STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 119 (M 47)

Handleiding voor het gebruik van Gumbel-papier

Ph. van Elteren.

1953

MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 119 (M 47)

Handleiding voor het gebruik van Gumbel-papier

door

Ph. van Elteren

1953

Handleiding voor het gebruik van Gumbel-papier.

1. De grootste van N onderling onafhankelijke waarnemingen van een stochastische variabele $\underline{z}$[1]) met verdelingsfunctie $F(x)$ is zelf een stochastische variabele, die wij hier aanduiden met $\underline{x}$. Onder bepaalde voorwaarden voor $F(x)$ is de verdelingsfunctie $\mathcal{G}(x)$ van $\underline{x}$ voor grote N bij benadering van het dubbelexponentiële type, dit wil zeggen:

(1) $$\mathcal{G}(x) \approx e^{-e^{-\alpha(x-\beta)}} ,$$

waarin α en β bepaalde van $F(x)$ en N afhankelijke constanten zijn.

Gumbel heeft nu voor verdelingen van het dubbelexponentiële type grafiekenpapier geconstrueerd, overeenkomend met het bekende "waarschijnlijkheidspapier" voor de normale verdeling. Wij zullen dit grafiekenpapier hier Gumbel-papier noemen. Indien men beschikt over n onderling onafhankelijke waarnemingen van $\underline{x}$ (weer te verkrijgen uit n onderling onafhankelijke steekproeven van de uitgebreidheid N van $\underline{z}$) kan men hiermee nagaan of $\mathcal{G}(x)$ redelijkerwijze dubbelexponentieel is en in dat geval de parameters α en β (zie (1)) schatten.

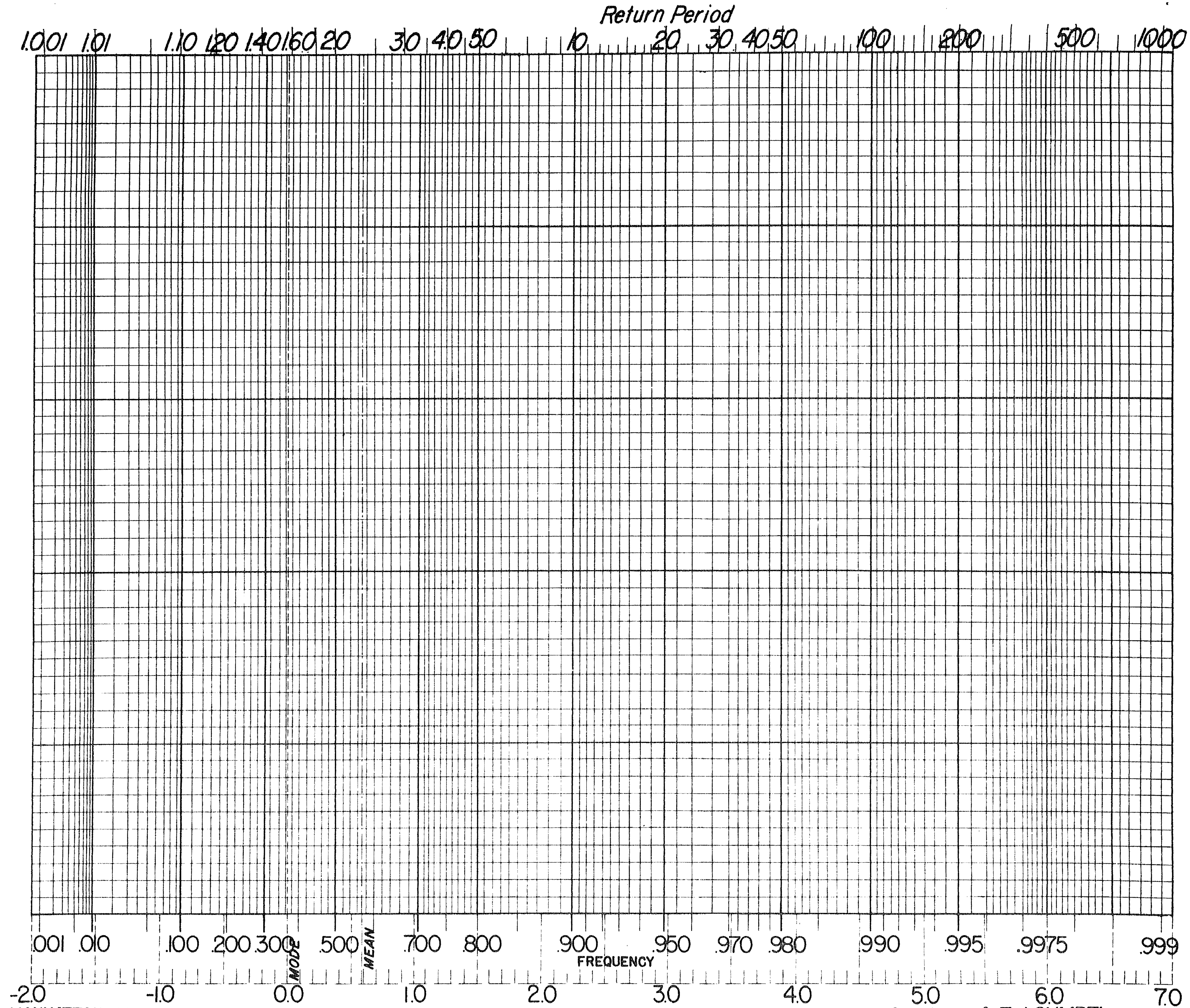
2. Het Gumbel-papier is als volgt ingericht. Geheel onderaan vindt men een lineaire schaalverdeling, waarbij "reduced variate" vermeld staat. Bij elke waarde y op deze schaal vindt men op de "frequency"-schaal verticaal hierboven de waarde van $H(y) = e^{-e^{-y}}$. De verticale lijnen op het papier corresponderen met deze "frequency"-schaal, terwijl de horizontale lijnen aequidistant zijn.

