MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig

Adviseur voor Statistische Consultatie: Prof. Dr J. Hemelrijk

S 227

Toetsing van de onafhankelijkheid van het hoogwater te Hoek van Holland en de waterafvoer van de Rijn bij Lobith.

door

Prof. Dr J. Hemelrijk

1957

Toetsing van de onafhankelijkheid van het hoogwater te Hoek van Holland en de waterafvoer van de Rijn bij Lobith.

door Prof. Dr J. Hemelrijk [*]

I Inleiding

Daar de waterstanden in plaatsen aan de benedenrivieren bepaald worden door zowel de waterstanden van de kust als de waterafvoeren van de bovenrivieren, is het van belang, na te gaan of er tussen deze beide factoren enig verband bestaat. In het bijzonder interesseert ons het voorkomen van combinaties van hoge hoogwaterstanden [1] aan de kust en grote rivierafvoeren. In dit rapport wordt de onafhankelijkheid nagegaan van de H.W.'s te Hoek van Holland en de afvoeren van de Rijn bij Lobith.

Het H.W. te Hoek van Holland noemen we $\underline{x}$ [2], de afvoer van de Rijn bij Lobith $\underline{y}$.

Deze grootheden bezitten kansverdelingen, die geschat kunnen worden op grond van waarnemingen en als $\underline{x}$ en $\underline{y}$ <u>onafhankelijk</u> zijn, d.w.z. als de kans op het optreden van een waarde van $\underline{x}$ niet wordt beïnvloed door de waarden die $\underline{y}$ aanneemt, kan de <u>simultane verdeling</u> van $\underline{x}$ en $\underline{y}$, d.i. de kansverdeling van paren $(\underline{x},\underline{y})$ berekend worden uit de verdelingen van $\underline{x}$ en $\underline{y}$.[3]

Op grond van de simultane verdeling kunnen dan uitspraken worden gedaan over de kansen op optreden van bepaalde combinaties van $\underline{x}$ en $\underline{y}$.

Getoetst wordt in dit rapport de hypothese, dat $\underline{x}$ en $\underline{y}$ onafhankelijk zijn zowel in ieder jaar afzonderlijk, als over een hele beschouwde periode van jaren.

---

*) De berekeningen, in dit rapport vermeld, zijn gebaseerd op gegevens van Rijkswaterstaat en uitgevoerd door F.W. STEUTEL, assistent van de Statistische Afdeling van het Mathematisch Centrum, die ook bij de opstelling van dit rapport zijn assistentie heeft verleend.

1) In het vervolg afgekort tot H.W.

2) Stochastische grootheden, d.w.z. grootheden, die een kansverdeling bezitten, worden onderscheiden van de waarden, die zij bij een experiment aannemen, door onderstreping van hun symbolen.

3) In geval van afhankelijkheid van $\underline{x}$ en $\underline{y}$ is de simultane verdeling niet zonder meer uit de verdelingen van $\underline{x}$ en $\underline{y}$ af te leiden.

We beschouwen hier alleen de "gevaarlijke" maanden, dat zijn die, waarin de hoge H.W.'s voornamelijk optreden, daar we ons speciaal voor de hoge waterstanden interesseren; t.w. de volgende perioden:

november t/m januari
november t/m februari } over de jaren 1900 t/m 1952.
oktober t/m maart

Beperking van het onderzoek tot deze gedeelten van het jaar geeft een beter inzicht in de situatie en betekent bovendien een aanzienlijke werkbesparing.

II Gebruikte methoden

1 De periode 1900 t/m 1952 als geheel

Laat het waarnemingsmateriaal over een periode bestaan uit N paren (x_i, y_i) $(i = 1,2,\ldots,N)$.

We kunnen deze waarnemingen als een "puntenwolk" uitzetten als in fig. 1.

