$$u = \log\left(\frac{1}{v}\right),$$

dan volg direk dat

$$\begin{aligned} \lim_{n \rightarrow \infty} E\left\{z\left(\frac{M_n - b_n}{a_n}\right)\right\} &= \int_{-\infty}^{\infty} z(u) \exp(-e^{-u}) e^{-u} du \\ &= \int_{-\infty}^{\infty} z(u) dG_\gamma(u) \end{aligned}$$



waar

$$G_\gamma(u) = \exp[-e^{-u}].$$

Aangesien hierdie geld vir enige begrensde, kontinue funksie z , volg dit

vanaf die Helly-Bray stelling dat $\frac{M_n - b_n}{a_n}$ in verdeling na die stogastiese veranderlike

Y_γ konvergeer, oftewel dat $\frac{M_n - b_n}{a_n} \xrightarrow{D} Y_\gamma$ waar $Y_\gamma \sim G_\gamma$. Hiermee is die bewys voltooi.

△

Opmerking 2.3.1

Volgens die Fisher-Tippett stelling kan die verdelingsfunksie van genormeerde maksima na slegs 'n enkele klas van limietverdelings konvergeer as $n \rightarrow \infty$, naamlik die klas

$$G_\gamma(x) = \exp[-(1 + \gamma x)^{-1/\gamma}], \quad 1 + \gamma x > 0, \quad (2.8)$$

waar γ 'n vormparameter is. Hierdie funksie G_γ staan as die veralgemeende ekstreemwaarde-verdeling bekend. Vervat in hierdie klas van verdelings is (soos geïndekseer deur γ) die volgende drie tradisionele klasse van ekstreemwaarde-verdelings:

$$\begin{aligned} \gamma > 0: \text{ Frechet} \quad G_\gamma\left(\frac{x-1}{\gamma}\right) &= \exp\left\{-\left(1 + \gamma\left(\frac{x-1}{\gamma}\right)\right)^{-1/\gamma}\right\} = \exp\left(-x^{-1/\gamma}\right) \\ \gamma = 0: \text{ Gumbel} \quad G_{\gamma=0}(x) &= \exp(-e^{-x}) \\ \gamma < 0: \text{ Weibull} \quad G_\gamma\left(-\frac{x+1}{\gamma}\right) &= \exp\left\{-\left(1 + \gamma\left(-\frac{x+1}{\gamma}\right)\right)^{-1/\gamma}\right\} = \exp\left\{-(-x)^{-1/\gamma}\right\} \end{aligned} \quad (2.9)$$

Die parameter γ staan as die ekstreemwaarde-indeks bekend. Dit blyk dat die waarde van γ suggereer na watter klas in G_γ die genormeerde maksimum konvergeer.

Opmerking 2.3.2

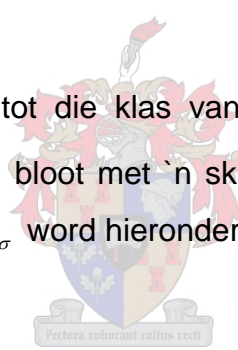
Die verdeling G_γ kan verder veralgemeen word deur die invoer van 'n skuifparameter μ en 'n skaalparameter $\sigma > 0$. Dit lei tot die volgende uitgebreide vorm van die veralgemeende ekstreemwaarde-verdeling:

$$G_{\gamma,\mu,\sigma}(x) = \exp\left[-\left(1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{1}{\gamma}}\right] = G_\gamma\left(\frac{x-\mu}{\sigma}\right),$$

mits

$$1 + \gamma\left(\frac{x-\mu}{\sigma}\right) > 0, \quad -\infty < \mu < \infty, \quad \sigma > 0 \quad \text{en} \quad -\infty < \gamma < \infty.$$

Hier behoort G_γ en $G_{\gamma,\mu,\sigma}$ albei tot die klas van veralgemeende ekstreemwaarde verdelings. Die verdelings verskil bloot met 'n skuif- en 'n skaalparameter. Die nut van hierdie uitgebreide vorm $G_{\gamma,\mu,\sigma}$ word hieronder verder toegelig.



Indien die steekproefgrootte n groot genoeg is, dan kan ons aanvaar dat die limietresultaat benaderd geldig is. Dit impliseer dan die volgende: Vir $F \in MDA(G_\gamma)$ bestaan daar konstantes $a_n > 0$ en b_n sodat vir groot n

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \approx G_\gamma(x)$$

oftewel

$$P(M_n \leq a_n x + b_n) \approx G_\gamma(x)$$

sodat

$$\begin{aligned} P(M_n \leq y) &\approx G_\gamma\left(\frac{y-b_n}{a_n}\right) \\ &= G_{\gamma,b_n,a_n}(y). \end{aligned}$$

Hier verteenwoordig b_n ons skuifparameter en $a_n > 0$ ons skaalparameter in die uitgebreide $G_{\gamma, \mu, \sigma}$.

△

Die Fisher-Tippett stelling is uiters kragtig en van fundamentele belang in die studiegebied van ekstreme waardes. Onthou dat M_n verdeel is volgens F^n , maar dat hierdie kennis onbruikbaar is omdat die verdeling F tipies onbekend is. Die Fisher-Tippett stelling bied 'n enkele, breë klas van limietverdelings $G_{\mu, \sigma, \gamma}$ waarna maksima – op gepaste wyse genormaliseer en met sekere beperkings op die verdeling F - in verdeling konvergeer. Eerder as om die onbekende verdeling F te beraam, reduceer inferensie onder Fisher-Tippett tot die beraming van slegs drie modelparameters, naamlik μ, σ en γ . In hierdie vereenvoudiging lê die krag van die Fisher-Tippett stelling. Die volgende waarskuwing is egter van toepassing: hou in gedagte dat die verdeling $G_{\mu, \sigma, \gamma}$ slegs by benadering geldig is, naamlik vir groot n . 'n Baie relevante vraag is dan: Hoe goed is die benadering $F^n \approx G_{\mu, \sigma, \gamma}$ vir eindige n ? Anders gestel, hoe groot moet n wees sodat die gebruik van die limietverdeling in inferensies aanvaarbare, akkurate resultate sal lewer?

Die Fisher-Tippett stelling sê dat as die verdelingsfunksie van genormeerde maksima met onderliggende verdeling F na 'n nie-ontaarde verdeling konvergeer, dan is G_γ hierdie limietverdeling. Dit laat die volgende vraag ontstaan: Aan watter vereistes moet die verdeling F voldoen om te verseker dat $F \in MDA(G_\gamma)$? Die antwoord op hierdie vraag is dat F streng stygend moet wees (Beirlant et al, 2004), sowel as regs-kontinu (Embrechts et al, 1997). Die implikasie van hierdie voorwaardes is dat al die gebruikelike kontinue verdelings wel na die familie van ekstreemwaardeverdelings konvergeer. Die tempo van konvergensie as die steekproefgrootte toeneem is egter nie ewe vinnig vir al hierdie verdelings nie. Indien F byvoorbeeld normaal verdeel is, vind konvergensie nie baie vinnig plaas nie (Galambos, 1978).

Dit bring die volgende vraag oor die normaliseringskonstantes $a_n > 0$ en b_n na vore: Hoe moet hierdie konstantes gekies word vir konvergensie en watter lid van die ekstreemwaarde-familie word as limiet verkry? Die konstantes is 'n funksie van die verdeling F , wat tipies onbekend is. Gevolglik moet die skuifparameter b_n en die skaalparameter a_n beraam word. Ter illustrasie gee ons $a_n > 0$ en b_n vir 'n paar bekende keuses van F .

Voorbeeld 2.1: Eksponensiaalverdeling.

As X_1, X_2, \dots 'n ry onafhanklike eksponensiaal(λ) stogastiese veranderlikes is, dan is $F(x) = 1 - e^{-\lambda x}$ vir $x > 0$.

Kies in

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n),$$

$$a_n = \frac{1}{\lambda} \text{ en } b_n = \frac{\log n}{\lambda}.$$



Dan is

$$\begin{aligned} P\left(\frac{M_n - b_n}{a_n} \leq x\right) &= F^n\left(\frac{x}{\lambda} + \frac{\log n}{\lambda}\right) \\ &= \{1 - e^{-\lambda[\frac{x}{\lambda} + \frac{\log n}{\lambda}]}\}^n \\ &= \{1 - e^{-(x + \log n)}\}^n \\ &= \left\{1 - \frac{e^{-x}}{n}\right\}^n \\ &\rightarrow \exp(-e^{-x}) \end{aligned}$$

as $n \rightarrow \infty$. Ons het dus aangetoon dat $F \in MDA(\text{Gumbel})$.

△

Voorbeeld 2.2: Uniforme verdeling.

As X_1, X_2, \dots 'n ry onafhanklike uniform(0,1) stogastiese veranderlikes is, dan is $F(x) = x$ vir $0 \leq x \leq 1$.

Kies in

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n),$$

$$a_n = \frac{1}{n} \text{ en } b_n = 1.$$

Dan is

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(x+1)$$

$$= \left(\frac{x}{n} + 1\right)^n$$

$$\rightarrow e^x$$

as $n \rightarrow \infty$. Ons het dus gewys dat $F \in MDA(\text{Weibull})$ waar $\gamma = -1$.

Δ

Voorbeeld 2.3: Veralgemeende Pareto-verdeling.

As X_1, X_2, \dots 'n ry onafhanklike veralgemeende Pareto(γ, β) stogastiese veranderlikes is, dan is $F(x) = 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{-1/\gamma}$, $\gamma \neq 0, \beta > 0, x \geq 0$.

Kies in

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n),$$

$$a_n = \frac{\beta n^\gamma}{\gamma} \text{ en } b_n = \frac{\beta n^\gamma - \beta}{\gamma}.$$

Dan is

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \left\{1 - \left(\frac{\beta}{\beta + \beta n^\gamma x + \beta n^\gamma - \beta}\right)^{1/\gamma}\right\}^n$$

$$= \left\{1 - \left[\frac{1}{n^\gamma (x+1)}\right]^{1/\gamma}\right\}^n$$

$$= \left\{1 - \left[\frac{1}{n} \left(\frac{1}{x+1}\right)^{1/\gamma}\right]\right\}^n$$

$$\rightarrow e^{-\left(\frac{1}{x+1}\right)^{1/\gamma}}$$

$$= e^{-(x+1)^{-1/\gamma}}$$

as $n \rightarrow \infty$. Daar is dus aangetoon dat $F \in MDA(\text{Frechet})$.

△

Voorbeeld 2.4: Standaard Normaalverdeling.

As X_1, X_2, \dots 'n ry onafhanklike normaal(0,1) stogastiese veranderlikes is, dan is

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2} dy \quad \text{vir } x \in \mathfrak{R}.$$

Volgens Mill se verhouding,

$$1 - \Phi(x) \approx \frac{\phi(x)}{x} \quad \text{as } x \rightarrow \infty. \quad (2.10)$$

Hier is $\phi(x) = \frac{d\Phi(x)}{dx}$, die digtheidsfunksie van die standaard normaalverdeling. Deur

van (2.10) gebruik te maak, volg dat

$$\lim_{u \rightarrow \infty} \frac{1 - \Phi(u + \frac{z}{u})}{1 - \Phi(u)} = e^{-z}. \quad (2.11)$$

Ons stel belang in $\lim_{n \rightarrow \infty} \Phi^n(a_n x + b_n)$. Kies ons

$$b_n = \Phi^{-1}\left(1 - \frac{1}{n}\right) \quad \text{en} \quad a_n = b_n^{-1},$$



dan volg uit (2.11),

$$\lim_{n \rightarrow \infty} \frac{1 - \Phi(a_n x + b_n)}{1 - \Phi(b_n)} = e^{-x},$$

oftewel

$$\lim_{n \rightarrow \infty} n[1 - \Phi(a_n x + b_n)] = e^{-x}.$$

Gevolglik,

$$\begin{aligned} \lim_{n \rightarrow \infty} \Phi^n(a_n x + b_n) &= \lim_{n \rightarrow \infty} \left(1 - \frac{e^{-x}}{n}\right)^n \\ &= \exp(-e^{-x}) \end{aligned}$$

Daar is dus aangetoon dat $\Phi \in MDA(\text{Gumbel})$.

△

2.4 Maksimum domeine van aantrekking

Indien die onderliggende verdeling F streng stygend sowel as regs-kontinu is, dan is $F \in MDA(G_\gamma)$. In hierdie paragraaf word aangetoon dat die eienskappe van die stert-kwantiefunksie U sowel as die stert-verdelingsfunksie \bar{F} bepaal in watter domein van aantrekking (Frechet, Gumbel of Weibull) die verdeling F lê. Die Von Mises stelling verskaf telkens 'n voldoende voorwaarde om te bepaal of 'n verdeling tot enige van hierdie drie domeine behoort.

2.4.1 Die Frechet-geval: $\gamma > 0$

'n Kernvoorbeeld van 'n verdelingsfunksie wat in die maksimum domein van aantrekking van die Frechet verdeling lê, is die Pareto-verdeling met stert-verdelingsfunksie

$$\bar{K}(x) = x^{-\alpha}, \quad x > 1.$$

Hier staan $\alpha > 0$ as die Pareto-indeks bekend. Vir hierdie verdeling is

$$Q(p) = (1-p)^{-\frac{1}{\alpha}} \text{ en } U(x) = x^\gamma \text{ met } \gamma = \frac{1}{\alpha}. \text{ Dan is}$$

$$\frac{U(xu) - U(x)}{a(x)} = \frac{x^\gamma (u^\gamma - 1)}{a(x)}.$$

Kies ons hierin $a(x) = \gamma x^\gamma$, dan is

$$\frac{U(xu) - U(x)}{a(x)} = \frac{u^\gamma - 1}{\gamma} = h_\gamma(u)$$

sodat voorwaarde (C_γ) - sien paragraaf 2.3 - bevredig word.

Daar is egter 'n breër klas van verdelings wat (C_γ) bevredig vir $\gamma > 0$. Indien ons $U(x) = x^\gamma \ell_U(x)$ neem, waar $\ell_U(x)$ 'n stadig-variërende funksie is, dan word voorwaarde (C_γ) bevredig:

$$\frac{U(xu) - U(x)}{a(x)} = \frac{(xu)^\gamma \ell_U(xu) - x^\gamma \ell_U(x)}{a(x)}$$

$$\begin{aligned}
&= \frac{\ell_U(x)x^\gamma}{a(x)} \left(\frac{\ell_U(xu)}{\ell_U(x)} u^\gamma - 1 \right) \\
&\approx \frac{u^\gamma - 1}{\gamma}
\end{aligned}$$

vir die keuse $a(x) = \gamma x^\gamma \ell_U(x) = \gamma U(x)$ as $x \rightarrow \infty$.

Verdelings waarvoor $U(x) = x^\gamma \ell_U(x)$ staan bekend as Pareto-tipe verdelings. Let op dat vir hierdie verdelings is U reëlmatig-variërend met indeks γ , aangesien

$$\lim_{x \rightarrow \infty} \frac{U(xt)}{U(x)} = \lim_{x \rightarrow \infty} \frac{(xt)^\gamma \ell_U(xt)}{x^\gamma \ell_U(x)} = t^\gamma \quad \text{vir alle } t > 0.$$

Dit kan bewys word dat die voorwaarde (C_γ) vir $\gamma > 0$ ekwivalent is aan die volgende voorwaarde op die stert-verdelingsfunksie \bar{F} :

$$\frac{\bar{F}(xw)}{\bar{F}(x)} \rightarrow w^{-\frac{1}{\gamma}} \quad \text{vir } w > 0 \text{ as } x \rightarrow \infty.$$

Dit is ekwivalent daaraan dat \bar{F} reëlmatig-variërend is met indeks $-\gamma^{-1}$, oftewel dat \bar{F} geskryf kan word as $\bar{F}(x) = x^{-\alpha} \ell_F(x)$ vir $\alpha = \gamma^{-1}$ en $\ell_F(x)$ 'n stadig-variërende funksie.

Die stellings $U(x) = x^\gamma \ell_U(x)$ en $\bar{F}(x) = x^{-\alpha} \ell_F(x)$ is ten volle ekwivalent sodat die definisie van 'n Pareto-tipe verdeling in terme van die stert-verdelingsfunksie \bar{F} sowel as die stert-kwantiefunksie U geformuleer kan word. Ons het aangetoon dat alle Pareto-tipe verdelings in die maksimum domein van aantrekking van die Frechet verdeling lê.

Verdeling	Stert-verdelingsfunksie $\bar{F}(x)$	Ekstreemwaarde- indeks γ
Pareto	$x^{-\alpha}$ $x > 1, \alpha > 0$	$\frac{1}{\alpha}$
Veralgemeende Pareto(γ, β)	$(1 + \frac{\gamma}{\beta}(x))^{-\frac{1}{\gamma}}$ $x > 0, \gamma, \beta > 0$	γ
Burr(η, τ, γ)	$(\frac{\eta}{\eta + x^\tau})^\lambda$ $x > 0, \eta, \tau, \gamma > 0$	$\frac{1}{\lambda\tau}$
Loggamma(λ, α)	$\int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} w^{-\lambda-1} (\log w)^{\alpha-1} dw$ $x > 1, \lambda, \alpha > 0$	$\frac{1}{\lambda}$
Frechet(α)	$1 - \exp(-x^{-\alpha})$ $x > 0, \alpha > 0$	$\frac{1}{\alpha}$

Tabel 2.4.1: `n Lys van verdelings in die Frechet-domein.

Ons gee `n nou `n resultaat wat `n voldoende voorwaarde verskaf dat `n verdeling in die maksimum domein van aantrekking van die Frechet-klas lê, in terme van die gevaarfunksie

$$r(x) \equiv \frac{f(x)}{\bar{F}(x)}.$$

Stelling 2.4.1: Von Mises Stelling vir die Frechet-klas

Laat $x_F = \infty$. Indien $\lim_{x \rightarrow \infty} xr(x) = \alpha > 0$, dan geld $F \in MDA(\text{Frechet})$.

Δ

2.4.2 Die Weibull-geval: $\gamma < 0$

Soos vir die Frechet-geval begin ons hierdie paragraaf met 'n eenvoudige voorbeeld.

Ons bekyk die stert-verdelingsfunksie

$$\bar{F}(x) = \left(1 - \frac{x}{x_F}\right)^\beta, \quad \beta > 0$$

- wat tot die uniforme verdeling reduceer as $\beta = 1$ - gedefinieer op $(0, x_F)$, waar

$0 < x_F < \infty$. Vir hierdie verdeling is $Q(p) = x_F[1 - (1-p)^{\frac{1}{\beta}}]$ en $U(x) = x_F(1 - x^{-\frac{1}{\beta}})$, vir $x \in [1, \infty)$. Dan is

$$\frac{U(xu) - U(x)}{a(x)} = \frac{x_F}{a(x)} \left\{ \left(1 - (xu)^{-\frac{1}{\beta}}\right) - \left(1 - x^{-\frac{1}{\beta}}\right) \right\}$$

$$= \frac{x_F x^{-\frac{1}{\beta}}}{a(x)} (1 - u^{-\frac{1}{\beta}})$$

$$= \frac{x_F x^{-\frac{1}{\beta}}}{\beta a(x)} h_{-\beta^{-1}}(u)$$

Kies ons hierin $a(x) = \frac{x_F x^{-\frac{1}{\beta}}}{\beta}$, dan word voorwaarde (C_γ) bevredig met $\gamma = -\frac{1}{\beta} < 0$.

'n Breër klas van verdelings bestaan egter waarvoor (C_γ) bevredig word vir $\gamma < 0$.

Indien ons $U(x) = x_F - x^\gamma \ell_U(x)$, $x \rightarrow \infty$, kies waar $\ell_U(x)$ 'n stadig-variërende funksie is, dan word voorwaarde (C_γ) bevredig.

$$\frac{U(xu) - U(x)}{a(x)} = \frac{x^\gamma \ell_U(x)}{a(x)} \left(1 - u^\gamma \frac{\ell_U(xu)}{\ell_U(x)} \right)$$

$$\approx -\gamma \frac{x^\gamma \ell_U(x)}{a(x)} h_\gamma(u).$$

Dit is van die vorm (C_γ) vir die keuse $\frac{a(x)}{x_F - U(x)} \rightarrow -\gamma$ as $x \rightarrow \infty$.

Soos in die geval van die Frechet-domein kan voorwaarde (C_γ) vir $\gamma < 0$ ook in

terme van die stert-verdelingsfunksie \bar{F} geïnterpreteer word. Dit kan bewys word dat die volgende twee stellings ekwivalent is:

$$U(x) = x_F - x^\gamma \ell_U(x), \quad x \rightarrow \infty \text{ en } \bar{F}(x) = x^\gamma \ell_F(x), \quad x \rightarrow \infty.$$

Hier is $\ell_F(x)$ 'n stadig-variërende funksie. Dus, enige verdeling waarvan die stert-verdelingsfunksie \bar{F} of die stert-kwantiel funksie U soos hierbo in terme van 'n stadig-variërende funksie uitgedruk kan word, behoort tot die maksimum domein van aantrekking van die Weibull klas. In tabel 2.4.2 word 'n aantal voorbeelde van sulke verdelings gegee.

Verdeling	Stert-verdelingsfunksie $\bar{F}(x_F - \frac{1}{x})$	Ekstremewaarde- indeks γ
Uniform(0,1)	$\frac{1}{x}$ $x > 1$	-1
Beta(p,q)	$\int_{1-\frac{1}{x}}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1} (1-u)^{q-1} du$ $x > 0; p, q > 0$	$-\frac{1}{q}$
Omgekeerde Burr	$(\frac{\beta}{\beta + x^\tau})^\lambda$ $x > 0; \lambda, \beta, \tau > 0$	$-\frac{1}{\lambda^\tau}$
Weibull	$1 - \exp(-x^{-\alpha})$ $x > 0; \alpha > 0$	$-\frac{1}{\alpha}$

Tabel 2.4.2: 'n Lys van verdelings in die Weibull-domein.

Soos vir die Frechet klas kan 'n voldoende voorwaarde dat 'n verdeling F in die Weibull domein lê in terme van die gevaarfunksie $r(x) = \frac{f(x)}{F(x)}$ gegee word.

Stelling 2.4.2: Von Mises Stelling vir die Weibull-klas

Laat $x_F < \infty$. Indien $\lim_{x \rightarrow x_F} (x_F - x)r(x) = \alpha > 0$, dan geld $F \in MDA(\text{Weibull})$.

Δ

2.4.3 Die Gumbel-geval: $\lambda = 0$

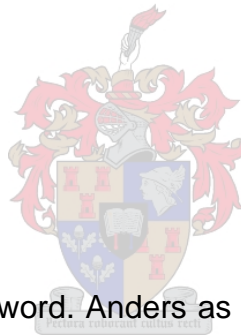
Beskou as kernvoorbeeld die eksponensiaalverdeling $\bar{F}(x) = e^{-\lambda x}$, $x > 0$ vir $\lambda > 0$. Vir

hierdie verdeling is $Q(p) = -\frac{1}{\lambda} \log(1-p)$ en $U(x) = \frac{1}{\lambda} \log x$. Dan is

$$\begin{aligned} \frac{U(xu) - U(x)}{a(x)} &= \frac{1}{\lambda} \frac{\log(xu) - \log(x)}{a(x)} \\ &= \frac{\log u}{\lambda a(x)}. \end{aligned}$$

Kies ons hierin $a(x) = \frac{1}{\lambda}$, dan is

$$\frac{U(xu) - U(x)}{a(x)} = \log u,$$



sodat voorwaarde (C_0) bevredig word. Anders as vir die Frechet- en Weibull-gevalle kan ons nie hier ander verdelings in die Gumbel domein beskou as van dieselfde tipe as ons kernvoorbeeld (hier die eksponensiaalverdeling) nie.

Die karakterisering van die maksimum domein van aantrekking van die Gumbel klas is ook meer kompleks as vir die ander twee gevalle. De Haan (1970) het die volgende oplossing vir hierdie probleem aangebied: 'n Verdeling $F \in MDA(\text{Gumbel})$ as en slegs as vir elke $v > 0$,

$$\frac{\bar{F}(x + b(x)v)}{\bar{F}(x)} \rightarrow e^{-v} \text{ as } x \rightarrow x_F, x_F \in (0, \infty),$$

met b 'n funksie wat voldoen aan

$$\lim_{x \rightarrow x_+} \frac{b(x + vb(x))}{b(x)} = 1.$$

Die Von Mises stelling verskaf weer eens 'n voldoende voorwaarde om te bepaal of 'n verdeling tot die domein van die Gumbel-klas behoort.

Stelling 2.4.3: Von Mises stelling vir die Gumbel-klas

Indien die gevaarfunksie $r(x)$ positief en differensieerbaar is in 'n omgewing van x_F

en $\lim_{x \rightarrow x_F} \frac{dr(x)}{dx} = 0$, dan geld $F \in MDA(\text{Gumbel})$.

Δ

Verdeling	Stert-verdelingsfunksie $\bar{F}(x)$
Benktander II	$x^{-(1-\beta)} \exp(-\frac{\alpha}{\beta} x^\beta)$ $x > 0; \alpha, \beta > 0$
Weibull	$\exp(-\lambda x^\tau)$ $x > 0; \lambda, \tau > 0$
Eksponensiaal	$\exp(-\lambda x)$ $x > 0; \lambda > 0$
Gamma	$\int_x^\infty \frac{\lambda^m}{\Gamma(m)} u^{m-1} \exp(-\lambda u) du$ $x > 0; \lambda, m > 0$
Logisties	$\frac{1}{1 + \exp(x)}$ $x \in \mathfrak{R}$

Tabel 2.4.3: 'n Lys van verdelings in die Gumbel-domein.

2.5 Parameter-inferensie

Gegewe dat die breë klas van verdelings $G_{\mu,\sigma,\gamma}$ as limietverdeling bevestig is, is die logiese volgende taak die beraming van die onbekende parameters μ , σ en γ . Die beraamde waarde van die ekstreemwaarde-indeks γ is natuurlik van besondere belang omdat dit suggereer na watter een van die ekstreemwaarde-verdelings die genormeerde maksimum konvergeer. In hierdie paragraaf word eerstens aandag geskenk aan 'n aantal algemene oorwegings wanneer inferensie in die ekstreemwaarde-konteks plaasvind. Vervolgens word die metodes van maksimum aanneemlikheid sowel as profiel-aanneemlikheid bespreek. Die paragraaf word afgesluit met metodes om die pasgehalte van die beraamde model te evalueer.

2.5.1 Algemene oorwegings

Veronderstel die volgende situasie geld: ons het blokmaksima data

$$\begin{aligned}\underline{X}_1 &= (X_{11}, \dots, X_{1n}) \\ \underline{X}_2 &= (X_{21}, \dots, X_{2n}) \\ \vdots & \quad \quad \quad \vdots \\ \underline{X}_m &= (X_{m1}, \dots, X_{mn})\end{aligned}$$



in die vorm van m blokke met n waarnemings elk. Laat vir $i = 1, 2, \dots, m$,

$$M_n^{(i)} = \max(X_{i1}, \dots, X_{in}),$$

en vorm die m -dimensionele vektor

$$\underline{y} = \begin{bmatrix} M_n^{(1)} \\ M_n^{(2)} \\ \vdots \\ M_n^{(m)} \end{bmatrix}$$

Dit word aanvaar dat $M_n^{(1)}, \dots, M_n^{(m)}$ onafhanklik is en afkomstig uit 'n benaderde ekstreemwaarde-verdeling met digtheidsfunksie $g_{\underline{\theta}}$. Ons stel belang in die beraming van die onbekende parametervektor $\underline{\theta}' = (\mu, \sigma, \gamma)$.

Op hierdie waarnemings $M_n^{(1)}, \dots, M_n^{(m)}$ word 'n ekstreemwaarde-verdeling gepas; daarna word die onbekende parametervektor $\underline{\theta}' = (\mu, \sigma, \gamma)$ beraam. Die keuse van die blokgrötte n (en dus van die aantal blokke m) is van kritieke belang wanneer die beskikbare steekproefdata in blokke opgedeel word. Die volgende afruiling tussen sydigheid en variansie duik naamlik op in hierdie situasie. Blokke met min waarnemings (klein n) impliseer dat die limietverdeling van paragraaf 2.3 nie 'n goeie benadering vir die verdeling F^n sal wees nie. Indien die limietverdeling nie geld nie, is die beraming van die parameters μ , σ en γ natuurlik aan 'n groot mate van sydigheid onderworpe. Klein blokke beteken egter ook meer blokke; meer waarnemings $M_n^{(1)}, \dots, M_n^{(m)}$ waarop die ekstreem-model gepas word. Dit lei weer tot 'n kleiner variansie van die parameterberamings. Kleiner blokke verhoog dus sydigheid, maar verlaag variansie; groter blokke verlaag sydigheid (die limietresultate geld meer eksak), maar het hoër variansie (minder waarnemings). Beide sydigheid en hoër variansie word in statistiek as ongewens beskou: die praktisyn sal by elke ekstreemwaarde-probleem 'n optimale balans tussen hierdie twee faktore moet probeer verkry. Praktiese oorwegings dikkeer dikwels die keuse van 'n blokgrötte. In ondersoekte waar weerstoestande 'n belangrike rol speel, is een jaar meestal 'n sinvolle keuse. Indien daaglikse maksimum temperature byvoorbeeld ondersoek word, sou 'n opdeling van die data in kwartale onvanpas wees. Die somertemperature is tipies heelwat hoër as die temperature in die ander seisoene sodat hierdie opdeling die aanname dat die maksima onafhanklik en identies verdeel is, sou verbreek. Inferensie gebaseer op sulke blokmaksima sal nie betroubaar wees nie.

2.5.2 Maksimum aanneemlikheid beraming

Die metode van maksimum aanneemlikheid is een van die mees konvensionele metodes in parameterberaming en kan ook in die ekstreemteorie aangewend word om die parameters μ , σ en γ te beraam. Die prosedure behels die maksimering van die logaritme van die aanneemlikheidsfunksie

$$l(\underline{\theta}; \underline{y}) = \log \left\{ \prod_{i=1}^m g_{\underline{\theta}}(M_n^{(i)}) I \left\{ 1 + \gamma \left(\frac{M_n^{(i)} - \mu}{\sigma} \right) > 0 \right\} \right\}. \quad (2.12)$$

met betrekking tot die vektor $\underline{\theta}$ ten einde die maksimum aanneemlikheids (MA) beramer

$$\hat{\underline{\theta}}' = (\hat{\gamma}, \hat{\mu}, \hat{\sigma})$$

te verkry.

'n Potensiële probleem gekoppel aan die gebruik van maksimum aanneemlikheid vir die ekstreemverdeling G_γ het te make met die reëlmatigheidsvoorwaardes wat vereis word voordat die gewone asimptotiese eienskappe van die maksimum aanneemlikheids beramers geld. In die ekstreemwaarde situasie word daar nie aan al hierdie voorwaardes voldoen nie sodat die asimptotiese eienskappe nie outomaties geld nie. Smith (1985), volgens Coles (2001), het die volgende resultate in hierdie verband gevind:

- indien $\gamma > -0.5$, word die reëlmatigheidsvoorwaardes bevredig; MA beramers besit die asimptotiese eienskappe
- indien $-1 < \gamma < -0.5$, besit die MA beramers nie die asimptotiese eienskappe nie, maar kan in die algemeen verkry word
- indien $\gamma < -1$, is dit onwaarskynlik dat die MA beramers verkry sal kan word.

Die geval $\gamma \leq -0.5$ stem ooreen met 'n verdeling met 'n baie kort boonste stert. Hierdie situasie word selde in die modellering van ekstreemwaardes aangetref sodat die teoretiese beperkings van die MA beramers ons gewoonlik nie in die praktyk behoort te belemmer nie.

Daar is geen analitiese oplossing vir die maksimeringsprobleem nie, maar die maksimering kan maklik uitgevoer word met behulp van standaard numeriese optimerings algoritmes. Daar moet egter omsigtig te werk gegaan word ten einde te verseker dat die waardes $\hat{\gamma}$, $\hat{\mu}$ en $\hat{\sigma}$ wel aan die reëlmatigheidsvoorwaardes voldoen. Indien hierdie voorwaardes bevredig word, besit die maksimum aanneemlikheids beramer asimptotiese eienskappe wat die basis vorm vir inferensie ten opsigte van die onbekende parametervektor $\underline{\theta}$.

Stelling 2.5.1

Veronderstel $\underline{\theta}$ is 'n onbekende parametervektor met MA beramer $\hat{\underline{\theta}}$. Dan, vir $m \rightarrow \infty$, geld dat

$$\sqrt{m} \cdot (\hat{\underline{\theta}} - \underline{\theta}) \xrightarrow{D} N(0, I(\underline{\theta})^{-1}), \quad \gamma > -0.5, \quad (2.13)$$

waar $I(\underline{\theta})^{-1}$ die inverse van die Fisher informasiematriks is, met

$$I(\underline{\theta})_{jk} = E_{\underline{\theta}} \left\{ -\frac{\partial^2}{\partial \theta_j \partial \theta_k} l(\underline{\theta}; Y) \right\}.$$

Δ

Hierdie Fisher informasiematriks is onbekend, dus benader ons $I(\underline{\theta})$ met die waargenome Fisher informasiematriks, $I(\hat{\underline{\theta}})$, geëvalueer by $\underline{\theta} = \hat{\underline{\theta}}$:

$$I(\hat{\underline{\theta}})_{jk} = -\frac{\partial^2}{\partial \theta_j \partial \theta_k} l(\underline{\theta}; y) \Big|_{\underline{\theta} = \hat{\underline{\theta}}}.$$



Vertrouensintervalle vir die onbekende $\underline{\theta}$ kan gevind word deur van die asimptotiese meerveranderlike normaalverdeling van die maksimum aanneemlikheids beramer $\hat{\underline{\theta}}$ gebruik te maak. Ons weet dat vir groot m , is $\hat{\theta}_j$ by benadering $N(\theta_j, \frac{I(\underline{\theta})_{jj}^{-1}}{m})$ verdeel. Hieruit volg by benadering dat

$$P\left\{ -z_{1-\alpha/2} \leq \frac{\hat{\theta}_j - \theta_j}{\sqrt{\frac{I(\underline{\theta})_{jj}^{-1}}{m}}} \leq z_{1-\alpha/2} \right\} \approx 1 - \alpha.$$

Ons verkry dus die volgende benaderde $(1 - \alpha)100\%$ vertrouensinterval vir $\hat{\theta}_j$:

$$\text{Vert}_{(1-\alpha)100\%}(\theta_j) = \hat{\theta}_j \pm z_{1-\alpha/2} \sqrt{\frac{I(\hat{\underline{\theta}})_{jj}^{-1}}{m}} \quad (2.14)$$

Indien MA beramings van die parameters μ, σ en γ gevind kan word, kan die kwantiele van die verdeling $G_{\mu, \sigma, \gamma}$ van die maksima beraam word. Ons stel in die statistiek van ekstreme waardes in die besonder in stertkwantiele belang. Een so 'n ekstreme kwantiel is R_k , die k -blok terugkeervlak ("return level"), gedefinieer deur

$$P(M_n > R_k) = \frac{1}{k}. \quad (2.15)$$

R_k is dus die kwantielwaarde sodanig dat dit slegs 1 keer per k blokke oorskry word, waar k tipies groot is.

Uit die definisie van R_k volg dat

$$P(M_n \leq R_k) = F^n(R_k) = 1 - \frac{1}{k}$$

Die verdeling F is egter onbekend, dus kan ons nie hieruit oplos vir R_k nie. Gevolglik moet R_k beraam word. Die benadering

$$F^n \approx G_{\mu, \sigma, \gamma}$$

geld indien n groot genoeg is, dus volg by benadering vir $\gamma \neq 0$:

$$G_{\mu, \sigma, \gamma}(R_k) \approx 1 - \frac{1}{k}$$



oftewel

$$\begin{aligned} R_k &\approx G_{\mu, \sigma, \gamma}^{-1}\left(1 - \frac{1}{k}\right) \\ &= \mu + \frac{\sigma \{(-\log(1 - \frac{1}{k}))^{-\gamma} - 1\}}{\gamma}. \end{aligned} \quad (2.16)$$

Indien $\gamma = 0$, dan is

$$\begin{aligned} R_k &\approx G_{\mu, \sigma, \gamma}^{-1}\left(1 - \frac{1}{k}\right) \\ &= \mu - \sigma \log\left\{-\log\left(1 - \frac{1}{k}\right)\right\}. \end{aligned} \quad (2.17)$$

Aangesien die werklike waarde van θ onbekend is, is die kwantiele R_k ook

onbekend. R_k kan as volg beraam word met behulp van die maksimum aanneemlikheids beramer $\hat{\theta}$:

$$\hat{R}_k = \hat{\mu} + \frac{\hat{\sigma}\{(-\log(1-\frac{1}{k}))^{-\hat{\gamma}} - 1\}}{\hat{\gamma}}.$$

Indien ons die volgende herparametrisering uitvoer,

$$\mu = R_k - \frac{\sigma\{(-\log(1-\frac{1}{k}))^{-\gamma} - 1\}}{\gamma}, \quad (2.18)$$

dan word die ekstreemwaarde-model uitgedruk in terme van die parameters R_k , σ en γ . Die asimptotiese normaliteit van die maksimum aanneemlikheids beramers kan nou aangewend word om benaderde vertrouensintervalle vir R_k te konstrueer.

2.5.3 Inferensie met behulp van die profiel aanneemlikheidsfunksie

'n Alternatiewe metode om die onsekerheid van die MA beramers te kwantifiseer, is om dit in terme van die deviansie funksie, gegee deur

$$D(\theta) = 2\{l(\hat{\theta}) - l(\theta)\},$$

te evalueer. Modelle met 'n lae deviansie is modelle met 'n hoë mate van aanneemlikheid in die sin dat hulle waardes van θ "naby" aan die waarde $\hat{\theta}$ is en gevolglik is 'n vertrouensgebied van die vorm

$$C = \{\theta : D(\theta) \leq c\}$$

van toepassing. Indien die verdeling van $D(\theta)$ bekend was, sou dit moontlik gewees het om c op so 'n wyse te kies dat die gebied C 'n vooraf-gespesifiseerde waarskynlikheid het om die werklike waarde van die parameter in te sluit. Hierdie verdeling van $D(\theta)$ is nie eksak bekend nie, maar die volgende asimptotiese resultaat geld wel:

Stelling 2.5.2

Veronderstel $\underline{\theta}$ is 'n onbekende parametervektor van dimensie d met MA beramer $\hat{\underline{\theta}}$. Dan, vir n groot genoeg en onder sekere reëlmatigheidsvoorwaardes, geld by benadering dat

$$D(\underline{\theta}) \sim \chi_d^2.$$

Δ

Dit volg uit stelling 2.5.2 dat 'n benaderde vertrouensgebied vir $\underline{\theta}$ as volg verkry kan word:

$$C_\alpha = \{\underline{\theta} : D(\underline{\theta}) \leq c_{1-\alpha}\},$$

waar $c_{1-\alpha}$ die $100(1-\alpha)$ -de kwantiel van die χ_d^2 -verdeling is.

Volgens Coles (2001) is hierdie benadering normaalweg meer akkuraat as die beraming gebaseer op die benaderde normaliteit van die MA beramer. Die χ^2 -benadering is egter aansienlik meer berekeningsintensief.

Indien inferensie oor 'n spesifieke lid van die parametervektor verlang word, kan die asimptotiese normaliteit volgens stelling 2.5.1 aangewend word. 'n Alternatief hiervoor is om gebruik te maak van die sogenaamde profiel log-aanneemlikheidsfunksie.

Definisie 2.5.1

Veronderstel $\underline{\theta}$ is 'n onbekende parametervektor. Die log-aanneemlikheid van $\underline{\theta}$ kan geskryf word as $l(\theta_i, \underline{\theta}_{-i})$, waar $\underline{\theta}_{-i}$ die vektor $\underline{\theta}$ aandui met θ_i uitgelaat. Die profiel log-aanneemlikheidsfunksie vir θ_i word gedefinieer as

$$l_p(\theta_i) = \underset{\underline{\theta}_{-i}}{\text{maks}} l(\theta_i, \underline{\theta}_{-i}).$$

Die log-aanneemlikheidsfunksie word dus vir elke waarde van θ_i gemaksimeer met betrekking tot al die ander komponente ($\underline{\theta}_{-i}$) van die parametervektor; dit vorm die profiel log-aanneemlikheidsfunksie van θ_i .

Δ

Definisie 2.5.1 kan as volg veralgemeen word vir enige partisie van $\underline{\theta}$ in twee komponente:

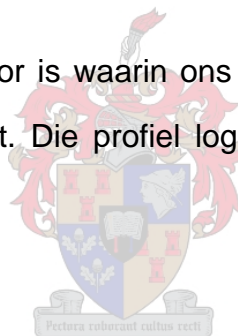
Definisie 2.5.2

Veronderstel die onbekende parametervektor $\underline{\theta}$ van dimensie d word as volg opgedeel in twee nie-oortrukkende dele $\underline{\theta}_1$ en $\underline{\theta}_2$:

$$\underline{\theta} = \begin{bmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \end{bmatrix},$$

waar $\underline{\theta}_1$ die k -dimensionele vektor is waarin ons belangstel en $\underline{\theta}_2$ die oorblywende $(d - k)$ komponente van $\underline{\theta}$ bevat. Die profiel log-aanneemlikheid van $\underline{\theta}_1$ word dan gedefinieer as

$$l_p(\underline{\theta}_1) = \underset{\underline{\theta}_2}{\text{maks}} l(\underline{\theta}_1, \underline{\theta}_2).$$



Indien $k = 1$, reduceer hierdie tot definisie 2.5.1.

Δ

Soortgelyk aan die wyse waarop definisie 2.5.2 'n veralgemening is van definisie 2.5.1, word stelling 2.5.3 verkry as 'n veralgemening van stelling 2.5.2.

Stelling 2.5.3

Veronderstel $\underline{\theta}$ is 'n onbekende parametervektor van dimensie d met MA beramer $\hat{\underline{\theta}}$. $\underline{\theta}$ word as volg opgedeel in twee nie-oortrukkende dele $\underline{\theta}_1$ en $\underline{\theta}_2$:

$$\underline{\theta} = \begin{bmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \end{bmatrix},$$

waar $\underline{\theta}_1$ 'n k -dimensionele vektor is en $\underline{\theta}_2$ van dimensie $(d - k)$ is. Dan, vir m groot genoeg en onder sekere reëlmatigheidsvoorwaardes, geld by benadering dat

$$D_p(\underline{\theta}_1) \equiv 2\{l(\hat{\underline{\theta}}) - l_p(\underline{\theta}_1)\} \sim \chi_k^2$$

Δ

Stelling 2.5.3 kan in die volgende twee situasies aangewend word. Ten eerste het ons vir 'n skalaarkomponent θ_i dat

$$C_\alpha = \{\theta_i : D_p(\theta_i) \leq c_{1-\alpha}\}$$

'n benaderde $(1 - \alpha)100\%$ vertrouensinterval is, waar $c_{1-\alpha}$ die $(1 - \alpha)$ -de kwantiel van die χ_k^2 -verdeling is.

Die tweede toepassing het te make met modelseleksie. Veronderstel $M_{\underline{\theta}}$ is 'n model met die parametervektor $\underline{\theta}$ en model $M_{\underline{\theta}_2}$ is die model wat verkry word deur k komponente van $\underline{\theta}$ te beperk tot sê nul. 'n Partisie van $\underline{\theta}$ soos in stelling 2.5.3 is dus ter sprake waar die eerste komponent $\underline{\theta}_1$ van dimensie k nul is in model $M_{\underline{\theta}_2}$. Laat $l_{\underline{\theta}}(M_{\underline{\theta}})$ die gemaksimeerde log-aanneemlikheid vir model $M_{\underline{\theta}}$ en $l_{\underline{\theta}_2}(M_{\underline{\theta}_2})$ die gemaksimeerde log-aanneemlikheid vir model $M_{\underline{\theta}_2}$ wees. Definieer die deviansie vir hierdie situasie as

$$D = 2\{l_{\underline{\theta}}(M_{\underline{\theta}}) - l_{\underline{\theta}_2}(M_{\underline{\theta}_2})\}.$$

Om te toets of $M_{\underline{\theta}_2}$ 'n aanvaarbare reduksie van model $M_{\underline{\theta}}$ verteenwoordig, is dit voldoende om te ondersoek of die berekende waarde van $D < c_{1-\alpha}$, waar $c_{1-\alpha}$ die $(1 - \alpha)$ -de kwantiel van die χ_k^2 -verdeling is. Hierdie word geformaliseer as stelling 2.5.4.

