STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

LEIDING: PROF. DR D. VAN DANTZIG
ADVISEUR VOOR STATISTISCHE CONSULTATIE: PROF. DR J. HEMELRIJK

Rapport S 247

De Gamma-verdeling; kort overzicht betreffende eigenschappen, schattings- en toetsings-methoden

door

Gerda Klerk-Grobben

april 1959

The Mathematical Centre at Amsterdam, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

1.1 De Γ-functie.

Voor gehele positieve waarden van n is $n!$ (n faculteit) gedefinieerd als:

$$n! = n(n-1)(n-2) \ldots 3.2.1.$$

Beschouw nu de integraal:

$$\int_0^\infty e^{-t} t^{z-1} dt ,$$

waarbij z een geheel en positief getal is. Door partiële integratie toe te passen vindt men:

$$\int_0^\infty e^{-t} t^{z-1} dt = -\int_0^\infty t^{z-1} de^{-t} = \left[-t^{z-1} e^{-t}\right]_{t=0}^{\infty} + \int_0^\infty e^{-t} dt^{z-1} =$$

$$= (z-1)\int_0^\infty e^{-t} t^{z-2} dt.$$

Herhaalt men dit, dan komt er dus:

$$\int_0^\infty e^{-t} t^{z-1} dt = (z-1)(z-2)\ldots 3.2.1 = (z-1)!$$

Men kan dus voor positieve, gehele waarden van z de faculteit van z voorstellen door een integraal. Stelt men in de laatste formule $z = 1$, dan vindt men

$$0! = \int_0^\infty e^{-t} dt = 1,$$

waaruit blijkt, dat de gebruikelijke afspraak $0!$ gelijk aan 1 te stellen, in dit kader past.

Voor het bestaan van de integraal is het niet nodig dat z geheel en positief is. Men kan bij voorbeeld voor z toelaten een willekeurig, reëel en positief getal, maar men kan dan de integraal niet zonder meer gelijk stellen aan $(z-1)!$, daar het faculteitsbegrip niet gedefinieerd is voor niet gehele waarden van z. Men noemt de integraal nu <u>de gamma-functie van het argument z</u> dus

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt ,$$

zodat voor gehele positieve z geldt:

$$\Gamma(z) = (z-1)!$$

Vervangen we in deze laatste uitdrukking z door $z+1$, dan staat er

$$\Gamma(z+1) = z!$$

en daar

$$z! = z(z-1)!$$

vindt men

$$\Gamma(z+1) = z\,\Gamma(z).$$

Ook deze betrekking blijft geldig als z geen geheel getal is.

Het is te bewijzen dat

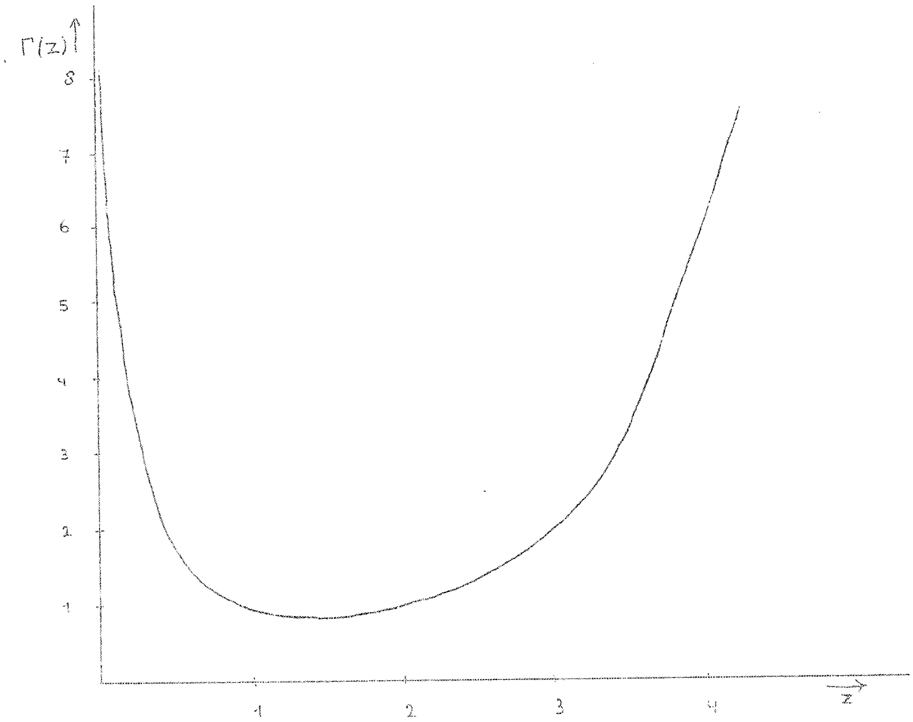
$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

terwijl reeds werd gezegd, dat

$$\Gamma(1) = 0! = 1.$$

De gamma-functie is grafisch aangegeven in figuur 1.

Voor verdere eigenschappen zie bijv.: E.T. WHITTAKER en G.N. WATSON (1946).



Figuur 1: de Γ-functie voor reële waarden van z.

1.2 De Γ-verdeling.

Uiteraard is nu (met vervanging van z door γ)

$$\frac{1}{\Gamma(\gamma)} \int_0^\infty e^{-t}\, t^{\gamma-1}\, dt = 1.$$

Dit wil zeggen dat men

$$\frac{1}{\Gamma(\gamma)} \quad e^{-t} \, t^{\gamma-1} \qquad\qquad (0 \leq t \leq \infty)$$

kan interpreteren als een kansdichtheid. De verdeling met deze dichtheid wordt wel de gamma-verdeling genoemd, maar de algemenere uitdrukking

$$g(x) = \frac{1}{\beta^{\gamma}\Gamma(\gamma)} \; e^{-\frac{t-\alpha}{\beta}} (t-\alpha)^{\gamma-1} \,, \qquad\qquad \alpha \leq t \leq \infty$$

heeft dezelfde naam, waarbij het beginpunt van de verdeling in plaats van bij o dus bij α ligt, en de eenheid door middel van de schaalparameter β is veranderd. Hierbij moet steed gelden $\beta > 0$ en $\gamma > 0$. Zie ook de figuren 2 tot en met 6.

1.3 Momenten.

Is $\alpha = 0$, dan geldt voor de momenten van de verdeling:

$$\mu_r = \mathcal{E}\, \underline{x}^r = \beta^r \frac{\Gamma(\gamma+r)}{\Gamma(\gamma)} \qquad\qquad (\alpha = 0)$$

Voor de verwachting en de variantie van $\underline{x}$ volgt hieruit:

$$\mathcal{E}\, \underline{x} = \beta\gamma \qquad\qquad \sigma^2\{\underline{x}\} = \gamma\beta^2 .$$

Is $\alpha \neq 0$, dan blijft de variantie onveranderd; het gemiddelde wordt

$$\mathcal{E}\, \underline{x} = \alpha + \beta\gamma .$$

1.4 Speciale Γ-verdelingen.

Tot de groep van de Γ-verdelingen behoren een paar zeer bekende verdelingen.

1.4.1 De exponentiële verdeling.

De exponentiële verdeling is een gamma-verdeling met $\gamma = 1$, nl.

$$g(x) = \frac{1}{\beta}\, e^{-\frac{x-a}{\beta}} \qquad\qquad (x \geq a)$$

De verwachting van $\underline{x}$ is $\alpha + \beta$; de variantie $\sigma^2\{\underline{x}\} = \beta^2$.