3. Indien men punten $(H(y), y)$ op Gumbel-papier uitzet ($H(y)$ horizontaal volgens de "frequency"-schaal, y verticaal met vrij te kiezen nulpunt en eenheid), zullen deze op een rechte lijn liggen. Als voldaan is aan de voor (1) vereiste voorwaarden, zullen punten $(\mathcal{G}(x), x)$ eveneens op een rechte lijn moeten liggen (Immers als $y = \alpha(x-\beta)$ geldt volgens (1): $H(y) \approx \mathcal{G}(x)$). In dat geval zal men, als $\mathcal{G}(x)$ bekend is, α en β gemakkelijk kunnen bepalen: β is de waarde van x als ordinaat behorend bij het nulpunt (mode) van de y-schaal (reduced-variate); zijn y_1 en y_2 de y-waarden corresponderend met twee gegeven ordinaten x_1 en x_2 van punten op de lijn, dan geldt:

(2) $$\alpha = \frac{y_2 - y_1}{x_2 - x_1}$$

---

1) Stochastische grootheden zijn in de tekst onderstreept


Return Period
1.001 1.01 1.10 1.20 1.40 1.60 2.0 3.0 4.0 5.0 10 20 30 40 50 100 200 500 1000
.001 .010 .100 .200 .300 .500 .700 .800 .900 .950 .970 .980 .990 .995 .9975 .999
MODE
MEAN
FREQUENCY
-2.0 -1.0 0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0
MANHATTAN COLLEGE SCHOOL of ENGINEERING

4. In de praktijk is $\mathcal{G}(x)$ echter niet bekend en dient men deze verdelingsfunctie te schatten uit een steekproef $x_1, \ldots, x_n$ van $\underline{x}$. Wij nemen hier aan dat de waarnemingen naar opklimmende grootte gerangschikt zijn. Wij zullen hier nu twee schattingsmethoden voor $\mathcal{G}(x_i)$ behandelen.

a) Men gebruikt als schatting voor $\mathcal{G}(x_i)$ de verwachting $\frac{i}{n+1}$

(Er geldt:

(3)
$$\mathcal{E}\,\mathcal{G}(\underline{x}_i) = \frac{i}{n+1}$$

zoals gemakkelijk bewezen kan worden.)

b) Naast deze eenvoudige methode beveelt Gumbel nog een andere methode aan, welke in de praktijk vaak wordt toegepast. Als β_i de modus van de verdeling van $\underline{x}_i$ is kan men $\mathcal{G}(\beta_n)$ gemakkelijk berekenen. Er blijkt te gelden:

(4)
$$\mathcal{G}(\beta_n) = e^{-\frac{1}{n}}$$

Eveneens kan men aantonen:

(5)
$$\frac{1}{\mathcal{G}(\beta_1)} + \frac{1}{\mathcal{G}(\beta_1)\ln \mathcal{G}(\beta_1)} - \frac{1}{\ln \mathcal{G}(\beta_1)} = n$$

$\mathcal{G}(\beta_n)$ kan zeer gemakkelijk met een logarithmentafel bepaald worden.