Stelling 2.5.4

Veronderstel $M_{\underline{\theta}_2}$ met parameter $\underline{\theta}_2$ is die submodel van $M_{\underline{\theta}}$ met parameter

$$\underline{\theta} = \begin{bmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \end{bmatrix}$$

onder die beperking dat $\underline{\theta}_1 = \underline{0}$. Laat $l_{\underline{\theta}}(M_{\underline{\theta}})$ en $l_{\underline{\theta}_2}(M_{\underline{\theta}_2})$ die gemaksimeerde log-aanneemlikheid vir onderskeidelik model $M_{\underline{\theta}}$ en $M_{\underline{\theta}_2}$ wees. Ons verwerp model $M_{\underline{\theta}_2}$ ten gunste van $M_{\underline{\theta}}$ indien

$$D = 2\{l_{\underline{\theta}}(M_{\underline{\theta}}) - l_{\underline{\theta}_2}(M_{\underline{\theta}_2})\} > c_{1-\alpha},$$

waar $c_{1-\alpha}$ die $(1-\alpha)$ -de kwantiel van die χ_k^2 -verdeling is.

△

2.5.4 Evaluering van die gepaste model

Die evaluering van enige gepaste model is 'n tweeledige taak. Ten eerste behels dit 'n bepaling van hoe goed die model die data waarop dit gemodelleer is, pas. Hiervoor bespreek ons in hierdie paragraaf 'n aantal grafiese metodes om die geldigheid van 'n veralgemeende ekstreemwaarde-model te bepaal. Tweedens stel die statistikus belang in die vermoë van die model om na die toekoms te ekstrapoleer. Hierdie gehalte van ekstrapolasie is aansienlik moeiliker om te peil. 'n Goeie passing van die model op die leerdata is nie 'n voldoende voorwaarde om akkurate ekstrapolasie te verseker nie, maar dit is wel 'n nodige voorwaarde. Die tegnieke wat hier volg, moet in hierdie konteks beskou word.

Die waarskynlikheidsstipping vergelyk die gepaste verdelingsfunksie met die empiriese verdelingsfunksie. Orden die blokmaksima-data $M_n^{(1)}, \dots, M_n^{(m)}$ as volg: $M_{(1)} \leq M_{(2)} \leq \dots \leq M_{(m)}$. Hierdie $M_{(1)}, M_{(2)}, \dots, M_{(m)}$ verteenwoordig dus die rangorde statistieke van die maksima-data. Die empiriese verdelingsfunksie in die punt $M_{(i)}$ word gegee deur

$$\tilde{G}(M_{(i)}) = \frac{i}{m}, \quad i = 1, 2, \dots, m$$

Die ooreenstemmende waarskynlikheidsberamings van die gepaste model is

$$\hat{G}(M_{(i)}) = \exp\{-[1 + \hat{\gamma}(\frac{M_{(i)} - \hat{\mu}}{\hat{\sigma}})]^{\frac{1}{\hat{\gamma}}}\}, \quad i = 1, 2, \dots, m$$

Indien die model die data goed pas, dan is

$$\hat{G}(M_{(i)}) \approx \tilde{G}(M_{(i)})$$

vir alle waardes van i , sodat 'n waarskynlikheidstipping bestaande uit die punte

$$\{(\tilde{G}(M_{(i)}), \hat{G}(M_{(i)})), \quad i = 1, 2, \dots, m\},$$

naby aan 'n reguitlyn deur die oorsprong behoort te lê. Enige wesentliche afwykings van lineariteit in die waarskynlikheidstipping dui op tekortkominge in die gepaste ekstreem-model.

'n Beperking van die waarskynlikheidstipping is dat beide $\tilde{G}(M_{(i)})$ en $\hat{G}(M_{(i)})$ na 1 konvergeer as $M_{(i)}$ toeneem. Ons stel juis belang in die akkuraatheid van die model vir groot waardes van die $M_{(i)}$ sodat die waarskynlikheidstipping in die area wat van die grootste belang is, die minste inligting bied. Hierdie tekortkoming word vermy deur van die kwantielstipping gebruik te maak. Die kwantielstipping bestaan uit die punte

$$\{(\hat{G}^{-1}(\frac{i}{m}), M_{(i)}), \quad i = 1, 2, \dots, m\}.$$

Die kwantiele soos deur die gepaste model beraam, word dus teenoor die werklike kwantiele gestip. Hier is

$$\hat{G}^{-1}(\frac{i}{m}) = \hat{\mu} + \frac{\hat{\sigma}\{(-\log(\frac{i}{m}))^{-\hat{\gamma}} - 1\}}{\hat{\gamma}}.$$

Soos by die waarskynlikheidstipping dui nie-lineariteit in die kwantielstipping op tekortkominge in die gepaste model.

Die terugkeervlak-stipping behels 'n stipping van die terugkeervlak R_k (sien vergelyking 2.15) teenoor die terugkeerperiode k . Stel $w_k = -\log(1 - \frac{1}{k})$ in vergelyking (2.16) en (2.17). Dan is

$$R_k = \begin{cases} \mu - \frac{\sigma}{\gamma} [1 - w_k^{-\gamma}] & \text{vir } \gamma \neq 0 \\ \mu - \sigma \log w_k & \text{vir } \gamma = 0 \end{cases} \quad (2.18)$$

Dit volg dat indien R_k op 'n logaritmiëse skaal teenoor w_k gestip word, oftewel R_k teenoor $\log w_k$, dan is die stipping lineêr vir $\gamma = 0$. Indien $\gamma > 0$ is die stipping konveks na onder en vir $\gamma < 0$ is die stipping konkaf na onder. Hierdie terugkeervlak-stipping suggereer dus of die werklike waarde van die vormparameter γ positief, negatief of naastenby nul is. Dit is gebruiklik om ook die bo- en die ondergrense van die vertrouensintervalle vir R_k by die terugkeervlak-stipping te voeg. Sien byvoorbeeld figuur 2.6.5. en 3.6.5.

Ten slotte kan 'n histogram van die maksima met die digtheidsfunksie van die gepaste model vergelyk word. Hier dui drastiese afwykings tussen die histogram en die beraamde digtheidsfunksie op 'n onvoldoende modelpassing.

2.6 Voorbeeld



In hierdie paragraaf word die teoretiese aspekte van die modellering van maksima as ekstreme waardes, soos behandel in paragraaf 2.4, geïllustreer aan die hand van 'n praktiese voorbeeld. Die datastel ter sprake bestaan uit die jaarlikse maksimum seevlak soos waargeneem by Port Pirie, net noord van Adelaide in Suid-Australië, gedurende die periode 1923-1987.

In die geval van die Port Pirie data mag dit van belang wees om te beraam watter maksimum seevlakke oor die volgende 10 of selfs 100 jaar in hierdie streek ondervind mag word.



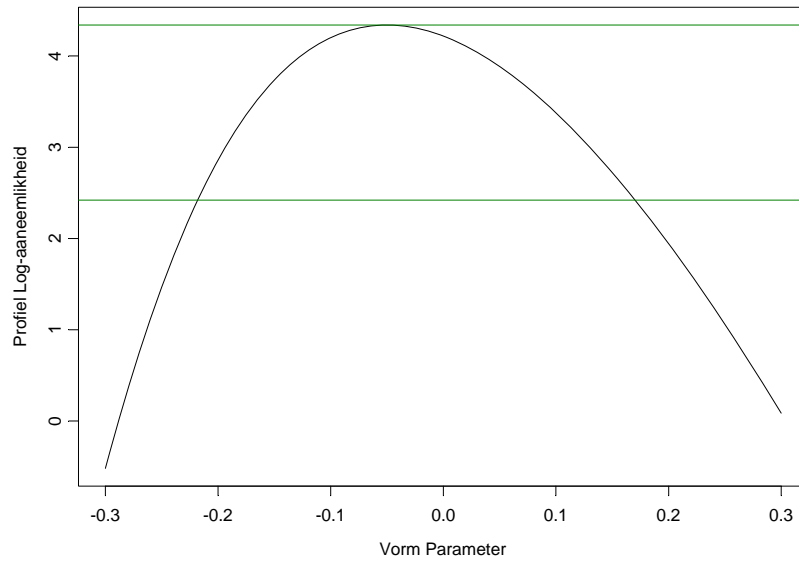
Figuur 2.6.1. Jaarlikse maksimum seevlakke by Port Pirie, Suid-Australië.

Uit figuur 2.6.1 wil dit voorkom of die variasie in die data rofweg konstant gebly het oor die periode onder beskouing, sodat ons met vertroue die datapunte by benadering kan beskou as onafhanklike waarnemings uit die veralgemeende ekstreemwaarde verdeling. Maksimering van die log-aanneemlikheidsfunksie (2.10) met behulp van S-Plus lei tot die volgende MA-beramings vir $\underline{\theta}$:

$$\hat{\underline{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\gamma}) = (3.87, 0.198, -0.050),$$

met standaardfoute 0.028, 0.020 en 0.098 vir μ , σ en γ onderskeidelik. Die maksimum van die log-aanneemlikheidsfunksie onder $\hat{\underline{\theta}}$ is -4.34, met drie vryheidsgrade.

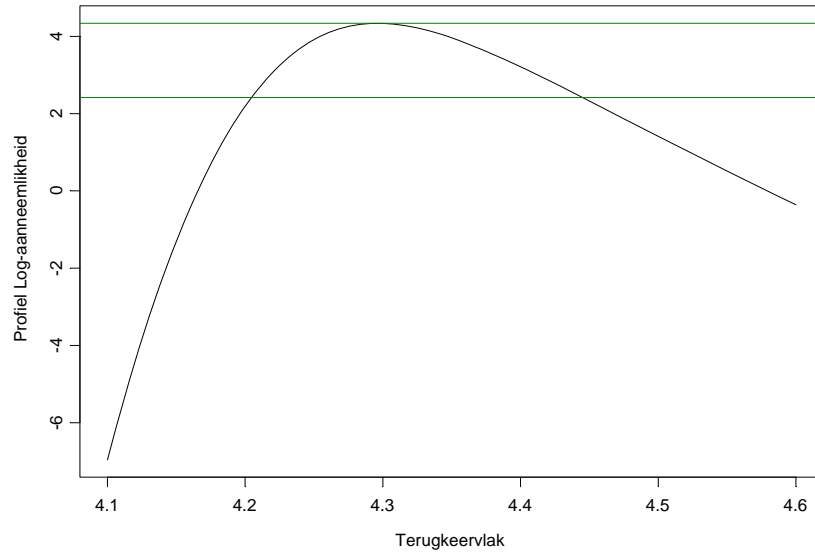
Deur gebruik te maak van die benaderde normaliteit van die MA-beramers (vergelykings 2.11 en 2.12) volg benaderde 95%-vertrouensintervalle as [3.82, 3.93] vir μ , [0.158, 0.238] vir σ en [-0.242, 0.142] vir γ . 'n Alternatiewe benadering om benaderde vertrouensintervalle vir die parameters μ , σ en γ te vind, is om van die profiel-aanneemlikheidsfunksie gebruik te maak. In figuur 2.6.2 word 'n benaderde 95% vertrouensinterval vir γ op hierdie manier gevind:



Figuur 2.6.2. Profiel aanneemlikheid vir γ in die Port Pirie seevlak voorbeeld.

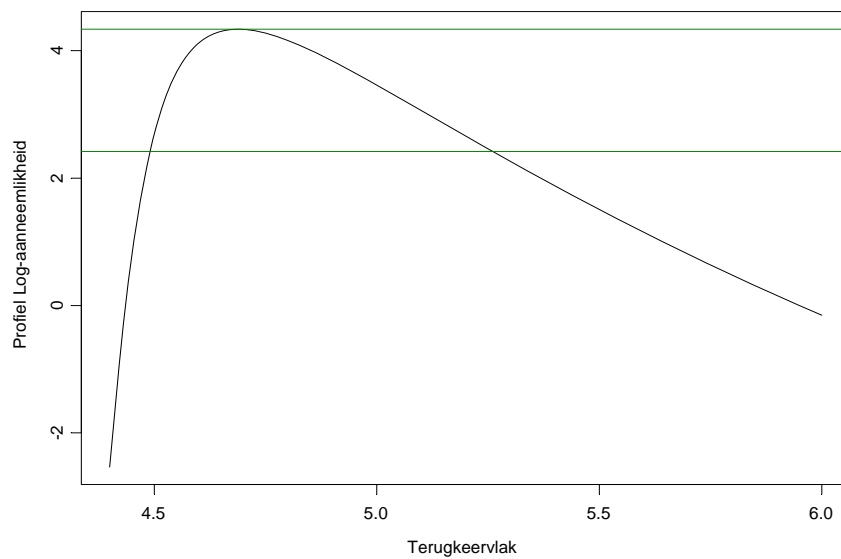
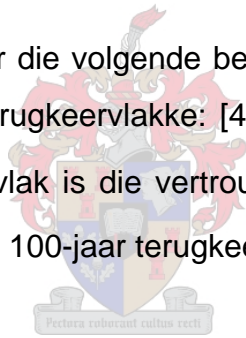
Figuur 2.6.2 vertoon die profiel aanneemlikheid vir γ . Dit lei tot 'n benaderde 95% vertrouensinterval vir γ , naamlik $[-0.21, 0.17]$. Hierdie vertrouensinterval is slegs effens verskillend van die vorige (maksimum aanneemlikheids) vertrouensinterval vir γ .

Soos verduidelik in paragraaf 2.5.2 (vergelyking 2.16) kan die veralgemeende ekstreemwaarde model ook uitgedruk word in terme van die k -blok terugkeervlak R_k , soos gedefinieer in (2.13). In die Port Pirie datastel werk ons natuurlik met blokke van een jaar elk. Veronderstel ons stel belang in die beraming van R_{10} en R_{100} , onderskeidelik die 10-jaar en 100-jaar terugkeervlakke. Die metode van maksimum aanneemlikheid lewer die volgende puntberamings, naamlik $\hat{R}_{10} = 4.30$ en $\hat{R}_{100} = 4.69$. Benaderde 95% vertrouensintervalle gebaseer op die asimptotiese normaliteit van die MA-beramers is $[4.19, 4.41]$ vir R_{10} en $[4.38, 5.00]$ vir R_{100} . Die metode van profiel aanneemlikheid kan ook aangewend word om benaderde vertrouensintervalle vir R_{10} en R_{100} af te lei:



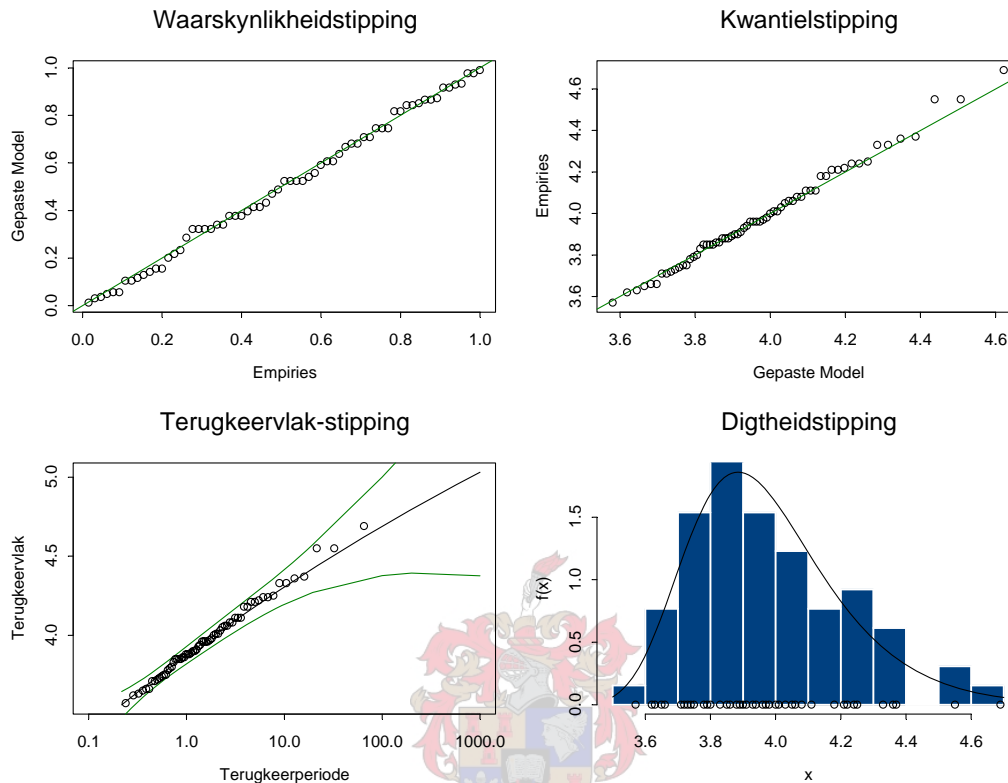
Figuur 2.6.3. Profiel aanneemlikheid vir die 10-jaar terugkeervlak R_{10} in die Port Pirie seevlak voorbeeld.

Figuur 2.6.3 en figuur 2.6.4 lewer die volgende benaderde 95% vertrouensintervalle vir die 10-jaar en die 100-jaar terugkeervlakke: $[4.12, 4.45]$ vir R_{10} en $[4.50, 5.27]$ vir R_{100} . Vir die 10-jaar terugkeervlak is die vertrouensintervalle soortgelyk vir beide metodes. Dit geld egter nie vir die 100-jaar terugkeervlak nie.



Figuur 2.6.4. Profiel aanneemlikheid vir die 100-jaar terugkeervlak R_{100} in die Port Pirie seevlak voorbeeld.

Die volgende grafika (soos bespreek in paragraaf 2.5.4) is diagnostiese maatstawwe wat gebruik word om te toets of die gepaste ekstreemwaarde-model die data goed pas.



Figuur 2.6.5. Diagnostiese maatstawwe vir die passing van `n veralgemeende ekstreemwaarde model op die Port Pirie seevlak data.

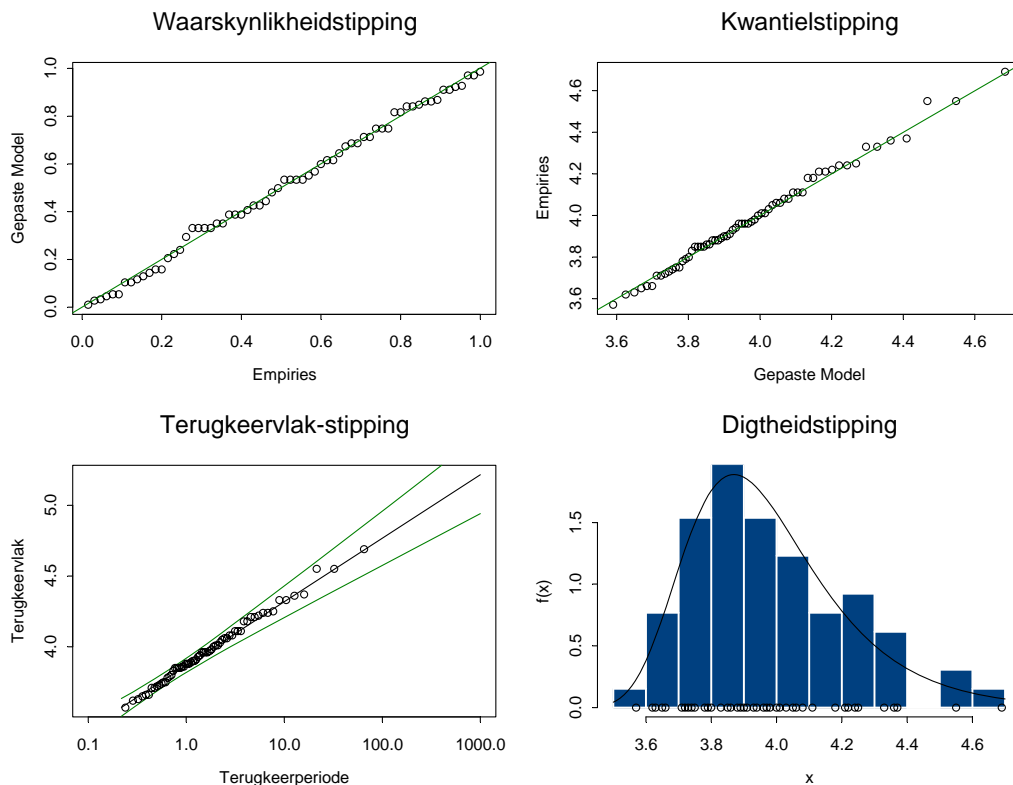
Die kwantielstipping sowel as die waarskynlikheidstipping verloop rofweg lineêr en bied dus geen rede om die gepaste model te verwerp nie. Die terugkeervlak-stipping is ook naastebly lineêr. Dit dui daarop dat die werklike waarde van γ naby aan nul is. Hierdie is konsekwent met ons MA-beraming $\hat{\gamma} = -0.050$. Uit die digtheidstipping blyk ook dat die beraamde digtheid nie veel van die waargenome histogram verskil nie. Al vier die diagnostiese stippings verleen dus getuienis wat die gepaste model ondersteun.

Die MA-beraming en die vertrouensintervalle van γ , sowel as die terugkeervlakstipping suggereer dat die waarde van γ moontlik nul is. Dit laat die vraag ontstaan of die Gumbel-verdeling (opmerking 2.3.1), daardie lid van die veralgemeende ekstreemwaarde-familie met $\gamma = 0$, nie die data ongeveer net so goed as die gepaste

model sal pas nie. Maksimering van die log-aanneemlikheidsfunksie (2.10), met die beperking dat $\gamma = 0$, lei daartoe dat MA-beramings van slegs die skuif- en skaalparameters gevind word. Ons verkry die volgende MA-beramings vir die skuif- en skaalparameter: $(\hat{\mu}, \hat{\sigma}) = (3.87, 0.195)$. Die standaardfoute is 0.03 vir μ en 0.019 vir σ . Die volgende benaderde 95% vertrouensintervalle geld: $[3.82, 3.92]$ vir μ en $[0.158, 0.232]$ vir σ . Die maksimum van die beperkte log-aanneemlikheidsfunksie is -4.22, met twee vryheidsgrade.

Die gebruikelike diagnostiese maatstawwe is gestip om te toets of die Gumbel-verdeling die data goed pas. Soos verwag bied die diagnostiese stippings nie veel rede om die Gumbel-model te verwerp nie. Hier ondersoek ons egter nie soseer of die Gumbel-verdeling die data goed pas nie, maar wel hoeveel swakker dit (as beperkte model) as die onbeperkte veralgemeende ekstreemwaarde verdeling vaar. Die aanneemlikheidsverhouding toetsstatistiek vir die dimensie-reduksie is in stelling 2.4.4 afgelei as $D = 2\{-4.22 - (-4.34)\} = 0.24$. Hierdie waarde is klein vergeleke met die χ_1^2 -verdeling. Die pasgehalte van die gereduseerde Gumbel-model is dus vergelykbaar met dié van die volle veralgemeende ekstreemwaarde model.

Volgens die diagnostiese stippings pas albei modelle die data redelik goed. Die aanneemlikheidsverhoudings toets toon ook dat die pasgehalte van die Gumbel-model nie betekenisvol swakker is as dié van die onbeperkte model nie. Watter model behoort dus gebruik te word? Die grootste verskil tussen die twee modelle is dat die vertrouensintervalle vir die modelparameters en terugkeervlak onder die volle model aansienlik wyer is as onder die Gumbel-model. Beskou byvoorbeeld die onderskeie terugkeervlak-stippings in figure 2.6.5 en 2.6.6: Die terugkeervlak-kurwes is soortgelyk, maar by veral hoë terugkeerperiodes is die vertrouensintervalle van die veralgemeende model heelwat wyer. Nouer vertrouensintervalle impliseer 'n laer mate van onsekerheid, sodat die Gumbel-model verkies behoort te word - mits die model vertrou kan word. Volgens Coles (2001) is die veiligste uitweg moontlik om te aanvaar dat daar inderdaard 'n mate van onsekerheid rakende die waarde van γ - en dus ook die korrektheid van die Gumbel-model - is en dat inferensie gebaseer op die onbeperkte model, verkies word.



Figuur 2.6.6. Diagnostiese maatstawwe vir die passing van 'n Gumbel-verdeling op die Port Pirie seevlak data.

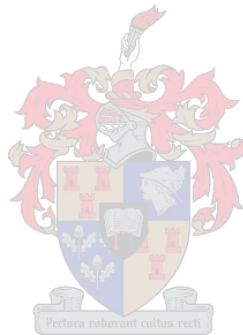
2.7 Gevolgtrekking



In hierdie hoofstuk het ons 'n limietverdeling afgelei waarna die verdeling van maksima, genormeer met geskikte konstantes, konvergeer as die steekproefgrootte na oneindig neig, mits die onderliggende verdeling aan sekere voorwaardes voldoen. Volgens die Fisher-Tippett stelling is die enigste sodanige limietverdeling die klas G_γ , oftewel $G_{\gamma, \mu, \sigma}$, waar μ 'n skuifparameter en σ 'n skaalparameter is. Onder Fisher-Tippett reduceer inferensie dus tot die beraming van slegs drie parameters: γ, μ en σ . Die metode van maksimum aanneemlikheid is aangewend om hierdie parameters te beraam. Die asimptotiese normaliteit van die MA beramers sowel as die profiel aanneemlikheid kan gebruik word om benaderde vertrouensintervalle vir die bogenoemde parameters, ekstreme kwantiele en terugkeervlakke te konstrueer.

Die volgende hoofstuk behels 'n teoretiese bespreking van drempelmodellering, 'n

meer moderne benadering tot die ekstreemwaarde-probleem as die maksimummetode. Uit die Fisher-Tippett stelling volg 'n unieke limietverdeling $G_{\gamma, \mu, \sigma}$ vir die steekproef maksimum indien sekere voorwaardes bevredig word. In hoofstuk vier word die volgende vraag ondersoek: Hoe goed is die benadering $F^n \approx G_{\gamma, \mu, \sigma}$ vir eindige n ? Die beraming van stertkwantiele en vertrouensintervalle vir stertkwantiele is een van die vernaamste doelwitte van 'n ekstreemwaarde-analise. Om hierdie rede word die akkuraatheid van maksimumgebaseerde inferensie ten opsigte van die onderliggende verdeling F in terme van die akkuraatheid van hierdie beramings beoordeel.



Hoofstuk 3: Drempelmodelle

3.1 Inleiding

In Hoofstuk 2 het die klem op die limietgedrag van maksima geval. Die beskikbare steekproef is in blokke opgedeel en die grootste waarneming in elke blok is gevind. Die volgende stap was om 'n ekstreemwaarde-model op hierdie reeks maksima te pas. 'n Potensiële tekortkoming van hierdie benadering is dat elke blok deur slegs 'n enkele waarneming, die blokmaksimum, verteenwoordig word. Indien selfs net een blok meer as 'n enkele ekstreme waarde bevat, dan word inligting vermors deur slegs die heel grootste waarneming in elke blok te beskou.

'n Alternatiewe benadering is om heeltemal met die gedagte van blok-opdeling weg te doen. Beskou eerder die volle stel steekproefdata as 'n geheel en definieer ekstreme waarnemings as daardie waarnemings in die steekproef wat 'n sekere hoë drempel oorskry. Vervolgens word 'n asimptotiese model op hierdie oorskrydings gepas. Dié benadering behels dus die bestudering van die stert van die steekproef, oftewel alle waarnemings regs van 'n sekere drempel u . 'n Onderskeid word getref tussen absolute oorskrydings van die vorm $(X - u)$ en relatiewe oorskrydings van die vorm $\frac{X}{u}$.

Paragraaf 3.2 word gewy aan 'n aantal belangrike definisies wat telkens in hoofstuk 3 gebruik word. In paragraaf 3.3 en 3.4 word absolute drempeloorskrydings ondersoek. Die Pickands-Balkema-De Haan stelling bewys dat die verdeling F_u van absolute oorskrydings van 'n drempel u na die veralgemeende Pareto-verdeling $H_{\gamma,\beta}$ konvergeer as $u \rightarrow \infty$, as en slegs F tot die maksimum domein van aantrekking van die veralgemeende ekstreemwaarde-verdeling G_γ behoort. Hierdie verband tussen die veralgemeende Pareto-verdeling en die veralgemeende ekstreemwaarde-verdeling word ook verder toegelig. Paragraaf 3.4 handel oor die modellering

van oorskrydings van die vorm $(X - u)$. Metodes om 'n geskikte drempel te kies word in paragraaf 3.4.1 bespreek. Die afruiling tussen sydigheid en variansie wat ontstaan by die keuse van 'n optimale drempel word oorweeg. In 3.4.2 word die metode van maksimum aanneemlikheid (MA) beskou en word die MA beramers van die parameters γ en β van die veralgemeende Pareto-verdeling $H_{\gamma,\beta}$ gevind. Hierdie resultate word in 3.4.3 aangewend in die beraming van ekstreme kwantiele en terugkeervlakke. Ten slotte word 'n aantal diagnostiese maatstawwe in 3.4.4 aangebied ten einde die pasgehalte van die veralgemeende Pareto-verdeling op die absolute oorskrydings te evalueer. Paragraaf 3.7 word gewy aan 'n praktiese voorbeeld van die modellering van absolute oorskrydings van 'n hoë drempel. Die datastel wat ondersoek word, is 'n tydreeks van 17531 metings van die daaglikse reënval in die suidweste van Engeland vanaf 1914 tot 1962.

Paragraaf 3.5 en 3.6 behandel oorskrydings van die vorm $\frac{X}{u}$. Volgens stelling 3.5.1 konvergeer die verdeling $F_{(u)}$ van relatiewe oorskrydings na die streng Pareto-verdeling as $u \rightarrow \infty$, mits die stert-verdelingsfunksie \bar{F} tot die klas van reëlmatig-variërende funksies behoort. Hierdie familie van verdelings waarvan die stert-verdelingsfunksie reëlmatig-variërend is, staan bekend as die swaarstert verdelings. In paragraaf 3.6 word die modellering van relatiewe oorskrydings ondersoek. Paragraaf 3.6.1 bespreek die alternatiewe metodologie om die aantal oorskryding vas te kies eerder as die drempelwaarde u . Dit word gedoen deur u gelyk te neem aan die $(k + 1)$ -de grootste rangorde statistiek van die steekproef. Die MA beramer van die α -parameter van die Pareto-verdeling word in 3.6.2 afgelei, beramers vir ekstreme kwantiele word in 3.6.3 gevind en grafiese tegnieke om die gehalte van die passing van die Pareto-verdeling op die relatiewe oorskrydings te beoordeel, word in 3.6.4 bespreek.

3.2 Definiesies en notasie

Die volgende definiesies sal gebruik word:

Definisie 3.2.1

Die stertverdelingsfunksie $\bar{F}_X(x)$ van 'n stogastiese veranderlike X met verdelingsfunksie $F_X(x)$ word gedefinieer as

$$\bar{F}_X(x) = P(X > x) = 1 - F_X(x).$$

Waar daar geen moontlikheid van verwarring bestaan nie, sal vir gerief bloot $\bar{F}(x)$ geskryf word.

Δ

Definisie 3.2.2

Laat X 'n stogastiese veranderlike met verdelingsfunksie $F(x)$ en regter eindpunt x_F wees. Vir 'n vaste $u < x_F$, word die oorskrydingsverdelingsfunksie gedefinieer as

$$F_u(y) = P(X - u < y \mid X > u), \quad y > 0.$$

Die funksie

$$e(u) = E(X - u \mid X > u)$$

staan bekend as die verwagte oorskrydings-funksie van X .

Δ

3.3 Die Pickands-Balkema-De Haan stelling

Breedweg kan die volgende onderskeid getref word ten opsigte van oorskrydings van 'n drempel u : Absolute oorskrydings is oorskrydings van die vorm $(X - u)$, terwyl relatiewe oorskrydings na oorskrydings van die vorm $\frac{X}{u}$ verwys. In hierdie paragraaf, sowel as in paragraaf 3.4, word 'n drempelmodel vir absolute oorskrydings ondersoek. Relatiewe oorskrydings word in paragraaf 3.5 en 3.6 behandel.

Ons stel belang in die volgende voorwaardelike waarskynlikheid:

$$P(X - u > y | X > u) = \frac{P(X > u + y)}{P(X > u)}, \quad y > 0,$$

oftewel

$$\bar{F}_u(y) = \frac{\bar{F}(u + y)}{\bar{F}(u)}, \quad y > 0 \quad (3.1)$$

Indien die onderliggende verdeling F bekend was, sou hierdie voorwaardelike verdeling F_u van die absolute drempeloorskrydings $(X - u)$ ook bekend gewees het. Tipies is F natuurlik onbekend. Soos in die geval van die ekstreemwaarde-verdeling vir maksima vir 'n groot genoeg waarde van die steekproefgrootte n , ondersoek ons of 'n asimptotiese model vir sulke drempeloorskrydings gevind kan word vir groot genoeg drempelwaardes. Volgens stelling 3.3.1 bied die familie van verdelings bekend as die veralgemeende Pareto-verdeling 'n natuurlike benadering vir hierdie voorwaardelike oorskrydingsverdeling. Die veralgemeende Pareto-familie word as volg gedefinieer:

Definisie 3.3.1

Die veralgemeende Pareto-verdeling is 'n twee-parameter verdeling met verdelingsfunksie

$$H_{\gamma,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}} & \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \gamma = 0 \end{cases} \quad (3.2)$$

waar $\beta > 0$. $H_{\gamma,\beta}(x)$ is gedefinieer op $x \geq 0$ as $\gamma \geq 0$ en op $0 \leq x \leq -\frac{\beta}{\gamma}$ as $\gamma < 0$.

Vir die geval $\gamma = 0$ werk ons met

$$H_{\gamma,\beta}(x) = \lim_{\gamma \rightarrow 0} \left\{ 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}} \right\} = 1 - \exp\left(-\frac{x}{\beta}\right).$$

Vervat in die veralgemeende Pareto-verdeling is:

- $\gamma > 0$: Pareto
- $\gamma = 0$: eksponensiaal
- $\gamma < 0$: Pareto tipe 2



Δ

Definisie 3.3.2

Die standaard veralgemeende Pareto-verdeling H_γ is 'n spesiale geval van die veralgemeende Pareto-verdeling met $\beta = 1$.

Δ

Stelling 3.3.1, die Pickands-Balkema-De Haan stelling, bewys dat 'n limietverdeling bestaan vir absolute oorskrydings van 'n drempel u , uit 'n onderliggende verdeling F , mits $F \in MDA(G_\gamma)$. Hierdie limietverdeling vir F_u is 'n veralgemeende Pareto-verdeling, naamlik $H_{\gamma,\beta(u)}$.

Stelling 3.3.1: Die Pickands-Balkema-De Haan stelling

$F \in MDA(G_\gamma)$ as en slegs as daar 'n positiewe, meetbare funksie $\beta(u)$ bestaan sodat

$$\lim_{u \rightarrow x_F} \sup_{0 \leq y \leq x_F - u} |F_u(y) - H_{\gamma, \beta(u)}(y)| = 0 \quad (3.3)$$

Δ

Die volgende resultaat, stelling 3.3.2, vorm die kern van die bewys van die Pickands-Balkema-De Haan stelling.

Stelling 3.3.2

$F \in MDA(G_\gamma)$ as en slegs as daar 'n positiewe, meetbare funksie $a(\cdot)$ bestaan sodat

$$\lim_{u \rightarrow x_F} \frac{\bar{F}(u + x \cdot a(u))}{\bar{F}(u)} = \begin{cases} (1 + \gamma x)^{\frac{1}{\gamma}} & \gamma \neq 0 \\ e^{-x} & \gamma = 0 \end{cases} \quad (3.4)$$

vir

$$1 + \gamma x > 0.$$

Δ

Ten einde stelling 3.3.2 te bewys, word die volgende twee hulpresultate gebruik. Sien Embrechts et al (1997) vir 'n bewys van lemma 3.3.3.

Lemma 3.3.3

$F \in MDA(G_\gamma)$ as en slegs as $\bar{F} \in \mathfrak{Z}_{-\alpha}$, met $\alpha = \frac{1}{\gamma}$ en $\gamma > 0$.

Δ

Lemma 3.3.4

’n Funksie $h \in \mathfrak{S}_\alpha$ vir ’n reële waarde van α as en slegs as

$$h(x) = c(x) \exp\left\{\int_z^x \frac{\delta(u)}{u} du\right\}$$

waar $x \geq z > 0$.

Hier is c en δ meetbare funksies, met

$$c(x) \rightarrow c_0 \in (0; \infty)$$

en

$$\delta(x) \rightarrow \alpha \text{ as } x \rightarrow \infty.$$

△

Hierdie hulpstellings stel ons nou in staat om stelling 3.3.2 te bewys.



Bewys van stelling 3.3.2

Laat $\alpha = \frac{1}{\gamma}$ vir $\gamma > 0$. Dan is $G_\gamma(x) = \exp\left\{-\left(1 + \frac{x}{\alpha}\right)^{-\alpha}\right\}$. Volgens lemma 3.3.3

impliseer dit dat $\bar{F} \in \mathfrak{S}_{-\alpha}$.

Dit volg dan uit lemma 3.3.4 dat, vir $z > 0$,

$$\bar{F}(x) = c(x) \exp\left\{-\int_z^x \frac{1}{a(t)} dt\right\}, \quad z < x < \infty,$$

waar

$$c(x) \rightarrow c > 0$$

en

$$\frac{a(x)}{x} \rightarrow \alpha^{-1} \text{ as } x \rightarrow \infty.$$

Gevolglik, indien $\gamma \neq 0$,

$$\lim_{u \rightarrow \infty} \frac{\overline{F}(u + xa(u))}{\overline{F}(u)} = \left(1 + \frac{x}{\alpha}\right)^{-\alpha}.$$

Indien $\gamma = 0$, werk ons met

$$\lim_{\alpha \rightarrow \infty} \left(1 + \frac{x}{\alpha}\right)^{-\alpha} = e^{-x}.$$

Hiermee is die bewys voltooi.

△

Stelling 3.3.2 stel ons in staat om ons hoofresultaat, die Pickands-Balkema-De Haan stelling, te bewys.

Bewys van stelling 3.3.1

Uit die definisie van konvergensie in verdeling weet ons dat $A_n \xrightarrow{D} A$ as en slegs as vir alle kontinuïteitspunte y van F_A ,

$$F_{A_n}(y) \rightarrow F_A(y) \quad \text{as } n \rightarrow \infty \quad (3.5)$$

met F_{A_n} die verdelingsfunksie van A_n en F_A die verdelingsfunksie van A .

Indien F_A kontinu is, impliseer $A_n \xrightarrow{D} A$ ook uniforme konvergensie, naamlik

$$\lim_{n \rightarrow \infty} \sup_x |F_{A_n}(x) - F_A(x)| = 0 \quad (3.6)$$

Omdat die veralgemeende Pareto-verdeling kontinu is, kan stelling 3.3.1 bewys word deur aan te toon dat $F \in MDA(G_\gamma)$ as en slegs as daar 'n positiewe, meetbare funksie $\beta(u)$ bestaan sodat $F_u(x\beta(u)) \rightarrow H_\gamma(x)$ as $u \rightarrow x_F$.

Die resultate van stelling 3.3.2 moet nou as volg geïnterpreteer word:

$F \in MDA(G_\gamma)$ as en slegs as daar 'n positiewe, meetbare funksie $\beta(\cdot)$ bestaan sodat

$$\lim_{u \rightarrow x_F} P\left(\frac{X-u}{\beta(u)} > x \mid X > u\right) = \begin{cases} (1+\gamma x)^{-\frac{1}{\gamma}} & \gamma \neq 0 \\ e^{-x} & \gamma = 0 \end{cases}$$

Stelling 3.3.2 bewys dus dat

$$F_u(x\beta(u)) \rightarrow H_\gamma(x) \quad \text{as } u \rightarrow x_F, \quad (3.7)$$

as en slegs as $F \in MDA(G_\gamma)$, vir $\beta(u)$ 'n positiewe, meetbare funksie. Hier verteenwoordig H_γ die standaard veralgemeende Pareto-verdeling.

Stel nou in (3.7) $y = x\beta(u)$.

Dan volg dit dat

$$\lim_{u \rightarrow x_F} \sup_{0 \leq y \leq x_F - u} |F_u(y) - H_{\gamma, \beta(u)}(y)| = 0.$$

Hiermee is die stelling bewys.

△

Die volgende resultaat, opmerking 3.3.1, is 'n meer intuïtiewe metode om die limietverdeling van drempeloorskrydings te verkry. Hoewel die bewys hiervan nie wiskundig streng is nie, bied dit waardevolle insig in die verband tussen die veralgemeende ekstreemwaarde-verdeling en die veralgemeende Pareto-verdeling. In die besonder lewer hierdie resultaat 'n uitdrukking vir die β -parameter van die verdeling $H_{\gamma, \beta(u)}$ in terme van die parameters van die verdeling $G_{\gamma, \mu, \sigma}$. Opmerking 3.3.1 beklemtoon ook die feit dat daar by die

benaderde veralgemeende Pareto-verdeling vir absolute drempeloorskrydings nie slegs vereis word dat die drempel $u \rightarrow x_F$ nie, maar ook dat die steekproefgrootte $n \rightarrow \infty$.

Opmerking 3.3.1

Veronderstel X_1, X_2, \dots, X_n is 'n reeks onafhanklike stogastiese veranderlikes elk met verdelingsfunksie F waar $F \in MDA(G_\gamma)$. Vir groot genoeg n volg dat

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \approx G_\gamma(x) \quad (3.8)$$

waar

$$M_n = \text{maks}(X_1, X_2, \dots, X_n),$$

met

$$G_\gamma(x) = \exp[-(1 + \gamma x)^{-\frac{1}{\gamma}}], \quad 1 + \gamma x > 0.$$

Dan, vir groot genoeg drempelwaardes u , is die verdelingsfunksie van $(X - u)$ gegee dat $X > u$, by benadering

$$H_{\gamma, \beta(u)}(y) = 1 - \left(1 + \frac{\gamma}{\beta(u)} y\right)^{-\frac{1}{\gamma}} \quad (3.9)$$

gedefinieer op $\{y : y > 0 \text{ en } (1 + \frac{\gamma}{\beta(u)} y) > 0\}$, waar

$$\beta(u) = \sigma + \gamma(u - \mu). \quad (3.10)$$

Bewys

Dit volg vanaf (3.8) en opmerking 2.3.2 dat vir groot genoeg n ,

$$F^n(x) \approx G_{\gamma, \mu, \sigma}(x),$$

waar γ, μ en σ van n kan afhang.

Dus,

$$n \log F(x) \approx -\left[1 + \gamma \left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}. \quad (3.11)$$

Vir groot waardes van x gee 'n Taylor uitbreiding dat

$$\log F(x) \approx -\{1 - F(x)\}$$

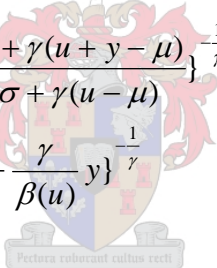
Deur x te vervang met 'n groot genoeg drempelwaarde u en dan te vervang in (3.11) volg dat

$$1 - F(u) \approx \frac{1}{n} \left[1 + \gamma \left(\frac{u - \mu}{\sigma} \right) \right]^{-1/\gamma}.$$

Netso, vir $y > 0$ en 'n groot genoeg u ,

$$1 - F(u + y) \approx \frac{1}{n} \left[1 + \gamma \left(\frac{u + y - \mu}{\sigma} \right) \right]^{-1/\gamma}.$$

Dus,

$$\begin{aligned} P(X > u + y \mid X > u) &\approx \frac{n^{-1} \left[1 + \gamma \left(\frac{u + y - \mu}{\sigma} \right) \right]^{-1/\gamma}}{n^{-1} \left[1 + \gamma \left(\frac{u - \mu}{\sigma} \right) \right]^{-1/\gamma}} \\ &= \left\{ \frac{\sigma + \gamma(u + y - \mu)}{\sigma + \gamma(u - \mu)} \right\}^{-1/\gamma} \\ &= \left\{ 1 + \frac{\gamma}{\beta(u)} y \right\}^{-1/\gamma} \end{aligned}$$


waar

$$\beta(u) = \sigma + \gamma(u - \mu).$$

Let op : Hier is μ en σ eintlik funksies van, naamlik $\mu = b_n$ en $\sigma = a_n$. Dit is dus aangetoon dat vir groot genoeg n en u ,

$$F_u(y) \approx H_{\gamma, \beta(u)}(y) = 1 - \left(1 + \frac{\gamma}{\beta(u)} y \right)^{-1/\gamma}$$

$F \in MDA(G_\gamma)$ as en slegs as daar 'n positiewe, meetbare funksie $\beta(u)$ bestaan sodat

$$\lim_{u \rightarrow x_F} \sup_{0 \leq y \leq x_F - u} |F_u(y) - H_{\gamma, \beta(u)}(y)| = 0 \quad (3.3)$$

△

Opmerking 3.3.2

Die Pickands-Balkema-De Haan stelling impliseer dat indien genormeerde maksima die limietverdeling G_γ het, dan het die drempeloorskrydings 'n ooreenstemmende limietverdeling binne die veralgemeende Pareto-familie.