1.4.2 De χ^2-verdeling.

De χ^2-verdeling met ν vrijheidsgraden is een gamma-verdeling met

$$\alpha = 0$$
$$\beta = 2$$
$$\gamma = \frac{\nu}{2} .$$

De verdelingsdichtheid is dus:

$$g(x) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)}\, e^{-x/2}\, x^{\nu/2 - 1},$$

de momenten:

$$\mathcal{E}\,\underline{x} = \nu \qquad\qquad \sigma^2\{\underline{x}\} = 2\nu .$$

2. Het verband tussen de Γ-verdeling en de POISSON-verdeling.

2.1 Beschouw de grootheden $\underline{x}_1, \underline{x}_2, \ldots,$ die onderling onafhankelijk verdeeld zijn volgens een exponentiële verdeling met $\alpha = 0$, dus met verdelingsdichtheid

$$g(x) = \frac{1}{\beta}\, e^{-x/\beta} . \qquad\qquad (x \geq 0)$$

Verder is een positief getal T gegeven. Vormt men nu de opeenvolgende sommen van de kansgrootheden, dus

$$\begin{array}{l} \underline{x}_1 \\ \underline{x}_1 + \underline{x}_2 \\ \underline{x}_1 + \underline{x}_2 + \underline{x}_3 \\ \vdots \\ \underline{x}_1 + \qquad\qquad + \underline{x}_i \\ \underline{x}_1 + \qquad\qquad + \underline{x}_i + \underline{x}_{i+1} \\ \vdots \end{array}$$

en zoekt men die waarde van i waarvoor geldt:

$$\underline{x}_1 + \cdots + \underline{x}_i \leq T$$

en

$$\underline{x}_1 + \cdots + \underline{x}_i + \underline{x}_{i+1} > T ,$$

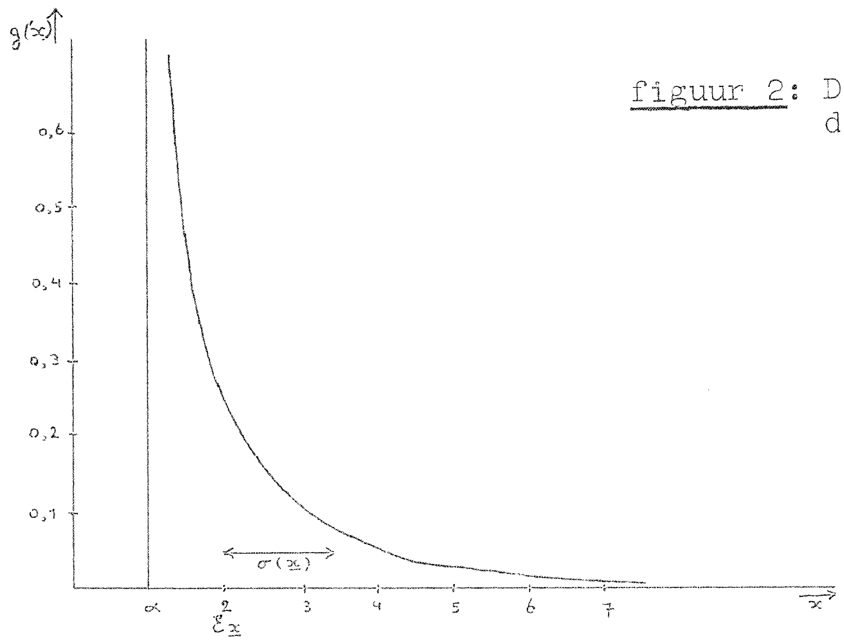
dan heeft $\underline{i}$ een POISSON-verdeling met gemiddelde T/β .

Een toepassing hiervan vindt men in de theorie van de wachttijdproblemen. Noemt men $\underline{x}_1$ de tijd tussen het ogenblik waarop een loket opengaat en het tijdstip, waarop de eerste klant binnen komt; $\underline{x}_2$ de tijd tussen de aankomsttijdstippen van de eerste en de tweede klant, enzovoorts. Dan is $\underline{i}$ het aantal klanten, die bij het loket aankomen vóór of op het tijdstip T . Zijn de $\underline{x}_i$ (i=1, ...) onderling onafhankelijke kansgrootheden alle met dezelfde exponentiële verdeling, dan heeft het aantal klanten, dat vóór of op tijdstip T aankomt een POISSON-verdeling.

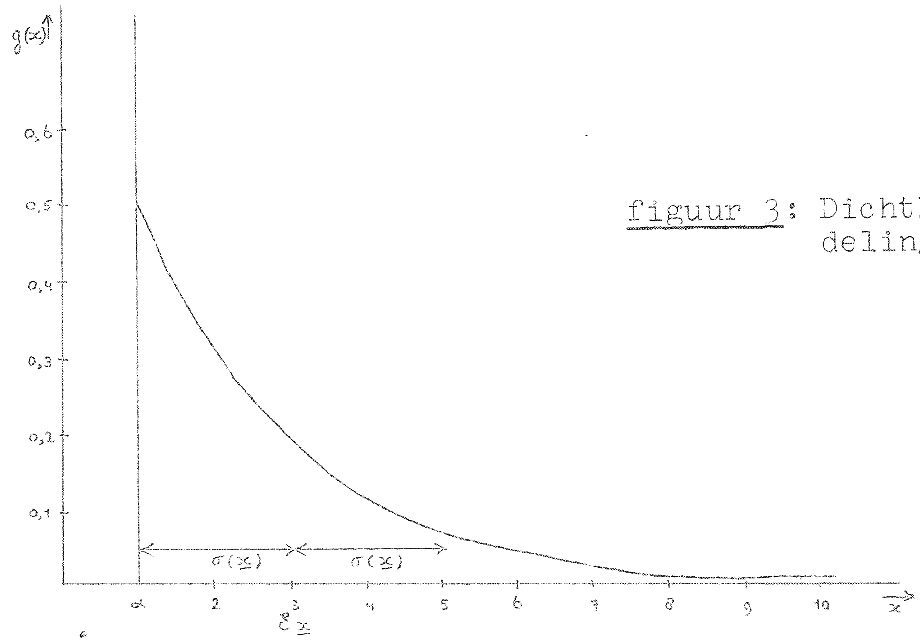
Omgekeerd geldt ook dat als de $\underline{i}$, gevonden op de genoemde wijze, voor elke waarde van T een POISSON-verdeling bezit met gemiddelde λ , en alle $\underline{x}_i$ zijn niet-negatieve, onderling onafhankelijke kansgrootheden met dezelfde verdeling, dat aan alle $\underline{x}_i$ een exponentiële verdeling bezitten met dichtheid

$$\frac{T}{\lambda} e^{-\frac{\lambda x}{T}} . \qquad (x \geq 0)$$

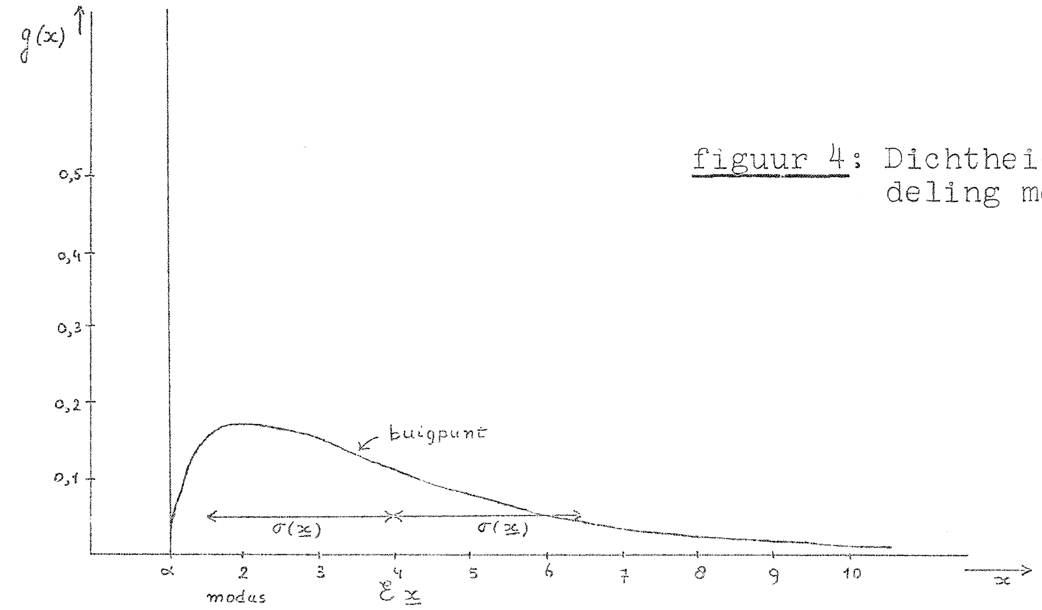
Zie L.A. GOODMAN (1951) en S. NABEYA (1950).