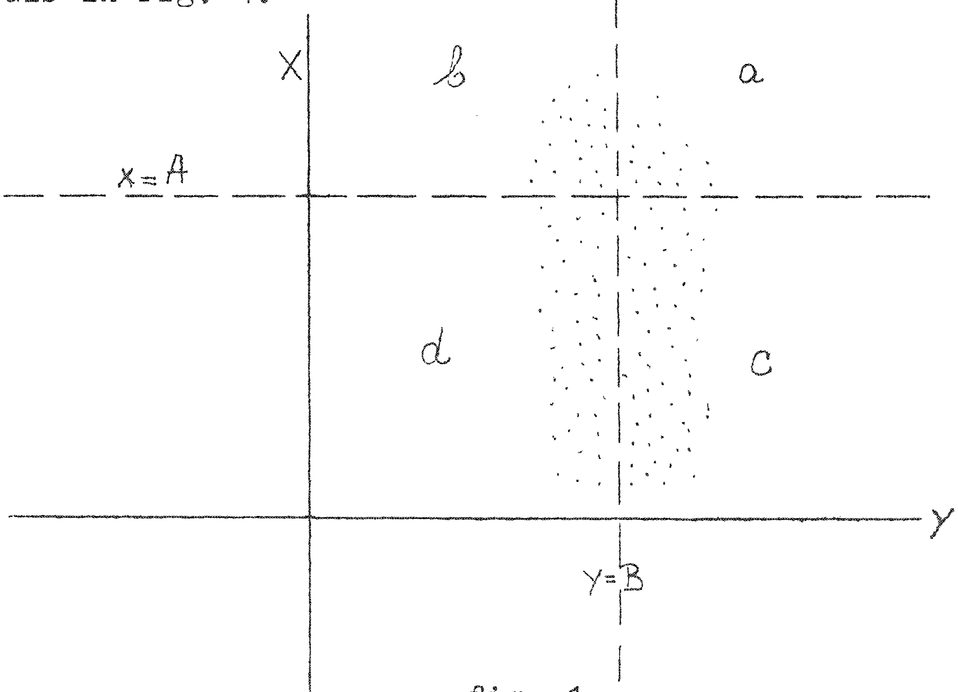

fig. 1

We verdelen nu deze puntenwolk door een horizontale lijn, x=A en een verticale lijn y = B in vier delen:
$(x>A, y>B)$, $(x>A, y<B)$, $(x<A, y>B)$ en $(x<A, y<B)$, bestaande uit resp. a, b, c en d punten. zoals in tabel I aangegeven.

	y<B	y>B	
x>A	b	a	m
x<A	d	c	N-m
	N-r	r	N

$$a+b+c+d = N$$
$$a+b = m$$
$$a+c = r$$

Tabel I

Zijn er m waarnemingen met x>A en r met y>B, dan is onder de hypothese H_o: "$\underline{x}$ en $\underline{y}$ zijn onafhankelijk", $\underline{a}$ een stochastische grootheid met een bekende (hypergeometrische) verdeling, n.l.

$$(1) \qquad P\{\underline{a} \leq a | H_o\} = \sum_{i=0}^{a} \frac{\binom{r}{i}\binom{N-r}{m-i}}{\binom{N}{m}} \,.$$

Keuze van A en B

In het onderhavige geval is de lijn x=A gekozen met voor A het grenspeil, d.i. het peil, dat gemiddeld één maal in twee jaar wordt overschreden, hetgeen bij ca. 730 waarnemingen per jaar impliceert, dat m<<N. Voor B is ongeveer het gemiddelde gekozen van de waterafvoeren in de beschouwde winterperioden, zodat $r \approx N-r \approx \frac{1}{2}N$ is en dus ook geldt: m<<r en m<<N-r.

Voor grote N, m<<N en m<<N-r kan men (1) benaderen door de binomiale verdeling:

$$(2) \qquad P\{\underline{a} \leq a | H_o\} = \sum_{i=0}^{a} \binom{m}{i}\left(\frac{r}{N}\right)^{i}\left(\frac{N-r}{N}\right)^{m-i}$$

met gemiddelde $\mathcal{E}(\underline{a}|H_o) = \frac{mr}{N}$ en variantie $\sigma^2(\underline{a}|H_o) = \frac{mr(N-r)}{N^2}$.

Op grond van deze verdeling kunnen we de rechteroverschrijdingskans, $k_r(a)$, behorende bij de gevonden waarde a van $\underline{a}$, berekenen. Deze is gedefinieerd als:

$$(3) \qquad K_r(a) = P\{\underline{a} \geq a | H_o\} = 1 - P\{\underline{a} < a | H_o\} \,,$$

dus hier:

$$(3') \qquad K_r(a) = 1 - \sum_{i=0}^{a-1} \binom{m}{i}\left(\frac{r}{N}\right)^{i}\left(\frac{N-r}{N}\right)^{m-i} \,,$$

welke kansen men kan vinden in een tabel van de binomiale verdeling.[4)]

Is de zo gevonden overschrijdingskans kleiner dan een van tevoren gekozen onbetrouwbaarheidsdrempel α (dikwijls wordt gebruikt $\alpha = 0{,}05$), dan verwerpt men de hypothese H_0 en concludeert: combinaties van grote $\underline{x}$ en grote $\underline{y}$ komen vaker voor dan met de onafhankelijkheid van deze grootheden te rijmen is, m.a.w. $\underline{x}$ en $\underline{y}$ zijn niet onafhankelijk.