Anders gestel, daar bestaan konstantes a_n en b_n sodat $\frac{M_n - b_n}{a_n} \xrightarrow{D} G_\gamma$ as en

slegs as daar 'n positiewe funksie $\beta(u)$ bestaan sodat $|F_u(y) - H_{\gamma, \beta(u)}(y)| \rightarrow 0$

as $u \rightarrow \infty$.

Die spesifieke lid van die hierdie familie waarna die oorskryingsverdeling konvergeer, word bepaal deur die waarde van die vormparameter γ , dieselfde γ wat ook die veralgemeende ekstreemwaarde-verdeling geïndekseer het. Gevolglik speel γ dieselfde dominerende rol by die bepaling van die vorm van die veralgemeende Pareto-familie as by die ekstreemwaarde-verdeling. Ons het ook in opmerking 3.3.1 aangetoon dat die waarde van die β -parameter van die veralgemeende Pareto-familie uniek gedefinieer is in terme van die ekstreemwaarde-parameters μ en σ (3.10). Die parameters van die veralgemeende Pareto-verdeling $H_{\gamma, \beta(u)}$ van drempeloorskrydings word dus uniek bepaal deur die parameters van die veralgemeende ekstreemwaarde-verdeling $G_{\gamma, \mu, \sigma}$ van die maksima.

△

Hoewel die onderliggende verdeling F tipies onbekend is, beskou ons die oorskryingswaarskynlikheid (3.1) vir dieselfde bekende verdelings wat in paragraaf 2.3 ondersoek is, naamlik die eksponensiaal-, uniforme, veralgemeende Pareto- en standaard normaalverdeling.

Voorbeeld 3.1: Eksponensiaalverdeling.

As X_1, X_2, \dots 'n ry onafhanklike eksponensiaal(λ) stogastiese veranderlikes is, dan $F(x) = 1 - e^{-\lambda x}$ vir $x > 0$. Dit volg dat

$$\bar{F}_u(y) = \frac{\bar{F}(u+y)}{\bar{F}(u)} = \frac{e^{-\lambda(u+y)}}{e^{-\lambda u}} = e^{-\lambda y}$$

vir alle $y > 0$. Dus is die verdeling vir oorskrydings uit die eksponensiaalpopulasie weer eens die eksponensiaalverdeling, daardie lid van die veralgemeende Pareto-familie wat ooreenstem met die waarde $\gamma = 0$. Dit volg ook vanuit die geheueloosheid van die eksponensiaalverdeling. Let op dat vir hierdie situasie is $\beta = \frac{1}{\gamma}$.

△

Voorbeeld 3.2: Uniforme verdeling.

As X_1, X_2, \dots, X_n ry onafhanklike uniform(0,1) stogastiese veranderlikes is, dan is $F(x) = x$ vir $0 \leq x \leq 1$. Dit volg dat

$$\bar{F}_u(y) = \frac{\bar{F}(u+y)}{\bar{F}(u)} = \frac{1-(u+y)}{1-u} = 1 - \frac{y}{1-u}$$

vir $0 \leq y \leq 1-u$. Dit stem ooreen met die veralgemeende Pareto-verdeling met $\gamma = -1$ en $\beta = u-1$.



△

Voorbeeld 3.3: Veralgemeende Pareto-verdeling.

As X_1, X_2, \dots, X_n ry onafhanklike veralgemeende Pareto(γ, β) stogastiese veranderlikes is, dan is $F(x) = 1 - (1 + \frac{\gamma x}{\beta})^{-\frac{1}{\gamma}}$, $\beta > 0, x \geq 0$. Dit volg dat

$$\bar{F}_u(y) = \frac{\bar{F}(u+y)}{\bar{F}(u)} = \frac{(1 + \frac{(y+u)\gamma}{\beta})^{-\frac{1}{\gamma}}}{(1 + \frac{u\gamma}{\beta})^{-\frac{1}{\gamma}}} = \left\{1 + \frac{y}{\beta + u\gamma}\right\}^{-\frac{1}{\gamma}}$$

vir $y \geq 0$ as $\gamma \geq 0$ of vir $0 \leq y \leq -\frac{\beta + u\gamma}{\gamma}$ as $\gamma < 0$. Dit stem ooreen met die veralgemeende Pareto-verdeling $H_{\gamma, \beta + u\gamma}(y)$. Die verdeling vir oorskrydings uit die veralgemeende Pareto-verdeling is dus weer 'n veralgemeende Pareto-verdeling, net geherparametriseer.

△

Voorbeeld 3.4: Standaard Normaalverdeling.

As X_1, X_2, \dots, X_n 'n ry onafhanklike normaal(0,1) stogastiese veranderlikes is, dan

$$\text{is } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2} dy \text{ vir } x \in \mathbb{R}.$$

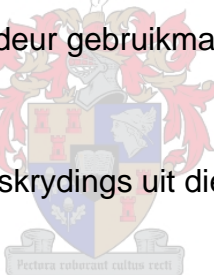
Anders as in die voorafgaande drie voorbeelde, kan ons die verdeling van drempeloorskrydings uit die normaalverdeling nie eksak herlei nie. Ons beskou dus die limietgedrag van

$$\begin{aligned} \bar{F}_u(x\beta(u)) &= P\left(\frac{X-u}{\beta(u)} > x \mid X > u\right) \\ &= \frac{1 - \Phi(u + x\beta(u))}{1 - \Phi(u)} \quad \text{as } u \rightarrow \infty. \end{aligned}$$

Kies ons hierin $\beta(u) = u^{-1}$, volg deur gebruikmaking van vergelyking (2.11) dat

$$\lim_{u \rightarrow \infty} \bar{F}_u(x\beta(u)) = e^{-x}.$$

Dus is die limietverdeling vir oorskrydings uit die standaard normaal-verdeling, die eksponensiaalverdeling.



△

3.4 Die modellering van absolute drempeloorskrydings

Die modellering van absolute drempeloorskrydings is in beginsel redelik eenvoudig: Beskou die beskikbare steekproefdata X_1, X_2, \dots, X_n , onafhanklik van mekaar en identies verdeel volgens die onderliggende verdeling F . Bepaal 'n drempelwaarde u . Definieer as ekstreme waardes daardie waarnemings in die steekproef wat die drempel oorskry, naamlik $\{x_i : x_i > u\}$. Noem hierdie waarnemings $x_{(1)}, \dots, x_{(k)}$ en definieer as drempeloorskrydings $y_j = x_{(j)} - u$ vir $j = 1, \dots, k$. Volgens stelling 3.3.1 mag ons die $\{y_j\}$ beskou as onafhanklike waarnemings van 'n stogastiese veranderlike waarvan die verdeling deur 'n lid van die veralgemeende Pareto-familie benader kan word.

Inferensie bestaan dan uit die pas van die veralgemeende Pareto-verdeling, die evaluering van die model en ekstrapolasie.

3.4.1 Die keuse van 'n drempel

Die benadering om oorskrydings te modelleer verskil van die metodologie in hoofstuk 2 in die sin dat ons hier ekstreme waardes in terme van 'n drempel definieer, eerder as om elke blokmaksimum as 'n ekstreme waarneming te beskou. Die keuse van die drempel is egter analoog aan die keuse van die blokagrootte. Weer eens word die praktisyn deur 'n afruiling tussen sydigheid en variansie gekonfronteer. Indien 'n te lae drempel gekies word, geld die asimptotiese model nie en ontstaan sydigheid. Indien ons 'n drempelwaarde kies wat te hoog is, word die model op te min waarnemings gepas en besit die beramers hoë variansie. Die uitdaging vir die praktisyn is om deur die keuse van die drempel 'n optimale balans tussen sydigheid en variansie te verkry. Volgens Coles (2001) is die standaard praktyk om die drempel so klein as moontlik te kies met die beperking dat die limietverdeling steeds 'n redelike benadering vir die oorskrydings moet wees. Dus, minimeer die variansie met die beperking dat die sydigheid binne perke bly.

Een metode wat aangewend kan word om 'n gepasde keuse van die drempel u te bepaal, is gebaseer op die verwagte waarde van die veralgemeende Pareto-verdeling. Indien 'n stogastiese veranderlike Y die veralgemeende Pareto-verdeling met parameters γ en β het, dan is

$$E(Y) \approx \frac{\beta}{1-\gamma}, \quad (3.12)$$

mits $\gamma < 1$.

Veronderstel nou die veralgemeende Pareto-verdeling is 'n geldige benadering vir die oorskrydings van 'n drempel u_0 , soos voortgebring deur 'n onafhanklike en identies-verdeelde reeks steekproefdata X_1, X_2, \dots, X_n . Noem hierdie oorskrydings $x_{(1)}, \dots, x_{(k)}$. Dit volg uit (3.12) dat

$$E(X - u_0 | X > u_0) \approx \frac{\beta(u_0)}{1-\gamma}, \quad \gamma < 1.$$

Indien die keuse van u_0 as drempel tot 'n geldige benaderde veralgemeende Pareto-verdeling vir die oorskrydings lei, dan moet hierdie limietverdeling ook geld vir enige drempel $u > u_0$. Ons het dus dat vir enige $u > u_0$,

$$E(X - u | X > u) \approx \frac{\beta(u)}{1 - \gamma},$$

mits $\gamma < 1$.

Let op dat ons hier oorskrydings van 'n drempel $u > u_0$ uit die steekproef $x_{(1)}, \dots, x_{(k)}$ ondersoek. Indien ons die $x_{(1)}, \dots, x_{(k)}$ as die oorspronklike steekproef beskou, dan is $x_{(1)}, \dots, x_{(k)}$ onafhanklik van mekaar, elk met 'n benaderde onderliggende verdeling $H_{\gamma, \beta(u_0)}$. Noem nou die oorskrydings uit $x_{(1)}, \dots, x_{(k)}$ ten opsigte van $u > u_0$, $z_{(1)}, \dots, z_{(r)}$. Dit volg uit voorbeeld 3.3 dat $z_{(1)}, \dots, z_{(r)}$ onafhanklik is, elk met benaderde verdeling $H_{\gamma, \beta(u_0) + \gamma(u - u_0)}$. Dus,

$$E(Z - u | Z > u) \approx \frac{\beta(u_0) + \gamma(u - u_0)}{1 - \gamma}. \quad (3.13)$$

Dit blyk dat die verwagte oorskrydings-funksie $e(u) = E(X - u | X > u)$ 'n lineêre funksie is van die drempelwaarde u . Verder is $E(X - u | X > u)$ die verwagte waarde van daardie oorskrydings van die drempel u . Die steekproefgemiddelde van die drempeloorskrydings bied 'n empiriese beramer vir $e(u)$. Volgens (3.13) behoort hierdie beramings lineêr in u te verander vir alle u waarvoor die veralgemeende Pareto-benadering geldig is.

Die volgende prosedure kan nou gevolg word om uit die data 'n gepasde keuse vir u te maak. Verkry 'n grafiese voorstelling van die punte

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^n (x_{(i)} - u) \right) : u < \text{maks}(x_{(i)}) \right\},$$

waar $x_{(1)}, \dots, x_{(n_u)}$ daardie waarnemings is wat u oorskry. Hierdie grafiek staan as die gemiddelde oorskryding-stipping bekend. Hierdie stipping behoort ongeveer lineêr te verloop bokant 'n drempelwaarde u_0 wat lei tot 'n geldige veralgemeende Pareto-benadering vir die oorskrydingsverdeling. Die praktisyen soek dus die kleinste u_0 sodat die gemiddelde oorskryding-stipping lineêr is vir alle $u > u_0$.

'n Tweede prosedure behels dat die veralgemeende Pareto-model vir verskeie voorafbepaalde waardes van u gepas word. Bokant 'n drempel u_0 waarvoor die veralgemeende Pareto-verdeling by benadering geld, behoort die beraming van γ rofweg konstant te wees, terwyl die beraming van

$$\beta(u) = \beta(u_0) + \gamma(u - u_0)$$

ongeveer lineêr behoort toe te neem vir $u > u_0$. Weer eens soek ons die kleinste u_0 wat aan hierdie twee vereistes voldoen.

Die praktisyen moet nooit verwag dat 'n unieke keuse vir u hom sal voordoen nie. Embrechts et al (1997) beveel aan dat 'n kombinasie van gemiddelde oorskryding-stippings (eerste metode), die vergelyking van die stabiliteit van parameter-beramings by verskillende waardes van u (tweede metode) en gesonde verstand ingespan word om u te kies.

3.4.2 Maksimum aanneemlikheid beraming

Nadat 'n gepaste drempelwaarde vasgestel is, kan die parameters van die veralgemeende Pareto-verdeling met behulp van maksimum aanneemlikheid beraam word. Veronderstel ons werk met waarnemings y_1, \dots, y_k , die k oorskrydings van 'n gekose drempel u . Volgens vergelyking (3.2) is

$$H_{\gamma, \beta}(y) = 1 - \left(1 + \frac{\gamma}{\beta} y\right)^{-\frac{1}{\gamma}}.$$

Dit volg hieruit dat

$$h_{\gamma, \beta}(y) = \frac{1}{\beta} \left(1 + \frac{\gamma}{\beta} y\right)^{-\frac{1}{\gamma} - 1}$$

waar

$$h_{\gamma, \beta}(y) = \frac{dH_{\gamma, \beta}(y)}{dy}.$$

Die log-aanneemlikheidsfunksie volg vir $\gamma \neq 0$ as:

$$l(\gamma, \beta) = -k \log \beta - \left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^k \log\left(1 + \frac{\gamma}{\beta} y_i\right) \quad (3.14)$$

mits $\left(1 + \frac{\gamma}{\beta} y_i\right) > 0$ vir $i = 1, \dots, k$.

In die geval $\gamma = 0$ volg die aanneemlikheidsfunksie as

$$l(\beta) = -k \log \beta - \frac{1}{\beta} \sum_{i=1}^k y_i. \quad (3.15)$$

Soos in die geval van vergelyking (2.10) kan die aanneemlikheidsfunksie (3.14) nie analities gemaksimeer word nie. Numeriese tegnieke moet dus weer eens ingespan word om die maksimum aanneemlikheids beramers $\hat{\gamma}$ en $\hat{\beta}$ te vind.

Die veralgemeende Pareto-model word op die k oorskrydings y_1, \dots, y_k gepas. Vir groot genoeg waardes van k geld die asimptotiese normaliteit (sien vergelyking 2.11) van die maksimum aanneemlikheids (MA) beramer $\hat{\underline{\tau}} = (\hat{\gamma}, \hat{\beta})$ soos gegee in stelling 2.4.1. Op dieselfde wyse as in paragraaf 2.4.2 kan benaderde $(1 - \alpha)100\%$ vertrouensintervalle vir γ en β gevind word deur van die asimptotiese normaliteit gebruik te maak. Sien vergelyking (2.12).

3.4.3 Kwantielberaming

In hierdie hoofstuk val die klem op waarnemings regs van 'n sekere drempelwaarde, dit wil sê op waardes wat ver in die stert aan die regterkant lê. Ons stel dus ook belang in die beraming van die funksie $\bar{F}(u + y)$ vir $y > 0$. Indien vergelyking (3.1) herskryf word in die vorm

$$\bar{F}(u + y) = \bar{F}(u) \cdot \bar{F}_u(y), \quad (3.16)$$

dan volg dit dat $\bar{F}(u+y)$ beraam kan word deur beraamers vir $\bar{F}(u)$ en $\bar{F}_u(y)$ te vind.

'n Natuurlike beraamer vir $\bar{F}(u)$ word gegee deur die empiriese stert-verdelingsfunksie

$$\hat{\bar{F}}(u) = \frac{1}{n} \sum_{i=1}^n I(X_i > u) = \frac{n_u}{n}. \quad (3.17)$$

Uit stelling 3.3.1 kan ons (vir groot genoeg waardes van u) die volgende beraamer vir $\bar{F}_u(y)$ motiveer, naamlik

$$\hat{\bar{F}}_u(y) = \bar{H}_{\hat{\gamma}, \hat{\beta}}(y). \quad (3.18)$$

Die volgende beraamer vir die stert-verdelingsfunksie $\bar{F}(u+y)$ vir $y > 0$ is dus afgelei:

$$\hat{\bar{F}}(u+y) = \frac{n_u}{n} \left[1 + \frac{\hat{\gamma}}{\hat{\beta}} y \right]^{-\frac{1}{\hat{\gamma}}} \quad (3.19)$$

Hieruit volg direk 'n beraamer vir die (100)p-de kwantiel x_p :

$$\hat{x}_p = u + \frac{\hat{\beta}}{\hat{\gamma}} \left\{ \left[(1-p) \frac{n}{n_u} \right]^{-\hat{\gamma}} - 1 \right\} \quad (3.20)$$

Een ekstreme stertkwantiel wat van besondere belang is, is R_k , die terugkeervlak wat as volg gedefinieer word:

$$P(X > R_k) = \frac{1}{k}. \quad (3.21)$$

R_k is dus die kwantielwaarde sodanig dat dit slegs 1 keer per k gevalle oorskry word, waar k tipies groot is. Dit volg uit (3.21) en (3.16) dat

$$R_k \approx \begin{cases} u + \frac{\beta}{\gamma} \left\{ \left[k \frac{n_u}{n} \right]^\gamma - 1 \right\} & \text{vir } \gamma \neq 0 \\ u + \beta \log \left(k \frac{n_u}{n} \right) & \text{vir } \gamma = 0 \end{cases}. \quad (3.22)$$

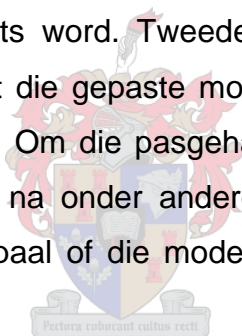
Dus kan ons R_k beraam met:

$$\hat{R}_k \approx \begin{cases} u + \frac{\hat{\beta}}{\hat{\gamma}} \left\{ \left[k \frac{n_u}{n} \right]^{\hat{\gamma}} - 1 \right\} & \text{vir } \gamma \neq 0 \\ u + \hat{\beta} \log\left(k \frac{n_u}{n}\right) & \text{vir } \gamma = 0 \end{cases}$$

Soos in paragraaf 2.4.2 (vergelyking 2.14) kan ons die model herparametriseer (vergelyking 3.22) in terme van die terugkeervlak R_k ten einde vertrouensintervalle vir R_k te vind.

3.4.4 Evaluering van die gepaste model

Dit is in paragraaf 2.2.4 gestel dat die beoordeling van enige gepaste model uit twee dele bestaan. Eerstens moet die gehalte van die passing van die model op die leerdata getoets word. Tweedens moet 'n model goed kan veralgemeen. Dit beteken dat die gepaste model ook goed moet vaar waar nuwe datastelle ter sprake is. Om die pasgehalte van ons model te bepaal, kyk ons in hierdie paragraaf na onder andere waarskynlikheidstippings en kwantielstippings. Om te bepaal of die model goed sal veralgemeen, is 'n aansienlik moeiliker taak.



Veronderstel ons werk met drempeloorskrydings y_1, y_2, \dots, y_k van 'n drempel u waarop 'n beraamde model $\hat{H}_{\hat{\gamma}, \hat{\beta}}$ gepas is. Orden die oorskrydings as volg: $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(k)}$. Hierdie $y_{(1)}, y_{(2)}, \dots, y_{(k)}$ verteenwoordig dus die rangorde statistieke van die oorskrydings.

Die empiriese verdelingsfunksie (definisie 2.2.8) in die punt $y_{(i)}$ word gegee deur

$$\tilde{H}(y_{(i)}) = \frac{i}{k}, \quad i = 1, 2, \dots, k.$$

Die ooreenstemmende waarskynlikheidsberamings van die gepaste model is

$$\hat{H}_{\hat{\gamma}, \hat{\beta}}(y_{(i)}) = 1 - \left(1 + \frac{\hat{\gamma}}{\hat{\beta}} y_{(i)}\right)^{-\frac{1}{\hat{\gamma}}}, \quad i = 1, 2, \dots, k.$$

Indien die model die data goed pas sal,

$$\hat{H}_{\hat{\gamma}, \hat{\beta}}(y_{(i)}) \approx \tilde{H}(y_{(i)})$$

vir alle waardes van i , sodat die waarskynlikheidstipping (definisie 2.2.9) van die pare

$$\{(\tilde{H}(y_{(i)}); \hat{H}_{\hat{\gamma}, \hat{\beta}}(y_{(i)})), \quad i = 1, \dots, k\}$$

naastenby lineêr behoort te wees. Wesenlike afwykings van lineariteit dui op 'n swak modelpassing.

'n Kwantielstipping (definisie 2.2.10) van die pare

$$\{(\hat{H}_{\hat{\gamma}, \hat{\beta}}^{-1}(\frac{i}{k}); y_{(i)}), \quad i = 1, \dots, k\},$$

waar

$$H_{\hat{\gamma}, \hat{\beta}}^{-1}(\frac{i}{k}) = u + \frac{\hat{\beta}}{\hat{\gamma}} [(1 - \frac{i}{k})^{-\hat{\gamma}} - 1], \quad i = 1, \dots, k,$$

sal ongeveer lineêr verloop. Indien dit nie geval is nie, kan dit vertolk word as getuienis dat die model $\hat{H}_{\hat{\gamma}, \hat{\beta}}$ die data nie goed pas nie.

Die terugkeervlak-stipping behels 'n stipping van die terugkeervlak R_k teenoor die terugkeerperiode k . Dit volg vanaf vergelyking (3.21) dat 'n stipping van R_k teenoor $\log k$ lineêr is vir $\gamma = 0$. Vir $\gamma > 0$ is die stipping konveks na onder en vir $\gamma < 0$ is dit konkaf na onder. Hierdie terugkeervlak-stipping suggereer dus of die werklike waarde van die vormparameter γ positief, negatief of naastenby nul is. Dit is gebruikelik om ook die bo- en die ondergrense van die vertrouensintervalle vir R_k by die terugkeervlak-stipping te voeg. Sien byvoorbeeld figuur 2.5.5. en 3.5.5.

'n Verdere metode om te toets hoe goed die model die data pas, is om die digtheidsfunksie van die gepaste veralgemeende Pareto-model met 'n histogram van die drempeloorskrydings te vergelyk.

3.5 'n Limietverdeling vir relatiewe oorskrydings

Paragraaf 3.3 en 3.4 het die teoretiese aspekte, sowel as die modellering van absolute drempeloorskrydings van die vorm $X - u$ ondersoek. In sommige situasies mag dit egter meer sinvol wees om relatiewe oorskrydings van 'n hoë drempel u te beskou. In hierdie paragraaf word die bestaan van 'n asimptotiese verdeling vir relatiewe oorskrydings van die vorm $\frac{X}{u}$ ondersoek en die modellering van sulke oorskrydings word in paragraaf 3.6 behandel. In die res van hierdie werkstuk konsentreer ons egter verder slegs op absolute oorskrydings.

Ons stel belang in die volgende voorwaardelike waarskynlikheid:

$$\bar{F}_{(u)}(x) = P\left(\frac{X}{u} > x \mid X > u\right) = \frac{\bar{F}(xu)}{\bar{F}(u)}, \quad x \geq 1. \quad (3.23)$$

Hier verteenwoordig $F_{(u)}$ die oorskrydings-verdelingsfunksie vir relatiewe oorskrydings van 'n drempel u . Indien die verdeling F bekend was, was die waarskynlikheid in (3.23) ook bekend. Tipies is dit egter nie die geval nie. Soos in die geval van die benaderde veralgemeende Pareto-verdeling vir absolute drempeloorskrydings vir 'n hoë drempelwaarde, ondersoek ons of 'n asimptotiese model vir relatiewe oorskrydings gevind kan word vir groot genoeg drempelwaardes. Dit blyk dat die Pareto-verdeling (definisie 3.5.1) as sodanige limietverdeling optree.

Definisie 3.5.1

Die Pareto-verdeling word gedefinieer as

$$\bar{K}(x) = x^{-\alpha} \quad (3.24)$$

vir $\alpha > 0$ en $x > 1$.

Hier staan α as die Pareto-indeks bekend.

△

Stelling 3.5.1

Veronderstel X is 'n stogastiese veranderlike met verdelingsfunksie $F(x)$, gedefinieer op $x \geq 0$. Indien die stert-verdelingsfunksie \bar{F} reëlmatig-variërend

is met indeks $-\alpha$, $\alpha > 0$, dan konvergeer die oorskrydings-verdelingsfunksie $F_{(u)}$ na die Pareto-verdelingsfunksie $\bar{K}(x) = x^{-\alpha}$, indien $u \rightarrow x_F$.

Bewys

Uit definisie 3.2.3 volg dat indien $\bar{F} \in \mathfrak{S}_{-\alpha}$, $\alpha > 0$, dan is

$$\lim_{u \rightarrow \infty} \frac{\bar{F}(xu)}{\bar{F}(u)} = x^{-\alpha}.$$

Dus,

$$F_{(u)}(x) \rightarrow K(x), \quad \text{as } u \rightarrow x_F.$$

Hiermee is die bewys voltooi.

△

Opmerking 3.5.1

Hierdie verdelings waarvan die stert tot die familie van reëlmatig-variërende funksies behoort, staan bekend as swaarstert verdelings. Voorbeelde hiervan sluit die Burr-, Loggamma- en Frechet-verdelings in. Stelling 3.5.1 het aangetoon dat vir swaarstert verdelings, die verdeling van relatiewe oorskrydings oor 'n drempel u na die Pareto-verdeling konvergeer indien u groot word.

△

3.6 Die modellering van relatiewe drempeloorskrydings

Veronderstel ons het 'n steekproef van waarnemings x_1, \dots, x_n uit 'n verdeling F met 'n swaar stert, soos verduidelik in opmerking 3.5.1. Die eerste stap in die modellering van relatiewe oorskrydings is, soos in paragraaf 3.4, die bepaling van 'n geskikte drempelwaarde. Daarna word 'n benaderde Pareto-verdeling op die relatiewe oorskrydings van hierdie drempel gepas.

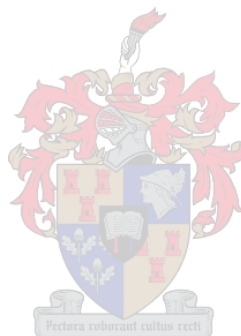
3.6.1 Die bepaling van 'n drempel

In paragraaf 3.4.1 word die afruiling tussen sydigheid en variansie wat by die kies van die drempel vir absolute oorskrydings ter sprake is, verduidelik.

Dieselfde situasie geld by die modellering van relatiewe oorskrydings. By groot drempels is daar min oorskrydings waarop die Pareto-verdeling gepas word, sodat hoë variansie bestaan. By lae drempels geld die asimptotiese verdeling weer nie, sodat sydigheid ontstaan. Sydigheid sowel as hoë variansie is ongewens; die statistikus sal dus deur die gekose drempel 'n optimale balans tussen hierdie twee faktore moet bewerkstellig.

Orden die waarnemings x_1, \dots, x_n as volg: $x_{(1)} \leq \dots \leq x_{(n)}$. Ons kies 'n arbitrêre heelgetal k , waar $2 \leq k \leq n-1$ en beskou k relatiewe oorskrydings. Die keuse van die drempel u word $x_{(k+1)}$, die $k+1$ -de grootste rangorde statistiek van die steekproef. Dus, eerder as om die drempel vas te pen soos by die modellering van absolute oorskrydings (sien paragraaf 3.4.1), word die aantal oorskrydings hier vasgepen. Dit lei tot die nuwe datastel z_1, \dots, z_k , georden as $z_{(1)} \leq \dots \leq z_{(k)}$, waar

$$z_{(i)} = \frac{x_{(i)}}{x_{(k+1)}}.$$



Hier is $z_i \geq 1$ vir $i = 1, \dots, k$.

3.6.2 Maksimum aanneemlikheid beraming

Nadat die drempel vasgestel is, word die Pareto-verdeling op die relatiewe oorskrydings z_1, \dots, z_k gepas. Soos gewoonlik maak ons van die metode van maksimum aanneemlikheid gebruik om die onbekende parameter α van die Pareto-verdeling te beraam.

Volgens vergelyking (3.24) is

$$\bar{K}(z) = z^{-\alpha} \quad , \quad \alpha > 0 \text{ en } z > 1.$$

Dit volg hieruit dat

$$k(z) = \alpha \cdot z^{-(\alpha+1)} \quad , \quad \alpha > 0$$

waar

$$k(z) = \frac{dK(z)}{dz}.$$

Die log-aanneemlikheidsfunksie volg as:

$$l(\alpha) = k \log \alpha - (\alpha + 1) \sum_{i=1}^k \log z_i. \quad (3.25)$$

Gelykstelling van $\frac{dl(\alpha)}{d\alpha}$ aan nul lewer die MA beramer van α :

$$\hat{\alpha} = \frac{k}{\sum_{i=1}^k \log z_i}. \quad (3.26)$$

Die Pareto-model word op die k relatiewe oorskrydings z_1, \dots, z_k gepas. Die z_i het elk die Pareto-verdeling, wat impliseer dat $\log z_i$ eksponensiaal verdeel is. Dit beteken dat $\hat{\alpha} = \frac{1}{k} \sum_{i=1}^k \log z_i$ die gamma-verdeling het. Hierdie feit kan nou gebruik word om eksakte vertrouensintervalle vir α te herlei.

Vir groot genoeg waardes van k geld die asimptotiese normaliteit (sien vergelyking 2.11) van die maksimum aanneemlikheids (MA) beramer $\hat{\theta} = \hat{\alpha}$ soos gegee in stelling 2.4.1. Op dieselfde wyse as in paragraaf 2.4.2 kan benaderde $(1-\alpha)100\%$ vertrouensintervalle vir α gevind word deur van die asimptotiese normaliteit gebruik te maak. Sien vergelyking (2.12).

3.6.3 Kwantielberaming

Ons stel belang in die beraming van ekstreme kwantiele, daardie waardes wat baie ver in die regterstert lê. Ten einde 'n beramer vir sulke ekstreme kwantiele af te lei, vind ons eers 'n beramer vir $\bar{F}(x)$, waar $x \geq u$. Indien vergelyking (3.23) herskryf word in die vorm

$$\bar{F}(x) = \bar{F}(u) \cdot \bar{F}_{(u)}\left(\frac{x}{u}\right), \quad x \geq u$$

waar $x \geq u$ geneem word, dan volg dat $\bar{F}(x)$ beraam kan word deur beramers vir $\bar{F}(u)$ en $\bar{F}_{(u)}\left(\frac{x}{u}\right)$ te vind.

Die drempel word gekies as $u = x_{(k+1)}$, die $k+1$ -de rangorde statistiek van die oorspronklike steekproef. Gevolglik is

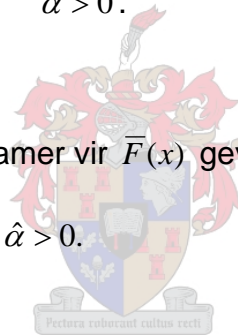
$$\hat{\bar{F}}(u) = \frac{k}{n}.$$

Verder is dit in paragraaf 3.5 afgelei dat die voorwaardelike verdeling van relatiewe oorskrydings uit 'n onderliggende verdeling met 'n swaar stert benader mag word met die Pareto-verdeling vir 'n groot genoeg drempelwaarde u . Dus,

$$\hat{\bar{F}}_{(u)}\left(\frac{x}{x_{(k+1)}}\right) = \left(\frac{x}{x_{(k+1)}}\right)^{-\hat{\alpha}}, \quad \hat{\alpha} > 0.$$

Ons het dus die volgende beramer vir $\bar{F}(x)$ gevind:

$$\hat{\bar{F}}(x) = \frac{k}{n} \cdot \left(\frac{x}{x_{(k+1)}}\right)^{-\hat{\alpha}}, \quad \hat{\alpha} > 0. \quad (3.27)$$



Uit (3.27) volg direk 'n beramer vir die $(100)p$ -de kwartiel x_p :

$$\hat{x}_p = x_{(k+1)} \left((1-p) \cdot \frac{n}{k} \right)^{\frac{1}{\hat{\alpha}}}. \quad (3.28)$$

Een ekstreme stertkwantiel wat van besondere belang is, is R_m , die terugkeervlak wat as volg gedefinieer word:

$$P(X > R_m) = \frac{1}{m}. \quad (3.29)$$

R_m is dus die kwantielwaarde sodanig dat dit slegs 1 keer per m gevalle oorskry word, waar m tipies groot is. Dit volg uit (3.27) dat

$$R_m \approx x_{(k+1)} \left(\frac{n}{mk} \right)^{\frac{1}{\hat{\alpha}}} \quad \text{vir groot } u. \quad (3.29)$$

Dus,

$$\hat{R}_m \approx x_{(k+1)} \left(\frac{n}{mk} \right)^{\frac{1}{\hat{\alpha}}}.$$

Soos in paragraaf 2.4.2 (vergelyking 2.14) kan ons die model herparametriseer (vergelyking 3.29) in terme van die terugkeervlak R_m ten einde vertrouensintervalle vir R_m te vind.

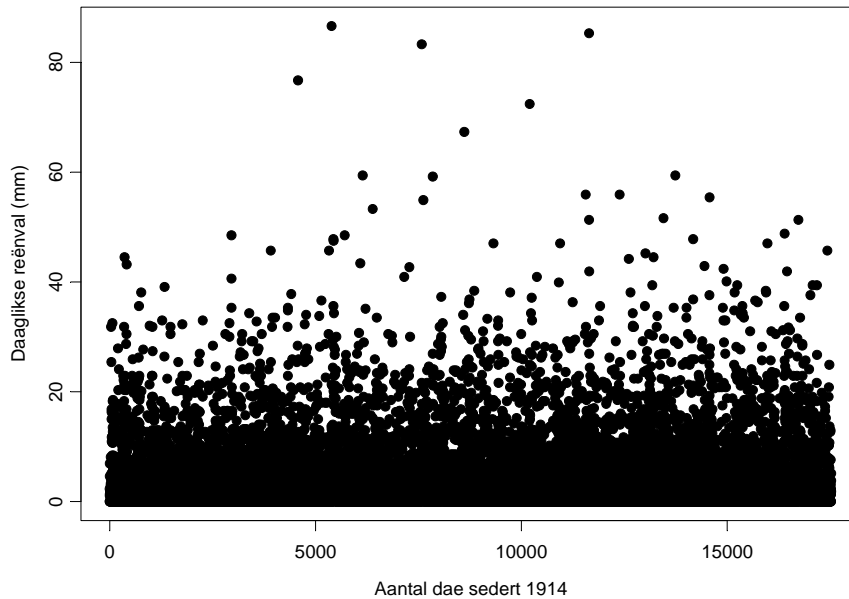
3.6.4 Evaluering van die gepaste model

’n Ondersoek na die gehalte van die passing van die Pareto-model op die relatiewe drempeloorskrydings verloop soortgelyk aan die evaluering van die pasgehalte van die veralgemeende Pareto-model op absolute oorskrydings. Metodes om laasgenoemde te ondersoek word volledig in paragraaf 3.4.4 bespreek. Die verskil is dat ons hier met die relatiewe oorskrydings z_1, \dots, z_k werk, eerder as met die absolute oorskrydings y_1, y_2, \dots, y_k van paragraaf 3.4.4. Ons gebruik egter dieselfde diagnostiese maatstawwe, naamlik die waarskynlikheidstipping (definisie 2.2.9), kwantielstipping (definisie 2.2.10) en digtheidstipping. Die Pareto-verdelingsfunksie $K(x) = 1 - x^{-\alpha}$, $\alpha > 0$, vervang natuurlik telkens die veralgemeende Pareto-verdelingsfunksie

$$H_{\gamma, \beta}(x) = 1 - \left(1 + \gamma \frac{x}{\beta} \right)^{-\frac{1}{\gamma}}.$$

3.7 Voorbeeld

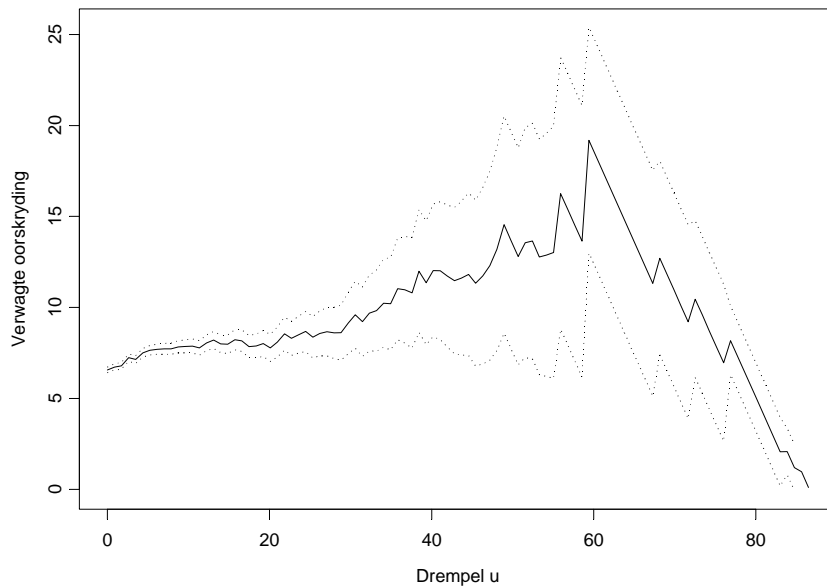
Ons illustreer in hierdie paragraaf die modellering van absolute drempeloorskrydings aan die hand van ’n praktiese voorbeeld. Die datastel wat ondersoek word, bestaan uit ’n tydreeks van daaglikse reënvalsyfers (17531 waarnemings) soos gemeet op ’n bepaalde ligging in die suidweste van Engeland gedurende die periode 1914-1962. Dae waarop uiters hoë neerslae van reën voorkom kan vloede veroorsaak en gevaar vir menselewens inhou. Ons poog vervolgens om hierdie soort rare gebeurtenisse te modelleer.



Figuur 3.7.1. Daaglikse reënval in suidweste van Engeland vanaf 1914-1962.

Die eerste stap by die modellering van enige soort drempeloorskrydings is die keuse van `n drempel. Soos bespreek in paragraaf 3.4.1 is daar twee maniere om `n gepaste drempel vas te stel. Ten eerste kan `n verwagte oorskrydingstipping verkry word. Ons soek die kleinste drempel u_0 sodat die verwagte oorskrydingstipping regs van u_0 ongeveer lineêr verloop.

Indien die vertrouensintervalle (stippellyne) in ag geneem word, lyk dit of die verwagte oorskrydingstipping `n kurwe vorm vanaf $u = 0$ tot $u = 30$, naastebly lineêr verloop tot by $u = 60$ en daarna vinnig na nul konvergeer. Dit sou suggereer dat ons die keuse $u_0 = 60$ maak. Daar is egter slegs 6 oorskrydings van hierdie drempel, heeltmal te min vir sinvolle inferensie. Dit is waarskynlik veiliger om redeneer dat daar sprake van lineariteit is bo $u = 30$ en dus aanvanklik $u_0 = 30$ as drempel te kies.



Figuur 3.7.2. Verwagte oorskryding-stipping vir die daaglikse reënvaldata.

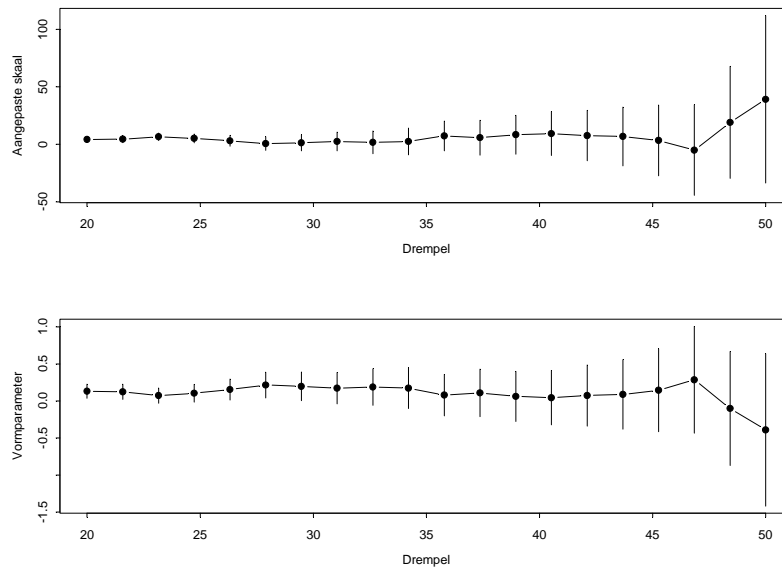
Die tweede benadering tot die keuse van 'n drempel is om die veralgemeende Pareto-model by verskillende drempelwaardes op die data te pas. Ons verlang die laagste drempelwaarde u_0 sodanig dat die parameterberaming van γ stabiel is en dié van β lineêr toeneem in u vir alle $u > u_0$.

Dit blyk uit figuur 3.7.3 dat die beraming van β rofweg lineêr begin toeneem in u vanaf $u = 30$. Gevolglik blyk $u_0 = 30$ 'n redelike drempel te wees.

By die gekose drempel $u_0 = 30$ is daar 152 oorskrydings. Die volgende taak is om nou 'n veralgemeende Pareto-verdeling op hierdie oorskrydings te pas. Maksimering van die log-aanneemlikheidsfunksie (3.14) lei tot die volgende parameterberamings vir γ en β :

$$(\hat{\gamma}, \hat{\beta}) = (0.184, 7.44).$$

Die standaardfoute is 0.101 vir γ en 0.959 vir β . Die gemaksimeerde log-aanneemlikheid is -485.1 by 2 vryheidsgrade.



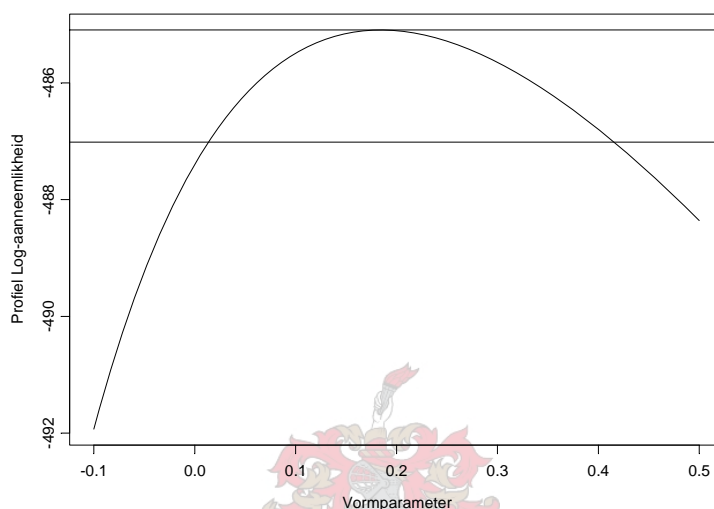
Figuur 3.7.3. Parameterberamings vir die skaal- en vormparameter by verskillende drempelwaardes vir die daaglikse reënvaldata.

Deur gebruik te maak van die asimptotiese normaliteit van die MA-beramers (vergelykings 2.11 en 2.12) volg benaderde 95%-vertrouensinterval vir γ as $0.184 \pm 1.96 \times 0.101 = [-0.014, 0.383]$. Die profiel aanneemlikheid metode (figuur 3.7.4) lewer die volgende benaderde 95% vertrouensinterval vir γ : $[0.019, 0.418]$. Die twee stelde benaderde vertrouensintervalle vir γ is soortgelyk: die normaliteitsgebaseerde vertrouensinterval lê grotendeels in die positiewe domein terwyl die profiel aanneemlikheids vertrouensinterval glad nie in die negatiewe waardes bevat nie. Dit is getuienis ten gunste daarvan dat die werklike waarde van γ positief is.

Aangesien daar in die reënvaldata 152 oorskrydings van die drempel $u_0 = 30$ is uit 'n totale datastel van 17531 datapunte, beraam ons die oorskrydingswaarskynlikheid as $\hat{F}(u_0 = 30) = \frac{152}{17531} = 0.00867$.

By enige ondersoek na ekstreemwaardes is ekstreme kwantiele van groot belang. Die metode van maksimum aanneemlikheid lewer die volgende beraming van die 100-jaar terugkeervlak $R_{100 \times 365}$: $\hat{R}_{100 \times 365} = 106.3$ met 'n

standaardfout van 20.77. Met behulp van die asimptotiese normaliteit van die MA-beramers vind ons die volgende benaderde 95% vertrouensinterval vir $R_{100 \times 365}$: [65.6 , 147.0]. Hierdie ondergrens van 65.6 is volgens die waargenome daaglikse reënvaldata 6 keer oorskry in 48 jaar van waarneming. Dit suggereer dat die ondergrens van hierdie vertrouensinterval dalk onrealisties laag is.