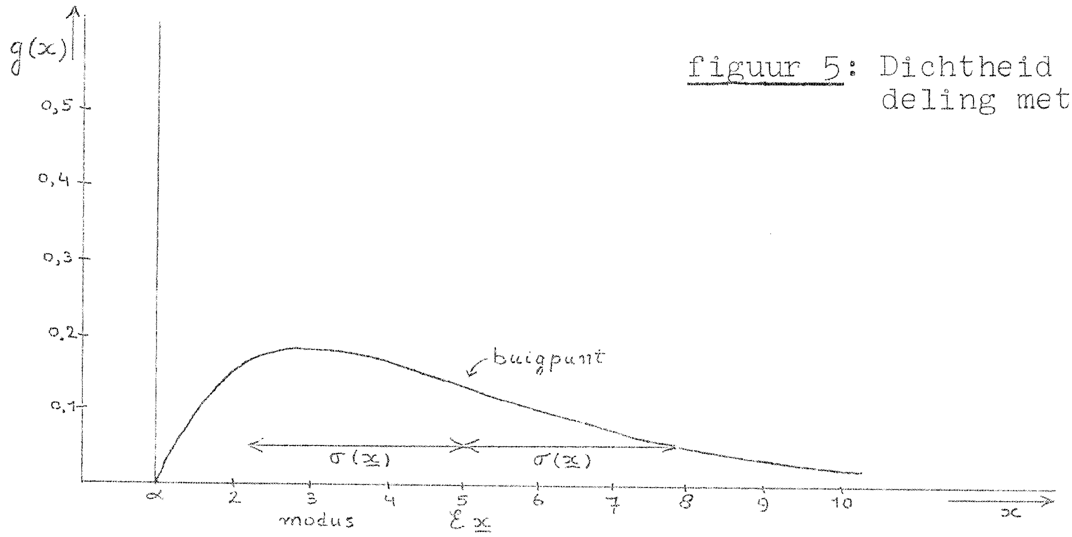
figuur 2: Dichtheid van de Γ-verdeling met $\alpha = 1$
$\beta = 2$
$\gamma = 0,5$



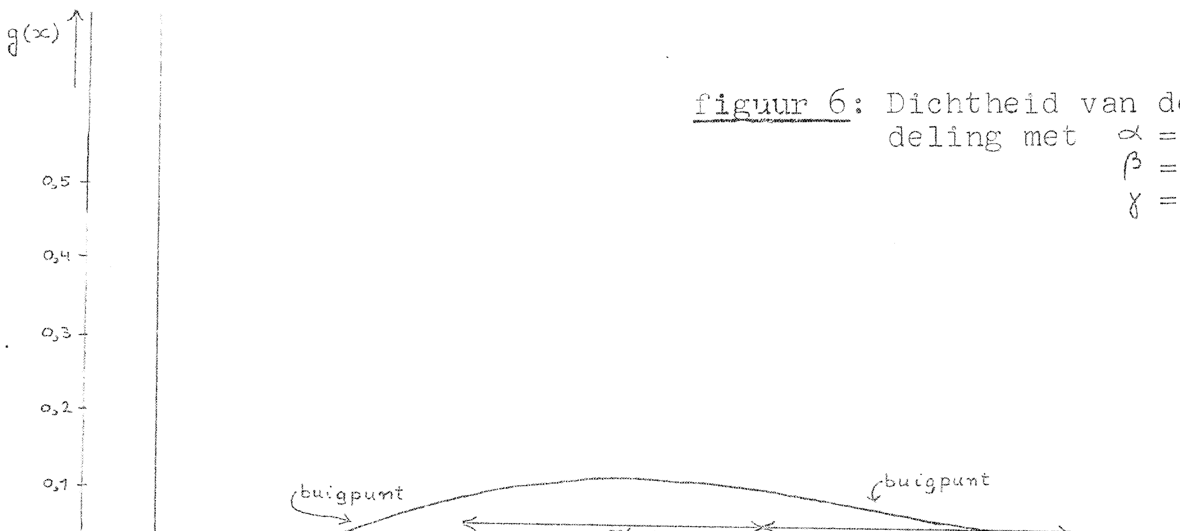
figuur 3: Dichtheid van de Γ-verdeling met $\alpha = 1$
$\beta = 2$
$\gamma = 1$



figuur 4: Dichtheid van de Γ-verdeling met $\alpha = 1$
$\beta = 2$
$\gamma = 1{,}5$



figuur 5: Dichtheid van de Γ-verdeling met $\alpha = 1$
$\beta = 2$
$\gamma = 2$



figuur 6: Dichtheid van de Γ-verdeling met $\alpha = 1$
$\beta = 2$
$\gamma = 4$

2.2 Tussen de POISSON-verdeling en de Γ -verdeling bestaat nog een ander verband, waardoor het mogelijk is overschrijdingskansen bij POISSON-verdeelde grootheden direct in een tabel van de Γ -verdeling op te zoeken en wel in die van een bijzondere Γ -verdeling, nl. de χ^2-tabel. Is nl. $\underline{n}$ een kansgrootheid met gemiddelde λ , dan is

$$P[\underline{n} \geq n] = \sum_{i=n}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} = P[\underline{x} \leq 2\lambda],$$

waarin $\underline{x}$ een kansgrootheid is met een χ^2-verdeling met $2n$ vrijheidsgraden. Zie M.G. KENDALL (1947), deel I, pag. 122 en paragraaf 4 van dit rapport.

Voorbeeld: $\lambda = 3$, $n = 4$

$$P[\underline{n} \geq 4] = \sum_{i=4}^{\infty} e^{-3} \frac{3^i}{i!} = P[\underline{x} \leq 6],$$

hetgeen opgezocht in de tabel van de χ^2 -verdeling bij 2 x 4 = 8 vrijheidsgraden de waarde 0,353 geeft.

3.1 Gedaante van de verdelingsdichtheid.

Het beginpunt van de verdelingsdichtheid is altijd bij $x = \alpha$. De schaal langs de x-as wordt bepaald door de keuze van β . De gedaante van de dichtheid is afhankelijk van γ , en hier moeten vijf gevallen onderscheiden worden:

$$\begin{cases} 0 < \gamma < 1 \\ \gamma = 1 \\ 1 < \gamma < 2 \\ \gamma = 2 \\ \gamma > 2 \end{cases}$$

Voor elk van deze gevallen, geldt dat de verwachting $\mathcal{E}\underline{x} = \alpha + \beta\gamma$ en de spreiding $\sigma\{\underline{x}\} = \beta\sqrt{\gamma}$.