Met behulp van vergelijking (5) kan men tabel I (zie blz. 3) vervaardigen, waarin $\mathcal{G}(\beta_1)$ als functie van $\frac{1}{n}$ gegeven is.

Deze functie is vrijwel lineair in $\frac{1}{n}$, zodat men lineair naar $\frac{1}{n}$ kan interpoleren. Voor voldoend grote n geldt bij benadering:

(6)
$$\mathcal{G}(\beta_n) \approx 1 - \frac{1}{n}$$

(7)
$$\mathcal{G}(\beta_1) \approx \frac{1}{n}$$

De schatting voor $\mathcal{G}(x_i)$ wordt nu verkregen door lineaire interpolatie tussen $\mathcal{G}(\beta_1)$ en $\mathcal{G}(\beta_n)$ en is dus:

$$\frac{n-i}{n-1}\,\mathcal{G}(\beta_1) + \frac{i-1}{n-1}\,\mathcal{G}(\beta_n)$$

5. Nadat men volgens een van deze twee methoden schattingen van $\mathcal{G}(x_i)$ voor $i = 1, \ldots, n$ verkregen heeft, gaat men "op het oog" na of er redelijkerwijze een rechte lijn bij de punten $(\mathcal{G}(x_i), x_i)$ aan te passen is. Indien dat het geval is kan men met behulp van de aangepaste lijn de coëfficiënten α en β schatten volgens de methode beschreven in 3.

Tabel I

$\mathcal{G}(\beta_1)$ als functie van $\frac{1}{n}$ [1])

$\frac{1}{n}$	$\mathcal{G}(\beta_1)$	$\frac{1}{n}$	$\mathcal{G}(\beta_1)$	$\frac{1}{n}$	$\mathcal{G}(\beta_1)$
0,081580	0,055	0,0195[illegible]6	0,015	0,0087514	0,007
0,0[illegible]3219	0,050	0,018205	0,014	0,0074467	0,006
0,06[illegible]025	0,045	0,016[illegible]24	0,013	0,0061561	0,005
0,0[illegible]7000	0,040	0,01[illegible]4[illegible]2	0,012	0,0048804	0,004
0,0[illegible]147	0,035	0,0140[illegible]0	0,011	0,0036216	0,003
0,041472	0,030	0,01273[illegible]	0,010	0,0023826	0,002
0,033[illegible]82	0,025	0,0113[illegible]8	0,009	0,0011691	0,001
0,026605	0,020	0,01003[illegible]	0,008	0,0000000	0,000

1) $\mathcal{G}(\beta_1)$ is de waarde van de verdelingsfunctie van de kleinste van n waarnemingen uit een dubbelexponentiële verdeling in de modus β_1 van de kleinste van n waarnemingen uit deze verdeling. De tabel is ontleend aan E.J.Gumbel (1945) (zie literatuurlijst), p. [illegible]4.

6. Het Gumbel-papier bevat aan de bovenzijde ook nog een schaalverdeling, waarbij vermeld staat "Return-period". Bij een gegeven y (onderste schaal), kan men hierop aflezen $\frac{1}{1-H(y)}$, dus als $\mathcal{G}(x)$ bekend of geschat is en voldaan is aan de voor (1) vereiste voorwaarden, bij een gegeven x ook $\frac{1}{1-\mathcal{G}(x)}$. Deze grootheid is de verwachting van het aantal waarnemingen, dat men moet verrichten na een waarneming, waarbij x overschreden wordt tot en met de volgende overschrijding van x. In de statistische theorie van overstromingen correspondeert "return-period" van x met de verwachting van de (stochastische) periode waarmee de overschrijdingen van een waterpeil x optreden. Deze grootheid is overigens van weinig betekenis en het verdient geen aanbeveling er mee te werken, omdat zij, zoals in de praktijk gebleken is, tot begripsverwarring aanleiding kan geven.

Literatuur.

D. van Dantzig (1947), Kadercursus Mathematische Statistiek VI par. 2, Mathematisch Centrum, Amsterdam, Statistische Afdeling

E.J Gumbel (1941), The return period of flood flows, Ann. Math. Stat. 12 (1941), p. 163-190.

E.J.Gumbel (1945), Simplified plotting of statistical observations, Trans. Am. Geophys. Union 26 I (1945), p. 69-82.