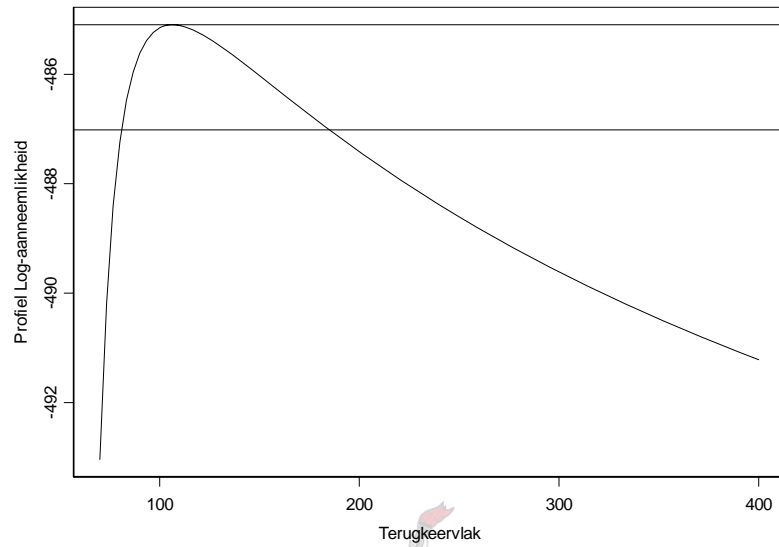


Figuur 3.7.4. Profiel aanneemlikheid vir γ in die drempeloorskrydingsmodel vir die daaglikse reënvaldata.

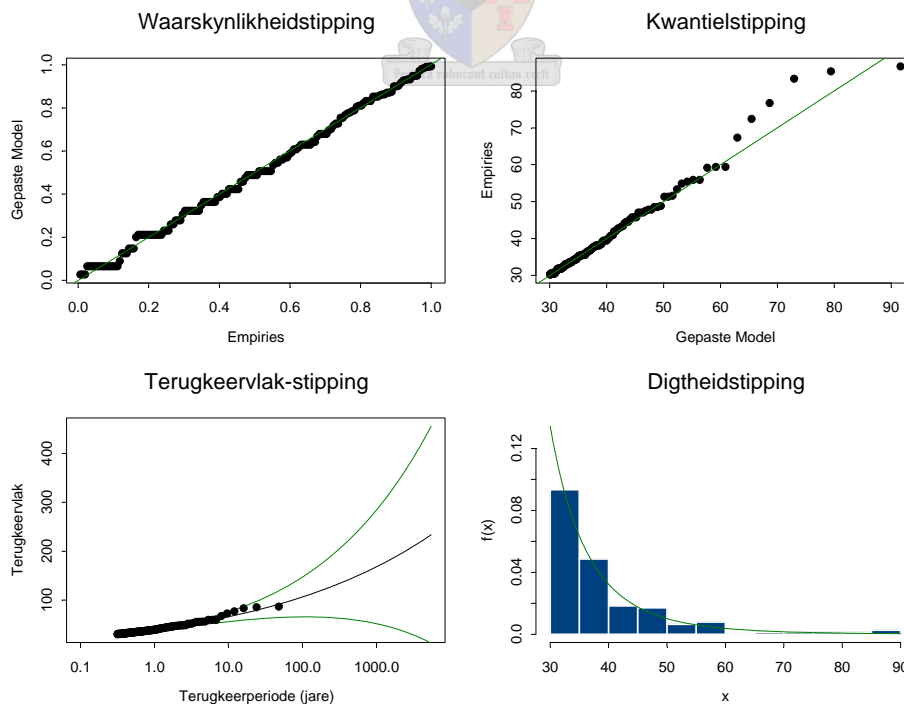
Die gebruik van die profiel aanneemlikheid in figuur 3.7.5 lewer die volgende benaderde 95% vertrouensinterval vir $R_{100 \times 365}$: [81.6 , 185.7]. Hierdie interval sluit die onrealistiese onderste gedeelte van die normaliteitsvertrouensinterval uit en het ook 'n hoër bogrens, wat 'n groter mate van onsekerheid weerspieël. Gevolglik verkies ons die benaderde 95% vertrouensinterval vir $R_{100 \times 365}$ soos verkry met behulp van profiel aanneemlikheid.

Die gebruikelike diagnostiese maatstawwe (sien paragraaf 3.4.4) is gestip ten einde die gehalte van die passing van die veralgemeende Pareto-model op die data te evalueer. Afgesien van die waargenome afwyking van lineariteit ten opsigte die grootste ses datapunte in die kwantielstipping, bied die diagnostiese stippings geen verdere betekenisvolle getuienis aan om die gepaste veralgemeende Pareto-model te verwerp nie. Ons maak die

gevolgtrekking dat die veralgemeende Pareto-model die data bevredigend goed pas.



Figuur 3.7.5. Profiel aanneemlikheid vir die 100-jaar terugkeervlak $R_{100 \times 365}$ in die drempeloorskrydings-model vir die daaglikse reënvaldata.



Figuur 3.7.6. Diagnostiese maatstawwe vir die drempeloorskrydings-model gepas op die daaglikse reënvaldata.

3.8 Gevolgtrekking

Maksima word in hoofstuk 2 as ekstreme waardes beskou. Dit het daartoe gelei dat die beskikbare steekproef in blokke van gelyke grootte opgedeel is, waarna die veralgemeende ekstreemwaarde-verdeling G_γ op die blokmaksima gepas is. Elke blok word deur slegs een (die grootste) waarneming verteenwoordig, sodat inligting verlore gaan indien enige blok meer as `n enkele ekstreme waarde bevat. In hierdie hoofstuk spreek ons dié probleem aan deur die steekproef in sy geheel te neem en alle oorskrydings – hetsy in `n absolute of relatiewe sin - van `n hoë drempel as ekstreme waarnemings te beskou. Die Pickands-Balkema-De Haan stelling bewys dat die voorwaardelike verdeling van absolute oorskrydings na die veralgemeende Pareto-verdeling konvergeer as $u \rightarrow \infty$. Dit word bewys dat die voorwaardelike verdeling van relatiewe oorskrydings uit `n onderliggende verdeling wat aan sekere vereistes voldoen, na die streng Pareto-verdeling konvergeer. Soos in hoofstuk 2 reduseer inferensie dus tot die beraaming van die parameters van hierdie verdeling. Maksimum aanneemlikheid word gebruik om hierdie parameters te beraam, terwyl die asimptotiese normaliteit van die MA beraamers, sowel as die metode van profiel aanneemlikheid aangewend word om benaderde vertrouensintervalle vir die parameters, ekstreme kwantiele en terugkeervlakke te konstrueer.

Volgens die Pickands-Balkema-De Haan stelling geld die benadering $F_u \approx H_{\gamma, \beta(u)}$ by hoë waardes van die drempel u . Hierdie benadering word gebruik vir inferensie ten opsigte van die onbekende onderliggende verdeling F . In hoofstuk vyf word die volgende vraag ondersoek: Hoe goed is die benadering $F_u \approx H_{\gamma, \beta(u)}$ by eindige drempelwaardes? Ten einde hierdie vraag te ondersoek, word van `n simulasiestudie gebruik gemaak. Met behulp van bogenoemde benadering beraam ons stertkwantiele van die onderliggende verdeling sowel as vertrouensintervalle vir hierdie kwantiele. Ons evalueer die geldigheid van die benadering $F_u \approx H_{\gamma, \beta(u)}$ aan die hand van die akkuraatheid van hierdie beraamings.



Hoofstuk 4: Eindige steekproefgedrag

van maksima-inferensie

4.1 Inleiding

Die vorige twee hoofstukke is gewy aan 'n bespreking van die teoretiese onderbou van die modellering van ekstreme waardes. Hierdie ekstreemwaarde-teorie bestaan breedweg uit twee benaderings, naamlik die klassieke maksima-metode (hoofstuk 2) en die meer moderne drempelteorie (hoofstuk 3).

Die basis van die maksimateorie is die Fisher-Tippett stelling. Laat $M_n = \max(X_1, X_2, \dots, X_n)$, waar X_1, X_2, \dots, X_n 'n steekproef van grootte n van 'n stogastiese veranderlike X met 'n onbekende verdelingsfunksie F voorstel. Die strekking van die stelling is dat indien gepaste normeringskonstantes $a_n > 0$ en b_n gevind kan word sodat $\frac{M_n - b_n}{a_n}$ na 'n verdeling G konvergeer as $n \rightarrow \infty$, dan moet G noodwendig van die vorm $G_{\gamma, \mu, \sigma}$ wees, met

$$G_{\gamma, \mu, \sigma}(x) = \exp\left[-\left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-1/\gamma}\right], \quad \text{vir } 1 + \gamma \frac{x - \mu}{\sigma} > 0.$$

Aangesien X_1, X_2, \dots, X_n onafhanklik en identies verdeel is, volg dat die eksakte verdelingsfunksie van M_n , naamlik F^n , deur die veralgemeende ekstreemwaarde verdeling $G_{\gamma, \mu, \sigma}$ benader kan word vir groot waardes van n , oftewel

$$F^n(y) \approx G_{\gamma, \mu, \sigma}(y) \tag{4.1}$$

as $n \rightarrow \infty$.

Inferensie ten opsigte van die onbekende onderliggende verdeling F word gebaseer op vergelyking 4.1. Hierdie vergelyking geld egter slegs by benadering in die geval

van eindige n . In hierdie hoofstuk word die volgende vraag ondersoek: Hoe goed is die benadering $F^n \approx G_{\gamma, \mu, \sigma}$ vir eindige n ? Die beraming van stertkwantiele van F en vertrouensintervalle vir hierdie stertkwantiele is een van die vernaamste doelwitte van 'n ekstreemwaarde-analise. Die kwessie van die geldigheid van die benadering $F^n \approx G_{\gamma, \mu, \sigma}$ waar n eindige waardes aanneem, sal dus in terme van die akkuraatheid van hierdie beramings geëvalueer word. Ons maak van 'n simulasiestudie gebruik om die gedrag van maksima-inferensie by eindige steekproewe te ondersoek.

In die volgende paragraaf word die simulasië-ontwerp uiteengesit en die belangrikste faktore uitgelig. Dit sluit onder andere die keuses van onderliggende verdelings in. Paragraaf 4.3 bevat 'n bespreking van die resultate van die simulasiestudie. Ons maak eers 'n aantal algemene waarnemings ten opsigte van die afvoer, waarna die simulasië-resultate vir die gekose verdelings uit elk van die drie maksimum domeine van aantrekking, in meer detail bestudeer word. Ons sluit af met paragraaf 4.4, waarin die gevolgtrekkings uit die simulasië-analise opgesom en 'n aantal slotopmerkings gemaak word.

4.2 Simulasië-ontwerp



Ons ondersoek na die geldigheid van inferensie oor die onbekende verdeling F gebaseer op die verdeling $G_{\gamma, \mu, \sigma}$ by eindige waardes van n , geskied met behulp van 'n simulasiestudie. 'n Belangrike faktor is die keuse van die onderliggende verdeling. Hier kies ons kontinue verdelings uit elk van die drie maksimum domeine van aantrekking van die verdeling $G_{\gamma, \mu, \sigma}$. 'n Verdere faktor is die steekproefgrootte wat gebruik word, sowel as die aantal blokke waarin dit opgedeel word. Elke lopie van die simulasië-eksperiment verloop as volg: Vir elke keuse van 'n verdeling F , totale steekproefgrootte N en aantal blokke m , word N waarnemings uit F gegenereer. Hierdie N waarnemings word dan in m blokke van gelyke aantal waarnemings (n) opgedeel, waarna die maksimum in elke blok gevind word. Die veralgemeende ekstreemwaarde-verdeling $G_{\gamma, \mu, \sigma}$ word op hierdie blokmaksima gepas. Maksimum aanneemlikheid (MA) – soos bespreek in definisie 2.2.7 en paragraaf 2.5.2 - word gebruik om die parameters van die verdeling $G_{\gamma, \mu, \sigma}$ te beraam. Ons gebruik hierdie

beramings om puntberamings vir stertkwantiele van die onderliggende verdeling F te verkry. Die asimptotiese normaliteit van die MA-beramers (paragraaf 2.5.2) sowel as die chikwadraat-gebaseerde profiel-aanneemlikheid metode (paragraaf 2.5.3) word aangewend om benaderde vertrouensintervalle vir die stertkwantiele te bereken. Ons beoordeel die prestasie van hierdie vertrouensintervalle in terme van die lengte en oordekkingswaarskynlikheid (OW) daarvan.

Die kontinue verdelings onder beskouing word gegroepeer volgens die maksimum domein van aantrekking binne $G_{\gamma,\mu,\sigma}$ waartoe elke verdeling behoort. Ons ondersoek die volgende onderliggende verdelings:

- Frechet-domein, $\gamma > 0$: Veralgemeende Pareto($\gamma = 0.5, \beta = 1$), Streng Pareto($\gamma = 0.5$), Burr(1, 1.5, 1)
- Gumbel-domein, $\gamma = 0$: Standaard Normaal, Standaard Logisties, Gamma (vormparameter = 3)
- Weibull-domein, $\gamma < 0$: Beta (vormparameters 5 en 3 onderskeidelik)

Die simulasiestudie is in die sagtewarepakket S-Plus uitgevoer. Die funksies en subroetines wat aangewend is om die ekstreemwaarde-analises deur te voer, is ontleen aan 'n bylae tot Coles (2001). In daardie bylae word volledige roetines verskaf vir maksima- en drempelgebaseerde ekstreme analises.

Elke eksperiment word 500 keer uitgevoer. Die grootte van die steekproef wat gegenereer word, is die produk van die gekose blokgrootte en die aantal blokke. Hier neem ons alle kombinasies van sowel blokgroottes en blokaantalle van 50, 100, 200 en 500. Dit laat ons met 'n kleinste steekproefgrootte van 2 500 en 'n grootste kombinasie van 250 000 waardes. Met behulp van maksimum aanneemlikheid word die volgende stertkwantiele van die onderliggende verdeling beraam: 0.99 en 0.999. Asimptotiese normaliteit sowel as profiel-aanneemlikheid word gebruik om benaderde 95%-vertrouensintervalle vir hierdie stertkwantiele te bereken.

Ten opsigte van die kwantielberamings word gekyk na die mate waarin die beraamde kwantiel van die werklike kwantiel verskil. Hier bereken ons telkens `n gemiddelde, mediaan en standaardafwyking vir die kwantielafwyking, sowel as mediaan absolute afwyking en `n mediaan absolute fout. By die 95%-vertrouensintervalle stel ons in die oordekkingswaarskynlikheid sowel as die lengte van die vertrouensinterval belang.

Laat \hat{q}_i die beraamde kwantiel op die i -de simulasielopie en q die teoretiese kwantiel van die onderliggende verdeling F aandui. Ten einde die moontlike sydigheid van ons beramings te ondersoek, bereken ons na 500 herhalings die gemiddelde kwantielafwyking, gegee deur

$$\frac{1}{500} \sum_{i=1}^{500} (q - \hat{q}_i) \quad (4.2)$$

sowel as die mediaan van die afwykings tussen die werklike en die beraamde kwantiel, naamlik

$$\text{mediaan}_i (q - \hat{q}_i). \quad (4.3)$$

As `n eerste maatstaf van variasie bereken ons die standaardafwyking van die kwantielafwykings, gegee deur

$$\sqrt{\frac{1}{500} \sum_{i=1}^{500} (\hat{q}_i - \bar{\hat{q}})^2}. \quad (4.4)$$

Hier is $\bar{\hat{q}}$ die gemiddelde kwantielberaming, oftewel

$$\bar{\hat{q}} = \frac{1}{500} \sum_{i=1}^{500} \hat{q}_i. \quad (4.5)$$

Ons beskou ook twee verdere maatstawwe van variasie, naamlik die mediaan absolute afwyking (MAD) en die mediaan absolute fout (MAE). Hierdie maatstawwe word as volg gedefinieer:

$$MAD = \text{mediaan}_i \{ |q - \hat{q}_i| \} \quad (4.6)$$

$$MAE = \text{mediaan}_i \{ | \text{mediaan}_j (\hat{q}_j) - \hat{q}_i | \} \quad (4.7)$$

By elke lopie van die simulasië-eksperiment word 'n benaderde 95%-vertrouensinterval vir elk van die stertkwantiele onder beskouing bereken. Na afloop van die 500 herhalings word die beraamde oordeckingswaarskynlikheid (OW) gedefinieer as die proporsie van die lopies waarby die teoretiese kwantiel q binne die beraamde vertrouensinterval geval het. Indien ons die benaderde ondergrens en bogrens van die beraamde vertrouensinterval op die i -de simulasielopie onderskeidelik VI_{-L_i} en VI_{-U_i} noem, dan volg dat

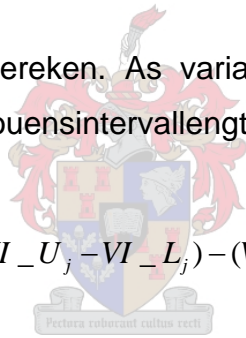
$$OW = \frac{1}{500} \sum_{i=1}^{500} I(VI_{-L_i} \leq q \leq VI_{-U_i}). \quad (4.8)$$

Die lengte van ons verkreë vertrouensintervalle moet saam met die oordeckingswaarskynlikhede oorweeg word. In hierdie verband word die gemiddelde vertrouensintervallengte, gegee deur

$$\frac{1}{500} \sum_{i=1}^{500} (VI_{-U_i} - VI_{-L_i}), \quad (4.9)$$

sowel as die mediaan lengte bereken. As variasiemaatstawwe bereken ons die standaardafwyking van die vertrouensintervallengtes en die mediaan absolute fout, naamlik

$$MAE = \underset{i}{\text{mediaan}}\{|\underset{j}{\text{mediaan}}(VI_{-U_j} - VI_{-L_j}) - (VI_{-U_i} - VI_{-L_i})|\} \quad (4.10)$$



4.3 Bespreking van resultate

In hierdie paragraaf word die resultate van die simulasiestudie, soos in paragraaf 4.2 omskryf, geanaliseer. Die volledige stel afvoer word in Aanhangsel 4A gegee. Ten eerste word 'n aantal algemene opmerkings oor die uitslae gemaak, waarna ons in die subparagrafe 4.3.1 tot 4.3.3 in meer detail na die resultate vir elk van die gekose verdelings kyk. Daar sal telkens van uittreksels uit Aanhangsel 4A gebruik gemaak word om waarnemings en gevolgtrekkings rakende die simulasië-uitslae te belig.

Met behulp van ons simulasiestudie is benaderde 95%-vertrouensintervalle (VI 's) ten opsigte van ekstreme stertkwantiele soos $q_{0.99}$ en $q_{0.999}$, die 0.99- en 0.999-kwantiele van 'n onderliggende verdeling F bereken. Die akkuraatheid van hierdie inferensie

ten opsigte van F , soos gebaseer op die benadering $G_{\gamma, \mu, \sigma}$, word beoordeel aan die hand van oordekkingswaarskynlikheid sowel as die lengte van die verkreeë vertrouensintervalle: Van 'n goeie prosedure word verwag dat sy OW naby aan 0.95 is en dat sy VI-lengte kort is. Die akkuraatheid van die puntberamings van $q_{0.99}$ en $q_{0.999}$ word ook ondersoek. Hier verwag ons van 'n goeie prosedure dat die beraamde kwantiel nie veel van die teoretiese kwantiel sal afwyk nie en ook nie oormatig sal varieer nie. 'n Verdere kwessie wat ondersoek word, behels 'n vergelyking tussen die prestasie van die metodes van maksimum aanneemlikheid en profiel aanneemlikheid in die berekening van 'n 95%-vertrouensinterval vir $q_{0.99}$ en $q_{0.999}$.

'n Eerste algemene waarneming uit die simulasieresultate is dat sowel die gemiddelde as die mediaan vertrouensintervallengte by 'n vaste blokgrötte afneem as ons die aantal blokke m , en dus die steekproefgrötte N , laat toeneem. 'n Toename in die aantal blokke – weer by 'n vaste blokgrötte – het ook laer variasie tot gevolg: Sowel die standaardafwyking as die MAE neem dan af. Ten opsigte van oordekking blyk dit dat die normaalgebaseerde vertrouensintervalle tipies 'n OW van minder as 0.95 het. Die OW's is egter bestendig: Oor al die verdelings onder beskouing is vir 95%-vertrouensintervalle vir beide $q_{0.99}$ en $q_{0.999}$ OW's van 0.90 en hoër aangeteken. Die swakste normaalgebaseerde OW was 0.87, soos behaal vir $q_{0.999}$ in die geval van 'n betaverdeling. Die profielgebaseerde vertrouensintervalle behaal ongeveer ewe gereeld OW's van meer en minder as 0.95. Hulle is egter aansienlik minder bestendig: Heelwat OW's van meer as 0.97 is aangeteken, maar ook baie OW's van minder as 0.90, met 'n swakste waarde van 0.64 vir $q_{0.999}$ van die veralgemeende Pareto-verdeling.

Ten opsigte van die kwantielberamings word waargeneem dat variasie maatstawwe soos die standaardafwyking (vergelyking 4.4), MAE (4.6) en MAD (4.7) sonder uitsondering afneem wanneer die aantal blokke by 'n vaste blokgrötte toeneem. Die MAE- en MAD-waardes is deurgaans ongeveer dieselfde; geen opmerklike verband word hier waargeneem nie. Die standaardafwyking sowel as MAE en MAD van die kwantiel $q_{0.999}$ daal as ons die blokgrötte by vaste aantal blokke laat toeneem.

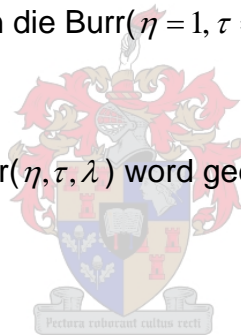
Ons bespreek vervolgens die simulasiresultate in groter detail vir elk van die verdelings onder beskouing. Hierdie verdelings word gegroepeer volgens die maksimum domein van aantrekking binne $G_{\gamma,\mu,\sigma}$ waartoe die verdeling behoort. By sommige kombinasies van die blokgrötte en die aantal blokke kon die S-Plus roetines in die geval van die Burr-verdeling nie funksioneer nie. In sulke gevalle is telkens van 'n tilde (~) gebruik gemaak om aan te dui dat die simulatie-eksperiment nie deurgevoer kon word nie.

4.3.1 Die Frechet-domein: $\gamma > 0$

Die belangrikste faktor in ons simulatie-ontwerp is die keuse van 'n onderliggende verdeling F . In die Frechet-domein word die veralgemeende Pareto($\gamma=0.5$ en $\beta=1$)-, streng Pareto($\gamma=0.5$)- en die Burr($\eta=1, \tau=1.5, \lambda=1$)- verdelings beskou.

Die verdelingsfunksie van die Burr(η, τ, λ) word gedefinieer as

$$F(x) = 1 - \left(\frac{\eta}{\eta + x^\tau} \right)^\lambda,$$



waar

$$x > 0; \eta, \tau, \lambda > 0.$$

Hier is die ekstreemwaarde-indeks $\gamma = \frac{1}{\lambda\tau}$. Ons keuse van $\tau=1.5$ en $\lambda=1$ stem dus met 'n γ -waarde van 1.5^{-1} ooreen.

Die bogenoemde verdelings is lede van 'n klas swaarstert verdelings wat tot die Frechet maksimum domein van aantrekking behoort. Onder dié drie verdelings het die Burr-verdeling die swaarste stert. Die klas van swaarstert verdelings is van groot belang in finansiële toepassings.

'n Vergelyking tussen die vertrouensintervalle vir $q_{0,99}$ en $q_{0,999}$, soos verkry met die normaal- en die profielmetode, toon dat die normaal-vertrouensinterval vir $q_{0,99}$ deurgaans effens korter is, terwyl die OW's ongeveer dieselfde is. Dit sou impliseer dat die normaalgebaseerde intervale hier beter vaar. Vir die meer ekstreme kwantiel $q_{0,999}$ is die gemiddelde profielgebaseerde VI by kleiner blokgroottes soos 50 en 100 drasties korter as die normaal-VI. Die profiel-VI's behaal egter baie swak OW's: By 'n blokgruotte van 50 is al die profiel-OW's onder 0.90, met 'n laagste waarde van 0.49 in die geval van die burr verdeling. Die OW's van die Burr-verdeling is nog heelwat laer as dié van die veralgemeende en streng Pareto-verdelings, wat alreeds onaanvaarbaar laag is.

Tabelle 4.3.1, 4.3.2 en 4.3.3 verskaf die oordeckingswaarskynlikhede van die normaal- (links) en profielgebaseerde (regs) VI's vir $q_{0,99}$ (bo) en $q_{0,999}$ (onder) van die veralgemeende Pareto-, streng Pareto- en Burr-verdeling.

Veralgemeende Pareto											
Normaliteit	blokgruotte (n)					Profiel	blokgruotte (n)				
	$q_{0,99}$	50	100	200	500		$q_{0,99}$	50	100	200	500
	50	0.93	0.93	0.92	0.89		50	0.92	0.96	0.93	0.92
aantal	100	0.95	0.95	0.93	0.91	aantal	100	0.94	0.96	0.97	0.98
blokke	200	0.96	0.95	0.95	0.94	blokke	200	0.96	0.93	0.93	0.94
	500	0.96	0.96	0.95	0.94		500	0.96	0.97	0.94	0.92

Normaliteit	blokgruotte (n)					Profiel	blokgruotte (n)				
	$q_{0,999}$	50	100	200	500		$q_{0,999}$	50	100	200	500
	50	0.91	0.91	0.93	0.94		50	0.64	0.78	0.96	0.87
aantal	100	0.92	0.95	0.94	0.93	aantal	100	0.70	0.89	0.94	0.97
blokke	200	0.96	0.93	0.93	0.94	blokke	200	0.80	0.96	0.99	0.97
	500	0.95	0.95	0.95	0.93		500	0.85	0.96	0.97	0.99

Tabel 4.3.1: OW's vir 95%-VI's vir $q_{0,99}$ en $q_{0,999}$ van die veralgemeende Pareto-verdeling

Pareto

Normaliteit	blokgrootte (n)					Profiel	blokgrootte (n)				
	$q_{0.99}$	50	100	200	500		$q_{0.99}$	50	100	200	500
	50	0.94	0.95	0.93	0.90		50	0.95	0.95	0.94	
aantal	100	0.93	0.95	0.94	0.91	aantal	100	0.95	0.95	0.94	0.97
blokke	200	0.93	0.94	0.95	0.95	blokke	200	0.96	0.94	0.99	0.94
	500	0.95	0.97	0.95	0.94		500	0.98	0.93	0.94	0.92

Normaliteit	blokgrootte (n)					Profiel	blokgrootte (n)				
	$q_{0.999}$	50	100	200	500		$q_{0.999}$	50	100	200	500
	50	0.91	0.89	0.90	0.93		50	0.67	0.82	0.93	
aantal	100	0.92	0.93	0.96	0.95	aantal	100	0.68	0.88	0.93	0.92
blokke	200	0.94	0.94	0.95	0.96	blokke	200	0.83	0.88	0.96	0.95
	500	0.95	0.96	0.96	0.94		500	0.88	0.97	0.97	0.93

Tabel 4.3.2: OW's vir 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ van die streng Pareto-verdeling.



Burr

Normaliteit	blokgrootte (n)					Profiel	blokgrootte (n)				
	$q_{0.99}$	50	100	200	500		$q_{0.99}$	50	100	200	500
	50	~	0.93	0.93	~		50	~	0.90	0.94	~
aantal	100	~	0.96	0.94	0.92	aantal	100	~	0.94	0.92	0.95
blokke	200	~	0.97	0.96	0.94	blokke	200	~	0.92	0.96	0.95
	500	~	0.95	0.94	0.97		500	~	0.93	0.97	0.93

	blokgrootte (n)						blokgrootte (n)				
	$q_{0.999}$	50	100	200	500		$q_{0.999}$	50	100	200	500
	50	~	0.90	0.92	~		50	~	0.49	0.77	~
aantal	100	~	0.94	0.91	0.94	aantal	100	~	0.67	0.76	0.91
blokke	200	~	0.96	0.94	0.95	blokke	200	~	0.70	0.85	0.97
	500	~	0.91	0.94	0.95		500	~	0.72	0.93	0.90

Tabel 4.3.3: OW's vir 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ van die Burr-verdeling.

In tabelle 4.3.4, 4.3.5 en 4.3.6 word die gemiddelde VI-lengte vir hierdie verdelings volgens sowel die normaal- as profielmetode gegee. In die geval van die 0.999-

kwantiel is die profiel-VI's is duidelik heelwat korter, maar vertoon onaanvaarbare lae oordekkings. 'n Soortgelyke patroon kan ook waargeneem word deur die mediaan lengtes te beskou.

Veralgemeende Pareto

Normaliteit	blokgrötte (n)					$q_{0.99}$	blokgrötte (n)				
	$q_{0.99}$	50	100	200	500		$q_{0.99}$	50	100	200	500
	50	7.45	6.24	5.57	10.50	50	7.13	6.39	5.98	11.27	
aantal	100	5.24	4.47	4.02	7.15	aantal	100	5.32	4.39	7.37	
blokke	200	3.74	3.12	2.84	5.05	blokke	200	3.76	3.11	5.09	
	500	2.35	1.98	1.80	3.12	500	2.34	1.98	1.82	3.16	

Normaliteit	blokgrötte (n)					Profiel	blokgrötte (n)				
	$q_{0.999}$	50	100	200	500		$q_{0.999}$	50	100	200	500
	50	71.78	48.95	35.62	23.57	50	28.37	26.49	25.38	22.90	
aantal	100	47.13	34.39	24.45	16.83	aantal	100	22.93	21.32	21.01	16.85
blokke	200	33.06	23.52	16.81	11.86	blokke	200	18.27	17.56	16.56	11.83
	500	20.26	14.91	10.77	7.47	500	13.57	13.45	11.01	7.51	

Tabel 4.3.4: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die veralgemeende Pareto-verdeling.



Pareto

Normaliteit	blokgrötte (n)					Profiel	blokgrötte (n)				
	$q_{0.99}$	50	100	200	500		$q_{0.99}$	50	100	200	500
	50	3.72	3.17	2.79	5.19	50	3.55	3.18	2.92	~	
aantal	100	2.61	2.22	2.03	3.56	aantal	100	2.65	2.20	2.92	4.01
blokke	200	1.86	1.56	1.42	2.52	blokke	200	1.88	1.56	1.42	2.53
	500	1.17	0.99	0.90	1.56	500	1.18	1.00	0.90	1.59	

Normaliteit	blokgrötte (n)					Profiel	blokgrötte (n)				
	$q_{0.999}$	50	100	200	500		$q_{0.999}$	50	100	200	500
	50	36.69	25.50	17.65	11.73	50	14.37	13.70	12.80	~	
aantal	100	23.42	17.29	12.41	8.39	aantal	100	11.38	10.80	12.80	8.41
blokke	200	16.46	11.83	8.58	5.90	blokke	200	9.06	8.86	8.36	5.95
	500	9.87	7.38	5.42	3.73	500	6.77	6.88	5.41	3.73	

Tabel 4.3.5: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$, streng Pareto-verdeling.

Burr						Burr					
Normaliteit		blokgrötte (n)				Profiel		blokgrötte (n)			
	$q_{0.99}$	50	100	200	500		$q_{0.99}$	50	100	200	500
	50	~	9.01	8.08	~		50	~	9.09	8.41	~
aantal	100	~	6.38	5.76	10.48	aantal	100	~	6.50	5.90	11.36
blokke	200	~	4.55	4.07	7.31	blokke	200	~	4.53	4.12	7.73
	500	~	2.83	2.61	4.61		500	~	2.84	2.61	4.16

Burr						Burr					
		blokgrötte (n)						blokgrötte (n)			
	$q_{0.999}$	50	100	200	500		$q_{0.999}$	50	100	200	500
	50	~	111.79	76.63	~		50	~	41.47	39.80	~
aantal	100	~	72.89	51.21	34.59	aantal	100	~	35.40	31.80	32.66
blokke	200	~	50.75	36.57	24.91	blokke	200	~	26.88	25.74	24.95
	500	~	30.91	22.48	15.63		500	~	19.31	19.57	15.53

Tabel 4.3.6: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die Burr-verdeling.

Uit die tabelle hierbo is dit duidelik dat die OW's van die profiel-VI's ten opsigte van $q_{0.999}$, veral by die kleiner steekproefgroottes onaanvaarbaar laag is. Hierdie lae OW's gaan met 'n korter gemiddelde VI-lengte gepaard. In 'n poging om hierdie lae OW's tot die verlangde vlak van 0.95 te kalibreer, is die berekende VI's wyer gemaak. Dit is bereik deur in die profielgebaseerde VI-berekening van meer ekstreme chi-kwadraat kwantiele soos 0.999, oftewel $\chi_1^2(0.999)$ en 0.9999, oftewel $\chi_1^2(0.9999)$, gebruik te maak.

In tabelle 4.3.7 tot 4.3.9 word die OW's (bo) en die gemiddelde VI-lengtes (onder) vir $q_{0.999}$ gegee. Gebruikmaking van die hoër chi-kwadraat kwantiele lei tot langer gemiddelde VI's met 'n hoër OW. Hierdie VI's bly egter heelwat korter as die normaal-VI's; die OW's styg ook nie veel nie en bly deurgaans onder 0.95. Die volgende interessante waarneming kan uit tabel 4.3.7 ten opsigte van die streng Pareto-verdeling gemaak word: By die kombinasie van 100 blokke en 100 waarnemings per blok word 'n OW van 0.93 met behulp van die aangepaste profielmetode bereik. Dit is presies dieselfde OW wat deur die ooreenstemmende (95%-) normaal-VI bereik word (sien tabel 4.3.2). Die verskil is die gemiddelde profiel-lengte van 14.18, terwyl die normaal-VI teen 'n gemiddelde lengte van 17.29 heelwat langer is. Dit laat die vraag ontstaan of die profiel-VI's – indien gekalibreer om die OW's van die normaal-VI's te ewenaar – moontlik korter as die normaal-VI's mag blyk te wees. Ten einde hierdie kwessie te ondersoek sal daar van uiters

ekstreme chi-kwadraat kwantiele - heelwat meer ekstreem as $\chi_1^2(0.999)$ en $\chi_1^2(0.9999)$ - in die profielmetode gebruik gemaak moet word. Die rede hiervoor is dat die profiel-lengtes (en dus die OW's) baie stadig toeneem namate hoër chi-kwadraat kwantiele gebruik word. Selfs by $\chi_1^2(0.9999)$ is die meerderheid van die OW's steeds tussen 0.69 en 0.84. In die praktyk is dit waarskynlik veiliger om in die geval van swaarstert verdelings eerder die normaalmetode te gebruik om VI's vir ekstreme kwantiele soos $q_{0.999}$ te bereken.

Veralgemeende Pareto

		blokgrootte (n)					blokgrootte (n)		
		$\chi_1^2(0.999)$	50	100	$\chi_1^2(0.9999)$	50	100		
aantal	50	0.69	0.84	aantal	50	0.70	0.81		
blokke	100	0.72	0.93	blokke	100	0.72	0.93		

		$\chi_1^2(0.999)$	blokgrootte (n)		$\chi_1^2(0.9999)$	blokgrootte (n)			
			50	100		50	100		
aantal	50	37.99	34.60	aantal	50	39.97	35.98		
blokke	100	29.50	27.88	blokke	100	31.73	30.20		

Tabel 4.3.7: Aangepaste profiel-OW's en gemiddelde VI-lengtes vir $q_{0.999}$ van die veralgemeende Pareto-verdeling

Pareto

		blokgrootte (n)					blokgrootte (n)		
		$\chi_1^2(0.999)$	50	100	$\chi_1^2(0.9999)$	50	100		
aantal	50	0.69	0.84	aantal	50	0.70	0.81		
blokke	100	0.72	0.93	blokke	100	0.72	0.93		

		blokgrootte (n)					blokgrootte (n)		
		$\chi_1^2(0.999)$	50	100	$\chi_1^2(0.9999)$	50	100		
aantal	50	18.03	17.20	aantal	50	19.352	18.88		
blokke	100	14.62	14.18	blokke	100	15.83	15.03		

Tabel 4.3.8: Aangepaste profiel-OW's en gemiddelde VI-lengtes vir $q_{0.999}$ van die streng Pareto-verdeling

Burr

blokgrötte (n)			blokgrötte (n)		
$\chi_1^2(0.999)$			$\chi_1^2(0.9999)$		
aantal	50	0.64	aantal	50	0.61
blokke	100	0.68	blokke	100	0.65

blokgrötte (n)			blokgrötte (n)		
$\chi_1^2(0.999)$			$\chi_1^2(0.9999)$		
aantal	50	56.58	aantal	50	60.97
blokke	100	44.83	blokke	100	49.14

Tabel 4.3.9: Aangepaste profiel-OW's en gemiddelde VI-lengtes vir $q_{0.999}$ van die streng Pareto-verdeling

Ten opsigte van die puntberaming van die sterkwantiele onder beskouing, blyk dit dat vir die Burr-, veralgemeende en streng Pareto-verdelings, $q_{0.99}$ tipies onderberaam word by die kleiner blokgröttes, maar oorberaam word by die groter blokgröttes. Ten opsigte van die oor- of onderberaming van $q_{0.999}$ van die veralgemeende en streng Pareto-verdelings, verkry ons gemengde resultate. Geen opmerklike verband kan hier waargeneem word nie. $q_{0.999}$ van die Burr-verdeling word by groot blokgröttes soos 500 tipies onderberaam. Dié resultate is in tabelle 4.3.10 en 4.3.11 vervat.

Veralgemeende Pareto					Pareto						
blok grootte (n)					blok grootte (n)						
	$q_{0.99}$	50	100	200	500		$q_{0.99}$	50	100	200	500
	50	0.12	0.01	-0.14	-0.69		50	0.16	0.00	-0.08	-0.25
aantal	100	0.16	-0.12	-0.06	-0.44	aantal	100	0.04	-0.02	-0.04	-0.21
blokke	200	0.04	0.03	-0.07	-0.19	blokke	200	0.03	0.00	-0.02	-0.09
	500	0.03	0.03	-0.05	-0.10		500	0.03	0.01	-0.01	-0.05

blok grootte (n)					blok grootte (n)						
	$q_{0.999}$	50	100	200	500		$q_{0.999}$	50	100	200	500
	50	-0.37	1.47	0.11	0.49		50	0.24	0.00	0.18	0.45
aantal	100	0.31	-0.19	-0.02	0.28	aantal	100	0.40	-0.18	-0.06	0.07
blokke	200	-0.84	-0.10	0.29	-0.11	blokke	200	-0.08	0.08	-0.08	-0.01
	500	-0.08	-0.20	-0.14	0.13		500	0.36	0.07	-0.15	0.08

Tabel 4.3.10: Mediaan afwyking vir $q_{0.99}$ en $q_{0.999}$ vir die veralgemeende en streng Pareto-verdelings.

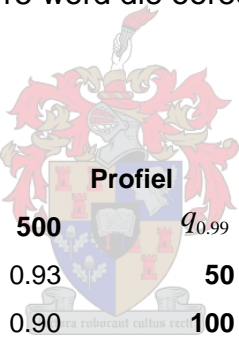
Burr											
Mediaan	blok grootte (n)				Gemiddeld	blok grootte (n)					
	$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500	
	50	~	0.11	-0.13	~	50	~	0.02	-0.28	~	
aantal	100	~	0.02	-0.01	-0.45	aantal	100	~	0.02	-0.15	-0.67
blokke	200	~	-0.16	0.16	-0.26	blokke	200	~	-0.12	0.05	-0.19
	500	~	0.14	0.03	-0.01		500	~	0.10	-0.01	-0.02

blok grootte (n)					blok grootte (n)						
	$q_{0.999}$	50	100	200	500		$q_{0.999}$	50	100	200	500
	50	~	0.22	1.15	~	50	~	-4.17	-1.55	~	
aantal	100	~	-0.13	1.39	1.75	aantal	100	~	-1.58	0.19	1.01
blokke	200	~	-0.07	-0.13	0.20	blokke	200	~	-0.09	-0.20	0.24
	500	~	0.72	1.08	0.12		500	~	0.49	0.45	0.14

Tabel 4.3.11: Mediaan en gemiddelde afwykings vir $q_{0.99}$ en $q_{0.999}$ vir die Burr verdeling.

4.3.2 Die Gumbel-domein: $\gamma = 0$

Die verdelings in die Gumbel-domein wat ondersoek word is die logistiese, normaal- en gammaverdeling. Vir die normaal- en logistiese verdeling (albei simmetries) vaar die profiel- en normaalmetode ongeveer dieselfde. In beide gevalle word OW's rondom 0.95 verkry; die profielgebaseerde OW's is egter wyer verspreid rondom 0.95. In die geval van die normaalverdeling duik 'n onaanvaarbare lae OW van 0.80 (sien tabel 4.3.12) vir 'n VI vir $q_{0,999}$ by die profielmetode op. Ook ten opsigte van die gemiddelde en mediaan lengte verskil die VI's nie veel nie. Dit wil voorkom of die MA-gebaseerde VI's vir $q_{0,999}$ by 'n klein blokgruotte effens korter is as die profiel-VI's. Dit gaan egter hier oor die algemeen met 'n laer OW gepaard. Tabele 4.3.12 en 4.3.13 bevat die oordekkingswaarskynlikhede vir die normaal- en logistiese verdeling. In tabelle 4.3.14 en 4.3.15 word die ooreenstemmende VI-lengtes gegee.



Normaal					Profiel				
Normaliteit	blokgruotte (n)				Normaliteit	blokgruotte (n)			
$q_{0,99}$	50	100	200	500	$q_{0,99}$	50	100	200	500
50	0.93	0.94	0.95	0.93	50	0.96	0.91	0.97	0.93
100	0.96	0.95	0.95	0.90	100	0.96	0.95	0.92	0.96
200	0.96	0.95	0.94	0.94	200	0.97	0.98	0.93	0.93
500	0.94	0.94	0.93	0.91	500	0.97	0.94	0.93	0.91

Normaal					Profiel				
Normaliteit	blokgruotte (n)				Normaliteit	blokgruotte (n)			
$q_{0,999}$	50	100	200	500	$q_{0,999}$	50	100	200	500
50	0.87	0.90	0.94	0.95	50	0.80	0.94	0.93	0.92
100	0.90	0.94	0.92	0.95	100	0.92	0.89	0.96	0.95
200	0.93	0.93	0.94	0.96	200	0.93	0.92	0.92	0.95
500	0.91	0.93	0.95	0.94	500	0.94	0.90	0.91	0.96

Tabel 4.3.12: OW's vir 95%-VI's vir $q_{0,99}$ en $q_{0,999}$ van die normaalverdeling.

Logisties

Normaliteit	blokgrootte (n)				Profiel	blokgrootte (n)				
	$q_{0.99}$	50	100	200		500	$q_{0.99}$	50	100	200
50	0.93	0.94	0.94	0.90	50	0.93	0.93	0.95	0.95	
100	0.94	0.93	0.94	0.94	100	0.94	0.92	0.92	0.97	
200	0.95	0.93	0.96	0.93	200	0.92	0.95	0.98	0.91	
500	0.96	0.94	0.93	0.95	500	0.97	0.97	0.95	0.96	

Normaliteit	blokgrootte (n)				Profiel	blokgrootte (n)				
	$q_{0.999}$	50	100	200		500	$q_{0.999}$	50	100	200
50	0.90	0.90	0.90	0.94	50	0.95	0.92	0.96	0.94	
100	0.93	0.93	0.93	0.96	100	0.93	0.93	0.91	0.93	
200	0.94	0.93	0.94	0.94	200	0.98	0.95	0.91	0.94	
500	0.94	0.95	0.94	0.97	500	0.96	0.99	0.98	0.95	

Tabel 4.3.13: OW's vir 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ van die logistiese verdeling.