De maximale waarde van de dichtheid $g(x)$ (de modus) ligt voor $\gamma \leq 1$ bij $x = \alpha$; voor $\gamma > 1$ is de modus bij $x = \alpha + (\gamma - 1)\beta$.

Bij $\gamma \leq 1$ bezit de verdeling geen buigpunten; bij $1 < \gamma \leq 2$ één buigpunt, nl. bij $x = \alpha + (\gamma-1)\beta + \beta\sqrt{\gamma - 1}$; bij $\gamma > 2$ twee buigpunten, en wel bij $x = \alpha + (\gamma-1)\beta \pm \beta\sqrt{\gamma-1}$, dus op gelijke afstanden van de modus. Zie de figuren 2 tot en met 6.

3.2 Limietverdeling.

De figuren 2-6 doen vermoeden, dat voor $\gamma \to \infty$ de Γ-verdeling tot de normale verdeling nadert. Inderdaad blijkt

$$\underline{u} = \frac{\underline{x} - \alpha - \beta\gamma}{\beta\sqrt{\gamma}}$$

voor $\gamma \to \infty$ een normale verdeling met gemiddelde 0 en spreiding 1 te bezitten. Voor grote waardenvan γ geldt dit dus bij benadering.

4. Tabellen.

Voor de Γ-verdeling bestaan er uitgebreide tabellen, nl. die van K. PEARSON (1946), van L.R. SALVOSA (1930) en die van de χ^2-verdeling, bijvoorbeeld E.S. PEARSON en H.O. HARTLEY, tabel 7 (1954).

4.1 De tabellen van K. PEARSON.

Is $\underline{x}$ een Γ-verdeelde grootheid met parameters α, β en γ, en wil men opzoeken de kans dat deze $\underline{x}$ kleiner (of groter) is dan een gegeven getal x, dan berekent men

$$u = \frac{x - \alpha}{\beta\sqrt{\gamma}}$$

en

$$p = \gamma - 1$$

en leest in de tabel bij de gevonden waarde van u en p een getal af, dat gelijk is aan $P[\underline{x} \leq x]$.

Voorbeeld

$\alpha = 6$; $\beta = 8$; $\gamma = 25$.

Men heeft een waarneming van $\underline{x}$ gedaan, deze is gelijk aan 274.

Wat is de rechter overschrijdingskans $P[\underline{x} \geq x]$?

Hier is dus

$$u = \frac{274 - 6}{8\sqrt{25}} = 6{,}7$$

$$p = 24 .$$

Opzoeken geeft $P[\underline{x} \leq x] = 0{,}9456$

en dus $P[\underline{x} \geq x] = 1 - 0{,}9456 = 0{,}0544$.

4.2 De tabellen van SALVOSA.

Wil men hier $P[\underline{x} \leq x]$ opzoeken, waarbij $\underline{x}$ een Γ-verdeling bezit met parameters α, β, γ, dan berekent men

$$t = \frac{x - \alpha - \beta\gamma}{\beta\sqrt{\gamma}}$$

en

$$\alpha_3 = \frac{2}{\sqrt{\gamma}} ,$$

en leest in de met "areas" aangeduide tabel af bij de gevonden waarde van t en α_3 . Het gevonden getal is $P[\underline{x} \leq x]$.

Men kan ook de verdelingsdichtheid g(x) opzoeken. Bij de waarde van t en α_3 in de tabel, die met "ordinates" is aangegeven, vindt men $\beta \sqrt{\gamma} \cdot g(x)$.

<u>Voorbeeld</u> $\alpha = 6$; $\beta = 8$; $\gamma = 25$.

Gegeven is een waarneming van $\underline{x}$, nl. 274. Wat is de rechteroverschrijdingskans $P[\underline{x} \geq 274]$ en wat is de hoogte van de verdelingsdichtheid bij x = 274 ?

$$t = \frac{274 - 6 - 200}{8\sqrt{25}} = 1,7 ,$$

$$\alpha_3 = \frac{2}{\sqrt{25}} = 0,4 .$$

Opzoeken in de "areas" -tabel geeft 0,9456, dus $P[\underline{x} \geq x] =$ 1 - 0,9456 = 0,0544. Opzoeken in de "ordinates"-tabel geeft 0,09088, zodat

$$g(x) = \frac{0,09088}{8\sqrt{25}} = 0,00227.$$

4.3 <u>De tabellen van de χ^2-verdeling.</u>

Is $\underline{x}$ een Γ -verdeelde grootheid met parameters α, β, γ, en men wil in de χ^2-tabel de kans opzoeken dat $\underline{x}$ groter is dan een gegeven getal x, dan berekent men

$$\chi^2 = 2\,\frac{x - \alpha}{\beta} \qquad \nu = 2\gamma$$

en zoekt in de tabel van de χ^2 -verdeling, de rechteroverschrijdingskans behorende bij de gevonden waarde χ^2 en wel voor 2γ vrijheidsgraden.

<u>Voorbeeld</u> $\alpha = 6$; $\beta = 8$; $\gamma = 25$.

Gegeven is een waarneming van $\underline{x}$ en wel x = 274. Wat is de rechteroverschrijdingskans $P[\underline{x} \geq x]$?

Hier is dus

$$\chi^2 = 2\,\frac{274 - 6}{8} = 67 \text{ met } 2 \text{ x } 25 = 50 \text{ vrijheidsgraden.}$$

In de tabel van PEARSON en HARTLY vindt men voor 50 vrijheidsgraden

$\chi^2 = 66$ overschrijdingskans 0,06418,
$\chi^2 = 68$ " 0,04596.

Lineaire interpolatie geeft $\frac{1}{2}$ (0,06418 + 0,04596) = 0,0550, hetgeen dus tengevolge van de interpolatie iets verschilt van de exacte waarde.

4.4 De normale benadering.

Zoals in 3.2 is aangegeven, kan men ook een normale benadering toepassen om het in 4.1, 4.2 en 4.3 gegeven voorbeeld de overschrijdingskans te berekenen. In de notatie van paragraaf 3.2 is dus voor dit voorbeeld

$$u = \frac{274 - 6 - 200}{8\sqrt{25}} = 1,7 .$$

De kans $P[\underline{x} \geq x]$ is nu bij benadering gelijk aan de kans, dat een normaal verdeelde grootheid met gemiddelde 0 en spreiding 1 groter is dan 1,7. Men vindt nu in de tabel van de normale verdeling 0,0446.

5. Eigenschappen van Γ-verdeelde grootheden.

5.1 Heeft $\underline{x}$ een Γ-verdeling met parameters α, β, γ en is a een constante dan heeft $\underline{x}$ + a een Γ-verdeling met parameters $\alpha + a, \beta, \gamma$. Is b een positieve constante, dan heeft b$\underline{x}$ een Γ-verdeling met parameters $b\alpha, b\beta$ en γ.

Een toepassing hiervan is dat direct volgt dat $\frac{2(\underline{x}-\alpha)}{\beta}$ een Γ-verdeling bezit met parameters 0, 2 en γ, dus zoals uit paragraaf 1.4.2 blijkt, heeft $\frac{2(\underline{x}-\alpha)}{\beta}$ een χ^2-verdeling met 2γ vrijheidsgraden. Deze eigenschap werd in paragraaf 4.3 gebruikt.

5.2 Zijn $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ onderling onafhankelijk verdeeld, en bezit $\underline{x}_i$ (i=1, ..., n) een Γ-verdeling met parameters $\alpha_i, \beta, \gamma_i$, dan is $\sum_{i=1}^{n} \underline{x}_i$ ook verdeeld volgens een Γ-verdeling en wel met parameters $\sum_{i=1}^{n} \alpha_i, \beta, \sum_{i=1}^{n} \gamma_i$.