2 Combinaties van de resultaten der jaren afzonderlijk

Op de boven beschreven wijze wordt gezocht naar afhankelijkheid over de gehele periode van 52 jaar, doch de eventuele verschillen, b.v. in aantal stormen, tussen de jaren worden daarbij niet in de beschouwingen betrokken. Wenst men te onderzoeken of binnen de jaren als afzonderlijke perioden beschouwd, onafhankelijkheid bestaat, dan dient men op andere wijze te werk te gaan. Men stelt dan tabel 1 op voor ieder der jaren afzonderlijk en uit de 52 op deze wijze verkregen tabelletjes - die ieder apart te weinig hoge H.W.'s bevatten om toetsing mogelijk te maken - wordt op hieronder te beschrijven wijze een gecombineerde toetsingsgrootheid berekend.

De waarnemingen leveren van ieder der k jaren afzonderlijk:
een waarde a_i van $\underline{a}_i$,
een waarde $\frac{m_i r_i}{N_i}$ van $\mathcal{E}(\underline{a}_i|H_0)$,
een waarde $\frac{m_i r_i (N_i - r_i)}{N_i^2}$ van $\sigma^2(\underline{a}_i|H_0)$.

Onder de hypothese $H_0^!$, dat $\underline{x}_i$ en $\underline{y}_i$ onafhankelijk zijn voor alle i, dus in ieder der jaren afzonderlijk, is de grootheid

$$\underline{z} = \frac{\sum_{i=1}^{k}\left(\underline{a}_i - \frac{m_i r_i}{N_i}\right)}{\sqrt{\sum_{i=1}^{k} \frac{m_i r_i (N - r_i)}{N_i^2}}}$$

bij benadering normaal verdeeld met gemiddelde 0 en spreiding 1.
Men berekent dus

$$z = \frac{\sum_{i=1}^{k}\left(a_i - \frac{m_i r_i}{N_i}\right)}{\sqrt{\sum_{i=1}^{k} \frac{m_i r_i (N_i - r_i)}{N_i^2}}}$$

, zoekt $k_r(z)$ op in een tabel van de

---

4) bijv. Tables of the binomial probability distribution, National bureau of Standards, Appl.Math. Series 6 (1950)

normale verdelingen verwerpt H'_o indien $k_r(z) < \alpha$ is.

Opmerking. Daar door de keuze van het deelpunt B $r \approx N-r \approx \frac{1}{2}N$ is, is de verdeling der a_i ongeveer symmetrisch, zodat ook voor vrij kleine $m = \sum m_i$ de normale benadering bruikbaar is.

III Resultaten van het onderzoek

We kunnen de resultaten van het onderzoek volgens II,1 en II,2 als volgt samenvatten:

1 Periode 1900 t/m 1952 als geheel beschouwd

	nov.t/m jan.	nov. t/m febr.	okt. t/m maart
N	9248	12085	18313
r	3348	4701	6567
m	15	16	20
r/n	0,3620	0,3890	0,3568
a	5	6	7
$k_r(a)$	0,68	0,64	0,62

2 Jaren afzonderlijk en dan gecombineerd

	nov. t/m jan.	nov. t/m febr.	okt. t/m maart
$\sum a_i$	5	6	7
$\sum \frac{m_i r_i}{N_i}$	6,2018	6,7614	7,9391
$\sum \frac{m_i r_i (N_i - r_i)}{N_i^2}$	2,4449	2,7153	3,5793
z	-0,77	-0,46	-0,50
$k_r(z)$	0,78	0,68	0,69

IV Interpretatie der resultaten

Uit de gevonden overschrijdingskansen blijkt, dat ér geen enkele aanleiding is om H_o of H'_o te verwerpen, zodat we mogen concluderen, dat noch over de perioden als geheel, noch in de jaren afzonderlijk enige afhankelijkheid van grote x-waarden en grote y-waarden blijkt.

Men kan dus voor eventuele verdere berekeningen de H.W.'s te Hoek van Holland en de rivierafvoeren bij Lobith als onafhankelijk beschouwen.