Normaal

Normaliteit	blokgrootte (n)				Profiel	blokgrootte (n)				
	$q_{0.99}$	50	100	200		500	$q_{0.99}$	50	100	200
aantal blokke	50	0.29	0.23	0.21	0.35	50	0.29	0.23	0.22	0.35
aantal blokke	100	0.20	0.16	0.15	0.23	100	0.20	0.16	0.15	0.25
aantal blokke	200	0.14	0.11	0.11	0.16	200	0.14	0.12	0.11	0.16
aantal blokke	500	0.09	0.07	0.07	0.10	500	0.08	0.07	0.07	0.10

Normaliteit	blokgrootte (n)				Profiel	blokgrootte (n)				
	$q_{0.999}$	50	100	200		500	$q_{0.999}$	50	100	200
aantal blokke	50	0.57	0.41	0.31	0.23	50	0.68	0.46	0.32	0.22
aantal blokke	100	0.39	0.29	0.22	0.16	100	0.42	0.30	0.22	0.16
aantal blokke	200	0.27	0.20	0.15	0.11	200	0.29	0.21	0.16	0.11
aantal blokke	500	0.17	0.13	0.10	0.07	500	0.17	0.13	0.10	0.07

Tabel 4.3.14: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die normaalverdeling.

Logisties

Normaliteit					Profiel						
	$q_{0.99}$	blokgrootte (n)					$q_{0.99}$	blokgrootte (n)			
		50	100	200	500			50	100	200	500
	50	0.76	0.62	0.56	0.98		50	0.76	0.61	0.60	1.07
aantal	100	0.54	0.44	0.40	0.66	aantal	100	0.54	0.44	0.42	0.69
blokke	200	0.38	0.31	0.29	0.46	blokke	200	0.38	0.31	0.29	0.47
	500	0.24	0.20	0.18	0.29		500	0.24	0.20	0.18	0.29

Normaliteit					Profiel						
	$q_{0.999}$	blokgrootte (n)					$q_{0.999}$	blokgrootte (n)			
		50	100	200	500			50	100	200	500
	50	2.01	1.40	1.03	0.75		50	2.43	1.49	1.10	0.76
aantal	100	1.42	0.99	0.73	0.53	aantal	100	1.53	1.08	0.76	0.54
blokke	200	0.98	0.70	0.52	0.38	blokke	200	1.02	0.74	0.54	0.38
	500	0.61	0.44	0.33	0.24		500	0.63	0.45	0.33	0.24

Tabel 4.3.15: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die logistiese verdeling.

Die gedrag van die gammaverdeling is heel anders as dié van die ander twee lede van die Gumbel-domein. Hier is die profiel-OW's merkbaar hoër as vir die normaalvertrouensintervalle (sien tabel 4.3.16); soms selfs so hoog as 0.98 en 0.99. Die afruiling ter sprake is egter dat die gemiddelde profielinterval (sien tabel 4.3.17) effens langer is as in die normaalgeval.

Gamma

Normaliteit					Profiel						
	$q_{0.99}$	blokgrootte (n)					$q_{0.99}$	blokgrootte (n)			
		50	100	200	500			50	100	200	500
	50	0.93	0.92	0.94	0.92		50	0.97	0.95	0.96	0.98
	100	0.94	0.95	0.96	0.94		100	0.97	0.97	0.97	0.97
	200	0.95	0.94	0.93	0.94		200	0.93	0.98	0.98	0.94
	500	0.95	0.95	0.95	0.93		500	0.99	0.99	0.92	0.94

Normaliteit					Profiel						
	$q_{0.999}$	blokgrootte (n)					$q_{0.999}$	blokgrootte (n)			
		50	100	200	500			50	100	200	500
	50	0.90	0.90	0.94	0.93		50	0.99	0.99	0.94	0.95
	100	0.92	0.94	0.94	0.93		100	0.97	0.92	0.96	0.97
	200	0.92	0.95	0.95	0.95		200	0.97	0.95	0.96	0.97
	500	0.92	0.96	0.95	0.95		500	0.96	0.96	0.98	0.97

Tabel 4.3.16: OW's vir 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ van die gammaverdeling.

Gamma

Normaliteit	$q_{0.99}$	blokgrötte (n)				Profiel	$q_{0.99}$	blokgrötte (n)			
		50	100	200	500			50	100	200	500
	50	0.96	0.78	0.71	1.22		50	0.97	0.79	0.75	1.31
aantal	100	0.68	0.55	0.51	0.84	aantal	100	0.68	0.56	0.51	0.85
blokke	200	0.48	0.39	0.35	0.57	blokke	200	0.49	0.39	0.36	0.71
	500	0.30	0.25	0.23	0.35		500	0.30	0.25	0.23	0.57

Normaliteit	$q_{0.999}$	blokgrötte (n)				Profiel	$q_{0.999}$	blokgrötte (n)			
		50	100	200	500			50	100	200	500
	50	2.38	1.66	1.21	0.91		50	2.94	1.87	1.31	0.92
aantal	100	1.67	1.18	0.87	0.64	aantal	100	1.84	1.25	0.90	0.65
blokke	200	1.14	0.83	0.62	0.45	blokke	200	1.24	0.83	0.63	0.55
	500	0.72	0.53	0.39	0.29		500	0.75	0.53	0.39	0.45

Tabel 4.3.17: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die gammaverdeling.

Ten opsigte van puntberamings van $q_{0.99}$ en $q_{0.999}$ blyk dat in die geval van die normaalverdeling (tabel 4.3.18) beide $q_{0.99}$ en $q_{0.999}$ by klein blokgröttes tipies onderberaam en by groter blokgröttes oorberaam word. Vir die logistiese verdeling word $q_{0.999}$ tipies onderberaam. Hierdie kwantiel word ook by die gammaverdeling onderberaam, terwyl $q_{0.99}$ vir hierdie verdeling oorberaam word. Sien tabel 4.3.19 vir die resultate rakende die logistiese en gammaverdeling.

Normaal


Mediaan		blokgrötte (n)				Gemiddelde		blokgrötte (n)			
$q_{0,99}$		50	100	200	500	$q_{0,99}$		50	100	200	500
	50	-0.0010	0.0070	-0.0010	0.0017		50	-0.001	0.003	-0.002	0.002
aantal	100	0.0030	0.0010	-0.0005	-0.0033	aantal	100	0.001	0.001	0.001	-0.001
blokke	200	0.0020	0.0040	0.0010	-0.0046	blokke	200	0.001	0.004	0.001	-0.004
	500	0.0003	0.0050	0.0036	-0.0065		500	0.000	0.006	0.003	-0.007

Mediaan		blokgrötte (n)				Gemiddelde		blokgrötte (n)			
$q_{0,999}$		50	100	200	500	$q_{0,999}$		50	100	200	500
	50	0.017	0.007	0.001	-0.003		50	0.016	0.013	0.001	-0.004
aantal	100	0.001	0.007	-0.001	0.002	aantal	100	0.012	0.001	-0.002	-0.002
blokke	200	0.001	0.000	-0.002	-0.001	blokke	200	0.001	-0.001	-0.003	-0.002
	500	-0.008	-0.006	-0.004	-0.00006		500	-0.007	-0.005	-0.003	-0.001

Tabel 4.3.18: Mediaan en gemiddelde afwyking vir $q_{0,99}$ en $q_{0,999}$ vir die normaalverdeling.

Logisties

		blokgrötte (n)			
$q_{0,99}$		50	100	200	500
	50	0.015	-0.019	-0.012	-0.020
aantal	100	0.000	-0.008	-0.005	-0.013
blokke	200	0.002	0.003	-0.001	-0.007
	500	0.000	-0.001	0.001	-0.007



Gamma

		blokgrötte (n)			
$q_{0,99}$		50	100	200	500
	50	0.000	-0.012	-0.028	-0.026
aantal	100	-0.003	-0.004	-0.008	0.017
blokke	200	-0.007	-0.010	-0.003	-0.024
	500	0.008	-0.028	-0.004	-0.013

		blokgrötte (n)						blokgrötte (n)			
$q_{0,999}$		50	100	200	500	$q_{0,999}$		50	100	200	500
	50	0.068	0.024	0.048	0.001		50	0.080	0.060	0.049	0.002
aantal	100	0.012	0.026	0.008	0.005	aantal	100	0.040	0.044	0.018	-0.001
blokke	200	0.035	0.016	0.006	0.008	blokke	200	0.025	0.002	0.017	0.007
	500	0.011	0.020	0.003	0.006		500	0.028	0.049	-0.001	-0.002

Tabel 4.3.19: Mediaan afwyking vir $q_{0,99}$ en $q_{0,999}$ vir die logistiese en gammaverdeling.

4.3.3 Die Weibull-domein: $\gamma < 0$

In die Weibull-domein word die betaverdeling beskou. Vir hierdie verdeling is van die mees uiteenlopende simulاسie-afvoer verkry: Ten opsigte van VI's vir $q_{0,99}$ is OW's van tussen 0.84 en 0.97 met behulp van die profielmetode verkry; vir vertrouensintervalle met betrekking tot $q_{0,999}$ is OW's van tussen 0.86 en 0.94 met die normaalmetode verkry. Sien hiervoor tabel 4.3.20. Veral by die profielmetode was daar aansienlike variasie in OW's oor die verskillende kombinasies van blokaantalle en blokgroottes. By die betaverdeling vaar die normaal-VI's die swakker in terme van oordeckingswaarskynlikheid.

Beta									
Normaliteit	blokgrоotte (n)				Profiel	blokgrоotte (n)			
$q_{0,99}$	50	100	200	500	$q_{0,99}$	50	100	200	500
50	0.91	0.93	0.90	0.94	50	0.89	0.91	0.84	0.84
100	0.95	0.94	0.95	0.94	100	0.87	0.95	0.92	0.87
200	0.93	0.93	0.95	0.95	200	0.93	0.95	0.95	0.97
500	0.92	0.90	0.87	0.94	500	0.93	0.96	0.94	0.94
Normaliteit	blokgrоotte (n)				Profiel	blokgrоotte (n)			
$q_{0,999}$	50	100	200	500	$q_{0,999}$	50	100	200	500
50	0.86	0.89	0.92	0.93	50	0.95	0.92	0.87	0.83
100	0.89	0.93	0.93	0.94	100	0.90	0.94	0.95	0.92
200	0.92	0.94	0.94	0.92	200	0.99	0.90	0.93	0.93
500	0.93	0.92	0.91	0.92	500	0.94	0.99	0.93	0.99

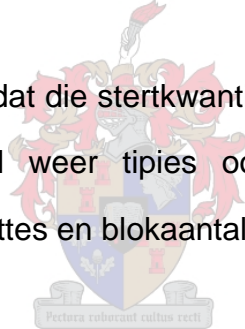
Tabel 4.3.20: OW's vir 95%-VI's vir $q_{0,99}$ en $q_{0,999}$ van die betaverdeling.

Soos die OW's variëer die VI-lengtes aansienlik. `n Lae/hoë OW gaan hier deurgaans met `n korter/langer gemiddelde intervallengte gepaard. Sien tabel 4.3.21.

Beta											
Normaliteit		blokgrötte (n)				Profiel		blokgrötte (n)			
	$q_{0,99}$	50	100	200	500		$q_{0,99}$	50	100	200	500
	50	0.019	0.015	0.005	0.023		50	0.018	0.014	0.013	0.015
aantal	100	0.013	0.011	0.010	0.015	aantal	100	0.013	0.010	0.009	0.011
blokke	200	0.009	0.007	0.007	0.011	blokke	200	0.010	0.008	0.007	0.009
	500	0.006	0.005	0.000	0.006		500	0.006	0.005	0.004	0.006
Normaliteit		blokgrötte (n)				Profiel		blokgrötte (n)			
	$q_{0,999}$	50	100	200	500		$q_{0,999}$	50	100	200	500
	50	0.018	0.013	0.005	0.008		50	0.023	0.015	0.010	0.008
aantal	100	0.013	0.009	0.008	0.006	aantal	100	0.014	0.010	0.008	0.006
blokke	200	0.009	0.007	0.006	0.004	blokke	200	0.010	0.007	0.006	0.004
	500	0.006	0.005	0.000	0.003		500	0.006	0.005	0.004	0.003

Tabel 4.3.21: Gemiddelde VI-lengtes vir $q_{0,99}$ en $q_{0,999}$ van die betaverdeling.

Dit is duidelik (sien tabel 4.3.22) dat die stertkwantiel $q_{0,99}$ sonder uitsondering effens onderberaam word. $q_{0,999}$ word weer tipies oorberaam; dit afgesien van die kombinasies van kleiner blokgröttes en blokaantalle waar $q_{0,999}$ onderberaam word.



Beta											
Mediaan		blokgrötte (n)				Gemiddeld		blokgrötte (n)			
	$q_{0,99}$	50	100	200	500		$q_{0,99}$	50	100	200	500
	50	0.0003	0.0005	0.0009	0.0002		50	0.0005	0.0004	0.0009	0.0006
aantal	100	0.0005	0.0005	0.0008	0.0005	aantal	100	0.0007	0.0005	0.0006	0.0008
blokke	200	0.0005	0.0007	0.0008	0.0005	blokke	200	0.0006	0.0007	0.0008	0.0006
	500	0.0009	0.0009	0.0009	0.0005		500	0.0008	0.0009	0.0009	0.0005
Mediaan		blokgrötte (n)				Gemiddeld		blokgrötte (n)			
	$q_{0,999}$	50	100	200	500		$q_{0,999}$	50	100	200	500
	50	0.0012	0.0003	-0.0006	-0.0005		50	0.0014	0.0006	-0.0006	-0.0005
aantal	100	0.0001	0.0000	-0.0003	-0.0004	aantal	100	0.0004	-0.0001	-0.0002	-0.0005
blokke	200	-0.0003	-0.0004	-0.0005	-0.0006	blokke	200	-0.0002	-0.0004	-0.0005	-0.0005
	500	-0.0005	-0.0006	-0.0005	-0.0004		500	-0.0005	-0.0006	-0.0005	-0.0004

Tabel 4.3.23: Mediaan en gemiddelde afwyking vir $q_{0,99}$ en $q_{0,999}$, betaverdeling.

4.4 Gevolgtrekking

Ekstreemwaarde-teorie is 'n metodologie wat op die gebruik van asimptotiese argumente berus. Hierdie teorie bestaan uit twee breë benaderings, naamlik die klassieke blokmaksima-metode en die meer moderne drempelmodellering. Die hoofresultaat waarop die maksima-teorie berus, is die Fisher-Tippett stelling. Hiervolgens kan die verdelingsfunksie van M_n , naamlik F^n , deur die veralgemeende ekstreemwaarde verdeling $G_{\gamma,\mu,\sigma}$ benader kan word vir groot waardes van die steekproefgrootte n , oftewel $F^n \approx G_{\gamma,\mu,\sigma}$ as $n \rightarrow \infty$. Ons het in hierdie hoofstuk die volgende vraag ondersoek: Hoe goed is die benadering $F^n \approx G_{\gamma,\mu,\sigma}$ vir eindige n ?

Die beraming van ekstreme stertkwantiele van die onderliggende verdeling F en vertrouensintervalle vir hierdie stertkwantiele is een van die belangrikste doelwitte van 'n ekstreemwaarde-analise. Gevolglik is die kwessie van die geldigheid van die benadering $F^n \approx G_{\gamma,\mu,\sigma}$ waar n eindige waardes aanneem, in terme van die akkuraatheid van hierdie beramings geëvalueer. Daar is van 'n simulasiestudie gebruik gemaak om die gedrag van maksima-inferensie by eindige steekproewe te ondersoek. Van 'n goeie benadering verwag ons die beraamde kwantiele nie te veel van die werklike kwantiele sal afwyk nie en ook nie oormatig variëer nie. Ten opsigte van die (95%-) vertrouensintervalle verwag ons oordekkingswaarskynlikhede van ongeveer 0.95, sowel as 'n kort intervallengte.

Die simulatie-afvoer suggereer dat daar in die puntberaming van die stertkwantiele $q_{0.99}$ en $q_{0.999}$ wel 'n mate van sydigheid teenwoordig is. In die geval van swaarder stert verdelings soos die streng en die veralgemeende Pareto-verdelings kan die praktisyn byvoorbeeld 'n onderberaming van $q_{0.99}$ by klein blokgroottes soos 50 en 100 verwag, maar 'n oorberaming hiervan by groot blokgroottes soos 200 en 500. Hierdie resultate variëer grootliks onder die verdelings wat beskou is. Dit geld egter vir al hierdie verdelings dat die variasie in die puntberamings van $q_{0.99}$ en $q_{0.999}$ - soos gemeet aan die SA, MAE en MAD - by 'n vaste blokgroutte verminder kan word deur die aantal blokke (en dus die totale steekproefgrootte) te laat toeneem.

Twee metodes, naamlik die asimptotiese normaliteit van die MA-beramers sowel as die chikwadraatgebaseerde profielmetode, is gebruik om 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ te konstrueer. In terme van oordekkingswaarskynlikheid wil dit voorkom of die normaalmetode oor al die verdelings onder beskouing, die meer betroubare benadering is. Die normaal-OW's was weliswaar tipies effens kleiner as 0.95, maar stabiel bo 0.90. Die heel laagste normaal-OW was 0.87, soos aangeteken by die betaverdeling. Die profielgebaseerde OW's is ongeveer ewe geneig om bo en onder 0.95 te wees en variëer aansienlik. Profiel-OW's van so hoog as 0.99 is bereik, terwyl 'n laagste waarde van 0.49 by die swaarstert Burr-verdeling aangeteken is. Onaanvaarbare lae OW's is ook met die profielmetode vir die streng en veralgemeende Pareto-verdelings verkry. Dit wil suggereer dat ons gewoonlik meer vertrou in die normaal-VI's kan hê in die sin dat die OW van 'n beraamde 95%-VI tipies nie veel van 0.95 sal afwyk nie. Veral wanneer die onderliggende verdeling 'n baie swaar stert het, bied die profielmetode veral by kleiner blokgroottes OW's wat soms baie laer as 0.95 is. Soos bespreek in paragraaf 4.3.1, is dit veiliger om hier eerder die normaalmetode te gebruik.

'n Alternatiewe benadering tot die maksimummetode is drempelteorie, waarvolgens oorskrydings van 'n drempel – eerder as blokmaksima – as ekstreme waardes gedefinieer word. In die volgende hoofstuk evalueer ons die geldigheid van drempelgebaseerde inferensie ten opsigte van 'n onderliggende verdeling. Soos in hierdie hoofstuk, word die akkuraatheid van beramings van ekstreme kwantiele en VI's vir dié kwantiele weer eens met behulp van 'n simulasiestudie beoordeel.



Aanhangsel 4A

In hierdie aanhangsel word die volledige resultate van die simulasiestudie soos in paragraaf 4.2 uiteengesit, verskaf. Die belangrikste faktor in dié studie is die keuse van die onderliggende verdeling F . Om hierdie rede word die afvoer gegroepeer volgens die maksimum domein van aantrekking binne die veralgemeende ekstreemwaarde-verdeling $G_{\gamma, \mu, \sigma}$ waartoe die onderliggende verdelings onder beskouing, behoort. Ons gee die resultate vir die verdelings in die Frechet-domein in paragraaf 4A.1, vir die Gumbel-klas in paragraaf 4A.2 en vir die verdelings in die Weibull-domein, in paragraaf 4A.3.

Benaderde 95%-vertrouensintervalle (VI's) vir die stertkwantiele $q_{0.99}$ en $q_{0.999}$ van die onderliggende verdelings is met behulp van die asimptotiese normaliteit van die MA-beramers, sowel as die profiel aanneemlikheid metode, verkry. Ons stel in die oordekkingswaarskynlikheid (OW), sowel as maatstawwe van die lengte van die berekende VI's belang. Hierdie maatstawwe is die gemiddelde en mediaan lengtes, die standaardafwyking van die VI-lengtes en die MAD (paragraaf 4.2). Die afvoer rakende die berekende VI's word in die subparagrafe 4A.1.1, 4A2.1 en 4A3.1 gegee.

Benewens VI's vir $q_{0.99}$ en $q_{0.999}$ is ook puntberamings vir hierdie kwantiele verkry. In hierdie verband is die gemiddelde en mediaan kwantielafwyking bereken, sowel as die standaardafwyking daarvan, die MAD en MAE (paragraaf 4.2). Ons verskaf dié resultate in die subparagrafe 4A.1.2, 4A2.2 en 4A3.2.

Let op dat in die geval van kombinasies van blokgroottes en blokaantalle waar die simulatie nie suksesvol deurgevoer kon word nie, dit met 'n tilde (~) aangedui is.

4A.1 Die Frechet-domein: $\gamma > 0$

Ons gee in hierdie paragraaf die simulasieresultate vir die verdelings veralgemeende Pareto, streng Pareto en Burr. Hierdie swaarstert verdelings behoort tot die Frechet maksimum domein van aantrekking.

4A1.1 Vertrauensinterval beraming vir die Frechet-domein

OW's van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$: Asimptotiese Normaliteit

Veralgemeende Pareto					Pareto						
		blok grootte (n)						blok grootte (n)			
	$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500	
aantal	50	0.93	0.93	0.92	0.89	aantal	50	0.94	0.95	0.93	0.90
blokke	100	0.95	0.95	0.93	0.91	blokke	100	0.93	0.95	0.94	0.91
	200	0.96	0.95	0.95	0.94	blokke	200	0.93	0.94	0.95	0.95
	500	0.96	0.96	0.95	0.94	blokke	500	0.95	0.97	0.95	0.94

Pareto					
		blok grootte (n)			
	$q_{0.999}$	50	100	200	500
aantal	50	0.91	0.91	0.93	0.94
blokke	100	0.92	0.95	0.94	0.93
	200	0.96	0.93	0.93	0.94
	500	0.95	0.95	0.95	0.93

Burr

		blok grootte (n)			
	$q_{0.99}$	50	100	200	500
aantal	50	~	0.93	0.93	~
blokke	100	~	0.96	0.94	0.92
	200	~	0.97	0.96	0.94
	500	~	0.95	0.94	0.97

		blok grootte (n)			
	$q_{0.999}$	50	100	200	500
aantal	50	~	0.90	0.92	~
blokke	100	~	0.94	0.91	0.94
	200	~	0.96	0.94	0.95
	500	~	0.91	0.94	0.95

Gemiddelde (links) en Mediaan (regs) lengtes van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$:

Asimptotiese Normaliteit

Veralgemeende Pareto

		blokgrötte (n)						blokgrötte (n)					
		50	100	200	500			50	100	200	500		
$q_{0.99}$		50	7.45	6.24	5.57	10.50	$q_{0.99}$		50	7.33	6.19	5.52	10.12
aantal	50	7.45	6.24	5.57	5.05	4.50	aantal	50	7.33	6.19	5.52	5.00	4.50
blokke	100	5.24	4.47	4.02	3.12	2.75	blokke	100	5.21	4.43	4.00	3.11	2.75
	200	3.74	3.12	2.84	2.35	2.12		200	3.72	3.11	2.85	2.34	2.12
	500	2.35	1.98	1.80	1.56	1.42		500	2.34	1.98	1.79	1.56	1.42
$q_{0.999}$		50	71.78	48.95	35.62	23.57	$q_{0.999}$		50	65.86	44.52	33.94	23.10
aantal	50	71.78	48.95	35.62	24.45	16.83	aantal	50	65.86	44.52	33.94	23.07	16.61
blokke	100	47.13	34.39	24.45	16.81	11.86	blokke	100	44.31	33.02	23.87	16.61	11.84
	200	33.06	23.52	16.81	11.86	7.47		200	32.34	23.40	16.78	11.84	7.47
	500	20.26	14.91	10.77	7.47	4.47		500	19.97	14.77	10.73	7.47	4.47

Pareto

		blokgrötte (n)						blokgrötte (n)					
		50	100	200	500			50	100	200	500		
$q_{0.99}$		50	3.72	3.17	2.79	5.19	$q_{0.99}$		50	3.68	3.12	2.75	4.98
aantal	50	3.72	3.17	2.79	2.03	3.56	aantal	50	3.68	3.12	2.75	2.01	3.50
blokke	100	2.61	2.22	2.03	1.42	2.52	blokke	100	2.59	2.22	2.01	1.42	2.50
	200	1.86	1.56	1.42	0.90	1.56		200	1.85	1.56	1.42	0.90	1.56
	500	1.17	0.99	0.90	0.90	1.56		500	1.16	0.99	0.90	0.90	1.56
$q_{0.999}$		50	36.69	25.50	17.65	11.73	$q_{0.999}$		50	30.98	23.90	16.74	11.58
aantal	50	36.69	25.50	17.65	12.41	8.39	aantal	50	30.98	23.90	16.74	11.99	8.33
blokke	100	23.42	17.29	12.41	8.58	5.90	blokke	100	21.97	16.74	11.99	8.50	5.87
	200	16.46	11.83	8.58	5.90	3.73		200	15.80	11.77	8.50	5.87	3.74
	500	9.87	7.38	5.42	3.73	3.73		500	9.69	7.32	5.41	3.74	3.74

Burr

		blokgrötte (n)						blokgrötte (n)					
		50	100	200	500			50	100	200	500		
$q_{0.99}$		50	~	9.09	8.41	~	$q_{0.99}$		50	~	8.75	8.35	~
aantal	50	~	9.09	8.41	5.90	11.36	aantal	50	~	8.75	8.35	5.87	10.91
blokke	100	~	6.50	5.90	4.12	7.73	blokke	100	~	6.44	5.87	4.09	7.64
	200	~	4.53	4.12	2.61	4.16		200	~	4.48	4.09	2.62	4.59
	500	~	2.84	2.61	2.61	4.16		500	~	2.86	2.62	2.62	4.59
$q_{0.999}$		50	~	41.47	39.80	~	$q_{0.999}$		50	~	38.86	38.50	~
aantal	50	~	41.47	39.80	31.80	32.66	aantal	50	~	38.86	38.50	31.38	32.59
blokke	100	~	35.40	31.80	24.95	24.95	blokke	100	~	34.54	31.38	25.59	24.78
	200	~	26.88	25.74	19.57	15.53		200	~	26.46	25.59	19.49	15.49
	500	~	19.31	19.57	15.53	15.53		500	~	19.11	19.49	15.49	15.49

Standaardafwyking (links) en MAD (regs) van die lengtes van 95%-VI's vir

$q_{0,99}$ en $q_{0,999}$: Asimptotiese Normaliteit

Veralgemeende Pareto

		blokgrootte (n)						blokgrootte (n)					
		$q_{0,99}$	50	100	200	500			$q_{0,99}$	50	100	200	500
aantal	50		1.37	1.01	1.00	3.32	aantal	50		0.92	0.70	0.66	1.88
	100		0.72	0.50	0.53	1.49		100		0.45	0.32	0.35	0.86
	200		0.35	0.25	0.24	0.67		200		0.23	0.17	0.16	0.48
	500		0.14	0.10	0.09	0.27		500		0.09	0.07	0.06	0.17
blokke	50		35.97	19.72	10.78	4.26	blokke	50		18.84	11.66	6.98	2.67
	100		15.88	9.55	5.20	2.32		100		8.72	6.43	3.52	1.37
	200		7.56	4.40	2.35	1.07		200		5.04	2.95	1.52	0.65
	500		3.11	1.79	1.00	0.46		500		1.96	1.23	0.66	0.33

Pareto

		blokgrootte (n)						blokgrootte (n)					
		$q_{0,99}$	50	100	200	500			$q_{0,99}$	50	100	200	500
aantal	50		0.73	0.53	0.51	1.48	aantal	50		0.48	0.35	0.36	0.90
	100		0.35	0.26	0.25	0.66		100		0.24	0.18	0.17	0.45
	200		0.19	0.12	0.12	0.32		200		0.13	0.08	0.09	0.22
	500		0.02	0.05	0.05	0.13		500		0.04	0.03	0.03	0.09
blokke	50		21.20	11.33	5.87	2.29	blokke	50		9.76	6.94	3.98	1.59
	100		8.13	4.84	2.66	1.08		100		4.77	3.21	1.67	0.76
	200		3.94	2.22	1.29	0.57		200		2.54	1.57	0.85	0.35
	500		1.40	0.85	0.48	0.23		500		0.90	0.58	0.32	0.15

Burr

		blokgrootte (n)						blokgrootte (n)					
		$q_{0,99}$	50	100	200	500			$q_{0,99}$	50	100	200	500
aantal	50		~	1.61	1.50	~	aantal	50		~	1.21	1.00	~
	100		~	0.81	0.78	2.17		100		~	0.55	0.51	1.41
	200		~	0.47	0.55	1.04		200		~	0.32	0.29	0.71
	500		~	0.15	0.15	0.38		500		~	0.10	0.10	0.26
blokke (N)	50		~	55.96	28.11	~	blokke (N)	50		~	31.90	15.40	~
	100		~	22.17	19.46	5.64		100		~	14.62	8.51	3.67
	200		~	12.45	13.56	2.77		200		~	7.56	6.01	1.83
	500		~	4.51	2.51	1.10		500		~	2.84	1.68	0.71

OW's van 95%-VI's vir $q_{0,99}$ en $q_{0,999}$: Profielmetode

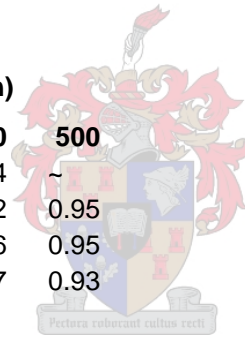
Veralgemeende Pareto					Pareto						
blokgrötte (n)					blokgrötte (n)						
	$q_{0,99}$	50	100	200	500		$q_{0,99}$	50	100	200	500
aantal	50	0.92	0.96	0.93	0.92	aantal	50	0.95	0.95	0.94	~
blokke	100	0.94	0.96	0.97	0.98	blokke	100	0.95	0.95	0.94	0.97
	200	0.96	0.93	0.93	0.94		200	0.96	0.94	0.99	0.94
	500	0.96	0.97	0.94	0.92		500	0.98	0.93	0.94	0.92

blokgrötte (n)					blokgrötte (n)						
	$q_{0,999}$	50	100	200	500		$q_{0,999}$	50	100	200	500
aantal	50	0.64	0.78	0.96	0.87	aantal	50	0.67	0.82	0.93	~
blokke	100	0.70	0.89	0.94	0.97	blokke	100	0.68	0.88	0.93	0.92
	200	0.80	0.96	0.99	0.97		200	0.83	0.88	0.96	0.95
	500	0.85	0.96	0.97	0.99		500	0.88	0.97	0.97	0.93

Burr

blokgrötte (n)					
	$q_{0,99}$	50	100	200	500
aantal	50	~	0.90	0.94	~
blokke	100	~	0.94	0.92	0.95
	200	~	0.92	0.96	0.95
	500	~	0.93	0.97	0.93

blokgrötte (n)					
	$q_{0,999}$	50	100	200	500
aantal	50	~	0.49	0.77	~
blokke	100	~	0.67	0.76	0.91
	200	~	0.70	0.85	0.97
	500	~	0.72	0.93	0.90



Gemiddelde (links) en Mediaan (regs) lengtes van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$:

Profielmetode

Veralgemeende Pareto

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500			$q_{0.99}$	50	100	200	500
aantal blokke	50	7.13	6.39	5.98	11.27	aantal blokke	50	6.98	6.33	5.82	10.80		
	100	5.32	4.39	4.09	7.37		100	5.26	4.37	4.11	7.23		
	200	3.76	3.11	2.86	5.09		200	3.72	3.12	2.87	4.99		
	500	2.34	1.98	1.82	3.16		500	2.34	1.98	1.83	3.14		
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
aantal blokke	50	28.37	26.49	25.38	22.90	aantal blokke	50	26.30	25.54	24.54	23.64		
	100	22.93	21.32	21.01	16.85		100	22.09	21.02	20.75	16.88		
	200	18.27	17.56	16.56	11.83		200	17.82	17.69	16.53	11.75		
	500	13.57	13.45	11.01	7.51		500	13.52	13.37	10.89	7.52		

Pareto

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500			$q_{0.99}$	50	100	200	500
aantal blokke	50	3.55	3.18	2.92	~	aantal blokke	50	3.48	3.15	2.91	~		
	100	2.65	2.20	2.92	4.01		100	2.63	2.17	2.91	3.88		
	200	1.88	1.56	1.42	2.53		200	1.90	1.55	1.42	2.49		
	500	1.18	1.00	0.90	1.59		500	1.18	0.99	0.90	1.59		
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
aantal blokke	50	14.37	13.70	12.80	~	aantal blokke	50	13.49	13.30	12.60	~		
	100	11.38	10.80	12.80	8.41		100	10.87	10.70	12.60	8.37		
	200	9.06	8.86	8.36	5.95		200	9.02	8.70	8.43	5.90		
	500	6.77	6.88	5.41	3.73		500	6.75	6.87	5.37	3.73		

Burr

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500			$q_{0.99}$	50	100	200	500
aantal blokke	50	~	9.09	8.41	~	aantal blokke	50	~	8.75	8.35	~		
	100	~	6.50	5.90	11.36		100	~	6.44	5.87	10.91		
	200	~	4.53	4.12	7.73		200	~	4.48	4.09	7.64		
	500	~	2.84	2.61	4.16		500	~	2.86	2.62	4.59		
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
aantal blokke	50	~	41.47	39.80	~	aantal blokke	50	~	38.86	38.50	~		
	100	~	35.40	31.80	32.66		100	~	34.54	31.38	32.59		
	200	~	26.88	25.74	24.95		200	~	26.46	25.59	24.78		
	500	~	19.31	19.57	15.53		500	~	19.11	19.49	15.49		

Standaardafwyking (links) en MAD (regs) van die lengtes van 95%-VI's vir

$q_{0.99}$ en $q_{0.999}$: Profielmetode.

Veralgemeende Pareto

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500		
aantal blokke	50	1.33	1.10	1.23	3.40	aantal blokke	50	0.85	0.71	0.07	1.75		
	100	0.69	0.46	0.45	1.31		100	0.45	0.28	0.30	0.71		
	200	0.36	0.24	0.24	0.71		200	0.21	0.16	0.15	0.47		
	500	0.14	0.10	0.10	0.22		500	0.09	0.06	0.06	0.15		
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
aantal blokke	50	8.31	5.42	4.16	4.04	aantal blokke	50	3.96	2.97	2.71	3.02		
	100	4.51	2.49	2.49	2.53		100	2.37	1.33	1.51	1.86		
	200	2.23	1.55	1.69	1.04		200	1.10	1.31	1.19	0.87		
	500	0.83	0.92	0.90	0.44		500	0.59	0.59	0.62	0.29		

Pareto

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500		
aantal blokke	50	0.62	0.46	0.52	~	aantal blokke	50	0.41	0.31	0.35	~		
	100	0.37	0.26	0.52	0.91		100	0.25	0.15	0.35	0.56		
	200	0.18	0.12	0.11	0.33		200	0.10	0.10	0.07	0.22		
	500	0.07	0.05	0.05	0.15		500	0.04	0.04	0.03	0.11		
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
aantal blokke	50	4.11	3.25	2.34	~	aantal blokke	50	1.92	1.72	1.69	~		
	100	2.29	1.44	2.34	1.16		100	1.22	0.91	1.69	0.71		
	200	0.94	0.82	0.82	0.62		200	0.58	0.62	0.47	0.36		
	500	0.40	0.46	0.42	0.22		500	0.28	0.36	0.29	0.14		

Burr

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500		
aantal blokke	50	~	1.79	1.58	~	aantal blokke	50	~	0.84	0.98	~		
	100	~	0.75	0.78	2.22		100	~	0.52	0.54	1.37		
	200	~	0.43	0.37	1.07		200	~	0.29	0.23	0.72		
	500	~	0.15	0.14	0.40		500	~	0.11	0.11	0.25		
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
aantal blokke	50	~	10.86	6.60	~	aantal blokke	50	~	5.31	3.96	~		
	100	~	7.09	4.23	4.59		100	~	4.46	2.87	2.61		
	200	~	3.68	2.31	2.26		200	~	2.17	1.70	1.62		
	500	~	1.41	1.41	1.15		500	~	0.93	0.94	0.84		

4A1.2 Kwantielberaming vir die Frechet-domein

Die Mediaan (links) en Gemiddelde kwantielafwyking (regs) ten opsigte van

$q_{0.99}$ en $q_{0.999}$:

Veralgemeende Pareto

blokgrötte (n)					blokgrötte (n)						
$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500		
aantal	50	0.12	0.01	-0.14	-0.69	aantal	50	0.04	-0.04	-0.35	0.70
blokke	100	0.16	-0.12	-0.06	-0.44	blokke	100	0.08	-0.13	-0.13	-0.38
	200	0.04	0.03	-0.07	-0.19		200	-0.01	-0.02	-0.09	-0.22
	500	0.03	0.03	-0.05	-0.10		500	0.03	-0.01	-0.07	-0.14
$q_{0.999}$	50	100	200	500	$q_{0.999}$	50	100	200	500		
aantal	50	-0.37	1.47	0.11	0.49	aantal	50	-2.54	-0.16	-0.66	0.18
blokke	100	0.31	-0.19	-0.02	0.28	blokke	100	-1.05	-0.89	-0.34	0.00
	200	-0.84	-0.10	0.29	-0.11		200	-1.45	-0.04	0.25	-0.19
	500	-0.08	-0.20	-0.14	0.13		500	-0.60	-0.30	-0.12	0.12

Pareto

blokgrötte (n)					blokgrötte (n)						
$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500		
aantal	50	0.16	0.00	-0.08	-0.25	aantal	50	0.02	-0.10	-0.14	-0.31
blokke	100	0.04	-0.02	-0.04	-0.21	blokke	100	0.03	-0.02	-0.08	-0.21
	200	0.03	0.00	-0.02	-0.09		200	0.03	0.00	-0.03	-0.11
	500	0.03	0.01	-0.01	-0.05		500	0.02	-0.01	-0.01	-0.04
$q_{0.999}$	50	100	200	500	$q_{0.999}$	50	100	200	500		
aantal	50	0.24	0.00	0.18	0.45	aantal	50	-1.48	-0.07	-0.19	0.24
blokke	100	0.40	-0.18	-0.06	0.07	blokke	100	-0.44	-0.42	-0.46	0.00
	200	-0.08	0.08	-0.08	-0.01		200	-0.60	-0.09	-0.11	-0.01
	500	0.36	0.07	-0.15	0.08		500	0.07	-0.04	-0.18	0.07

Burr

blokgrötte (n)					blokgrötte (n)						
$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500		
aantal	50	~	0.11	-0.13	~	aantal	50	~	0.02	-0.28	~
blokke	100	~	0.02	-0.01	-0.45	blokke	100	~	0.02	211.47	-0.67
	200	~	0.04	0.02	-0.26		200	~	0.06	0.33	-0.19
	500	~	0.14	0.03	-0.01		500	~	0.10	-0.01	-0.02
$q_{0.999}$	50	100	200	500	$q_{0.999}$	50	100	200	500		
aantal	50	~	0.22	1.15	~	aantal	50	~	-4.17	-1.55	~
blokke	100	~	-0.13	1.39	1.75	blokke	100	~	-1.58	-180.70	1.01
	200	~	0.23	1.34	0.20		200	~	0.17	0.67	0.24
	500	~	0.72	1.08	0.12		500	~	0.49	0.45	0.14

Die standaardafwyking van die kwantielafwyking ten opsigte van

$q_{0,99}$ en $q_{0,999}$:

Veralgemeende Pareto

		blokgrötte (n)				
		$q_{0,99}$	50	100	200	500
aantal blokke	50		1.87	1.61	1.51	2.90
	100		1.31	1.16	1.09	1.86
	200		0.91	0.80	0.47	1.33
	500		0.56	0.51	0.46	0.79

		blokgrötte (n)				
		$q_{0,999}$	50	100	200	500
aantal blokke	50		17.20	12.03	8.73	5.93
	100		12.02	8.55	6.04	4.48
	200		8.12	5.78	2.72	2.99
	500		5.32	3.79	2.73	1.95

Pareto

		blokgrötte (n)				
		$q_{0,99}$	50	100	200	500
aantal blokke	50		0.96	0.83	0.75	1.32
	100		0.65	0.56	0.55	0.91
	200		0.50	0.41	0.37	0.62
	500		0.29	0.24	0.23	0.39

		blokgrötte (n)				
		$q_{0,999}$	50	100	200	500
aantal blokke	50		9.71	6.88	4.94	3.00
	100		5.87	4.46	3.10	2.07
	200		4.31	2.93	2.13	1.48
	500		2.48	1.81	1.30	0.96

Burr

		blokgrötte (n)				
		$q_{0,99}$	50	100	200	500
aantal blokke	50		~	2.28	2.07	~
	100		~	1.68	4231.89	2.75
	200		~	1.12	1.03	1.91
	500		~	0.72	0.68	1.11

		blokgrötte (n)				
		$q_{0,999}$	50	100	200	500
aantal blokke	50		~	28.79	18.88	~
	100		~	17.94	3614.22	8.86
	200		~	13.92	11.78	6.45
	500		~	8.36	6.08	3.98

Die MAE (links) en MAD (regs) ten opsigte van $q_{0,99}$ en $q_{0,999}$:

Veralgemeende Pareto

		blokgrootte (n)						blokgrootte (n)			
		$q_{0,99}$						$q_{0,99}$			
		50	100	200	500			50	100	200	500
aantal	50	1.16	1.08	1.01	1.82	aantal	50	1.16	1.07	1.04	1.79
blokke	100	0.83	0.79	0.73	1.27	blokke	100	0.84	0.79	0.72	1.29
	200	0.63	0.54	0.47	0.93		200	0.61	0.52	0.47	0.91
	500	0.40	0.35	0.30	0.57		500	0.41	0.36	0.31	0.59

		blokgrootte (n)						blokgrootte (n)			
		$q_{0,999}$						$q_{0,999}$			
		50	100	200	500			50	100	200	500
aantal	50	10.27	8.25	5.58	3.79	aantal	50	10.50	7.88	5.59	3.71
blokke	100	7.01	5.48	4.15	2.95	blokke	100	6.75	5.58	4.12	2.86
	200	5.21	3.74	2.86	1.94		200	5.24	3.79	2.85	1.97
	500	3.41	2.69	1.80	1.32		500	3.40	2.55	1.86	1.30

Pareto

		blokgrootte (n)						blokgrootte (n)			
		$q_{0,99}$						$q_{0,99}$			
		50	100	200	500			50	100	200	500
aantal	50	0.65	0.56	0.46	0.86	aantal	50	0.62	0.56	0.47	0.89
blokke	100	0.44	0.38	0.36	0.62	blokke	100	0.45	0.39	0.36	0.61
	200	0.12	0.27	0.27	0.41		200	0.34	0.27	0.27	0.40
	500	0.20	0.17	0.16	0.28		500	0.20	0.17	0.16	0.27

		blokgrootte (n)						blokgrootte (n)			
		$q_{0,999}$						$q_{0,999}$			
		50	100	200	500			50	100	200	500
aantal	50	5.75	4.26	3.27	2.07	aantal	50	5.63	4.26	3.25	1.90
blokke	100	3.82	3.06	1.93	1.45	blokke	100	3.66	3.04	1.92	1.46
	200	2.96	1.95	1.49	1.02		200	2.95	1.99	1.51	1.02
	500	1.68	1.21	0.89	0.67		500	1.62	1.18	0.90	0.67

Burr

		blokgrootte (n)						blokgrootte (n)			
		$q_{0,99}$						$q_{0,99}$			
		50	100	200	500			50	100	200	500
aantal	50	~	1.54	1.34	~	aantal	50	~	1.51	1.34	~
blokke	100	~	1.12	1.09	1.75	blokke	100	~	1.14	1.09	1.73
	200	~	0.76	0.72	1.29		200	~	0.77	0.70	1.30
	500	~	0.51	0.46	0.78		500	~	0.47	0.46	0.78

		blokgrootte (n)						blokgrootte (n)			
		$q_{0,999}$						$q_{0,999}$			
		50	100	200	500			50	100	200	500
aantal	50	~	17.06	12.11	~	aantal	50	~	17.09	11.49	~
blokke	100	~	12.45	8.50	6.47	blokke	100	~	12.36	8.70	6.20
	200	~	9.33	8.61	4.45		200	~	9.21	8.55	4.54
	500	~	5.47	4.22	2.72		500	~	5.61	4.16	2.67

4A.2 Die Gumbel-domein: $\gamma = 0$

Ons gee in hierdie paragraaf die simulasieresultate vir die normaal-, logistiese en gammaverdeling.