Uit (5.1) volgt dan direct dat het gemiddelde $\frac{1}{n}\sum_{i=1}^{n} \underline{x}_i$ ook een Γ-verdeling bezit met parameters $\frac{1}{n}\sum_{i=1}^{n} \alpha_i, \frac{\beta}{n}, \sum_{i=1}^{n} \gamma_i$.

5.3 Zijn $\underline{x}_1$ en $\underline{x}_2$ onderling onafhankelijk verdeeld volgens een Γ-verdeling met parameters $0, \beta, \gamma_1$ resp. $0, \beta, \gamma_2$, dan zijn

$$\underline{x}_1 + \underline{x}_2 \quad \text{en} \quad \frac{\underline{x}_1}{\underline{x}_2}$$

onderling onafhankelijk verdeeld. Evenzo geldt dit voor

$$\underline{x}_1 + \underline{x}_2 \quad \text{en} \quad \frac{\underline{x}_1}{\underline{x}_1 + \underline{x}_2} .$$

Deze eigenschap kan tot meer variabelen en ingewikkelder functies van $\underline{x}_i$ worden uitgebreid: zie M.V. JAMBUNATHAN (1954), R.G. LAHA (1954) en (1956), E. LUKACS (1955) en (1956), E.J.G. PITMAN (1937) en K.C. SEAL (1957). Echter gelden deze eigenschappen niet als $\alpha \neq 0$.

5.4 Zijn $\underline{x}_1$ en $\underline{x}_2$ onderling onafhankelijke, Γ-verdeelde grootheden met parameters $0, \beta, \gamma_1$, resp. $0, \beta, \gamma_2$, dan heeft

$$\underline{y} = \frac{\underline{x}_1}{\underline{x}_1 + \underline{x}_2}$$

een B-verdeling, met verdelingsdichtheid:

$$\frac{\Gamma(\gamma_1 + \gamma_2)}{\Gamma(\gamma_1) \cdot \Gamma(\gamma_2)} \; y^{\gamma_1 - 1} (1-y)^{\gamma_2 - 1} \qquad 0 \leq y \leq 1.$$

De verdeling van $\underline{z} = \frac{\underline{x}_1}{\underline{x}_2}$ kan hieruit worden afgeleid, zie de in 5.3 aangehaalde artikelen en ook nog S. KULLBACK (1936).

5.5 <u>Exponentiële verdelingen.</u>

Zoals in 1.4.1 is aangegeven is een Γ-verdeling met parameters α, β en $\gamma = 1$ een exponentiële verdeling. Door E.J.G.PITMAN (1937) worden een aantal eigenschappen van exponentiëel verdeelde grootheden afgeleid, waarvan enkele vermeld zullen worden.

Zijn $\underline{x}_1, \ldots, \underline{x}_n$, n onderling onafhankelijke waarnemingen met eenzelfde exponentiële verdeling (parameters: α en β), dan heeft de kleinste waarde, die gevonden wordt bij deze n waarnemingen ook een exponentiële verdeling, maar met parameters α en $\frac{\beta}{n}$. Noemen we deze kleinste waarde $\underline{x}_0$, dan heeft verder

$$\sum_{i=1}^{n} (\underline{x}_i - \underline{x}_0)$$

een Γ-verdeling met parameters $0, \frac{\beta}{n}$ en $n-1$.

5.6 Verder is bewezen dat $\underline{x}_0$ en $\sum_{i=1}^{n} (\underline{x}_i - \underline{x}_0)$ onderling onafhankelijk zijn. Deze uitspraak kan nog gegeneraliseerd worden door in plaats van $\sum_{i=1}^{n} (\underline{x}_i - \underline{x}_0)$ een willekeurige functie van de $\underline{x}_i$ te nemen, die onafhankelijk is van de keuze van de oorsprong. Zie verder het betreffende artikel van PITMAN.

6. <u>Schattingen, toetsen en betrouwbaarheidsintervallen.</u>

Over schattingen en toetsen bij de algemene Γ-verdeling met onbekende parameters α, β en γ is in de litteratuur weinig bekend. Ligt de waarde van één of meer parameters vast dan is de situatie iets gunstiger, terwijl voor de exponentiële verdelingen ($\gamma = 1$) een hele reeks toetsen en betrouwbaarheidsintervallen beschikbaar zijn.

In dit hoofdstuk zullen de toetsingsgrootheden aangegeven worden met $\underline{T}_i$, waarbij i het nummer van de paragraaf, waarin de toets besproken wordt, is. Is $f(T_i)$ de verdelingsdichtheid van $\underline{T}_i$ onder de nulhypothese, H_0 , dan is $T_i(\varepsilon)$ die waarde waarvoor:

$$(6;1) \qquad \int_{-\infty}^{T_i(\varepsilon)} f(T_i)\, dT_i = \varepsilon ,$$

de overschrijdingskans, $P\left[\underline{T}_i \geq T_i(\varepsilon)\right]$, is dus $1-\varepsilon$. In elke paragraaf wordt uitgegaan van een reeks waarnemingen, $\underline{x}_1, \ldots, \underline{x}_n$, die onderling onafhankelijk volgens dezelfde Γ-verdeling met parameters α, β, γ verdeeld zijn; in de laatste twee paragrafen van twee dergelijke steekproeven uit verschillende Γ-verdelingen.

6.1 Zijn alle drie parameters onbekend dan kunnen schattingen voor de eerste drie momenten van de Γ-verdeling gebruikt worden om α, β, γ te schatten. In hoeverre echter deze schattingen "goed" zijn is niet eenvoudig te zeggen (zie ook par. 6.2). Schattingen voor de momenten zijn:

$$\underline{\bar{x}} = \sum_{i=1}^{n} \underline{x}_i / n \quad \text{voor} \quad \mathcal{E}\,\underline{x} = \alpha + \beta\gamma ,$$

$$\underline{s}^2 = \sum_{i=1}^{n} (\underline{x}_i - \underline{\bar{x}})^2 / n \quad \text{voor} \quad \sigma^2\{\underline{x}\} = \beta^2\gamma ,$$

en:

$$\underline{m}_3 = \sum_{i=1}^{n} (\underline{x}_i - \underline{\bar{x}})^3 / n \quad \text{voor} \quad \mathcal{E}(\underline{x} - \mathcal{E}\,\underline{x})^3 = 2\beta^3\gamma .$$

Hieruit is als schatting voor β, γ, α respectievelijk af te leiden:

$$\underline{b}_1 = \underline{m}_3 / (2\underline{s}^2) ,$$

$$\underline{c}_1 = 4(\underline{s}^2)^3 / \underline{m}_3^2 \qquad \text{en}$$

$$\underline{a}_1 = \underline{\bar{x}} - 2(\underline{s}^2)^2 / \underline{m}_3 .$$

J.A. GREENWOOD en D. DURAND (1958) geven een iteratie methode voor het berekenen van aannemelijkste schattingen, waarbij ze gebruik maken van de bepaling van aannemelijkste schattingen voor β en γ bij bekende α (zie volgende paragraaf).

6.2 Is de parameter $\underline{\alpha \text{ bekend}}$ dan kan men alle waarnemingen met α verminderen, zodat men dan waarnemingen $\underline{x}_1, \ldots, \underline{x}_n$ verkrijgt die een Γ-verdeling bezitten met parameters $\alpha = 0$, β en γ. Uit het gemiddelde en de spreiding van de steekproef volgen dan:

$$\underline{b}_2 = \underline{s}^2 / \underline{\bar{x}} \quad \text{en} \quad \underline{c}_2 = \underline{\bar{x}}^2 / \underline{s}^2$$

als schattingen voor β en γ.