4A2.1 Vertrauensinterval beraming vir die Gumbel-domein

OW's van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$: Asimptotiese Normaliteit

Normaal					Logisties						
blokgrootte (n)					blokgrootte (n)						
	$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500	
aantal	50	0.93	0.94	0.95	0.93	aantal	50	0.93	0.94	0.94	0.90
blokke	100	0.96	0.95	0.95	0.90	blokke	100	0.94	0.93	0.94	0.94
	200	0.96	0.95	0.94	0.94		200	0.95	0.93	0.96	0.93
	500	0.94	0.94	0.93	0.91		500	0.96	0.94	0.93	0.95

Normaal					Logisties						
blokgrootte (n)					blokgrootte (n)						
	$q_{0.999}$	50	100	200	500	$q_{0.999}$	50	100	200	500	
aantal	50	0.87	0.90	0.94	0.95	aantal	50	0.90	0.90	0.90	0.94
blokke	100	0.90	0.94	0.92	0.95	blokke	100	0.93	0.93	0.93	0.96
	200	0.93	0.93	0.94	0.96		200	0.94	0.93	0.94	0.94
	500	0.91	0.93	0.95	0.94		500	0.94	0.95	0.94	0.97

Gamma					
blokgrootte (n)					
	$q_{0.99}$	50	100	200	500
aantal	50	0.93	0.92	0.94	0.92
blokke	100	0.94	0.95	0.96	0.94
	200	0.95	0.94	0.93	0.94
	500	0.95	0.95	0.95	0.93

Gamma					
blokgrootte (n)					
	$q_{0.999}$	50	100	200	500
aantal	50	0.90	0.90	0.94	0.93
blokke	100	0.92	0.94	0.94	0.93
	200	0.92	0.95	0.95	0.95
	500	0.92	0.96	0.95	0.95

Gemiddelde (links) en Mediaan (regs) lengtes van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$:

Asimptotiese Normaliteit

Normaal					Normaal						
blokgrötte (n)					blokgrötte (n)						
$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500		
	50	0.28	0.43	0.69	1.18		50	0.28	0.42	0.65	1.08
aantal	100	0.20	0.29	0.45	0.74	aantal	100	0.20	0.29	0.44	0.72
blokke	200	0.14	0.20	0.31	0.50	blokke	200	0.14	0.20	0.30	0.49
	500	0.09	0.12	0.19	0.30		500	0.09	0.12	0.19	0.30
blokgrötte (n)					blokgrötte (n)						
$q_{0.999}$	50	100	200	500	$q_{0.999}$	50	100	200	500		
	50	0.29	0.23	0.21	0.35		50	0.28	0.23	0.21	0.34
aantal	100	0.20	0.16	0.15	0.23	aantal	100	0.20	0.16	0.15	0.23
blokke	200	0.14	0.11	0.11	0.16	blokke	200	0.14	0.12	0.10	0.16
	500	0.09	0.07	0.07	0.10		500	0.09	0.07	0.07	0.10

Logisties					Logisties						
blokgrötte (n)					blokgrötte (n)						
$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500		
	50	0.76	0.62	0.56	0.98		50	0.76	0.62	0.57	0.95
aantal	100	0.54	0.44	0.40	0.66	aantal	100	0.54	0.44	0.40	0.65
blokke	200	0.38	0.31	0.29	0.46	blokke	200	0.38	0.31	0.28	0.46
	500	0.24	0.20	0.18	0.29		500	0.24	0.20	0.18	0.28
blokgrötte (n)					blokgrötte (n)						
$q_{0.999}$	50	100	200	500	$q_{0.999}$	50	100	200	500		
	50	2.01	1.40	1.03	0.75		50	1.90	1.38	1.01	0.75
aantal	100	1.42	0.99	0.73	0.53	aantal	100	1.39	0.98	0.73	0.53
blokke	200	0.98	0.70	0.52	0.38	blokke	200	0.96	0.70	0.52	0.38
	500	0.61	0.44	0.33	0.24		500	0.60	0.44	0.33	0.24

Gamma					Gamma						
blokgrötte (n)					blokgrötte (n)						
$q_{0.99}$	50	100	200	500	$q_{0.99}$	50	100	200	500		
	50	0.96	0.78	0.71	1.22		50	0.95	0.77	0.70	1.17
aantal	100	0.68	0.55	0.51	0.84	aantal	100	0.67	0.55	0.51	0.82
blokke	200	0.48	0.39	0.35	0.57	blokke	200	0.48	0.39	0.35	0.56
	500	0.30	0.25	0.23	0.35		500	0.30	0.25	0.23	0.35
blokgrötte (n)					blokgrötte (n)						
$q_{0.999}$	50	100	200	500	$q_{0.999}$	50	100	200	500		
	50	2.38	1.66	1.21	0.91		50	2.27	1.61	1.19	0.90
aantal	100	1.67	1.18	0.87	0.64	aantal	100	1.61	1.16	0.87	0.64
blokke	200	1.14	0.83	0.62	0.45	blokke	200	1.12	0.82	0.62	0.45
	500	0.72	0.53	0.39	0.29		500	0.72	0.53	0.39	0.29

Standaardafwyking (links) en MAD (regs) van die lengtes van 95%-VI's vir

$q_{0,99}$ en $q_{0,999}$: Asimptotiese Normaliteit

Normaal						Normaal					
		blokgrootte (n)						blokgrootte (n)			
$q_{0,99}$		50	100	200	500	$q_{0,99}$		50	100	200	500
	50	0.055	0.115	0.211	0.465		50	0.035	0.067	0.134	0.240
aantal	100	0.024	0.048	0.090	0.190	aantal	100	0.017	0.030	0.056	0.110
blokke	200	0.011	0.023	0.046	0.083	blokke	200	0.007	0.015	0.028	0.054
	500	0.004	0.009	0.016	0.032		500	0.003	0.007	0.012	0.023
		blokgrootte (n)						blokgrootte (n)			
$q_{0,999}$		50	100	200	500	$q_{0,999}$		50	100	200	500
	50	0.033	0.027	0.034	0.090		50	0.020	0.020	0.021	0.053
aantal	100	0.016	0.013	0.015	0.040	aantal	100	0.011	0.008	0.010	0.026
blokke	200	0.007	0.006	0.008	0.018	blokke	200	0.005	0.005	0.005	0.011
	500	0.003	0.002	0.003	0.006		500	0.002	0.002	0.002	0.005

Logisties						Logisties					
		blokgrootte (n)						blokgrootte (n)			
$q_{0,99}$		50	100	200	500	$q_{0,99}$		50	100	200	500
	50	0.09	0.07	0.09	0.26		50	0.06	0.05	0.06	0.16
aantal	100	0.05	0.04	0.05	0.11	aantal	100	0.03	0.02	0.03	0.07
blokke	200	0.02	0.02	0.02	0.05	blokke	200	0.01	0.01	0.01	0.05
	500	0.01	0.01	0.01	0.02		500	0.01	0.00	0.01	0.01
		blokgrootte (n)						blokgrootte (n)			
$q_{0,999}$		50	100	200	500	$q_{0,999}$		50	100	200	500
	50	0.69	0.40	0.19	0.09		50	0.42	0.27	0.12	0.06
aantal	100	0.34	0.18	0.09	0.05	aantal	100	0.21	0.12	0.06	0.03
blokke	200	0.16	0.09	0.04	0.02	blokke	200	0.11	0.07	0.04	0.02
	500	0.06	0.03	0.02	0.01		500	0.04	0.02	0.01	0.01

Gamma						Gamma					
		blokgrootte (n)						blokgrootte (n)			
$q_{0,99}$		50	100	200	500	$q_{0,99}$		50	100	200	500
	50	0.11	0.09	0.12	0.31		50	0.07	0.06	0.08	0.19
aantal	100	0.06	0.05	0.05	0.15	aantal	100	0.04	0.03	0.04	0.09
blokke	200	0.03	0.02	0.03	0.07	blokke	200	0.02	0.02	0.02	0.05
	500	0.01	0.01	0.01	0.03		500	0.01	0.01	0.01	0.02
		blokgrootte (n)						blokgrootte (n)			
$q_{0,999}$		50	100	200	500	$q_{0,999}$		50	100	200	500
	50	0.79	0.44	0.23	0.11		50	0.48	0.28	0.15	0.07
aantal	100	0.40	0.21	0.12	0.05	aantal	100	0.24	0.14	0.08	0.04
blokke	200	0.18	0.10	0.05	0.03	blokke	200	0.12	0.05	0.04	0.02
	500	0.07	0.04	0.02	0.01		500	0.05	0.02	0.01	0.01

OW's van 95%-VI's vir $q_{0,99}$ en $q_{0,999}$: Profielmetode

Normaal					Logisties						
blokgrötte (n)					blokgrötte (n)						
$q_{0,99}$	50	100	200	500	$q_{0,99}$	50	100	200	500		
aantal	50	0.96	0.91	0.97	0.93	aantal	50	0.93	0.93	0.95	0.95
blokke	100	0.96	0.95	0.92	0.96	blokke	100	0.94	0.92	0.92	0.97
	200	0.97	0.98	0.93	0.93		200	0.92	0.95	0.98	0.91
	500	0.97	0.94	0.93	0.91		500	0.97	0.97	0.95	0.96

blokgrötte (n)					blokgrötte (n)						
$q_{0,999}$	50	100	200	500	$q_{0,999}$	50	100	200	500		
aantal	50	0.80	0.94	0.93	0.92	aantal	50	0.95	0.92	0.96	0.94
blokke	100	0.92	0.89	0.96	0.95	blokke	100	0.93	0.93	0.91	0.93
	200	0.93	0.92	0.92	0.95		200	0.98	0.95	0.91	0.94
	500	0.94	0.90	0.91	0.96		500	0.96	0.99	0.98	0.95

Gamma					
blokgrötte (n)					
$q_{0,99}$	50	100	200	500	
aantal	50	0.97	0.95	0.96	0.98
blokke	100	0.97	0.97	0.97	0.97
	200	0.93	0.98	0.98	0.94
	500	0.99	0.99	0.92	0.94

blokgrötte (n)					
$q_{0,999}$	50	100	200	500	
aantal	50	0.99	0.99	0.94	0.95
blokke	100	0.97	0.92	0.96	0.97
	200	0.97	0.95	0.96	0.97
	500	0.96	0.96	0.98	0.97

Gemiddelde (links) en Mediaan (regs) lengtes van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$:

Profielmetode

Normaal

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500			$q_{0.99}$	50	100	200	500
	50	0.29	0.23	0.22	0.35				50	0.29	0.23	0.22	0.34
aantal	100	0.20	0.16	0.15	0.25			aantal	100	0.20	0.16	0.15	0.25
blokke	200	0.14	0.12	0.11	0.16			blokke	200	0.14	0.12	0.11	0.16
	500	0.08	0.07	0.07	0.10				500	0.08	0.07	0.07	0.01

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
	50	0.68	0.46	0.32	0.22				50	0.65	0.44	0.32	0.22
aantal	100	0.42	0.30	0.22	0.16			aantal	100	0.42	0.31	0.22	0.16
blokke	200	0.29	0.21	0.16	0.11			blokke	200	0.29	0.21	0.17	0.11
	500	0.17	0.13	0.10	0.07				500	0.17	0.13	0.10	0.07

Logisties

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500			$q_{0.99}$	50	100	200	500
	50	0.76	0.62	0.60	1.07				50	0.76	0.62	0.59	1.04
aantal	100	0.54	0.44	0.42	0.69			aantal	100	0.53	0.44	0.41	0.67
blokke	200	0.38	0.31	0.29	0.47			blokke	200	0.38	0.31	0.29	0.47
	500	0.24	0.20	0.18	0.29				500	0.24	0.20	0.18	0.29

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
	50	2.43	1.40	1.10	0.76				50	2.43	1.40	1.10	0.77
aantal	100	1.53	1.08	0.76	0.54			aantal	100	1.53	1.05	0.75	0.54
blokke	200	1.02	0.74	0.54	0.38			blokke	200	0.99	0.74	0.53	0.38
	500	0.63	0.45	0.33	0.24				500	0.63	0.45	0.33	0.24

Gamma

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500			$q_{0.99}$	50	100	200	500
	50	0.97	0.79	0.75	1.31				50	0.98	0.78	0.73	1.28
aantal	100	0.68	0.56	0.51	0.85			aantal	100	0.68	0.56	0.51	0.83
blokke	200	0.49	0.39	0.36	0.71			blokke	200	0.49	0.39	0.36	0.70
	500	0.30	0.25	0.23	0.57				500	0.31	0.25	0.23	0.57

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
	50	2.94	1.87	1.31	0.92				50	2.78	1.80	1.27	0.91
aantal	100	1.84	1.25	0.90	0.65			aantal	100	1.82	1.22	0.89	0.64
blokke	200	1.24	0.83	0.63	0.55			blokke	200	1.23	0.82	0.63	0.55
	500	0.75	0.53	0.39	0.45				500	0.75	0.53	0.39	0.45

Standaardafwyking (links) en MAD (regs) van die lengtes van 95%-VI's vir

$q_{0,99}$ en $q_{0,999}$: Profielmetode

Normaal

		blokgrötte (n)						blokgrötte (n)					
		$q_{0,99}$	50	100	200	500			$q_{0,99}$	50	100	200	500
	50		0.032	0.028	0.034	0.084			50	0.024	0.017	0.021	0.056
aantal	100		0.016	0.014	0.017	0.038	aantal	100		0.011	0.009	0.012	0.026
blokke	200		0.009	0.006	0.008	0.018	blokke	200		0.006	0.005	0.005	0.013
	500		0.002	0.002	0.003	0.008		500		0.001	0.001	0.002	0.006
		blokgrötte (n)						blokgrötte (n)					
		$q_{0,999}$	50	100	200	500			$q_{0,999}$	50	100	200	500
	50		0.232	0.117	0.054	0.027			50	0.145	0.070	0.032	0.017
aantal	100		0.093	0.052	0.026	0.013	aantal	100		0.060	0.043	0.019	0.009
blokke	200		0.041	0.022	0.011	0.006	blokke	200		0.023	0.015	0.006	0.004
	500		0.011	0.009	0.005	0.003		500		0.012	0.006	0.004	0.002

Logisties

		blokgrötte (n)						blokgrötte (n)					
		$q_{0,99}$	50	100	200	500			$q_{0,99}$	50	100	200	500
	50		0.08	0.09	0.09	0.26			50	0.06	0.05	0.07	0.14
aantal	100		0.05	0.04	0.05	0.12	aantal	100		0.03	0.03	0.03	0.07
blokke	200		0.02	0.02	0.02	0.05	blokke	200		0.02	0.02	0.02	0.03
	500		0.01	0.01	0.01	0.02		500		0.01	0.01	0.01	0.02
		blokgrötte (n)						blokgrötte (n)					
		$q_{0,999}$	50	100	200	500			$q_{0,999}$	50	100	200	500
	50		0.74	0.55	0.20	0.10			50	0.60	0.19	0.14	0.06
aantal	100		0.38	0.19	0.12	0.05	aantal	100		0.30	0.12	0.08	0.03
blokke	200		0.15	0.09	0.05	0.02	blokke	200		0.10	0.06	0.03	0.01
	500		0.06	0.03	0.02	0.01		500		0.04	0.02	0.01	0.01

Gamma

		blokgrötte (n)						blokgrötte (n)					
		$q_{0,99}$	50	100	200	500			$q_{0,99}$	50	100	200	500
	50		0.096	0.080	0.121	0.290			50	0.060	0.060	0.077	0.170
aantal	100		0.050	0.047	0.048	0.130	aantal	100		0.039	0.029	0.029	0.077
blokke	200		0.027	0.022	0.029	0.095	blokke	200		0.019	0.014	0.020	0.058
	500		0.011	0.009	0.010	0.060		500		0.007	0.007	0.006	0.038
		blokgrötte (n)						blokgrötte (n)					
		$q_{0,999}$	50	100	200	500			$q_{0,999}$	50	100	200	500
	50		0.905	0.467	0.250	0.098			50	0.550	0.300	0.187	0.060
aantal	100		0.376	0.240	0.111	0.053	aantal	100		0.229	0.139	0.067	0.038
blokke	200		0.170	0.097	0.048	0.040	blokke	200		0.099	0.060	0.032	0.028
	500		0.070	0.035	0.020	0.027		500		0.050	0.023	0.012	0.017

4A2.2 Kwantielberaming vir die Gumbel-domein

Die Mediaan (links) en Gemiddelde kwantielafwyking (regs) ten opsigte van

$q_{0,99}$ en $q_{0,999}$:

Normaal					Normaal						
blokgrötte (n)					blokgrötte (n)						
	$q_{0,99}$	50	100	200	500		$q_{0,99}$	50	100	200	500
	50	-0.0010	0.0070	-0.0010	0.0017		50	-0.001	0.003	-0.002	0.002
aantal	100	0.0030	0.0010	-0.0005	-0.0033	aantal	100	0.001	0.001	0.001	-0.001
blokke	200	0.0020	0.0040	0.0010	-0.0046	blokke	200	0.001	0.004	0.001	-0.004
	500	0.0003	0.0050	0.0036	-0.0065		500	0.000	0.006	0.003	-0.007
	$q_{0,999}$	50	100	200	500		$q_{0,999}$	50	100	200	500
	50	0.0170	0.0070	0.0006	-0.0030		50	0.016	0.013	0.001	-0.004
aantal	100	0.0010	0.0070	-0.0009	0.0020	aantal	100	0.012	0.001	-0.002	-0.002
blokke	200	0.0010	0.0000	-0.0020	-0.0010	blokke	200	0.001	-0.001	-0.003	-0.002
	500	-0.0080	-0.0060	-0.0036	-0.0001		500	-0.007	-0.005	-0.003	-0.001

Logisties					Logisties						
blokgrötte (n)					blokgrötte (n)						
	$q_{0,99}$	50	100	200	500		$q_{0,99}$	50	100	200	500
	50	0.015	-0.019	-0.012	-0.020		50	-0.003	-0.022	-0.018	-0.016
aantal	100	0.000	-0.008	-0.005	-0.013	aantal	100	-0.007	-0.004	-0.010	-0.007
blokke	200	0.002	0.003	-0.001	-0.007	blokke	200	0.001	-0.002	-0.005	-0.009
	500	0.000	-0.001	0.001	-0.007		500	0.001	0.000	0.000	-0.004
	$q_{0,999}$	50	100	200	500		$q_{0,999}$	50	100	200	500
	50	0.068	0.024	0.048	0.001		50	0.048	0.032	0.029	0.009
aantal	100	0.012	0.026	0.008	0.005	aantal	100	0.001	0.023	0.013	0.005
blokke	200	0.035	0.016	0.006	0.008	blokke	200	0.017	0.014	-0.002	0.006
	500	0.011	0.020	0.003	0.006		500	0.010	0.008	0.003	0.003

Gamma					Gamma						
blokgrötte (n)					blokgrötte (n)						
	$q_{0,99}$	50	100	200	500		$q_{0,99}$	50	100	200	500
	50	0.000	-0.012	-0.028	-0.026		50	-0.016	-0.020	-0.027	0.014
aantal	100	-0.003	-0.004	-0.008	0.017	aantal	100	-0.003	-0.007	-0.011	0.013
blokke	200	-0.007	-0.010	-0.003	-0.024	blokke	200	-0.005	-0.010	-0.009	0.018
	500	0.008	-0.028	-0.004	-0.013		500	0.007	-0.027	-0.004	-0.008
	$q_{0,999}$	50	100	200	500		$q_{0,999}$	50	100	200	500
	50	0.080	0.060	0.049	0.002		50	0.040	0.044	0.044	-0.004
aantal	100	0.040	0.044	0.018	-0.001	aantal	100	0.024	0.020	0.005	0.002
blokke	200	0.025	0.002	0.017	0.007	blokke	200	0.022	0.003	0.012	0.001
	500	0.028	0.049	-0.001	-0.002		500	0.033	0.044	-0.003	0.004

Die standaardafwyking van die kwantielafwyking ten opsigte van

$q_{0,99}$ en $q_{0,999}$:

Normaal

		blokgröötte (n)				
		$q_{0,99}$	50	100	200	500
aantal blokke	50	0.07	0.06	0.05	0.09	
	100	0.05	0.04	0.04	0.06	
	200	0.04	0.03	0.03	0.04	
	500	0.02	0.02	0.02	0.03	
		blokgröötte (n)				
		$q_{0,999}$	50	100	200	500
aantal blokke	50	0.15	0.10	0.08	0.06	
	100	0.11	0.08	0.06	0.04	
	200	0.07	0.05	0.04	0.03	
	500	0.05	0.03	0.02	0.02	

Logisties

		blokgröötte (n)				
		$q_{0,99}$	50	100	200	500
aantal blokke	50	0.20	0.15	0.14	0.27	
	100	0.14	0.12	0.11	0.16	
	200	0.10	0.08	0.07	0.12	
	500	0.06	0.05	0.05	0.07	
		blokgröötte (n)				
		$q_{0,999}$	50	100	200	500
aantal blokke	50	0.49	0.37	0.28	0.19	
	100	0.37	0.26	0.19	0.13	
	200	0.25	0.18	0.13	0.10	
	500	0.15	0.11	0.09	0.06	

Gamma

		blokgröötte (n)				
		$q_{0,99}$	50	100	200	500
aantal blokke	50	0.26	0.21	0.18	0.31	
	100	0.18	0.14	0.12	0.22	
	200	0.13	0.10	0.10	0.15	
	500	0.07	0.18	0.06	0.09	
		blokgröötte (n)				
		$q_{0,999}$	50	100	200	500
aantal blokke	50	0.61	0.45	0.31	0.24	
	100	0.43	0.31	0.23	0.17	
	200	0.32	0.13	0.16	0.12	
	500	0.18	0.31	0.10	0.07	

Die MAE (links) en MAD (regs) ten opsigte van $q_{0,99}$ en $q_{0,999}$:

Normaal

		blokgrootte (n)				blokgrootte (n)				
		$q_{0,99}$				$q_{0,99}$				
		50	100	200	500	50	100	200	500	
aantal blokke	50	0.05	0.04	0.03	0.06	50	0.05	0.04	0.03	0.06
	100	0.04	0.03	0.02	0.04	100	0.03	0.03	0.02	0.04
	200	0.02	0.02	0.02	0.03	200	0.02	0.02	0.02	0.03
	500	0.01	0.01	0.01	0.02	500	0.01	0.01	0.01	0.02

		blokgrootte (n)				blokgrootte (n)				
		$q_{0,999}$				$q_{0,999}$				
		50	100	200	500	50	100	200	500	
aantal blokke	50	0.10	0.07	0.05	0.04	50	0.10	0.07	0.05	0.04
	100	0.07	0.05	0.04	0.03	100	0.07	0.05	0.04	0.03
	200	0.05	0.04	0.03	0.02	200	0.05	0.04	0.03	0.02
	500	0.04	0.02	0.02	0.01	500	0.03	0.02	0.02	0.01

Logisties

		blokgrootte (n)				blokgrootte (n)				
		$q_{0,99}$				$q_{0,99}$				
		50	100	200	500	50	100	200	500	
aantal blokke	50	0.14	0.11	0.10	0.17	50	0.14	0.11	0.10	0.17
	100	0.09	0.08	0.07	0.11	100	0.09	0.08	0.07	0.10
	200	0.06	0.06	0.05	0.08	200	0.06	0.06	0.05	0.08
	500	0.04	0.03	0.03	0.05	500	0.04	0.03	0.03	0.05

		blokgrootte (n)				blokgrootte (n)				
		$q_{0,999}$				$q_{0,999}$				
		50	100	200	500	50	100	200	500	
aantal blokke	50	0.34	0.26	0.18	0.13	50	0.34	0.26	0.17	0.13
	100	0.24	0.18	0.13	0.09	100	0.23	0.18	0.13	0.09
	200	0.17	0.12	0.08	0.07	200	0.16	0.12	0.08	0.07
	500	0.10	0.07	0.06	0.04	500	0.10	0.07	0.06	0.04

Gamma

		blokgrootte (n)				blokgrootte (n)				
		$q_{0,99}$				$q_{0,99}$				
		50	100	200	500	50	100	200	500	
aantal blokke	50	0.17	0.14	0.12	0.20	50	0.17	0.14	0.12	0.20
	100	0.13	0.10	0.08	0.15	100	0.13	0.10	0.08	0.15
	200	0.09	0.06	0.06	0.10	200	0.09	0.06	0.06	0.10
	500	0.05	0.12	0.04	0.06	500	0.05	0.12	0.04	0.06

		blokgrootte (n)				blokgrootte (n)				
		$q_{0,999}$				$q_{0,999}$				
		50	100	200	500	50	100	200	500	
aantal blokke	50	0.41	0.31	0.22	0.15	50	0.41	0.32	0.22	0.15
	100	0.28	0.21	0.15	0.11	100	0.28	0.21	0.15	0.11
	200	0.21	0.13	0.11	0.08	200	0.20	0.13	0.11	0.08
	500	0.12	0.22	0.07	0.05	500	0.12	0.23	0.07	0.05

4A.3 Die Weibull-domein: $\gamma < 0$

In hierdie domein ondersoek ons die betaverdeling.

4A3.1 Vertrauensinterval beraming vir die Weibull-domein

OW's van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$: Asimptotiese Normaliteit

		Beta			
		blokgrootte (n)			
$q_{0.99}$		50	100	200	500
aantal	50	0.91	0.93	0.90	0.94
blokke	100	0.95	0.94	0.95	0.94
	200	0.93	0.93	0.95	0.95
	500	0.92	0.90	0.87	0.94
		blokgrootte (n)			
$q_{0.999}$		50	100	200	500
aantal	50	0.86	0.89	0.92	0.93
blokke	100	0.89	0.93	0.93	0.94
	200	0.92	0.94	0.94	0.92
	500	0.93	0.92	0.91	0.92

Gemiddelde (links) en Mediaan (regs) lengtes van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$:

Asimptotiese Normaliteit

		Beta				Beta					
		blokgrootte (n)				blokgrootte (n)					
$q_{0.99}$		50	100	200	500	$q_{0.99}$	50	100	200	500	
aantal	50	0.019	0.015	0.005	0.023	aantal	50	0.019	0.015	0.005	0.022
blokke	100	0.013	0.011	0.010	0.015	blokke	100	0.013	0.010	0.010	0.015
	200	0.009	0.007	0.007	0.011		200	0.010	0.007	0.007	0.010
	500	0.006	0.005	0.000	0.006		500	0.006	0.005	0.005	0.006
		blokgrootte (n)				blokgrootte (n)					
$q_{0.999}$		50	100	200	500	$q_{0.999}$	50	100	200	500	
aantal	50	0.018	0.013	0.005	0.008	aantal	50	0.018	0.013	0.005	0.008
blokke	100	0.013	0.009	0.008	0.006	blokke	100	0.013	0.009	0.008	0.006
	200	0.009	0.007	0.006	0.004		200	0.009	0.007	0.006	0.004
	500	0.006	0.005	0.000	0.003		500	0.006	0.005	0.004	0.003

Standaardafwyking (links) en MAD (regs) van die lengtes van 95%-VI's vir

$q_{0.99}$ en $q_{0.999}$: Asimptotiese Normaliteit

Beta

		blokgrötte (n)						blokgrötte (n)			
		50	100	200	500			50	100	200	500
$q_{0.99}$	50	0.0020	0.0010	0.0002	0.0059	$q_{0.99}$	50	0.0010	0.0011	0.0001	0.0030
aantal	100	0.0009	0.0008	0.0013	0.0027	aantal	100	0.0006	0.0005	0.0008	0.0018
blokke	200	0.0005	0.0004	0.0005	0.0012	blokke	200	0.0003	0.0002	0.0003	0.0012
	500	0.0002	0.0002	0.0002	0.0004		500	0.0001	0.0001	0.0001	0.0003

		blokgrötte (n)						blokgrötte (n)			
		50	100	200	500			50	100	200	500
$q_{0.999}$	50	0.0050	0.0028	0.0002	0.0008	$q_{0.999}$	50	0.0030	0.0017	0.0001	0.0006
aantal	100	0.0026	0.0012	0.0006	0.0004	aantal	100	0.0017	0.0008	0.0006	0.0003
blokke	200	0.0012	0.0006	0.0055	0.0002	blokke	200	0.0008	0.0004	0.0003	0.0002
	500	0.0005	0.0002	0.0001	0.0008		500	0.0002	0.0001	0.0001	0.0001

OW's van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$: Profielmetode

Beta

		blokgrötte (n)			
		50	100	200	500
$q_{0.99}$	50	0.89	0.91	0.84	0.84
aantal	100	0.87	0.95	0.92	0.87
blokke	200	0.93	0.95	0.95	0.97
	500	0.93	0.96	0.94	0.94

		blokgrötte (n)			
		50	100	200	500
$q_{0.999}$	50	0.95	0.92	0.87	0.83
aantal	100	0.90	0.94	0.95	0.92
blokke	200	0.99	0.90	0.93	0.93
	500	0.94	0.99	0.93	0.99

Gemiddelde (links) en Mediaan (regs) lengtes van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$:

Profielmetode

Beta

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500			$q_{0.99}$	50	100	200	500
	50	0.0180	0.0142	0.0127	0.0149			50	0.0180	0.0020	0.0127	0.0146	
aantal	100	0.0130	0.0100	0.0090	0.0110			aantal	100	0.0130	0.0100	0.0090	0.0115
blokke	200	0.0095	0.0076	0.0069	0.0088			blokke	200	0.0096	0.0076	0.0070	0.0088
	500	0.0060	0.0048	0.0044	0.0060			500	0.0060	0.0048	0.0044	0.0060	

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
	50	0.0230	0.0150	0.0100	0.0080			50	0.0220	0.0038	0.0100	0.0080	
aantal	100	0.0140	0.0101	0.0078	0.0057			aantal	100	0.0140	0.0104	0.0078	0.0058
blokke	200	0.0099	0.0072	0.0056	0.0041			blokke	200	0.0100	0.0072	0.0056	0.0041
	500	0.0059	0.0045	0.0035	0.0025			500	0.0058	0.0045	0.0035	0.0025	

Standaardafwyking (links) en MAD (regs) van die lengtes van 95%-VI's vir

$q_{0.99}$ en $q_{0.999}$: Profielmetode



Beta

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.99}$	50	100	200	500			$q_{0.99}$	50	100	200	500
	50	0.0030	0.0013	0.0022	0.0026			50	0.0020	0.0144	0.0016	0.0018	
aantal	100	0.0014	0.0011	0.0010	0.0019			aantal	100	0.0008	0.0006	0.0006	0.0015
blokke	200	0.0007	0.0004	0.0005	0.0011			blokke	200	0.0004	0.0003	0.0004	0.0008
	500	0.0002	0.0002	0.0002	0.0007			500	0.0001	0.0001	0.0002	0.0004	

		blokgrootte (n)						blokgrootte (n)					
		$q_{0.999}$	50	100	200	500			$q_{0.999}$	50	100	200	500
	50	0.0070	0.0026	0.0022	0.0013			50	0.00450	0.01500	0.00130	0.00064	
aantal	100	0.0031	0.0015	0.0007	0.0005			aantal	100	0.00180	0.00080	0.00030	0.00033
blokke	200	0.0014	0.0006	0.0003	0.0002			blokke	200	0.00090	0.00042	0.00020	0.00010
	500	0.0005	0.0002	0.0001	0.0026			500	0.00031	0.00016	0.00008	0.00007	

4A3.2 Kwantielberaming vir die Weibull-domein

Die Mediaan (links) en Gemiddelde kwantielafwyking (regs) ten opsigte van

$q_{0,99}$ en $q_{0,999}$:

Beta					Beta						
blokgrötte (n)					blokgrötte (n)						
$q_{0,99}$	50	100	200	500	$q_{0,99}$	50	100	200	500		
	50	0.0003	0.0005	0.0009	0.0002		50	0.0005	0.0004	0.0009	0.0006
aantal	100	0.0005	0.0005	0.0008	0.0005	aantal	100	0.0007	0.0005	0.0006	0.0008
blokke	200	0.0005	0.0007	0.0008	0.0005	blokke	200	0.0006	0.0007	0.0008	0.0006
	500	0.0009	0.0009	0.0009	0.0005		500	0.0008	0.0009	0.0009	0.0005

blokgrötte (n)					blokgrötte (n)						
$q_{0,999}$	50	100	200	500	$q_{0,999}$	50	100	200	500		
	50	0.0012	0.0003	-0.0006	-0.0005		50	0.0014	0.0006	-0.0006	-0.0005
aantal	100	0.0001	0.0000	-0.0003	-0.0004	aantal	100	0.0004	-0.0001	-0.0002	-0.0005
blokke	200	-0.0003	-0.0004	-0.0005	-0.0006	blokke	200	-0.0002	-0.0004	-0.0005	-0.0005
	500	-0.0005	-0.0006	-0.0005	-0.0004		500	-0.0005	-0.0006	-0.0005	-0.0004

Die standaardafwyking van die kwantielafwyking ten opsigte van

$q_{0,99}$ en $q_{0,999}$:

Beta					
blokgrötte (n)					
$q_{0,99}$	50	100	200	500	
	50	0.005	0.004	0.001	0.006
aantal	100	0.005	0.002	0.002	0.004
blokke	200	0.003	0.002	0.002	0.003
	500	0.002	0.001	0.001	0.002

blokgrötte (n)					
$q_{0,999}$	50	100	200	500	
	50	0.005	0.004	0.001	0.002
aantal	100	0.004	0.003	0.002	0.002
blokke	200	0.003	0.002	0.001	0.001
	500	0.002	0.001	0.001	0.001

Die MAE (links) en MAD (regs) ten opsigte van $q_{0,99}$ en $q_{0,999}$:

Beta

		blokgrootte (n)						blokgrootte (n)				
		$q_{0,99}$						$q_{0,99}$				
		50	100	200	500			50	100	200	500	
		50	0.0032	0.0030	0.0011	0.0038		50	0.0032	0.0028	0.0008	0.0039
aantal	100	0.0024	0.0018	0.0017	0.0028		aantal	100	0.0024	0.0018	0.0015	0.0028
blokke	200	0.0016	0.0014	0.0013	0.0017		blokke	200	0.0016	0.0014	0.0012	0.0017
	500	0.0012	0.0010	0.0010	0.0011		500	0.0010	0.0007	0.0008	0.0011	

		blokgrootte (n)						blokgrootte (n)				
		$q_{0,999}$						$q_{0,999}$				
		50	100	200	500			50	100	200	500	
		50	0.0035	0.0025	0.0009	0.0014		50	0.0034	0.0026	0.0008	0.0015
aantal	100	0.0023	0.0017	0.0014	0.0011		aantal	100	0.0023	0.0017	0.0014	0.0010
blokke	200	0.0018	0.0011	0.0011	0.0008		blokke	200	0.0018	0.0011	0.0010	0.0007
	500	0.0011	0.0009	0.0007	0.0005		500	0.0011	0.0008	0.0007	0.0004	



Hoofstuk 5: Die gedrag van drempelinferensie by eindige drempelwaardes

5.1 Inleiding

Die klassieke benadering tot die studie van ekstreemwaarde-teorie is die maksimamemete van Fisher en Tippett. Hierdie teorie is in detail in hoofstuk twee uiteengesit. In die vorige hoofstuk is die geldigheid van maksima-inferensie by eindige steekproewe ondersoek. 'n Meer populêre moderne beskouing van ekstreemwaarde-teorie is drempelmodellering. Dié teorie word bespreek in hoofstuk drie. In hierdie hoofstuk word die klem geplaas op die akkuraatheid van drempelgebaseerde inferensie ten opsigte van die onbekende onderliggende verdeling F .

In drempelteorie word alle waarnemings van 'n stogastiese veranderlike X , met verdelingsfunksie F , wat 'n drempel u oorskry, as ekstreme waardes beskou. Om hierdie rede stel ons belang in die volgende voorwaardelike waarskynlikheid:

$$P(X - u > y | X > u) = \frac{P(X > u + y)}{P(X > u)}, \quad y > 0$$

oftewel

$$\bar{F}_u(y) = \frac{\bar{F}(u + y)}{\bar{F}(u)}, \quad y > 0$$

Indien die onderliggende verdeling F bekend was, sou hierdie voorwaardelike verdeling F_u van die absolute drempeloorskrydings $(X - u)$ ook bekend gewees het. Tipies is F natuurlik onbekend.

Die Pickands-Balkema-De Haan stelling vorm die kern van die teorie van drempelmodellering. Hiervolgens kan F_u benader word, vir groot u , deur 'n

veralgemeende Pareto-verdeling $H_{\gamma, \beta(u)}$. Dit wel sê, $F \in MDA(G_\gamma)$ as en slegs as daar 'n positiewe, meetbare funksie $\beta(u)$ bestaan sodat

$$\lim_{u \rightarrow x_F} \sup_{0 \leq y \leq x_F - u} |F_u(y) - H_{\gamma, \beta(u)}(y)| = 0.$$

Hieruit volg dat vir groot u geld

$$F_u(y) \approx H_{\gamma, \beta(u)}(y) = 1 - \left(1 + \frac{\gamma}{\beta(u)} y\right)^{-\frac{1}{\gamma}}. \quad (5.1)$$

Inferensie ten opsigte van die onbekende onderliggende verdeling F word gebaseer op vergelyking 5.1. By eindige waardes van die drempel u , geld 5.1 natuurlik slegs by benadering. Die vraag is dus: Hoe goed vaar die benadering $F_u(y) \approx H_{\gamma, \beta(u)}(y)$ vir eindige drempelgrootte? Aangesien 'n belangrike doelstelling van ekstreemwaarde analises die beraming van ekstreme stertkwantiele sowel as vertrouensintervalle vir hierdie kwantiele behels, sal die geldigheid van die benadering $F_u(y) \approx H_{\gamma, \beta(u)}(y)$ aan die hand van die akkuraatheid van hierdie beramings beoordeel word. Soos in hoofstuk vier word daar van 'n simulasiestudie gebruik gemaak om die gedrag van drempelinferensie by eindige steekproewe en drempels te ondersoek.

In paragraaf 5.2 word die ontwerp van ons simulasiestudie uiteengesit. Maatstawwe waarmee die sydigheid en die variasie van die punt- en intervalberamings van die ekstreme kwantiele, beoordeel sal word, word ook gedefiniëer. Paragraaf 5.3 bevat 'n aantal algemene waarnemings uit die simulatie-afvoer. Daarna word die simulatie-resultate vir die gekose verdelings uit die Frechet-, Gumbel- en Weibull-domein in die subparagrafe 5.3.1 tot 5.3.3 in groter besonderhede beskou. Die laaste paragraaf is 5.4, waarin die gevolgtrekkings rakende die simulatie-resultate opgesom word.

5.2 Simulasie-ontwerp

Die ondersoek na die geldigheid van inferensie oor die onbekende verdeling F gebaseer op die verdeling $H_{\gamma, \beta(u)}$ by eindige waardes van die drempel u (en n), geskied met behulp van 'n simulasiestudie. Die belangrikste faktor is die keuse van onderliggende verdeling F . Hier gebruik ons dieselfde verdelings as in hoofstuk vier (sien paragraaf 4.2). 'n Verdere faktor is die keuse van steekproefgrootte n wat gegenereer word, sowel as die keuse van 'n drempel u . Elke lopie van die simulatie-eksperiment verloop as volg: Vir elke keuse van 'n verdeling F , steekproefgrootte n en drempel u , word n waarnemings uit F gegenereer. Alle waardes wat die drempel u oorskry, word geïdentifiseer en die veralgemeende Pareto-verdeling $H_{\gamma, \beta(u)}$ word op hierdie oorskrydings gepas. Die parameters van hierdie verdeling word met behulp van maksimum aanneemlikheid (MA) - sien definisie 2.2.7 en paragraaf 2.5.2 - beraam. Hierdie MA-beramings word ook aangewend om puntberamings vir stertkwantiele van die onderliggende verdeling F te verkry. Benaderde vertrouensintervalle vir hierdie stertkwantiele word beraam met behulp van die asimptotiese normaliteit van die MA-beramers sowel as die profiel-aanneemlikheid metode, wat op 'n benaderde chikwadraat-verdeling gebaseer is. Ons evalueer die prestasie van hierdie vertrouensintervalle in terme van die lengte en oordeckingswaarskynlikheid (OW) daarvan.

Die kontinue verdelings onder beskouing word gegroepeer volgens die maksimum domein van aantrekking binne $G_{\gamma, \mu, \sigma}$ waartoe elke verdeling behoort.

Ons ondersoek die volgende onderliggende verdelings:

- Frechet-domein, $\gamma > 0$: Veralgemeende Pareto($\gamma = 0.5, \beta = 1$), Streng Pareto($\gamma = 0.5$), Burr(1, 1.5, 1)
- Gumbel-domein, $\gamma = 0$: Standaard Normaal, Standaard Logisties, Gamma (vormparameter = 3)

- Weibull-domein, $\gamma < 0$: Beta (vormparameters 5 en 3 onderskeidelik)

Elke eksperiment word 500 keer uitgevoer. Hier neem ons steekproefgroottes van 200, 500, 1 000, 2 000 en 5 000. Eerder as om 'n absolute drempelwaarde vir u vas te stel, kies ons u telkens as 'n rangorde statistiek (sien paragraaf 3.6.1) van die steekproef van grootte n . Ons kies u as die 0.50-, 0.80- of 0.90-kwantiel van die steekproef. Dit behels dat die aantal oorskrydings vasgepen word. So lei 'n keuse van $n=1000$ en u as die 0.80-kwantiel van hierdie steekproef tot 200 oorskrydings van die drempel, waarop die veralgemeende Pareto-verdeling gepas word. Soos in die vorige hoofstuk verkry ons puntberamings en benaderde 95%-vertrouensintervalle vir $q_{0,99}$ en $q_{0,999}$, die kwantiele 0.99 en 0.999 van die onderliggende verdeling F .

Laat \hat{q}_i die beraamde kwantiel op die i -de simulasieloope en q die teoretiese kwantiel van die onderliggende verdeling F aandui. As maatstawwe om die moontlike sydigheid van ons puntberamings ten opsigte van $q_{0,99}$ en $q_{0,999}$ te evalueer, gebruik ons soos in hoofstuk vier die gemiddelde kwantielafwyking, gegee deur

$$\frac{1}{500} \sum_{i=1}^{500} (q - \hat{q}_i) \quad (5.2)$$

sowel as die mediaan van die afwykings tussen die werklike en die beraamde kwantiel, naamlik

$$\text{mediaan}_i (q - \hat{q}_i). \quad (5.3)$$

As maatstawwe van variasie gebruik ons weer eens soos in hoofstuk vier die standaardafwyking sowel as mediaan absolute fout (MAE) en mediaan absolute afwyking (MAD) van die kwantielafwykings. Die standaardafwyking word gegee deur

$$\sqrt{\frac{1}{500} \sum_{i=1}^{500} (\hat{q}_i - \bar{\hat{q}})^2}, \quad (5.4)$$

waar $\bar{\hat{q}}$ die gemiddelde kwantielberaming is, oftewel

$$\bar{\hat{q}} = \frac{1}{500} \sum_{i=1}^{500} \hat{q}_i. \quad (5.5)$$

Die twee verdere maatstawwe van variasie, naamlik die MAD en die MAE word as volg gedefinieer:

$$MAD = \text{mediaan}\{ |q - \hat{q}_i| \} \quad (5.6)$$

$$MAE = \text{mediaan}\{ | \text{mediaan}(\hat{q}_j) - \hat{q}_i | \} \quad (5.7)$$

By elke herhaling van die simulatie-eksperiment word soos in hoofstuk vier 'n benaderde 95%-vertrouensinterval vir elk van die stertkwantiele onder beskouing bereken. In die vorige hoofstuk is die beraamde oordekkingswaarskynlikheid (OW) gedefinieer as die proporsie van die lopies waarby die teoretiese kwantiel q binne die beraamde vertrouensinterval geval het. Indien ons die benaderde ondergrens en bogrens van die beraamde vertrouensinterval op die i -de simulasieloope onderskeidelik VI_L_i en VI_U_i noem, dan volg dat

$$OW = \frac{1}{500} \sum_{i=1}^{500} I(VI_L_i \leq q \leq VI_U_i). \quad (5.8)$$

Die lengte van ons verkreeë vertrouensintervalle moet saam met die oordekkingswaarskynlikhede oorweeg word. In hierdie verband word die gemiddelde vertrouensintervallengte, gegee deur

$$\frac{1}{500} \sum_{i=1}^{500} (VI_U_i - VI_L_i), \quad (5.9)$$

sowel as die mediaan lengte bereken. As variasiemaatstawwe bereken ons die standaardafwyking van die vertrouensintervallengtes en die mediaan absolute fout, naamlik

$$MAE = \text{mediaan}\{ | \text{mediaan}(VI_U_j - VI_L_j) - (VI_U_i - VI_L_i) | \}. \quad (5.10)$$

5.3 Bespreking van resultate

In hierdie paragraaf word die resultate van die simulasiestudie, soos in paragraaf 5.2 uiteengesit, aangebied en geanaliseer. Die volledige stel afvoer word in Aanhangsel 5A gegee. Ons maak eers 'n aantal algemene opmerkings oor die uitslae. Daarna ondersoek ons in die subparagraawe 5.3.1 tot 5.3.2 in meer detail die resultate vir elk van die gekose verdelings. Daar sal telkens van uittreksels uit Aanhangsel 5A gebruik gemaak word om waarnemings en gevolgtrekkings rakende die simulasië-afvoer te belig.