In plaats van deze momentenmethode kan nu ook de methode der meest aannemelijke schattingen gebruikt worden. Hierbij blijkt dan dat voor de meest aannemelijke schattingen $\hat{\underline{b}}$ en $\hat{\underline{c}}$ geldt:

(6.2;1) $$\ln \hat{\underline{c}} - \left\{\frac{d}{d\gamma} \ln \Gamma(\gamma)\right\}_{\gamma=\hat{\underline{c}}} = \ln \frac{\sum_{i=1}^{n} x_i}{n} - \frac{\sum_{i=1}^{n} \ln x_i}{n}$$

(6.2;2) en $$\hat{\underline{b}} = \frac{\bar{x}}{\hat{\underline{c}}} .$$

Uit een tabel van de functie $\ln t - \frac{d}{dt} \ln \Gamma(t)$ kan, bij de uit de steekproef gevonden waarde van het rechterlid van (6.2;1), de bijbehorende $t = \hat{c}$ voor het linkerlid worden afgelezen. GREENWOOD en DURAND (1958) noemen het rechterlid van (6.2;1) y en de bijbehorende waarde voor t : $\hat{\underline{c}} = \phi(y)$ dus:

$$\underline{y} = \ln \frac{\sum_{i=1}^{n} x_i}{n} - \frac{\sum_{i=1}^{n} \ln x_i}{n}$$

en $$\hat{\underline{c}} = \phi(\underline{y}) .$$

(Teneinde een goede interpolatie mogelijk te maken, tabelleren zij $y\,\phi(y)$ in plaats van y).

Als benadering voor deze aannemelijkste schattingen kunnen:

$$\underline{c}^* = \frac{1}{2\left(\ln \bar{x} - \frac{\sum_{i=1}^{n} \ln x_i}{n}\right)}$$

en $$\underline{b}^* = \frac{\bar{x}}{\underline{c}^*}$$

gebruikt worden.

Benaderingsformules voor de varianties van de schattingen $\hat{\underline{b}}$ en $\hat{\underline{c}}$ van GREENWOOD en DURAND (1958) en de resultaten van een steekproefexperiment, uitgevoerd in de Statistische Afdeling van het Mathematisch Centrum, doen vermoeden dat al deze schattingen een grote spreiding om de te schatten parameter-waarden vertonen, vooral bij grote waarden van γ.

6.3 Bezitten alle $\underline{x}_i$ een Γ-verdeling, met parameters $\alpha = 0$, β en <u>bekende γ</u>, dan is

$$\hat{\underline{b}}_3 = \frac{\bar{x}}{\gamma}$$

een zuivere aannemelijkste <u>schatting</u> voor β en $\frac{1}{\underline{b}'} = \frac{n\gamma - 1}{n\bar{x}}$

een zuivere schatting voor $\frac{1}{\beta}$ (zie M.G. KENDALL (1947) II en C.R. RAO (1952)).

Uit eigenschap 5.2 volgt dat $\sum_{i=1}^{n} \underline{x}_i$ een Γ-verdeling bezit met parameters $\alpha = 0$, β, $n\gamma$ en dus $2\sum_{i=1}^{n} \underline{x}_i / \beta$ een χ^2-verdeling met $2n\gamma$ vrijheidsgraden. Hiervan kan gebruik gemaakt worden om de hypothese $H_0 : \beta = \beta_0$ te <u>toetsen</u>. Toetsingsgrootheid hiervoor is dan

$$\underline{T}_3 = \frac{2\sum_{i=1}^{n} x_i}{\beta_0} ,$$

die onder H_0 een χ^2-verdeling met $2n\gamma$ vrijheidsgraden bezit.

Bij een gevonden waarde T_3 zijn de overschrijdingskansen in een tabel van de χ^2-verdeling op te zoeken. Indien $2n\gamma$ geen geheel getal is zal men in het algemeen de overschrijdingskans door interpolatie uit de tabel moeten vinden. Men kan echter wel de exacte waarde vinden uit de tabellen van PEARSON en SALVOSA. Men gaat er dan van uit dat $\underline{T}_3$ onder H_0 een Γ-verdeling bezit met $\alpha = 0$, $\beta = 2$ en $n\gamma$ als parameters. Met het in paragraaf 4 gegeven recept is voor een gevonden waarde van T_3 dus de rechteroverschrijdingskans $K_r = P[\underline{T}_3 \geq T_3 \mid H_0]$ of de linkeroverschrijdingskans $K_l = P[\underline{T}_3 \leq T_3 \mid H_0]$ op te zoeken.
De tweezijdige overschrijdingskans kan als

$$k = 2 \min\ (K_r, K_l)$$

gedefinieerd worden(zie M.G. KENDALL (1947) II en ook G.D. BERNDT (1958) voor het onderscheidingsvermogen). Zijn $T_3(\varepsilon)$ en $T_3(1-\varepsilon)$ gedefinieerd als in (6;1) dan is het interval

$$\left(2\sum_{i=1}^{n} x_i / T_3(1-\varepsilon)\ ;\ 2\sum_{i=1}^{n} x_i / T_3(\varepsilon)\right)$$

een tweezijdig begrensd <u>betrouwbaarheidsinterval</u> voor β met onbetrouwbaarheid 2ε.

<u>Opmerking.</u>

In plaats van alle dezelfde γ mogen de grootheden $\underline{x}_i$ ook uit verdelingen met verschillende γ_i afkomstig zijn. In de formules moet dan overal $n\gamma$ door $\sum_{i=1}^{n}\gamma_i$ vervangen worden. β moet evenwel voor alle verdelingen dezelfde zijn.

6.4 Bezitten alle $\underline{x}_i$ een Γ-verdeling met onbekende α en met <u>bekende β en γ</u>, dan is

$$\underline{a}_4 = \bar{\underline{x}} - \beta\gamma$$

een zuivere <u>schatting</u> voor α. De grootheid $2(\sum_{i=1}^{n}\underline{x}_i - n\alpha)/\beta$ heeft een Γ-verdeling met parameters $\alpha = 0$, $\beta = 2$, $n\gamma$, dus een χ^2-verdeling met $2n\gamma$ vrijheidsgraden. Om de hypothese $H_0: \alpha = \alpha_0$ te <u>toetsen</u> kan dus

$$\underline{T}_4 = 2\left(\sum_{i=1}^{n}\underline{x}_i - n\alpha_0\right)/\beta$$

als toetsingsgrootheid gebruikt worden, welke onder H_0 de genoemde verdeling bezit. De overschrijdingskansen van een gevonden waarde kunnen weer als onder 6.3 bepaald worden.

Een <u>betrouwbaarheidsinterval</u> voor α met onbetrouwbaarheid 2ε is

$$\left(\bar{x} - \beta T_4(1-\varepsilon)/(2n)\ ;\ \bar{x} - \beta T_4(\varepsilon)/(2n)\right),$$

waarin $T_4(\varepsilon)$ en $T_4(1-\varepsilon)$ weer door (6;1) gedefinieerd zijn.

Opmerking.

Ook hier mogen de grootheden $\underline{x}_i$ uit verdelingen met verschillende γ_i afkomstig zijn, β en α moeten echter dezelfde zijn.

De volgende toetsen en betrouwbaarheidsintervallen hebben alle betrekking op exponentiële verdelingen, waarbij dus $\gamma = 1$ is. Er wordt verder dus alleen over de parameters α en β gesproken. Zoals in hoofdstuk 5 wordt ook nu $\underline{x}_0$ gedefinieerd als de kleinste waarde uit de steekproef $\underline{x}_1, \ldots, \underline{x}_n$.