Ten opsigte van die OW's van die verkreeë 95%-VI's blyk dit uit die simulasië-afvoer dat OW's van minder as 0.95 meestal verkry word. By die normaalgebaseerde VI's is dit veral baie selde dat 'n OW van 0.95 of meer verkry word. Die profielgebaseerde VI's het gewoonlik hoër OW's as die normaal-VI's. Ten opsigte van oordekking is die resultate van die drempelbenadering baie meer wispelturig. Die maksimummetode het 'n laagste OW van 0.49 vir die Burr-verdeling opgelewer (sien tabel 4.3.3), maar die laagste drempel-OW's is 0.12 en 0.15 by die normaalverdeling (sien tabel 5.3.13).

Alhoewel die profiel-OW's deurgaans hoër is, gaan hulle ook tipies met langer vertrouensintervalle gepaard. Verally by 'n kombinasie van 'n steekproefgrootte van 200 waarnemings en 'n keuse van die drempel as die mediaan van hierdie steekproef, is die gemiddelde profiel-VI tipies aansienlik langer as die normaal-VI's.

Dit blyk dat die variasie maatstawwe soos die standaardafwyking van die kwantielafwykings (vergelyking 5.4), die MAD (5.5) en die MAE (5.6) by 'n vaste kwantiel as drempelwaarde, afneem indien ons die steekproefgrootte n -en dus die aantal oorskrydings waarop die verdeling $H_{\gamma, \beta(u)}$ gepas word- laat toeneem.

Ons bespreek nou die simulasieresultate vir elk van die verdelings onder beskouing, in groter detail. By sommige kombinasies van die steekproefgrootte en die drempel - byvoorbeeld $n = 200$ met u as die 0.80-kwantiel van die steekproef - kon die S-Plus roetines van die bylae tot Coles (2001) nie funksioneer nie. In sulke gevalle is telkens van 'n tilde (~) gebruik gemaak om aan te dui dat die simulatie-eksperiment nie deurgevoer kon word nie.

5.3.1 Die Frechet-domein: $\gamma > 0$

Soos in die geval van die vorige hoofstuk, ondersoek ons weer die veralgemeende en streng Pareto-verdelings, sowel as die Burr-verdeling (gedefinieer in paragraaf 4.3.1) as voorbeelde van verdelings wat tot die Frechet maksimum domein van aantrekking behoort. Hierdie verdelings is lede van 'n klas swaarstert verdelings wat van groot belang in veral finansiële toepassings is.

Tabelle 5.3.1 en 5.3.2 toon dat die profielgebaseerde VI's in terme van oordekkingswaarskynlikheid beter as die normaalgebaseerde VI's vaar. Vir beide die veralgemeende en die streng Pareto-verdeling word OW's van ongeveer 0.95 verkry. By die Burr-verdeling (tabel 5.3.3) is die profiel-OW's vir VI's ten opsigte van $q_{0,999}$ duidelik hoër, terwyl die resultate vir $q_{0,99}$ meer gemeng is. Die profielmetode is ook baie bestendig: Die laagste OW is 0.91 en die hoogste OW is 0.98. Die normaalmetode lei afgesien van enkele uitsonderings tot OW's van laer as 0.95. Die heel swakste OW is 0.83. Dit is aangeteken vir 'n VI vir $q_{0,999}$ van die veralgemeende Pareto-verdeling by 'n steekproefgrootte van 500 en 'n drempel gelykstaande aan die 0.90-kwantiel van hierdie steekproef (dus 50 oorskrydings).

Veralgemeende Pareto

Normaliteit	steekproefgrootte (n)						Profiel	steekproefgrootte (n)					
	$q_{0.99}$	200	500	1000	2000	5000		$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.87	0.94	0.91	0.94	0.96		0.5	0.95	0.95	0.95	0.95	~
drempel	0.8	~	0.89	0.90	0.94	0.93	drempel	0.8	~	0.93	0.94	0.94	0.94
r.o.s	0.9	~	0.90	0.92	0.92	0.94	r.o.s	0.9	~	0.91	0.94	0.91	0.95

Normaliteit	steekproefgrootte (n)						Profiel	steekproefgrootte (n)					
	$q_{0.999}$	200	500	1000	2000	5000		$q_{0.999}$	200	500	1000	2000	5000
	0.5	0.83	0.91	0.90	0.94	0.95		0.5	0.96	0.94	0.96	0.94	~
drempel	0.8	~	0.85	0.88	0.93	0.93	drempel	0.8	~	0.94	0.95	0.95	0.96
r.o.s	0.9	~	0.83	0.85	0.90	0.93	r.o.s	0.9	~	0.92	0.94	0.91	0.98

Tabel 5.3.1: OW's vir 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ van die veralgemeende Pareto-verdeling.

Pareto

Normaliteit	steekproefgrootte (n)						Profiel	steekproefgrootte (n)					
	$q_{0.99}$	200	500	1000	2000	5000		$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.88	0.94	0.95	0.94	0.93		0.5	0.95	0.96	0.94	0.94	0.93
drempel	0.8	~	0.92	0.92	0.94	0.95	drempel	0.8	~	0.94	0.95	0.95	0.93
r.o.s	0.9	~	~	0.92	0.93	0.94	r.o.s	0.9	~	~	0.91	0.93	0.92

Normaliteit	steekproefgrootte (n)						Profiel	steekproefgrootte (n)					
	$q_{0.999}$	200	500	1000	2000	5000		$q_{0.999}$	200	500	1000	2000	5000
	0.5	0.85	0.91	0.92	0.94	0.93		0.5	0.95	0.97	0.94	0.94	0.94
drempel	0.8	~	0.87	0.88	0.91	0.94	drempel	0.8	~	0.95	0.96	0.95	0.96
r.o.s	0.9	~	~	0.84	0.89	0.92	r.o.s	0.9	~	~	0.94	0.95	0.94

Tabel 5.3.2: OW's vir 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ van die streng Pareto-verdeling.

Burr

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000		5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	~	~	~	~	~	0.5	~	~	~	~	~	
drempel	0.8	~	0.89	0.91	0.94	0.93	drempel	0.8	~	0.94	0.93	0.96	0.95
r.o.s	0.9	~	~	0.93	0.95	0.96	r.o.s	0.9	~	~	0.92	0.92	0.91

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000		5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	~	~	~	~	~	0.5	~	~	~	~	~	
drempel	0.8	~	0.84	0.86	0.92	0.89	drempel	0.8	~	0.92	0.92	0.96	0.97
r.o.s	0.9	~	~	0.84	0.87	0.90	r.o.s	0.9	~	~	0.94	0.93	0.92

Tabel 5.3.3: OW's vir 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ van die Burr-verdeling

Soos tabelle 5.3.4, 5.3.5 en 5.3.6 duidelik uitwys, gaan die hoër profiel-OW's meestal met 'n hoër gemiddelde VI-lengte gepaard. By die kleiner steekproefgroottes soos 200, 500 en 1000 is die profielgebaseerde VI's heelwat langer. In die geval van die VI's vir $q_{0.999}$ van die Burr-verdeling lei die kombinasie van 500 waarnemings met 'n 0.80-kwantiel as drempel, tot profiel-VI's wat amper twee keer so lank as die ooreenstemmende normaal-VI's is. Die gemiddelde VI-lengtes vir die profiel- en normaalmetodes is egter ongeveer dieselfde by groter steekproewe.

Veralgemeende Pareto

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000		5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	22.65	13.63	9.36	6.64	4.34	0.5	29.43	15.72	10.26	6.98	~	
drempel	0.8	~	14.05	10.07	7.38	4.50	drempel	0.8	~	17.64	11.17	7.29	4.48
r.o.s	0.9	~	14.92	10.55	7.45	4.59	r.o.s	0.9	~	17.62	10.88	7.05	4.37

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)					~	
	$q_{0.999}$	200	500	1000	2000		5000	$q_{0.999}$	200	500	1000		2000
	0.5	186.85	94.42	61.60	43.41	28.70	0.5	319.60	127.52	76.70	48.29	~	
drempel	0.8	~	118.93	81.14	59.88	34.95	drempel	0.8	~	219.48	110.32	63.42	36.96
r.o.s	0.9	~	169.19	103.51	69.62	41.13	r.o.s	0.9	~	381.33	149.36	79.14	43.96

Tabel 5.3.4: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die veralgemeende Pareto-verdeling.

Pareto

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000		5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	10.58	7.26	4.92	3.42	2.20	0.5	15.40	7.67	5.04	3.46	2.15	
drempel	0.8	~	7.34	4.98	3.68	2.26	drempel	0.8	~	8.96	5.55	3.71	2.24
r.o.s	0.9	~	~	5.39	3.73	2.31	r.o.s	0.9	~	~	5.41	3.62	2.19

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000		5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	77.40	51.22	32.97	22.66	14.67	0.5	171.07	60.63	36.38	23.87	14.38	
drempel	0.8	~	64.36	40.48	30.12	17.70	drempel	0.8	~	115.13	54.05	32.92	18.43
r.o.s	0.9	~	~	52.30	35.01	21.21	r.o.s	0.9	~	~	73.63	40.91	22.19

Tabel 5.3.5: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die streng Pareto-verdeling.

Burr

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000		5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	~	~	~	~	~	0.5	~	~	~	~	~	
drempel	0.8	~	20.87	14.46	10.28	6.41	drempel	0.8	~	26.54	15.36	10.65	6.59
r.o.s	0.9	~	~	16.35	11.62	7.29	r.o.s	0.9	~	~	16.10	10.49	6.41

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)					~	
	$q_{0.999}$	200	500	1000	2000		5000	$q_{0.999}$	200	500	1000		2000
	0.5	~	~	~	~	~	0.5	~	~	~	~	~	
drempel	0.8	~	253.81	158.59	109.94	66.30	drempel	0.8	~	473.25	199.80	127.45	74.15
r.o.s	0.9	~	~	204.80	136.53	81.22	r.o.s	0.9	~	~	335.01	161.32	89.81

Tabel 5.3.6: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die streng Pareto-verdeling.

Soos die tabelle hierbo uitwys, is die profiel-OW's ten opsigte van VI's vir $q_{0.999}$, opmerklik hoër as die normaal-OW's, veral by die kombinasies van steekproefgrootte 1 000 en drempelkeuses as onderskeidelik die 0.80- en 0.90-kwantiel van die steekproef. Soos verwag, is die normaal-VI's ook deurgaans heelwat korter. Ten einde die normaal-OW's te verhoog, is die gemiddelde lengte van hierdie VI's verhoog. Dit is vermag deur in die normaalmetode hoër standaard normaal kwantiele soos $\Phi_{0.99}$ en $\Phi_{0.999}$ te gebruik. In tabelle 5.3.7 en 5.3.8 word die afvoer van hierdie analise verskaf.

In hierdie twee onderstaande tabelle word die OW's sowel as die gemiddelde lengtes van die aangepaste normaal-VI's gegee. Hierdie OW's is deurgaans hoër en die VI's langer as die oorspronklike normaal-resultate (tabelle 5.3.1 tot 5.3.6). In daardie gevalle waar die langer normaal-VI's tot gelyke profiel- en normaal-OW's lei, kan die gemiddelde VI-lengtes van die twee metodes direk vergelyk word. Een voorbeeld hiervan is die 0.90-kwantiel drempel in die geval van die Pareto-verdeling. Die profiel- sowel as die aangepaste (indien $\Phi_{0.999}$

gebruik word) normaal-OW's is gelyk aan 0.94. Die gemiddelde profiel-VI is egter korter met 'n lengte van 73.63 teenoor die normaal-lengte van 79.51. In die geval 'n 0.80-kwantiel drempel in die geval van die Burr-verdeling lewer die profiel- en aangepaste normaalmetode (met $\Phi_{0.999}$) profiel- en normaal-OW's van 0.92. Die profiel-VI's is weer eens korter; gemiddeld het die profiel-VI's 'n lengte van 199.80, teenoor 247.25 vir die normaal-VI's.

Die profiel- en die normaalmetode lewer soortgelyke resultate wanneer ons die normaalkwantiel by laasgenoemde benadering gebruik. Dit wil voorkom of die profieltegniek effens beter vaar. In praktiese toepassing waar swaarstert verdeling ter sprake is, is dit sinvol om VI's vir ekstreme kwantiele soos $q_{0.999}$ met behulp van beide metodes te bereken en albei die resultate in ag te neem.

	Veralgemeende Pareto		Pareto		Burr			
	sp-grootte (n)		sp-grootte (n)		sp-grootte (n)			
	$\Phi_{0.99}$	1000	$\Phi_{0.99}$	1000	$\Phi_{0.99}$	1000		
drempel	0.8	0.94	drempel	0.8	0.93	drempel	0.8	0.90
r.o.s	0.9	0.90	r.o.s	0.9	0.90	r.o.s	0.9	0.90
	$\Phi_{0.999}$ sp-grootte (n)		$\Phi_{0.999}$ sp-grootte (n)		$\Phi_{0.999}$ sp-grootte (n)			
	1000		1000		1000			
drempel	0.8	0.97	drempel	0.8	0.97	drempel	0.8	0.92
r.o.s	0.9	0.92	r.o.s	0.9	0.94	r.o.s	0.9	0.93

Tabel 5.3.7: Aangepaste normaal-OW's vir $q_{0.999}$ van die veralgemeende en streng Pareto-verdeling, sowel as die Burr-verdeling

Veralgemeende Pareto			Pareto			Burr		
	$\Phi_{0.99}$ sp-grootte (n)			$\Phi_{0.99}$ sp-grootte (n)			$\Phi_{0.99}$ sp-grootte (n)	
	1000			1000		1000		
drempel	0.8	101.51	drempel	0.8	49.95	drempel	0.8	183.56
r.o.s	0.9	115.20	r.o.s	0.9	59.53	r.o.s	0.9	238.58
	$\Phi_{0.999}$ sp-grootte (n)			$\Phi_{0.999}$ sp-grootte (n)			$\Phi_{0.999}$ sp-grootte (n)	
	1000			1000		1000		
drempel	0.8	131.37	drempel	0.8	67.87	drempel	0.8	247.25
r.o.s	0.9	151.01	r.o.s	0.9	79.51	r.o.s	0.9	329.35

Tabel 5.3.8: Aangepaste normaal- gemiddelde VI-lengtes vir $q_{0.999}$ van die veralgemeende en streng Pareto-verdeling, sowel as die Burr-verdeling

Rakende die sydigheid van puntberamings van $q_{0.99}$ en $q_{0.999}$: Dit blyk uit tabel 5.3.9 dat beide stertkwantiele van die veralgemeende en streng Pareto-verdelings by kleiner steekproefgroottes onderberaam word. By 'n groot steekproef soos $n = 5000$ word $q_{0.99}$ en $q_{0.999}$ egter meestal oorberaam. Die grootste mediaan afwyking -'n oorberaming- geskied deurgaans by die kombinasie van 200 waarnemings en die 0.50-kwantiel van hierdie steekproef as drempelwaarde. In terme van die mediaan wil dit voorkom of beide $q_{0.99}$ en $q_{0.999}$ van die Burr-verdeling tipies onderberaam word. Soos tabel 5.3.10 egter aantoon, verkry ons gemengde resultate indien die gemiddelde kwantiel afwykings beskou word.

Veralgemeende Pareto							Pareto					
steekproefgrootte (n)							steekproefgrootte (n)					
	$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	2.05	0.28	0.08	0.35	-0.11	0.5	0.69	-0.03	-0.14	-0.01	-0.18
drempel	0.8	~	0.95	0.27	-0.38	-0.04	drempel	0.8	0.25	0.12	0.09	-0.08
r.o.s	0.9	~	0.71	-0.07	-0.18	-0.07	r.o.s	0.9	-0.62	-1.02	-0.68	-0.40

steekproefgrootte (n)							steekproefgrootte (n)					
	$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	14.13	1.52	1.17	2.21	-1.00	0.5	5.81	0.99	-0.78	0.36	-1.03
drempel	0.8	~	8.31	-0.32	-0.18	0.47	drempel	0.8	2.82	0.70	0.40	0.48
r.o.s	0.9	~	-0.11	-2.26	-0.60	-0.98	r.o.s	0.9	1.22	-0.51	-1.37	-0.41

Tabel 5.3.9: Mediaan afwykings vir $q_{0.99}$ en $q_{0.999}$ vir die veralgemeende en streng Pareto-verdelings.

Burr													
Mediaan	steekproefgrootte (n)					Gemiddelde	steekproefgrootte (n)						
	200	500	1000	2000	5000		200	500	1000	2000	5000		
	0.5	~	~	~	~		0.5	~	~	~	~	~	
drempel	0.8	~	1.24	0.23	0.24	0.16	drempel	0.8	~	0.02	0.03	-0.06	0.06
r.o.s	0.9	~	~	0.45	-0.04	0.05	r.o.s	0.9	~	~	0.01	-0.09	-0.10

steekproefgrootte (n)						steekproefgrootte (n)							
	200	500	1000	2000	5000		200	500	1000	2000	5000		
	0.5	~	~	~	~	0.5	~	~	~	~	~		
drempel	0.8	~	15.06	5.82	5.42	5.34	drempel	0.8	~	-6.75	0.20	0.51	3.47
r.o.s	0.9	~	~	8.87	1.64	2.13	r.o.s	0.9	~	~	-4.50	-2.38	0.80

Tabel 5.3.10: Mediaan en gemiddelde afwykings vir $q_{0.99}$ en $q_{0.999}$ vir die Burr-verdeling.

Die standaardafwyking (SA) van die kwantielafwykings (sien vergelyking 5.4) word in tabel 5.3.11 vir die veralgemeende en streng Pareto-verdelings gegee. Die heel grootste SA kom deurgaans voor by die kombinasie van 200 waarnemings en 'n 0.5-kwantiel as drempel. In tabel 5.3.12 word die SA en MAE vir die Burr-verdeling gegee. Hier vind ons die hoogste SA en MAE by die

kombinasie van 500 waarnemings met 'n drempelkwantiel van 0.80. Dit is verder duidelik dat die SA's afneem wanneer ons die totale steekproefgrootte n - en dus die aantal oorskrydings - by 'n vaste drempelkwantiel laat toeneem. Die keuse van hoër drempels by 'n vaste steekproefgrootte lei by $q_{0,999}$ meestal tot 'n toename in die SA. Soortgelyke resultate is ook vir ander maatstawwe van variasie soos die MAE (vergeljking 4.7) en MAD (4.6) verkry.

Veralgemeende Pareto						Pareto							
	$q_{0,99}$	steekproefgrootte (n)						$q_{0,99}$	steekproefgrootte (n)				
		200	500	1000	2000	5000			200	500	1000	2000	5000
	0.5	7.64	3.65	2.38	1.41	1.08		0.5	2.76	1.92	1.26	0.91	0.60
drempel	0.8	~	3.35	2.48	1.77	1.08	drempel	0.8	~	2.06	1.11	0.95	0.60
r.o.s	0.9	~	3.02	2.61	1.94	1.07	r.o.s	0.9	~	1.99	1.46	0.96	0.70

Veralgemeende Pareto						Pareto							
	$q_{0,999}$	steekproefgrootte (n)						$q_{0,999}$	steekproefgrootte (n)				
		200	500	1000	2000	5000			200	500	1000	2000	5000
	0.5	67.61	25.20	15.33	9.72	7.49		0.5	21.59	13.36	7.72	5.97	3.93
drempel	0.8	~	28.81	19.26	15.14	8.73	drempel	0.8	~	17.89	10.31	7.87	4.75
r.o.s	0.9	~	33.64	22.67	17.39	10.04	r.o.s	0.9	~	14.04	11.47	7.54	5.96

Tabel 5.3.11: Standaardafwykings van die mediaan afwykings vir $q_{0,99}$ en $q_{0,999}$ vir die veralgemeende en streng Pareto-verdelings.

Burr													
SA		steekproefgrootte (n)					MAE	steekproefgrootte (n)					
		200	500	1000	2000	5000		200	500	1000	2000	5000	
	0.5	~	~	~	~	~		0.5	~	~	~	~	~
drempel	0.8	~	5.59	3.59	2.76	1.82	drempel	0.8	~	3.51	2.14	1.82	1.01
r.o.s	0.9	~	~	3.95	2.79	1.75	r.o.s	0.9	~	~	2.56	1.75	1.17

Veralgemeende Pareto						Pareto							
	$q_{0,99}$	steekproefgrootte (n)						$q_{0,999}$	steekproefgrootte (n)				
		200	500	1000	2000	5000			200	500	1000	2000	5000
	0.5	~	~	~	~	~		0.5	~	~	~	~	~
drempel	0.8	~	77.55	42.60	29.07	18.72	drempel	0.8	~	35.87	24.53	19.07	11.52
r.o.s	0.9	~	~	59.39	37.11	21.29	r.o.s	0.9	~	~	31.31	21.81	15.33

Tabel 5.3.12: Standaardafwykings van die kwantiel afwykings en MAE's vir $q_{0,99}$ en $q_{0,999}$ van die Burr-verdeling.

5.3.2 Die Gumbel-domein: $\gamma = 0$

Die lede van die Gumbel-domein van maksimum aantrekking waarna gekyk word, is die standaard logistiese, standaard normaal- en gammaverdeling met 'n vormparameter gelyk aan 3. Vir die twee simmetriese verdelings onder beskouing, die normaal- en die logistiese verdeling, wys tabelle 5.3.13 en 5.3.14 baie duidelik dat sowel die normaal- en profielmetode ten opsigte van oordekkingswaarskynlikheid baie swak vaar in die beraming van 'n 95%-VI vir $q_{0,99}$ en $q_{0,999}$ indien 'n mediaan drempel gebruik word. Vir die normaalverdeling is die onderskeie OW's 0.15 en 0.12 by die 0.99-kwantiel; vir die logistiese verdeling is die onderskeie OW's 0.38 en 0.46 by die 0.999-kwantiel. Oor al die verdelings waarna gekyk word, is hierdie die heel laagste verkree OW's. Die OW's verbeter aansienlik by hoër drempelwaardes. Hier is die profiel oordekkings meestal hoër as vir die normaal-VI's. Ook by die gammaverdeling (tabel 5.3.15) word die laagste OW's verkry by 'n kombinasie van 5 000 datapunte en 'n mediaan drempel. Die laagste OW hier is 0.67. Oor al drie verdelings in die Gumbel-domein is die grootste verskil in OW tussen die normaal- en profielmetode by die kombinasie van 200 waardes en 'n mediaan drempel. Hier is die profiel-OW telkens aansienlik hoër as die normaal-OW.

Normaal											
Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)				
$q_{0,99}$	200	500	1000	2000	5000	$q_{0,99}$	200	500	1000	2000	5000
0.5	0.77	0.79	0.76	0.55	0.15	0.5	0.91	0.77	0.68	0.46	0.12
drempel 0.8	~	0.90	0.92	0.91	0.95	drempel 0.8	~	0.89	0.93	0.90	0.83
r.o.s 0.9	~	~	0.94	0.95	0.91	r.o.s 0.9	~	~	0.95	0.93	0.92

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)				
$q_{0,999}$	200	500	1000	2000	5000	$q_{0,999}$	200	500	1000	2000	5000
0.5	0.59	0.59	0.62	0.54	0.50	0.5	0.78	0.70	0.68	0.58	0.49
drempel 0.8	~	0.75	0.76	0.84	0.92	drempel 0.8	~	0.88	0.88	0.86	0.81
r.o.s 0.9	~	~	0.81	0.87	0.76	r.o.s 0.9	~	~	0.99	0.89	0.93

Tabel 5.3.13: OW's vir 95%-VI's vir $q_{0,99}$ en $q_{0,999}$, normaalverdeling.

Logisties

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)				
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	0.80	0.83	0.84	0.78	0.61	0.5	0.90	0.85	0.84	0.76	0.50
drempel 0.8	~	0.88	0.80	0.93	0.88	drempel 0.8	~	0.94	0.93	0.91	0.88
r.o.s 0.9	~	~	0.85	0.93	0.948	r.o.s 0.9	~	~	0.88	0.93	0.91

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)				
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	0.62	0.62	0.61	0.53	0.38	0.5	0.83	0.78	0.66	0.67	0.46
drempel 0.8	~	0.79	0.68	0.88	0.86	drempel 0.8	~	0.92	0.93	0.94	0.88
r.o.s 0.9	~	~	0.76	0.91	0.866	r.o.s 0.9	~	~	0.93	0.94	0.94

Tabel 5.3.14: OW's vir 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ van die logistiese verdeling.

Gamma

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)				
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	0.82	0.89	0.91	0.91	0.85	0.5	0.90	0.90	0.92	0.91	0.85
drempel 0.8	~	0.92	0.93	0.94	0.92	drempel 0.8	~	0.92	0.94	0.93	0.92
r.o.s 0.9	~	~	0.93	0.95	0.94	r.o.s 0.9	~	~	0.93	0.92	0.94

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)				
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	0.73	0.76	0.76	0.75	0.67	0.5	0.91	0.88	0.87	0.81	0.77
drempel 0.8	~	0.82	0.87	0.88	0.89	drempel 0.8	~	0.92	0.92	0.93	0.93
r.o.s 0.9	~	~	0.83	0.89	0.92	r.o.s 0.9	~	~	0.95	0.94	0.93

Tabel 5.3.15: OW's vir 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ van die gammaverdeling.

Behalwe vir 'n enkele uitsondering, blyk dit uit tabelle 5.3.16 tot 5.3.18 dat die langste VI deurgaans by die kombinasie van 200 waardes en die mediaan as drempel verkry word. By laer steekproefgroottes –veral by die bogenoemde kombinasie- is die profiel-VI's aansienlik langer. By die groter steekproewe is die

profiel- en normaal-VI's egter ongeveer ewe lank. Dit is dus weer 'n geval dat langer/korter VI-lengtes met hoër/laer OW's gepaard gaan.

Normaal

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000		5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.59	0.39	0.27	0.19	0.12		0.5	0.68	0.403	0.28	0.20	0.12
drempel	0.8	~	0.47	0.34	0.24	0.15	drempel	0.8	~	0.50	0.34	0.24	0.15
r.o.s	0.9	~	~	0.35	0.26	0.17	r.o.s	0.9	~	~	0.56	0.25	0.16

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000		5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	0.98	0.62	0.41	0.27	0.16		0.5	1.22	0.66	0.44	0.29	0.17
drempel	0.8	~	0.98	0.68	0.47	0.30	drempel	0.8	~	1.23	0.75	0.50	0.30
r.o.s	0.9	~	~	0.74	0.55	0.35	r.o.s	0.9	~	~	2.01	0.62	0.36

Tabel 5.3.16: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die normaalverdeling.



Logisties

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000		5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	1.64	1.05	0.76	0.53	0.34		0.5	1.94	1.12	0.77	0.54	0.34
drempel	0.8	~	1.27	0.90	0.67	0.42	drempel	0.8	~	1.40	0.95	0.66	0.41
r.o.s	0.9	~	~	0.98	0.69	0.34	r.o.s	0.9	~	~	0.96	0.67	0.42

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000		5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	3.37	2.07	1.51	0.99	0.62		0.5	4.41	2.31	1.53	1.05	0.64
drempel	0.8	~	3.61	2.37	1.77	1.10	drempel	0.8	~	4.54	2.77	1.88	1.11
r.o.s	0.9	~	~	2.88	1.96	0.63	r.o.s	0.9	~	~	3.67	2.31	1.33

Tabel 5.3.17: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die logistiese verdeling.

Gamma

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000		5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	2.33	1.47	1.04	0.76	0.47	0.5	2.92	1.6	1.11	0.75	0.47	
drempel	0.8	~	1.59	1.17	0.83	0.53	drempel	0.8	~	1.74	1.22	0.83	0.52
r.o.s	0.9	~	~	1.21	0.87	0.56	r.o.s	0.9	~	~	1.22	0.84	0.53

Normaliteit	steekproefgrootte (n)					Profiel	steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000		5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	5.42	3.22	2.25	1.63	0.99	0.5	7.60	3.69	2.49	1.63	1.01	
drempel	0.8	~	4.24	3.08	2.17	1.37	drempel	0.8	~	5.46	3.62	2.31	1.38
r.o.s	0.9	~	~	3.32	2.38	1.52	r.o.s	0.9	~	~	4.40	2.78	1.60

Tabel 5.3.18: Gemiddelde VI-lengtes vir $q_{0.99}$ en $q_{0.999}$ van die gammaverdeling.

Tabelle 5.3.19 en 5.3.20 wys duidelik hoe $q_{0.99}$ deurgaans oorberaam word, terwyl $q_{0.999}$ feitlik sonder uitsondering onderberaam word.



Normaal

Mediaan	steekproefgrootte (n)					Gemiddeld	steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000		5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.01	-0.07	-0.03	-0.10	-0.10	0.5	0.03	-0.06	-0.03	-0.11	-0.10	
drempel	0.8	~	-0.02	0.00	-0.02	-0.03	drempel	0.8	~	-0.01	-0.01	-0.02	-0.02
r.o.s	0.9	~	0.04	-0.01	-0.01	-0.03	r.o.s	0.9	~	0.03	-0.01	-0.01	-0.02

Mediaan	steekproefgrootte (n)					Gemiddeld	steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000		5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	0.34	0.16	0.18	0.06	0.05	0.5	0.33	0.15	0.17	0.04	0.04	
drempel	0.8	~	0.11	0.07	0.06	0.02	drempel	0.8	~	0.07	0.07	0.05	0.03
r.o.s	0.9	~	0.16	0.00	0.04	0.01	r.o.s	0.9	~	0.15	-0.01	0.02	0.00

Tabel 5.3.19: Mediaan en gemiddelde afwykings vir $q_{0.99}$ en $q_{0.999}$ vir die normaalverdeling.

Logisties						Gamma							
steekproefgrootte (n)						steekproefgrootte (n)							
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000		
	0.5	0.06	-0.09	-0.06	-0.08	-0.14	0.5	-0.09	0.05	-0.02	-0.05	-0.04	
drempel	0.8	~	0.06	-0.02	-0.05	-0.06	drempel	0.8	~	-0.02	-0.03	-0.01	-0.02
r.o.s	0.9	~	0.09	-0.05	-0.02	-0.03	r.o.s	0.9	~	-0.01	0.06	-0.01	-0.03

steekproefgrootte (n)						steekproefgrootte (n)							
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000		
	0.5	1.01	0.52	0.52	0.51	0.37	0.5	0.70	0.66	0.57	0.36	0.42	
drempel	0.8	~	0.55	0.31	0.12	0.07	drempel	0.8	~	0.28	0.18	0.25	0.12
r.o.s	0.9	~	0.35	0.19	0.08	-0.01	r.o.s	0.9	~	0.51	0.39	0.05	-0.08

Tabel 5.3.20: Mediaan afwykings vir $q_{0.99}$ en $q_{0.999}$ vir die logistiese en gammaverdeling.

5.3.3 Die Weibull-domein: $\gamma < 0$

In die geval van die beta(5,3)-verdeling wat tot die Weibull-domein van maksimum aantrekking behoort, kon die simulatie-eksperimente tipies nie suksesvol voltooi word nie.

5.4 Gevolgtrekking

Drempelmodellering is die meer moderne benadering tot ekstreemwaarde-teorie. Dit behels dat ons die volle steekproef beskou en waarnemings wat 'n sekere gekose drempel oorskry, as ekstreme waardes definieer. Volgens die kernresultaat van drempelteorie kan die voorwaardelike verdeling van hierdie oorskrydings by hoë drempelwaardes met die veralgemeende Pareto-verdeling benader word. Dit beteken dat vir groot drempelwaardes u geld die benadering

$F_u \approx H_{\gamma, \beta(u)}$. In hierdie hoofstuk het ons die volgende vraag ondersoek: Hoe goed is die benadering $F_u \approx H_{\gamma, \beta(u)}$ by eindige drempelwaardes?

Soos in hoofstuk vier het ons die geldigheid van die benadering $F_u \approx H_{\gamma, \beta(u)}$, waar u eindige waardes aanneem, in terme van die akkuraatheid van die drempelgebaseerde beraming van stertkwantiele sowel as VI's vir hierdie stertkwantiele van die onderliggende verdeling F , geëvalueer. Weer eens is van 'n simulasiestudie gebruik gemaak. Ten opsigte van die kwantielberamings is die sydigheid en variasie van die puntberamings ondersoek. By die vertrouensintervalle is die lengte sowel as die oordekkingswaarskynlikheid daarvan, geëvalueer.

Twee metodes, naamlik die asimptotiese normaliteit van die MA-beramers sowel as die chikwadraatgebaseerde profielmetode, is gebruik om 95%-VI's vir $q_{0.99}$ en $q_{0.999}$ te konstrueer. Volgens die simulasië-afvoer is die profielgebaseerde VI's by die drempelbenadering deurgaans langer as dié volgens die normaalmetode. Die normaal-VI's bereik slegs by uitsondering OW's van 0.95 en is dikwels minder as 0.90. By lae drempels vaar die normaal- sowel as die profiel-VI's maar swak in die geval van onderliggende verdelings uit die Gumbel-domein. Afgesien hiervan is die profiel-OW's egter meestal stabiel in 'n nou band rondom 0.95. Vir die swaarstert verdelings soos die Pareto's en die Burr-verdeling het die profiel-VI's veral by die kleiner steekproefgroottes heelwat hoër OW's. Hier is die normaal-OW's meestal onder 0.95, terwyl die profiel-OW's dikwels hoër as 0.95 is. 'n Verdere simulasië-analise (sien paragraaf 5.3.1) het getoon dat die normaal-OW's gekalibreer kan word tot waardes wat ongeveer die profiel-OW's ewenaar, deur van meer ekstreme kwantiele soos $\Phi_{0.999}$ in die normaalmetode gebruik te maak.

Soos in hoofstuk vier, is daar 'n mate van sydigheid in die beraming van $q_{0,99}$ en $q_{0,999}$ met behulp van drempelteorie. $q_{0,999}$ van die Burr-verdeling word byvoorbeeld in mediaan-terme sistematies onderberaam, terwyl ons vir die Pareto-verdelings gemengde resultate verkry. Dit is duidelik dat al die variasie maatstawwe onder beskouing, naamlik die SA, MAE en MAD afneem wanneer ons die totale steekproefgrootte – en dus die aantal oorskrydings by 'n vaste drempelkwantiel laat toeneem.

In die volgende hoofstuk word die hooftrekke van hierdie werkstuk opgesom en moontlike kwessies vir verdere ondersoek, uitgelig.



Aanhangsel 5A

In aanhangsel 5A gee ons die volledige afvoer van die simulاسie-analise, soos bespreek in paragraaf 5.2. Die heel belangrikste faktor in hierdie studie is die keuse van die onderliggende verdeling F . Gevolglik orden ons die resultate na aanleiding van die drie maksimum domein van aantrekking van die veralgemeende ekstreemwaarde-verdeling $G_{\gamma, \mu, \sigma}$, waartoe elk van die onderliggende verdelings behoort. Die afvoer vir die verdelings in die Frechet-domein word in paragraaf 5A.1 gegee, terwyl die resultate vir die Gumbel- en Weibull-domein in onderskeidelik paragraaf 5A.2 en 5A.3 aangebied word.

In die simulاسiestudie het die klem op die puntberaming van ekstreme stertkwantiele soos $q_{0.99}$ en $q_{0.999}$ van die onderliggende verdeling, sowel as op die berekening van benaderde 95%-vertrouensintervalle (VI's) vir hierdie kwantiele, geval. Die VI's is op twee maniere verkry: Met behulp van die profiel aanneemlikheid metode, sowel as die asimptotiese normaliteit van die MA-beramers. Die oordeckingswaarskynlikhede (OW's) vir hierdie VI's is bereken, sowel as die gemiddelde en mediaan intervallengte. As variاسiemaatstawwe is die standaardafwyking en die MAD (sien paragraaf 5.2) van hierdie lengtes bereken. Hierdie afvoer word in die subparagrafe 5A1.1 en 5A.2.1 verskaf. Ten opsigte van die puntberaming van $q_{0.99}$ en $q_{0.999}$ is die gemiddelde en mediaan afwykings bereken. As maatstawwe van die variasie in dié beramings is die standaardafwyking, MAD en MAE (sien paragraaf 5.2) van die bogenoemde afwykings verkry. Dié resultate kan in die subparagrafe 5A1.2 en 5A.2.2 gevind word.

In sommige gevalle kon die simulاسie-eksperiment nie suksesvol deurgevoer word nie. `n Tilde (~) is gebruik om sulke gevalle aan te dui.

5A.1 Die Frechet-domein: $\gamma > 0$

In die Frechet-domein ondersoek ons die volgende verdelings: veralgemeende Pareto, streng Pareto en die Burr-verdeling. Hierdie verdelings is lede van die klas van sogenaamde swaartert verdelings.

5A1.1 Vertrouensinterval beraming vir die Frechet-domein

OW's van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$: Asimptotiese Normaliteit

Veralgemeende Pareto						Pareto					
steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	0.87	0.94	0.91	0.94	0.96	0.5	0.88	0.94	0.95	0.94	0.93
drempel 0.8	0.84	0.89	0.90	0.94	0.93	drempel 0.8	0.86	0.92	0.92	0.94	0.95
r.o.s 0.9	~	0.90	0.92	0.92	0.94	r.o.s 0.9	~	0.88	0.92	0.93	0.94

Veralgemeende Pareto						Pareto					
steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	0.83	0.91	0.90	0.94	0.95	0.5	0.85	0.91	0.92	0.94	0.93
drempel 0.8	0.79	0.85	0.88	0.93	0.93	drempel 0.8	0.76	0.87	0.88	0.91	0.94
r.o.s 0.9	~	0.83	0.85	0.90	0.93	r.o.s 0.9	~	0.83	0.84	0.89	0.92

Burr

steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000
0.5	~	~	~	~	~
drempel 0.8	~	0.89	0.91	0.94	0.93
r.o.s 0.9	~	~	0.93	0.95	0.96

steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000
0.5	~	~	~	~	~
drempel 0.8	~	0.84	0.86	0.92	0.89
r.o.s 0.9	~	~	0.84	0.87	0.90

Gemiddelde (links) en Mediaan (regs) lengtes van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$:

Asimptotiese Normaliteit

Veralgemeende Pareto

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	22.65	13.63	9.36	6.64	4.34	0.5	17.37	13.54	9.64	6.50	4.36	
drempel	0.8	~	14.05	10.07	7.38	4.50	drempel	0.8	~	12.46	10.20	7.35	4.51
r.o.s	0.9	~	14.92	10.55	7.45	4.59	r.o.s	0.9	~	14.52	10.27	7.27	4.56

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	186.85	94.42	61.60	43.41	28.70	0.5	101.95	86.44	61.69	41.77	28.59	
drempel	0.8	~	118.93	81.14	59.88	34.95	drempel	0.8	~	89.86	80.04	57.04	34.79
r.o.s	0.9	~	169.19	103.51	69.62	41.13	r.o.s	0.9	~	137.59	94.53	63.26	40.21

Pareto

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	10.58	7.26	4.92	3.42	2.20	0.5	9.19	6.69	5.01	3.35	2.21	
drempel	0.8	~	7.34	4.98	3.68	2.26	drempel	0.8	~	6.82	4.89	3.54	2.24
r.o.s	0.9	~	7.29	5.39	3.73	2.31	r.o.s	0.9	~	7.30	5.17	3.73	2.29

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	77.40	51.22	32.97	22.66	14.67	0.5	52.80	42.70	33.29	21.92	14.65	
drempel	0.8	~	64.36	40.48	30.12	17.70	drempel	0.8	~	47.22	36.05	27.49	17.01
r.o.s	0.9	~	70.16	52.30	35.01	21.21	r.o.s	0.9	~	62.32	48.78	33.43	20.49

Burr

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	~	~	~	~	~	0.5	~	~	~	~	~	
drempel	0.8	~	20.87	14.46	10.28	6.41	drempel	0.8	~	18.33	14.11	9.94	6.33
r.o.s	0.9	~	~	16.35	11.62	7.29	r.o.s	0.9	~	~	15.57	11.51	7.25

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	~	~	~	~	~	0.5	~	~	~	~	~	
drempel	0.8	~	253.81	158.59	109.94	66.30	drempel	0.8	~	176.31	141.90	102.22	64.30
r.o.s	0.9	~	~	204.80	136.53	81.22	r.o.s	0.9	~	~	164.54	126.38	78.38

Standaardafwyking (links) en MAD (regs) van die lengtes van 95%-VI's vir

$q_{0.99}$ en $q_{0.999}$: Asimptotiese Normaliteit

Veralgemeende Pareto

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	16.30	4.65	2.13	0.93	0.44	0.5	5.53	2.87	1.32	0.54	0.28
drempel 0.8	~	5.25	2.52	1.38	0.53	drempel 0.8	~	2.99	1.71	1.08	0.41
r.o.s 0.9	~	5.34	2.91	1.60	0.57	r.o.s 0.9	~	3.75	1.93	1.16	0.37

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	248.00	49.96	20.77	9.92	4.47	0.5	47.68	28.66	14.70	5.89	3.03
drempel 0.8	~	84.17	36.51	20.94	7.28	drempel 0.8	~	38.79	23.53	12.02	4.87
r.o.s 0.9	~	124.31	54.02	28.38	10.61	r.o.s 0.9	~	59.72	36.69	19.37	6.62

Pareto

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	5.65	2.41	1.05	0.56	0.24	0.5	2.82	1.41	0.72	0.34	0.16
drempel 0.8	~	3.28	1.31	0.07	0.29	drempel 0.8	~	2.20	0.70	0.49	0.14
r.o.s 0.9	~	2.67	1.45	0.69	0.34	r.o.s 0.9	~	1.85	1.04	0.46	0.21

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	72.34	27.55	10.57	5.83	2.46	0.5	23.26	13.56	6.84	3.22	1.62
drempel 0.8	~	49.51	20.21	10.69	3.91	drempel 0.8	~	24.44	10.00	5.95	1.64
r.o.s 0.9	~	46.20	27.71	12.24	6.12	r.o.s 0.9	~	34.24	13.62	5.61	3.95

Burr

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	~	~	~	~	~	0.5	~	~	~	~	~
drempel 0.8	~	10.06	4.44	2.35	1.01	drempel 0.8	~	5.21	2.65	1.59	0.54
r.o.s 0.9	~	~	5.98	2.96	1.15	r.o.s 0.9	~	~	3.95	1.81	0.79

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	~	~	~	~	~	0.5	~	~	~	~	~
drempel 0.8	~	262.99	94.50	42.86	17.30	drempel 0.8	~	78.96	48.60	25.90	9.50
r.o.s 0.9	~	~	177.71	71.86	24.94	r.o.s 0.9	~	~	72.50	38.86	16.83

OW's van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$: Profielmetode

Veralgemeende Pareto							Pareto						
steekproefgrootte (n)							steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000	5000		$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.95	0.95	0.95	0.95	~		0.5	0.95	0.96	0.94	0.94	0.93
drempel	0.8	~	0.93	0.94	0.94	0.94	drempel	0.8	~	0.94	0.95	0.95	0.93
r.o.s	0.9	~	0.91	0.94	0.91	0.95	r.o.s	0.9	~	~	0.91	0.93	0.92

steekproefgrootte (n)							steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000	5000		$q_{0.999}$	200	500	1000	2000	5000
	0.5	0.96	0.94	0.96	0.94	~		0.5	0.95	0.97	0.94	0.94	0.94
drempel	0.8	~	0.94	0.95	0.95	0.96	drempel	0.8	~	0.95	0.96	0.95	0.96
r.o.s	0.9	~	0.92	0.94	0.91	0.98	r.o.s	0.9	~	~	0.94	0.95	0.94

Burr

steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000	5000
	0.5	~	~	~	~	~
drempel	0.8	~	0.94	0.93	0.96	0.95
r.o.s	0.9	~	~	0.92	0.92	0.91

steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000	5000
	0.5	~	~	~	~	~
drempel	0.8	~	0.92	0.92	0.96	0.97
r.o.s	0.9	~	~	0.94	0.93	0.92

Gemiddelde (links) en Mediaan (regs) lengtes van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$:

Profielmetode

Vergemeende Pareto

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	29.43	15.72	10.26	6.98	~	0.5	25.79	15.25	9.68	6.88	~	
drempel	0.8	~	17.64	11.17	7.29	4.48	drempel	0.8	~	16.45	10.86	7.15	4.41
r.o.s	0.9	~	17.62	10.88	7.05	4.37	r.o.s	0.9	~	15.70	10.54	6.81	4.33

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	319.60	127.52	76.70	48.29	~	0.5	221.75	111.36	66.30	45.56	~	
drempel	0.8	~	219.48	110.32	63.42	36.96	drempel	0.8	~	168.65	103.76	59.06	35.83
r.o.s	0.9	~	381.33	149.36	79.14	43.96	r.o.s	0.9	~	219.59	122.03	67.25	42.22