6.5 De waarnemingen $\underline{x}_i$ zijn dus nu uit een exponentiële verdeling met parameters α en β afkomstig. Volgens eigenschap 5.5. bezit dan $\sum_{i=1}^{n}(\underline{x}_i - \underline{x}_0)$ een Γ-verdeling met parameters $\alpha = 0$ β en $\gamma = n-1$, en dus is $\mathcal{E}\sum_{i=1}^{n}(\underline{x}_i - \underline{x}_0) = (n-1)\beta$.

De schatting

$$\underline{b}_5 = \sum_{i=1}^{n}(\underline{x}_i - \underline{x}_0)/(n-1)$$

is dan ook zuiver en bezit bovendien een minimale variantie (zie B. EPSTEIN en M. SOBEL (1954)). $2(n-1)\underline{b}/\beta$ heeft een Γ-verdeling met parameters $\alpha = 0$, $\beta = 2$, $\gamma = n-1$, dus een χ^2-verdeling met 2n-2 vrijheidsgraden. Om de hypothese $H_0: \beta = \beta_0$ te toetsen kan daarom de toetsingsgrootheid

$$\underline{T}_5 = 2(n-1)\,\underline{b}/\beta_0$$

gebruikt worden (onder H_0: χ^2-verdeeld met 2n-2 vrijheidsgraden). Hieruit volgt tevens als tweezijdig betrouwbaarheidsinterval voor β met onbetrouwbaarheid 2ε :

$$\left(\frac{2(n-1)\underline{b}}{\chi^2_{2n-2}(1-\varepsilon)}\ ;\ \frac{2(n-1)\underline{b}}{\chi^2_{2n-2}(\varepsilon)}\right).$$

B. EPSTEIN en M. SOBEL (1954) hebben de hier besproken toets uitgebreid voor 't geval van meer steekproeven met verschillende (onbekende) waarden voor α.

6.6 Zijn $\underline{x}_i$ als in 6.5. dan is

$$\underline{a}_6 = \underline{x}_0 - \sum_{i=1}^{n}(\underline{x}_i - \underline{x}_0)/\{(n-1)n\}$$

een zuivere schatting voor α met minimale variantie (zie EPSTEIN en SOBEL (1954)). Een toets voor $H_0: \alpha = \alpha_0$ en een betrouwbaarheidsinterval voor α volgen nu uit de eigenschappen 5.5 en 5.6. De toetsingsgrootheid

$$\underline{T}_6 = n(\underline{x}_0 - \alpha_0)/\sum_{i=1}^{n}(\underline{x}_i - \underline{x}_0)$$

heeft nl. onder H_0 een verdeling met verdelingsdichtheid

$$f(T_6) = \frac{n-1}{(1+T_6)^n} \qquad (0 < T_6 < \infty)$$

(zie ook M.V. JAMBUNATHAN (1954) en S. KULLBACK (1936) en eigenschap 5.4). Hieruit volgt voor een gevonden waarde T_6 :

$$P[\underline{T}_6 \leq T_6] = \int_0^{T_6} \frac{n-1}{(1+u)^n} du = 1 - \frac{1}{(1+T_6)^{n-1}} = k_\ell .$$

$T_6(\varepsilon)$, bepaald door

$$\int_0^{T_6(\varepsilon)} \frac{n-1}{(1+u)^n} du = \varepsilon$$

is nu $T_6(\varepsilon) = \varepsilon^{-\frac{1}{n+1}} - 1$. Een betrouwbaarheidsinterval met onbetrouwbaarheid ε (niet 2ε) is :

$$\left(\underline{x}_o - T_6(1-\varepsilon) \sum_{i=1}^{n} (\underline{x}_i - \underline{x}_o)/n \; ; \underline{x}_o\right).$$

Hoewel dit interval er als een tweezijdig begrensd interval uitziet is toch de onbetrouwbaarheid ε in plaats van 2ε, want

$$P[\alpha < \underline{x}_o] = 1.$$

6.7 Bij levensduurproeven worden de waarnemingen uit een exponentiële verdeling in volgorde van opklimmende grootte waargenomen (zie B. EPSTEIN en M. SOBEL (1953)). Wordt nu het experiment gestaakt nadat van r van de n voorwerpen, die beproefd worden, de levensduur bekend is, en heeft de lengte van de levensduur een exponentiële verdeling met parameters $\alpha = 0$ en β, dan is

$$\underline{b}_{r,n} = \left[\sum_{i=1}^{r} \underline{x}_{(i)} + (n-r)\, \underline{x}_{(r)}\right]/r$$

de aannemelijkste <u>schatting</u> voor β . Voor de <u>toets</u> voor $H_o: \beta = \beta_o$ kan

$$\underline{T}_7 = 2r\, \underline{b}_{r,n}/\beta_o$$

als toetsingsgrootheid gebruikt worden, want $\underline{T}_7$ heeft onder H_o een χ^2-verdeling met $2r$ vrijheidsgraden, dus onafhankelijk van n .

Als <u>betrouwbaarheidsinterval</u> met onbetrouwbaarheid 2ε voor β volgt hieruit:

$$\left(\frac{2r\,\underline{b}_{r,n}}{\chi^2_{2r}(1-\varepsilon)} \; ; \frac{2r\,\underline{b}_{r,n}}{\chi^2_{2r}(\varepsilon)}\right).$$

Ook voor het vergelijken van twee of meer exponentiële verdelingen zijn toetsen beschikbaar; zie hiervoor bijvoorbeeld S. KULLBACK (1936), P.V. SUKHATME (1936), B. EPSTEIN en M. SOBEL (1954). In de volgende paragraaf vermelden wij twee van deze toetsen.

6.8 Stel dat n onderling onafhankelijke waarnemingen $\underline{x}_1, \ldots, \underline{x}_n$ verricht zijn van een kansgrootheid $\underline{x}$, die een exponentiële verdeling met parameters α (onbekend) en β (bekend) bezit; voorts dat n onderling onafhankelijke waarnemingen $\underline{y}_1, \ldots, \underline{y}_n$ verricht zijn van een kansgrootheid $\underline{x}$, die een exponentiële verdeling bezit met parameters α' (onbekend) en β (bekend). In beide gevallen heeft β dus dezelfde bekende waarde. Om de hypothese

$$H_0 : \quad \alpha = \alpha' + \delta_0 ,$$

waarbij δ_0 een gegeven getal is, te toetsen, gebruikt men als toetsingsgrootheid :

$$\frac{n}{\beta}(\underline{x}_0 - \underline{y}_0 - \delta_0),$$

die we met $\underline{T}_8$ zullen aangeven. De tweezijdige overschrijdingskans van een gevonden waarde T_8 is volgens KULLBACK (1936) :

$$P[\,|\underline{T}_8| \geq |T_8|\,] = e^{-|T_8|}.$$

Is $T_8(\varepsilon) = -\ln \varepsilon$, dan is verder

$$\left(\underline{x}_0 - \underline{y}_0 - \frac{\beta}{n} T_8(\varepsilon)\; ;\; \underline{x}_0 - \underline{y}_0 + \frac{\beta}{n} T_8(\varepsilon)\right)$$

een betrouwbaarheidsinterval voor $\delta = \alpha - \alpha'$ met onbetrouwbaarheid ε.