Pareto

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	15.40	7.67	5.04	3.46	2.15	0.5	13.42	7.32	4.93	3.41	2.14	
drempel	0.8	~	8.96	5.55	3.71	2.24	drempel	0.8	~	8.11	5.09	3.72	2.22
r.o.s	0.9	~	~	5.41	3.62	2.19	r.o.s	0.9	~	~	5.46	3.47	2.18

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	171.07	60.63	36.38	23.87	14.38	0.5	115.94	53.19	34.06	23.12	14.25	
drempel	0.8	~	115.13	54.05	32.92	18.43	drempel	0.8	~	82.52	48.76	31.93	17.93
r.o.s	0.9	~	~	73.63	40.91	22.19	r.o.s	0.9	~	~	57.13	36.18	21.40

Burr

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	~	~	~	~	~	0.5	~	~	~	~	~	
drempel	0.8	~	26.54	15.36	10.65	6.59	drempel	0.8	~	24.71	15.00	10.74	6.45
r.o.s	0.9	~	~	16.10	10.49	6.41	r.o.s	0.9	~	~	14.88	10.26	6.38

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	~	~	~	~	~	0.5	~	~	~	~	~	
drempel	0.8	~	473.25	199.80	127.45	74.15	drempel	0.8	~	353.25	176.30	122.70	71.57
r.o.s	0.9	~	~	335.01	161.32	89.81	r.o.s	0.9	~	~	242.92	143.02	87.03

Standaardafwyking (links) en MAD (regs) van die lengtes van 95%-VI's vir

$q_{0.99}$ en $q_{0.999}$: Profielmetode

Veralgemeende Pareto

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	16.77	5.58	5.64	1.10	~	0.5	8.44	3.73	1.11	0.80	
drempel 0.8	~	7.72	3.12	1.47	0.56	drempel 0.8		3.69	2.26	1.13	0.40
r.o.s 0.9	~	8.80	3.34	1.67	0.57	r.o.s 0.9			2.15	0.71	0.27

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	353.11	80.31	99.57	12.23	~	0.5	136.44	41.52	11.95	7.27	
drempel 0.8	~	182.57	56.67	23.07	8.25	drempel 0.8		52.58	43.87	13.91	5.66
r.o.s 0.9	~	473.84	104.79	41.24	11.75	r.o.s 0.9			43.09	19.94	5.17

Pareto

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	8.62	2.33	1.19	0.57	0.22	0.5	4.70	1.80	0.82	0.34	0.18
drempel 0.8	~	4.12	1.59	0.69	0.27	drempel 0.8	~	3.19	0.85	0.44	0.23
r.o.s 0.9	~	~	1.85	0.78	0.31	r.o.s 0.9	~	~	0.96	0.51	0.21

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	176.47	30.44	14.10	6.27	2.29	0.5	53.83	23.13	11.31	4.67	1.70
drempel 0.8	~	106.84	27.67	10.76	3.98	drempel 0.8	~	58.68	10.53	5.61	3.46
r.o.s 0.9	~	~	55.16	19.55	6.33	r.o.s 0.9	~	~	17.46	11.95	4.49

Burr

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	~	~	~	~	~	0.5	~	~	~	~	~
drempel 0.8	~	12.31	4.99	2.07	0.88	drempel 0.8	~	7.26	3.25	1.33	0.59
r.o.s 0.9	~	~	6.40	2.44	1.03	r.o.s 0.9	~	~	3.85	1.46	0.67

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	~	~	~	~	~	0.5	~	~	~	~	~
drempel 0.8	~	421.00	117.95	42.84	17.57	drempel 0.8	~	191.50	65.45	25.20	11.11
r.o.s 0.9	~	~	338.65	82.19	28.33	r.o.s 0.9	~	~	124.48	35.05	17.99

5A1.2 Kwantielberaming vir die Frechet-domein

Die Mediaan (links) en Gemiddelde kwantielafwyking (regs) ten opsigte van

$q_{0,99}$ en $q_{0,999}$:

Veralgemeende Pareto

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0,99}$	200	500	1000	2000	5000	$q_{0,99}$	200	500	1000	2000	5000
0.5	2.05	0.28	0.08	0.35	-0.11	0.5	0.09	0.05	0.25	0.29	-0.13
drempel 0.8	~	0.95	0.27	-0.38	-0.04	drempel 0.8	~	0.40	0.21	-0.20	0.04
r.o.s 0.9	~	0.71	-0.07	-0.18	-0.07	r.o.s 0.9	~	0.13	-0.24	-0.36	-0.14

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0,999}$	200	500	1000	2000	5000	$q_{0,999}$	200	500	1000	2000	5000
0.5	14.13	1.52	1.17	2.21	-1.00	0.5	-8.80	-1.62	1.26	1.52	-0.08
drempel 0.8	~	8.31	-0.32	-0.18	0.47	drempel 0.8	~	-0.54	-0.73	-2.81	0.05
r.o.s 0.9	~	-0.11	-2.26	-0.60	-0.98	r.o.s 0.9	~	-8.75	-4.75	-3.58	-1.25

Pareto

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0,99}$	200	500	1000	2000	5000	$q_{0,99}$	200	500	1000	2000	5000
0.5	0.69	-0.03	-0.14	-0.01	-0.18	0.5	0.29	-0.31	-0.13	-0.04	-0.15
drempel 0.8		0.25	0.12	0.09	-0.08	drempel 0.8		-0.11	0.13	-0.04	-0.01
r.o.s 0.9		-0.62	-1.02	-0.68	-0.40	r.o.s 0.9		-0.82	-0.83	-0.67	-0.50

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0,999}$	200	500	1000	2000	5000	$q_{0,999}$	200	500	1000	2000	5000
0.5	5.81	0.99	-0.78	0.36	-1.03	0.5	-0.57	-2.72	-0.95	-0.41	-0.09
drempel 0.8		2.82	0.70	0.40	0.48	drempel 0.8		-2.33	-0.24	-1.50	-0.24
r.o.s 0.9		1.22	-0.51	-1.37	-0.41	r.o.s 0.9		-1.33	-3.01	-1.91	-1.01

Burr

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0,99}$	200	500	1000	2000	5000	$q_{0,99}$	200	500	1000	2000	5000
0.5	~	~	~	~	~	0.5	~	~	~	~	~
drempel 0.8	~	1.24	0.23	0.24	0.16	drempel 0.8	~	0.02	0.03	-0.06	0.06
r.o.s 0.9	~	~	0.45	-0.04	0.05	r.o.s 0.9	~	~	0.01	-0.09	-0.10

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0,999}$	200	500	1000	2000	5000	$q_{0,999}$	200	500	1000	2000	5000
0.5	~	~	~	~	~	0.5	~	~	~	~	~
drempel 0.8	~	15.06	5.82	5.42	5.34	drempel 0.8	~	-6.75	0.20	0.51	3.47
r.o.s 0.9	~	~	8.87	1.64	2.13	r.o.s 0.9	~	~	-4.50	-2.38	0.80

Die standaardafwyking van die kwantielafwyking ten opsigte van

$q_{0,99}$ en $q_{0,999}$:

Veralgemeende Pareto

		steekproefgrootte (n)					
		$q_{0,99}$	200	500	1000	2000	5000
	0.5		7.64	3.65	2.38	1.41	1.08
drempel	0.8	~	3.35	2.48	1.77	1.08	
r.o.s	0.9	~	3.02	2.61	1.94	1.07	

		steekproefgrootte (n)					
		$q_{0,999}$	200	500	1000	2000	5000
	0.5		67.61	25.20	15.33	9.72	7.49
drempel	0.8	~	28.81	19.26	15.14	8.73	
r.o.s	0.9	~	33.64	22.67	17.39	10.04	

Pareto

		steekproefgrootte (n)					
		$q_{0,99}$	200	500	1000	2000	5000
	0.5		2.76	1.92	1.26	0.91	0.60
drempel	0.8	~	2.06	1.11	0.95	0.60	
r.o.s	0.9	~	1.99	1.46	0.96	0.70	

		steekproefgrootte (n)					
		$q_{0,999}$	200	500	1000	2000	5000
	0.5		21.59	13.36	7.72	5.97	3.93
drempel	0.8	~	17.89	10.31	7.87	4.75	
r.o.s	0.9	~	14.04	11.47	7.54	5.96	

Burr

		steekproefgrootte (n)					
		$q_{0,99}$	200	500	1000	2000	5000
	0.5		~	~	~	~	~
drempel	0.8	~	5.59	3.59	2.76	1.82	
r.o.s	0.9	~	~	3.95	2.79	1.75	

		steekproefgrootte (n)					
		$q_{0,999}$	200	500	1000	2000	5000
	0.5		~	~	~	~	~
drempel	0.8	~	77.55	42.60	29.07	18.72	
r.o.s	0.9	~	~	59.39	37.11	21.29	

Die MAE (links) en MAD (regs) ten opsigte van $q_{0.99}$ en $q_{0.999}$:

Veralgemeende Pareto

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	3.64	2.72	1.46	1.08	0.82	0.5	3.04	2.71	1.54	0.95	0.85
drempel 0.8	~	2.33	1.61	1.38	0.75	drempel 0.8	~	1.89	1.63	1.31	0.73
r.o.s 0.9	~	2.36	1.84	1.19	0.73	r.o.s 0.9	~	2.41	1.80	1.12	0.73

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	22.33	15.35	10.29	6.95	4.65	0.5	16.16	13.93	10.44	5.93	5.08
drempel 0.8	~	18.13	13.01	9.23	6.51	drempel 0.8	~	14.92	12.84	9.23	6.04
r.o.s 0.9	~	18.41	14.82	11.50	6.11	r.o.s 0.9	~	18.31	16.25	12.10	5.80

Pareto

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	1.66	1.00	0.90	0.56	0.41	0.5	1.59	1.00	0.88	0.55	0.38
drempel 0.8	~	1.43	0.63	0.67	0.33	drempel 0.8	~	1.54	0.62	0.59	0.31
r.o.s 0.9	~	1.14	1.22	0.86	0.51	r.o.s 0.9	~	1.02	1.07	0.70	0.45

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	9.60	7.36	5.32	3.09	2.32	0.5	8.40	7.00	4.83	3.34	2.66
drempel 0.8	~	11.39	5.11	4.89	2.46	drempel 0.8	~	9.74	5.04	4.78	2.01
r.o.s 0.9	~	11.17	7.43	4.83	4.18	r.o.s 0.9	~	10.05	7.21	4.78	4.04

Burr

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
0.5	~	~	~	~	~	0.5	~	~	~	~	~
drempel 0.8	~	3.51	2.14	1.82	1.01	drempel 0.8	~	3.17	2.17	1.77	1.07
r.o.s 0.9	~	~	2.56	1.75	1.17	r.o.s 0.9	~	~	2.51	1.75	1.19

steekproefgrootte (n)						steekproefgrootte (n)					
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
0.5	~	~	~	~	~	0.5	~	~	~	~	~
drempel 0.8	~	35.87	24.53	19.07	11.52	drempel 0.8	~	29.74	23.61	19.02	10.26
r.o.s 0.9	~	~	31.31	21.81	15.33	r.o.s 0.9	~	~	28.45	21.91	14.98

5A.2 Die Gumbel-domein: $\gamma = 0$

Die verdelings in hierdie domein waarna ons ondersoek instel, is die simmetriese normaal- en logistiese verdelings, sowel as die gammaverdeling.

5A2.1 Vertrauensinterval beraming vir die Gumbel-domein

OW's van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$: Asimptotiese Normaliteit

Normaal							Logisties						
steekproefgrootte (n)							steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000	
	0.5	0.77	0.79	0.76	0.55	0.15	0.5	0.80	0.83	0.84	0.78	0.61	
drempel	0.8	0.82	0.90	0.92	0.91	0.95	0.8	0.61	0.88	0.80	0.93	0.88	
r.o.s	0.9	-	0.91	0.94	0.95	0.91	0.9	-	0.91	0.85	0.93	0.95	

Normaal							Logisties						
steekproefgrootte (n)							steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000	
	0.5	0.59	0.59	0.62	0.54	0.50	0.5	0.62	0.62	0.61	0.53	0.38	
drempel	0.8	0.72	0.75	0.76	0.84	0.92	0.8	0.81	0.79	0.68	0.88	0.86	
r.o.s	0.9	-	0.73	0.81	0.87	0.76	0.9	-	0.77	0.76	0.91	0.87	

Gamma						
steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.82	0.89	0.91	0.91	0.85
drempel	0.8	0.83	0.92	0.93	0.94	0.92
r.o.s	0.9	~	0.90	0.93	0.95	0.94

Gamma						
steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000	5000
	0.5	0.73	0.76	0.76	0.75	0.67
drempel	0.8	0.73	0.82	0.87	0.88	0.89
r.o.s	0.9	~	0.78	0.83	0.89	0.92

Gemiddelde (links) en Mediaan (regs) lengtes van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$:

Asimptotiese Normaliteit

Normaal

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.59	0.39	0.27	0.19	0.12	0.5	0.57	0.39	0.26	0.19	0.12	
drempel	0.8	~	0.47	0.34	0.24	0.15	drempel	0.8	~	0.45	0.34	0.25	0.16
	r.o.s	0.9	~	0.49	0.35	0.26	r.o.s	0.9	~	0.49	0.35	0.26	0.17

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	0.98	0.62	0.41	0.27	0.16	0.5	0.90	0.61	0.39	0.27	0.16	
drempel	0.8	~	0.98	0.68	0.47	0.30	drempel	0.8	~	0.91	0.68	0.47	0.30
	r.o.s	0.9	~	1.15	0.74	0.55	r.o.s	0.9	~	1.05	0.68	0.55	0.34

Logisties

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	1.64	1.05	0.76	0.53	0.34	0.5	1.67	1.06	0.75	0.52	0.33	
drempel	0.8	~	1.27	0.90	0.67	0.42	drempel	0.8	~	1.25	0.89	0.66	0.42
	r.o.s	0.9	~	1.31	0.98	0.69	r.o.s	0.9	~	1.28	0.98	0.69	0.34

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	3.37	2.07	1.51	0.99	0.62	0.5	3.25	2.03	1.45	0.98	0.63	
drempel	0.8	~	3.61	2.37	1.77	1.10	drempel	0.8	~	3.44	2.27	1.75	1.09
	r.o.s	0.9	~	4.14	2.88	1.96	r.o.s	0.9	~	3.36	2.75	1.89	0.63

Gamma

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	2.33	1.47	1.04	0.76	0.47	0.5	2.13	1.46	1.01	0.75	0.47	
drempel	0.8	~	1.59	1.17	0.83	0.53	drempel	0.8	~	1.58	1.16	0.84	0.53
	r.o.s	0.9	~	1.72	1.21	0.87	r.o.s	0.9	~	1.65	1.21	0.87	0.55

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	5.42	3.22	2.25	1.63	0.99	0.5	4.44	3.14	2.11	1.61	0.98	
drempel	0.8	~	4.24	3.08	2.17	1.37	drempel	0.8	~	4.11	2.88	2.15	1.36
	r.o.s	0.9	~	5.37	3.32	2.38	r.o.s	0.9	~	4.64	3.19	2.34	1.49

Standaardafwyking (links) en MAD (regs) van die lengtes van 95%-VI's vir

$q_{0.99}$ en $q_{0.999}$: Asimptotiese Normaliteit

Normaal

		steekproefgrootte (n)							steekproefgrootte (n)				
$q_{0.99}$		200	500	1000	2000	5000	$q_{0.99}$		200	500	1000	2000	5000
	0.5	0.19	0.08	0.04	0.02	0.01		0.5	0.11	0.06	0.03	0.01	0.00
drempel	0.8	~	0.12	0.05	0.03	0.01	drempel	0.8	~	0.07	0.04	0.02	0.01
r.o.s	0.9	~	0.12	0.06	0.03	0.01	r.o.s	0.9	~	0.08	0.04	0.02	0.01

		steekproefgrootte (n)							steekproefgrootte (n)				
$q_{0.999}$		200	500	1000	2000	5000	$q_{0.999}$		200	500	1000	2000	5000
	0.5	0.43	0.19	0.09	0.04	0.02		0.5	0.27	0.13	0.06	0.02	0.01
drempel	0.8	~	0.46	0.19	0.10	0.04	drempel	0.8	~	0.24	0.13	0.06	0.03
r.o.s	0.9	~	0.63	0.28	0.13	0.06	r.o.s	0.9	~	0.37	0.17	0.11	0.04

Logisties

		steekproefgrootte (n)							steekproefgrootte (n)				
$q_{0.99}$		200	500	1000	2000	5000	$q_{0.99}$		200	500	1000	2000	5000
	0.5	0.51	0.23	0.12	0.06	0.02		0.5	0.35	0.15	0.08	0.04	0.01
drempel	0.8	~	0.35	0.16	0.08	0.03	drempel	0.8	~	0.26	0.13	0.06	0.02
r.o.s	0.9	~	0.37	0.15	0.08	0.02	r.o.s	0.9	~	0.23	0.09	0.06	0.02

		steekproefgrootte (n)							steekproefgrootte (n)				
$q_{0.999}$		200	500	1000	2000	5000	$q_{0.999}$		200	500	1000	2000	5000
	0.5	1.48	0.67	0.37	0.17	0.07		0.5	1.02	0.42	0.26	0.12	0.05
drempel	0.8	~	1.79	0.79	0.37	0.16	drempel	0.8	~	1.15	0.56	0.25	0.09
r.o.s	0.9	~	2.86	1.07	0.54	0.06	r.o.s	0.9	~	1.33	0.64	0.36	0.04

Gamma

		steekproefgrootte (n)							steekproefgrootte (n)				
$q_{0.99}$		200	500	1000	2000	5000	$q_{0.99}$		200	500	1000	2000	5000
	0.5	0.95	0.30	0.17	0.07	0.03		0.5	0.54	0.24	0.09	0.05	0.02
drempel	0.8	~	0.38	0.21	0.09	0.03	drempel	0.8	~	0.28	0.16	0.06	0.03
r.o.s	0.9	~	0.41	0.21	0.11	0.04	r.o.s	0.9	~	0.31	0.14	0.06	0.03

		steekproefgrootte (n)							steekproefgrootte (n)				
$q_{0.999}$		200	500	1000	2000	5000	$q_{0.999}$		200	500	1000	2000	5000
	0.5	3.35	0.96	0.58	0.24	0.10		0.5	1.61	0.79	0.25	0.16	0.06
drempel	0.8	~	1.86	1.04	0.44	0.16	drempel	0.8	~	1.24	0.66	0.33	0.11
r.o.s	0.9	~	2.89	1.31	0.71	0.27	r.o.s	0.9	~	1.75	0.77	0.41	0.17

OW's van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$: Profielmetode

Normaal							Logisties						
steekproefgrootte (n)							steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000	5000		$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.91	0.77	0.68	0.46	0.12		0.5	0.90	0.85	0.84	0.76	0.50
drempel	0.8	~	0.89	0.93	0.90	0.83	drempel	0.8	~	0.94	0.93	0.91	0.88
r.o.s	0.9	~	~	0.95	0.93	0.92	r.o.s	0.9	~	~	0.88	0.93	0.91

steekproefgrootte (n)							steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000	5000		$q_{0.999}$	200	500	1000	2000	5000
	0.5	0.78	0.70	0.68	0.58	0.49		0.5	0.83	0.78	0.66	0.67	0.46
drempel	0.8	~	0.88	0.88	0.86	0.81	drempel	0.8	~	0.92	0.93	0.94	0.88
r.o.s	0.9	~	~	0.99	0.89	0.93	r.o.s	0.9	~	~	0.93	0.94	0.94

Gamma

steekproefgrootte (n)						
	$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.90	0.90	0.92	0.91	0.85
drempel	0.8	~	0.92	0.94	0.93	0.92
r.o.s	0.9	~	~	0.93	0.92	0.94

steekproefgrootte (n)						
	$q_{0.999}$	200	500	1000	2000	5000
	0.5	0.91	0.88	0.87	0.81	0.77
drempel	0.8	~	0.92	0.92	0.93	0.93
r.o.s	0.9	~	~	0.95	0.94	0.93

Gemiddelde (links) en Mediaan (regs) lengtes van 95%-VI's vir $q_{0.99}$ en $q_{0.999}$:

Profielmetode

Normaal

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.68	0.403	0.28	0.20	0.12	0.5	0.66	0.40	0.28	0.20	0.12	
drempel	0.8	~	0.50	0.34	0.24	0.15	drempel	0.8	~	0.49	0.33	0.24	0.15
	r.o.s	0.9	~	~	0.56	0.25	r.o.s	0.9	~	~	0.56	0.25	0.16

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	1.22	0.66	0.44	0.29	0.17	0.5	1.10	0.64	0.43	0.29	0.17	
drempel	0.8	~	1.23	0.75	0.50	0.30	drempel	0.8	~	1.39	0.72	0.50	0.30
	r.o.s	0.9	~	~	2.01	0.62	r.o.s	0.9	~	~	1.92	0.61	0.36

Logisties

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	1.94	1.12	0.77	0.54	0.34	0.5	1.89	1.10	0.77	0.54	0.34	
drempel	0.8	~	1.40	0.95	0.66	0.41	drempel	0.8	~	1.38	0.96	0.66	0.41
	r.o.s	0.9	~	~	0.96	0.67	r.o.s	0.9	~	~	0.95	0.67	0.42

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	4.41	2.31	1.53	1.05	0.64	0.5	4.02	2.25	1.49	1.05	0.63	
drempel	0.8	~	4.54	2.77	1.88	1.11	drempel	0.8	~	4.10	2.75	1.83	1.11
	r.o.s	0.9	~	~	3.67	2.31	r.o.s	0.9	~	~	3.40	2.20	1.30

Gamma

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	2.92	1.6	1.11	0.75	0.47	0.5	2.84	1.58	1.10	0.75	0.47	
drempel	0.8	~	1.74	1.22	0.83	0.52	drempel	0.8	~	1.67	1.19	0.83	0.51
	r.o.s	0.9	~	~	1.22	0.84	r.o.s	0.9	~	~	1.21	0.85	0.52

		steekproefgrootte (n)					steekproefgrootte (n)						
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	7.60	3.69	2.49	1.63	1.01	0.5	6.96	3.53	2.44	1.62	1.01	
drempel	0.8	~	5.46	3.62	2.31	1.38	drempel	0.8	~	4.94	3.42	2.30	1.37
	r.o.s	0.9	~	~	4.40	2.78	r.o.s	0.9	~	~	4.08	2.71	1.59

Standaardafwyking (links) en MAD (regs) van die lengtes van 95%-VI's vir

$q_{0.99}$ en $q_{0.999}$: Profiel

Normaal

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.22	0.08	0.04	0.20	0.01	0.5	0.13	0.03	0.03	0.02	0.00	
drempel	0.8	~	0.13	0.05	0.03	0.01	drempel	0.8	0.09	0.04	0.01	0.01	
r.o.s	0.9	~	~	0.12	0.03	0.01	r.o.s	0.9	~	~	0.04	0.02	0.00

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	0.56	0.19	0.09	0.05	0.02	0.5	0.39	0.05	0.07	0.05	0.01	
drempel	0.8	~	0.57	0.21	0.10	0.04	drempel	0.8	0.26	0.12	0.04	0.02	
r.o.s	0.9	~	~	0.93	0.18	0.06	r.o.s	0.9	~	~	0.22	0.14	0.04

Logisties

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.63	0.23	0.12	0.06	0.02	0.5	0.38	0.13	0.10	0.03	0.02	
drempel	0.8	~	0.35	0.15	0.08	0.03	drempel	0.8	~	0.18	0.11	0.05	0.01
r.o.s	0.9	~	~	0.18	0.09	0.03	r.o.s	0.9	~	~	0.16	0.07	0.01

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	2.20	0.70	0.36	0.01	0.07	0.5	1.14	0.37	0.30	0.13	0.05	
drempel	0.8	~	2.01	0.77	0.40	0.15	drempel	0.8	~	1.27	0.56	0.23	0.04
r.o.s	0.9	~	~	1.55	0.67	0.23	r.o.s	0.9	~	~	1.49	0.50	0.06

Gamma

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000
	0.5	0.94	0.34	0.16	0.07	0.03	0.5	0.44	0.19	0.09	0.04	0.03	
drempel	0.8	~	0.44	0.22	0.09	0.04	drempel	0.8	~	0.38	0.13	0.08	0.02
r.o.s	0.9	~	~	0.22	0.11	0.04	r.o.s	0.9	~	~	0.17	0.07	0.02

		steekproefgrootte (n)							steekproefgrootte (n)				
		$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000
	0.5	3.73	1.19	0.56	0.25	0.10	0.5	1.67	0.67	0.33	0.15	0.09	
drempel	0.8	~	2.53	1.13	0.45	0.18	drempel	0.8	~	2.05	0.77	0.42	0.11
r.o.s	0.9	~	~	1.89	0.78	0.28	r.o.s	0.9	~	~	1.27	0.39	0.14

5A2.2 Kwantielberaming vir die Gumbel-domein

Die Mediaan (links) en Gemiddelde kwantielafwyking (regs) ten opsigte van

$q_{0,99}$ en $q_{0,999}$:

Normaal

steekproefgrootte (n)						steekproefgrootte (n)							
	$q_{0,99}$	200	500	1000	2000	5000		$q_{0,99}$	200	500	1000	2000	5000
	0.5	0.006	-0.070	-0.029	-0.098	-0.097		0.5	0.025	-0.058	-0.032	-0.108	-0.100
drempel	0.8	~	-0.015	-0.001	-0.016	-0.025	drempel	0.8	~	-0.013	-0.010	-0.021	-0.023
r.o.s	0.9	~	0.039	-0.012	-0.010	-0.025	r.o.s	0.9	~	0.029	-0.007	-0.010	-0.015

steekproefgrootte (n)						steekproefgrootte (n)							
	$q_{0,999}$	200	500	1000	2000	5000		$q_{0,999}$	200	500	1000	2000	5000
	0.5	0.340	0.160	0.176	0.060	0.049		0.5	0.328	0.150	0.172	0.037	0.038
drempel	0.8	~	0.105	0.070	0.059	0.022	drempel	0.8	~	0.068	0.068	0.053	0.026
r.o.s	0.9	~	0.160	-0.005	0.039	0.006	r.o.s	0.9	~	0.145	-0.009	0.023	0.004

Logisties

steekproefgrootte (n)						steekproefgrootte (n)							
	$q_{0,99}$	200	500	1000	2000	5000		$q_{0,99}$	200	500	1000	2000	5000
	0.5	0.06	-0.09	-0.06	-0.08	-0.14		0.5	0.05	-0.11	-0.08	-0.09	-0.14
drempel	0.8		0.06	-0.02	-0.05	-0.06	drempel	0.8		0.07	-0.01	-0.06	-0.06
r.o.s	0.9		0.09	-0.05	-0.02	-0.03	r.o.s	0.9		0.05	-0.05	-0.02	-0.01

steekproefgrootte (n)						steekproefgrootte (n)							
	$q_{0,999}$	200	500	1000	2000	5000		$q_{0,999}$	200	500	1000	2000	5000
	0.5	1.01	0.52	0.52	0.51	0.37		0.5	0.79	0.43	0.48	0.49	0.38
drempel	0.8		0.55	0.31	0.12	0.07	drempel	0.8		0.49	0.29	0.07	0.09
r.o.s	0.9		0.35	0.19	0.08	-0.01	r.o.s	0.9		0.21	-0.07	0.06	-0.03

Gamma

steekproefgrootte (n)						steekproefgrootte (n)							
	$q_{0,99}$	200	500	1000	2000	5000		$q_{0,99}$	200	500	1000	2000	5000
	0.5	-0.09	0.05	-0.02	-0.05	-0.04		0.5	0.00	0.02	0.02	-0.02	-0.05
drempel	0.8	~	-0.02	-0.03	-0.01	-0.02	drempel	0.8	~	-0.05	-0.03	0.01	0.15
r.o.s	0.9	~	-0.01	0.06	-0.01	-0.03	r.o.s	0.9	~	0.05	0.05	-0.01	-0.02

steekproefgrootte (n)						steekproefgrootte (n)							
	$q_{0,999}$	200	500	1000	2000	5000		$q_{0,999}$	200	500	1000	2000	5000
	0.5	0.70	0.66	0.57	0.36	0.42		0.5	0.40	0.57	0.51	0.42	0.36
drempel	0.8	~	0.28	0.18	0.25	0.12	drempel	0.8	~	0.16	0.21	0.25	0.07
r.o.s	0.9	~	0.51	0.39	0.05	-0.08	r.o.s	0.9	~	0.28	0.27	0.11	-0.03

Die standaardafwyking van die kwantielafwyking ten opsigte van

$q_{0,99}$ en $q_{0,999}$:

Normaal

steekproefgrootte (n)

	$q_{0,99}$	200	500	1000	2000	5000
	0.5	0.185	0.131	0.101	0.077	0.600
drempel	0.8	~	0.139	0.110	0.070	0.043
r.o.s	0.9	~	0.115	0.050	0.057	0.050

steekproefgrootte (n)

	$q_{0,999}$	200	500	1000	2000	5000
	0.5	0.333	0.250	0.195	0.160	0.122
drempel	0.8	~	0.344	0.202	0.148	0.093
r.o.s	0.9	~	0.276	0.157	0.144	0.106

Logisties

steekproefgrootte (n)

	$q_{0,99}$	200	500	1000	2000	5000
	0.5	0.51	0.34	0.26	0.18	0.13
drempel	0.8		0.37	0.27	0.17	0.11
r.o.s	0.9		0.32	0.25	0.18	0.13

steekproefgrootte (n)

	$q_{0,999}$	200	500	1000	2000	5000
	0.5	1.20	0.78	0.57	0.38	0.28
drempel	0.8		1.01	0.67	0.46	0.35
r.o.s	0.9		1.12	0.81	0.45	0.43

Gamma

steekproefgrootte (n)

	$q_{0,99}$	200	500	1000	2000	5000
	0.5	0.62	0.32	0.34	0.22	0.19
drempel	0.8	~	0.48	0.30	0.19	0.15
r.o.s	0.9	~	0.48	0.32	0.25	0.14

steekproefgrootte (n)

	$q_{0,999}$	200	500	1000	2000	5000
	0.5	1.60	0.79	0.76	0.49	0.41
drempel	0.8	~	1.39	0.80	0.46	0.39
r.o.s	0.9	~	1.45	0.90	0.72	0.48

Die MAE (links) en MAD (regs) ten opsigte van $q_{0.99}$ en $q_{0.999}$

Normaal

steekproefgrootte (n)						steekproefgrootte (n)							
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000		
	0.5	0.12	0.09	0.07	0.10	0.10	0.5	0.12	0.08	0.07	0.05	0.04	
drempel	0.8	~	0.10	0.07	0.05	0.04	drempel	0.8	~	0.11	0.07	0.05	0.03
r.o.s	0.9	~	0.09	0.04	0.04	0.03	r.o.s	0.9	~	0.09	0.04	0.04	0.03

steekproefgrootte (n)						steekproefgrootte (n)							
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000		
	0.5	0.370	0.215	0.224	0.118	0.084	0.5	0.222	0.157	0.138	0.099	0.068	
drempel	0.8	~	0.223	0.150	0.091	0.067	drempel	0.8	~	0.236	0.152	0.092	0.061
r.o.s	0.9	~	0.240	0.091	0.094	0.080	r.o.s	0.9	~	0.170	0.087	0.086	0.075

Logisties

steekproefgrootte (n)						steekproefgrootte (n)							
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000		
	0.5	0.30	0.25	0.17	0.13	0.15	0.5	0.32	0.26	0.17	0.14	0.08	
drempel	0.8	~	0.25	0.16	0.10	0.10	drempel	0.8	~	0.28	0.16	0.11	0.07
r.o.s	0.9	~	0.24	0.17	0.13	0.10	r.o.s	0.9	~	0.27	0.16	0.12	0.09

steekproefgrootte (n)						steekproefgrootte (n)							
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000		
	0.5	1.14	0.77	0.61	0.51	0.37	0.5	0.59	0.55	0.35	0.27	0.16	
drempel	0.8	~	0.83	0.44	0.36	0.22	drempel	0.8	~	0.72	0.40	0.30	0.26
r.o.s	0.9	~	0.94	0.52	0.33	0.27	r.o.s	0.9	~	0.81	0.53	0.32	0.27

Gamma

steekproefgrootte (n)						steekproefgrootte (n)							
$q_{0.99}$	200	500	1000	2000	5000	$q_{0.99}$	200	500	1000	2000	5000		
	0.5	0.43	0.23	0.25	0.16	0.12	0.5	0.38	0.20	0.23	0.15	0.13	
drempel	0.8	~	0.31	0.17	0.11	0.10	drempel	0.8	~	0.31	0.16	0.11	0.09
r.o.s	0.9	~	0.26	0.24	0.18	0.09	r.o.s	0.9	~	0.26	0.21	0.18	0.08

steekproefgrootte (n)						steekproefgrootte (n)							
$q_{0.999}$	200	500	1000	2000	5000	$q_{0.999}$	200	500	1000	2000	5000		
	0.5	1.18	0.70	0.75	0.41	0.45	0.5	0.92	0.58	0.49	0.37	0.26	
drempel	0.8	~	0.89	0.46	0.37	0.29	drempel	0.8	~	0.95	0.38	0.26	0.27
r.o.s	0.9	~	1.07	0.63	0.43	0.29	r.o.s	0.9	~	0.66	0.62	0.42	0.30

A5.3 Die Weibull-domein: $\gamma < 0$

In die geval van die beta(5,3)-verdeling wat tot die Weibull-domein van maksimum aantrekking behoort, kon die simulatie-eksperimente tipies nie suksesvol voltooi word nie.



Hoofstuk 6: Samevatting

In die studie van ekstreemwaarde-teorie stel ons belang in tegnieke om rare, ontpiense gebeurtenisse wat slegs by uitsondering plaasvind, te beskryf. Dit behels die beraming van uitermate hoë kwantiele en verg dikwels dat gebeure wat nog nooit vantevore gerealiseer het nie, statisties gemodelleer word. Die akkurate modellering van sulke ekstreme waardes is tipies van kritieke belang: `n uiters groot beleggingsverlies kan tot die bankrotskap van `n finansiële instelling lei, terwyl ekstreme weerstoestande tot `n toename in die seevlak mag lei wat weer rampspoedige oorstromings van laagliggende kusgebiede tot gevolg mag hê.

In die geval van toepassings waar die praktisyn ten doel het om `n stogastiese proses by uiterste hoë vlakke te beskryf, kan die ekstreemteorie gebruik word. Hier bestaan twee alternatiewe: Die praktisyn kan die probleem met behulp van die maksimummetode, sowel as die drempelbenadering, aanspreek. Die maksimummetode behels ten eerste dat die beskikbare steekproef van onderliggende waarnemings in `n aantal blokke van gelyke grootte opgedeel word. Die toepassing ter sprake dikteer dikwels die optimale opdeling, terwyl die praktisyn ook `n afruiling tussen sydigheid en variansie in ag moet neem. Uit elke blok word die grootste waarde geneem en die veralgemeende ekstreemwaardeverdeling word op hierdie waarnemings gepas. Dit behels die verkryging van MA-beramings van die parameters van dié verdeling. Vervolgens kan die ekstreme kwantiele van belang in terme van bogenoemde parameters uitgedruk word ten einde MA-puntberamings vir hierdie kwantiele te verkry. Daar bestaan twee maniere waarop benaderde vertrouensintervalle vir dié kwantiele verkry kan word: die chi-kwadraatgebaseerde profielmetode, sowel as `n metode wat op die asimptotiese normaliteit van die MA-beramers berus. Die pasgehalte van die maksimummodel kan aan die hand van `n aantal diagnostiese stippings beoordeel word.

Drempelteorie - `n meer moderne benadering tot die ekstreemwaarde-probleem - is `n aanvulling tot die maksimummetode. Dit maak beter van die beskikbare steekproef gebruik deurdat dit alle waarnemings – eerder as slegs die blokmaksima – wat `n sekere drempel oorskry, as ekstreem beskou. Die eerste stap in `n drempelanalise is die keuse van hierdie drempel. Hier moet – soos by die maksima - praktiese oorwegings, sowel as `n variansie-sydigheid afruiling in gedagte gehou word. Op daardie waardes wat die drempel oorskry, word die veralgemeende Pareto-verdeling gepas. Soos in die geval van die maksimummetode, kan MA-beramings vir die parameters van hierdie verdeling, sowel as vir die ekstreme kwantiele onder beskouing, verkry word. Vervolgens stel ons tipies ook in benaderde vertrouensintervalle vir hierdie kwantiele belang. `n Finale stap behels die evaluering van die akkuraatheid van die gepaste drempelmodel met behulp van grafiese maatstawwe.

In die eerste hoofstuk van hierdie werkstuk is die aard van die ekstreemwaarde-probleem verduidelik. Historiese en hedendaagse voorbeelde van die toepassing van die ekstreemteorie is omskryf. In hoofstuk twee is die klassieke ekstreemwaarde-teorie, die maksimummetode, bespreek. Die hoofresultaat is die Fisher-Tippett stelling. Hierdie stelling het aanleiding gegee tot die herleiding van die algemene vorm van `n unieke klas van limietverdelings – indien sekere voorwaardes bevredig word sodat die limiet wel bestaan - vir die steekproef maksimum. Hierdie klas van verdelings staan as die veralgemeende ekstreemwaarde-verdeling bekend. Van besondere belang is die benadering $F^n \approx G_{\gamma, \mu, \sigma}$, waarvolgens ekstreme kwantiele van die onbekende onderliggende verdeling met behulp van die veralgemeende ekstreemwaarde-verdeling beraam kan word as $n \rightarrow \infty$.

Hoofstuk drie is `n teoretiese bespreking van die drempelbenadering tot ekstreme waardes. Hier was ons kernresultaat die Pickands-Balkema-De Haan stelling, waarvolgens die voorwaardelike verdeling van die drempeloorskrydings by hoë drempels deur die veralgemeende Pareto-verdeling benader kan word. In

die besonder bied hierdie stelling by groot drempels die benadering $F_u \approx H_{\gamma, \beta(u)}$. Hierdie benadering kan gebruik word vir drempelgebaseerde inferensie rakende die onderliggende verdeling.

In hoofstukke vier en vyf het ons onderskeidelik die geldigheid van die benaderings $F^n \approx G_{\gamma, \mu, \sigma}$ en $F_u \approx H_{\gamma, \beta(u)}$ by eindige steekproefgroottes en eindige drempelwaardes oorweeg. Dit is gedoen met behulp van 'n simulasië-analise deur die akkuraatheid van die beraming van stertkwantiele van die onderliggende verdeling, te evalueer. Vertrouensintervalle (VI's) vir hierdie kwantiele is met behulp van die profiel- sowel as die normaalmetode bereken. Klem is geplaas op die oordeckingswaarskynlikheid (OW), sowel as die gemiddelde lengte van hierdie vertrouensintervalle. Vir die verdelings onder beskouing wat tot die Frechet-domein behoort, het die profiel-VI's vir die 0.999-kwantiel by die maksimametode onaanvaarbare lae OW's gehad, terwyl dieselfde vir die normaalgebaseerde VI's by die drempelmetode gegeld het. Daarenteen was die OW's van die normaalgebaseerde VI's in die geval van die maksima stabiel tussen 0.90 en 0.95. By die drempelstudie was die profiel-OW's ook konsekwent in 'n nou band rondom 0.95. Hoewel ons die OW's van die normaal-VI's tot die vlakke van die profiel-VI's kon kalibreer in die drempelgeval, was die kalibrasie van die profiel-VI's by die maksimametode problematies en dit kon nie suksesvol deurgevoer word nie.

In toekomstige simulasiestudies sal dit sinvol wees om meer verdelings – veral nog swaarstert verdelings uit die Frechet-domein wat van soveel belang in finansiële toepassings is – te beskou ten einde 'n meer verteenwoordigende studie daar te stel. 'n Verdere belangrike kwessie is om vergelykbare steekproefgroottes by die maksima- en drempelanalyse te gebruik ten einde die gehalte van inferensie wat op elk van hierdie benaderings gebaseer is, te vergelyk. In hierdie werkstuk was die maksima-studie op aansienlik groter steekproewe as die drempelstudie gebaseer. Afgesien van die maksima- en drempelbenadering, kan meer rangorde statistieke as slegs die

steekproefmaksimum beskou word, byvoorbeeld die tweede grootste waarneming, en so voorts. Limietverdelings is vir hierdie rangorde statistieke herlei (Coles, 2001), sowel as vir die gesamentlike verdeling daarvan. Die opdeling van 'n beskikbare steekproef in die maksima-geval behoort ook in opvolg-ondersoeke meer aandag te geniet. In ons studie het ons byvoorbeeld 'n steekproef van 2 500 waarnemings op slegs een manier opgedeel, naamlik 50 blokke van 50 waardes elk. Alternatiewe opdelings soos (sê) 25 blokke van 100 waardes elk, behoort ook beskou te word. Ten opsigte van die keuse van 'n optimale drempel – oftewel 'n optimale aantal oorskrydings – bly hierdie kwessie problematies en verdere ondersoek behoort gedoen te word. Hier moet gekyk word na metodes waarvolgens die drempelkeuse op grond van die beskikbare data bepaal word ten einde variansie te minimeer terwyl sydigheid binne perke gehou word. In hierdie werkstuk is die drempel arbitrêr as verskillende rangorde statistieke van die data gekies. Dit moet ook beklemtoon word dat ons in hierdie werkstuk deurgaans aanvaar het dat die onderliggende waarnemings onafhanklik is. In praktiese toepassings met tydreeksdata is dit dikwels nie die geval nie en dit sal sinvol wees om die teoretiese sowel as empiriese aspekte van die modellering van sulke waardes, te ondersoek.



Ekstreemwaarde-teorie bestaan uit klasse van limietverdelings wat by eindige steekproewe en drempelwaardes slegs by benadering geld. Daarby word hierdie asimptotiese verdelings gebruik om te ekstrapoleer vanaf bekende waarnemings na vlakke wat nog nie vantevore aangeteken is nie. Voorstanders van die ekstreemteorie gee toe dat hierdie geen maklike taak is nie. Ekstreemwaarde-teorie kan nie noodwendig wonderwerke verrig nie, maar dit bly tans die beste benadering tot die modellering van 'n stogastiese proses by uiterste hoë vlakke.

Verwysings

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2004) *Statistics of Extremes*. Wiley.

Beirlant, J., Berning, T., Segers, J. (2003) *Large Quantile Estimation with the perturbed Pareto distribution*. Lesing aangebied by SASV konferensie, Johannesburg.

Billingsley, P. (1995) *Probability and Measure*. Wiley.

Chavez-Demoulin, A. en Roehrl, A. (2004) *Extreme Value Theory can save your neck*. Internet: www.approximity.com/papers/evt_wp.pdf.

Coles, S. (2001) *An Introduction to Statistical Modelling of Extreme Values*. Springer-Verlag.

De Haan, L. (1970) *On Regular variation and its Applications to the Weak Convergence of Sample Extremes*. Mathematical Centre Tract, 32, Amsterdam.

Embrechts, P. (2004) *Extremes in Economics and the Economics of Extremes*. Verskyn in *Extreme Values in Finance, Telecommunications and the Environment*. Redakteurs: Finkelstädt, B. en Rootzen, H. Chapman & Hall/CRC.

Embrechts, P. (2004) *From Dutch Dykes to Value at Risk: Extreme value theory and copulae as risk management tools*. Derde Nomura lesing.

Embrechts, P., Klüppelberg, C., en Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.

Fisher, R.A., and L.H.C. Tippett (1928) *Limiting forms of the frequency distribution of the largest and smallest member of a sample*, Proc. Cambridge Phil. Soc., **24**, 180-190.

Galambos, J. (1978) *The Asymptotic Theory of Extreme Order Statistics*. Wiley.

Gnedenko, B.V. (1943) *Sur la distribution limite du terme maximum d'une serie aleatoire*, Annals of Mathematics, **44**, 423-453.

Gumbel, E.J. (1958) *Statistics of Extremes*. Columbia University Press.

Hull, J.C. (2000) *Options, Futures and other Derivatives*. Prentice-Hall.

Klüppelberg, C. (2004) *Risk Management with Extreme Value Theory*. Verskyn in *Extreme Values in Finance, Telecommunications and the Environment*. Redakteurs: Finkelstädt, B. en Rootzen, H. Chapman & Hall/CRC.

Paneah, B. (2003) *Dynamic methods in the general theory of Cauchy type functional equations*. Israel Mathematical Conference Proceedings **161**.

Smith, R.L. (1985) *Maximum likelihood estimation in a class of non-regular cases*. Biometrika **72**, 67-90.