6.9 Stel dat n onderling onafhankelijke waarnemingen $\underline{x}_1, \ldots, \underline{x}_n$ zijn verricht van de kansgrootheid $\underline{x}$, die exponentieel verdeeld is met parameters α en β ; voorts dat m onderling onafhankelijke waarnemingen $y_1, \ldots, y_m$ zijn verricht van een grootheid $\underline{y}$ met een exponentiële verdeling met parameters α' en β' . De nulhypothese

$$H_0 : \quad \beta = k_0 \beta' ,$$

waarbij k_0 een gegeven getal is, kan getoetst worden met de toetsingsgrootheid:

$$\underline{T}_9 = \frac{\sum_{i=1}^{n}(\underline{x}_i - \underline{x}_0)}{\sum_{i=1}^{n}(\underline{x}_i - \underline{x}_0) + k_0 \sum_{i=1}^{m}(\underline{y}_i - \underline{y}_0)},$$

welke onder H_0 een B-verdeling bezit (zie par. 5.4) dus met verdelingsdichtheid:

$$B(T_9\,; n-1, m-1) = \frac{\Gamma(n+m-2)}{\Gamma(n-1)\Gamma(m-1)}\, T_9^{\,n-2}(1-T_9)^{m-2} , \qquad 0 \leq T_9 \leq 1.$$

Een tabel van de B -verdeling is gegeven door K. PEARSON (1934) en in E.S. PEARSON en H.O. HARTLEY (1954) (tabel 16).
Uit de kritieke waarden $T_9(\varepsilon)$ en $T_9(1-\varepsilon)$ volgt als tweezijdig begrensd betrouwbaarheidsinterval voor $k = \beta/\beta'$ met onbetrouwbaarheid 2ε :

$$\left(\frac{\sum_{i=1}^{n} (\underline{x}_i - \underline{x}_0)}{\sum_{j=1}^{m} (\underline{y}_j - \underline{y}_0)} \left(\frac{1}{T_{g(1-\varepsilon)}} - 1 \right) ; \frac{\sum_{i=1}^{n} (\underline{x}_i - \underline{x}_0)}{\sum_{j=1}^{m} (\underline{y}_j - \underline{y}_0)} \left(\frac{1}{T_{g(\varepsilon)}} - 1 \right) \right).$$

Opmerking.

In de laatste paragraaf mogen dus de waarden van α en α' verschillend en onbekend zijn; in par. 6.8. is vereist dat beide verdelingen dezelfde β bezitten, die bovendien bekend moet zijn, en dat de steekproeven evengroot zijn. Deze eisen kunnen iets verzwakt worden tot de eis dat $\frac{\beta}{n} = \frac{\beta'}{m}$ met bekende β en β'.

6.10 Het volgende schema geeft een overzicht van de situaties, wat betreft de parameters α, β, γ, waarin schattingen, toetsen of betrouwbaarheidsintervallen mogelijk zijn. De cijfers verwijzen naar de hierop betrekking hebbende paragrafen.

α	β	γ
schatting: 6.1.	schatting: 6.1.	schatting: 6.1.
bekend	schatting: 6.2.	schatting: 6.2
bekend	schatting voor β en $1/\beta$ toets voor $H_0: \beta = \beta_0$ betr.interval voor β } 6.3.	bekend (eventueel verschillende γ_i)
schatting toets voor $H_0: \alpha = \alpha_0$ betr.interval voor α } 6.4.	bekend	bekend (eventueel verschillende γ_i)
(eventueel meer steekproeven met verschillende onbekende α_i)	schatting voor β toets voor $H_0: \beta = \beta_0$ betr.interval voor β } 6.5.	$\gamma = 1$
schatting voor α toets voor $H_0: \alpha = \alpha_0$ betr.interval } 6.6.	onbekend	$\gamma = 1$
Bij levensduurproeven: uit kleinste r waarnemingen:		
bekend	schatting voor β toets voor $H_0: \beta = \beta_0$ betr.interval voor β } 6.7.	$\gamma = 1$
Bij vergelijking van twee of meer verdelingen:		
toets voor $H_0: \alpha = \alpha' + \delta_0$ betr.interval voor $\delta = \alpha - \alpha'$ } 6.8.	$\frac{\beta}{n} = \frac{\beta'}{n'}$ bekend	$\gamma = \gamma' = 1$
α en α' onbekend	toets voor $H_0: \beta = k_0 \beta'$ betr.interval voor $k = \beta/\beta'$ } 6.9.	$\gamma = \gamma' = 1$

Litteratuur.

In deze litteratuurlijst zijn ook artikelen opgenomen, welke in dit rapport niet ter sprake zijn gekomen.

H. Akaike (1956) On the distribution of the product of two Γ-distributed variables. Ann.Inst.Stat.Math., Tokyo, 8, p.53-54.

G.D. Berndt (1958) Powerfunctions of the Γ-distribution. Ann.Math.Stat., 29, pp.302-306.

D.G. Chapman (1956) Estimating the parameters of a truncated gamma distribution. Ann. 27 pp. 498-506.

H. Cramèr (1946) Mathematical Methods of Statistics, Princeton, University Press (2e druk).

D. van Dantzig (1947) Kadercursus Mathematische Statistiek, hoofdstuk 3 Verdelingen, Amsterdam, Math.Centrum.

B. Epstein & M. Sobel (1953) Life testing, J.A.S.A., 48, pp. 486-502.

B. Epstein & M. Sobel (1954) Some theorems relevant to life testing from an exponential distribution, Ann.Math.Stat., 25, pp. 373-381.

L.A. Goodman (1951-'52) On the Poisson-Gamma distribution problem. Ann.Inst.Stat.Math., Tokyo, 3, pp. 123-125.

J.A. Greenwood & D. Durand (1958) Aids for fitting the Pearson type III-curve; nog niet gepubliceerd: abstract: Ann.Math.Stat., 29, p. 1280.

M.V. Jambunathan(1954) Some properties of β- and γ-distributions, Ann.Math.Stat., 25, pp. 401-405.

M.G. Kendall (1947) The advanced theory of statistics I and II, Ch. Griffin & Co., Ltd., London.

A.S. Krishnamoorthy & M.Parthasarathy (1951) A multivariate Gamma-type distribution. Ann.Math.Stat., 22, pp. 549-557.

S. Kullback (1936) The distribution laws of the difference and quotient of variables independently distributed in Pearson type III laws. Ann Math.Stat., 7, pp. 51-53.

R.G. Laha (1954) On a characterisation of the gamma distribution, Ann.Math.Stat., 25, pp. 784-787.

R.G. Laha (1956) On some properties of the normal and gamma distribution. Proc.Am.Math.Soc., 7, pp. 172-174.

E. Lukacs (1955) A characterisation of the gamma distribution. Ann.Math.Stat., 26, pp. 319-324.

E. Lukacs (1956) Characterisation of populations by properties of suitable statistics. Third Berkeley Symp. on Math.Stat. & Probab.II, pp.195-215.

Seiji Nabeya (1950) On a relation between exponential law and Poisson's law. Ann.Inst.Stat.Math., Tokyo, 2, pp.13-16.

A.C. Olshen (1938) Transformations of the Pearson type III distribution, Ann.Math.Stat., 9, pp. 176-200.

E.S. Pearson & H.O.Hartley (1954) Biometrika tables for Statisticians vol I Cambridge University Press.

K. Pearson (1934) Tables of the incomplete B-function, Biometrika Office, University College, London.

K. Pearson (1946) Tables of the incomplete Γ-function, Cambridge, University Press.

E.J.G. Pitman (1937) The "closest" estimates of statistical parameters, Proc. Cambridge Phil.Soc., 33, pp. 212-222.

C.R. Rao (1952) Advanced Statistical Methods in Biometric Research, Chapman & Hall, London.

L.R. Salvosa (1930) Tables of Pearson's type III function, Ann.Math.Stat., 1, pp. 191-198.

K.C. Seal (1957) On a characterisation of gamma distributions. Calcutta Stat.Assoc.Bulletin, 7, pp. 60-72.

P.V. Sukhatme (1936) Analysis of k samples from exponential populations. Stat.Res.Mem., 1, pp. 44-112.

E.T. Whittaker & G.N. Watson (1946) A course of modern analysis, Cambridge, University Press.