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 165 (M 58)

Uittreksel van het artikel van A.C. Berry:

The accuracy of the Gaussian approximation to the sum of independent variables.
(Trans. Am. Math. Soc. 49 (1941) p. 122-136)

door

R. Doornbos en D.J. Stoker

1954

Uittreksel van het artikel van A.C. Berry

<u>The accuracy of the Gaussian approximation to the sum of independent variables.</u>

(Trans.Am.Math.Soc. <u>49</u> (1941)p. 122-136)

$\underline{X}_1, \ldots, \underline{X}_n$ zijn onderling onafhankelijke stochastische grootheden met gemiddelden $\alpha_1, \ldots, \alpha_n$. De absolute gereduceerde momenten van de 2e en 3e orde bestaan en worden aangegeven met $\mu_2(\underline{X}_k)$ en $\mu_3(\underline{X}_k)$ $\quad (k = 1, \ldots, n)$.

Stel

$$\lambda(\underline{X}_k) = \begin{cases} \mu_3(\underline{X}_k)/\mu_2(\underline{X}_k) & \text{als } \mu_2(\underline{X}_k) \neq 0 \\ 0 & \text{als } \mu_2(\underline{X}_k) = 0, \end{cases}$$

$$\alpha = \alpha_1 + \ldots\ldots + \alpha_n ,$$

$$\sigma = \left(\mu_2(\underline{X}_1) + \ldots\ldots + \mu_2(\underline{X}_n)\right)^{1/2},$$

$$\mathcal{G}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt ,$$

$$\Lambda = \max\left(\lambda(\underline{X}_1), \ldots\ldots, \lambda(\underline{X}_n)\right),$$

$$F(x) = P\left[\underline{X}_1 + \ldots\ldots + \underline{X}_n \leq x\right],$$

$$M = \sup_{-\infty < x < \infty} \left| F(x) - \mathcal{G}\left(\frac{x-\alpha}{\sigma}\right) \right| ,$$

$$\varepsilon = \Lambda / \sigma .$$

De eerste stelling van Berry luidt dan:

<u>Stelling 1</u>: $\boxed{M \leq (1{,}95)\,\varepsilon}$ 1)

Deze stelling wordt bewezen voor het geval (R), d.w.z.

(1) $\alpha_k = 0 \qquad (k = 1, \ldots\ldots, n)$,

(2) $\sigma = 1$,

(3) $\varepsilon < (1{,}95)^{-1}$.

We zien dat als aan (3) niet voldaan is stelling 1 triviaal is, want dan houdt de stelling in $M \leq 1$. Als de variabelen van stelling 1 niet voldoen aan (1) en (2) voeren wij in

$$\underline{X}'_k = \sigma^{-1}(\underline{X}_k - \alpha_k) \qquad (k = 1, \ldots\ldots, n).$$

Hiervoor geldt $M' = M$ en $\varepsilon' = \varepsilon$, dus stelling 1 is een gevolg van stelling 1(R).

----

1) Berry geeft hier de waarde 1,88. Uit stelling I bij de dissertatie van D.J. Stoker (1955) blijkt, dat deze waarde door de hier gegevene vervangen moet worden.

Het bewijs van stelling 1(R) verloopt als volgt: Eerst wordt bewezen:

Lemma 1:

Er bestaat een getal a zodat een van de ongelijkheden

(4) $\quad F(x+a) - G(x+a) \geqq \frac{\delta - x}{\sqrt{2\pi}}$,

(4') $\quad F(x+a) - G(x+a) \leqq \frac{-\delta - x}{\sqrt{2\pi}}$

geldt in het interval $-\delta \leqq x \leqq \delta$, waarin
$\delta = M\,(\pi/2)^{1/2}$.

Dit lemma wordt bewezen door de existentie aan te tonen van een eindig getal b, waarvoor geldt:

$F(b+) - G(b) = M$ of

$F(b-) - G(b) = -M$.

Als we nu $a = b + \delta$ nemen, dan blijkt a aan lemma 1 te voldoen.

Het bewijs van stelling 1(R) maakt gebruik van de karakteristieke functies

$$\varphi(t) = \int_{-\infty}^{\infty} e^{ixt}\, d\,F(x) \quad \text{en}$$

$$\psi(t) = \int_{-\infty}^{\infty} e^{ixt}\, d\,G(x) = e^{-t^2/2} .$$

Wij beschouwen nu het verschil:

$$\varphi(t) - \psi(t) = \int_{-\infty}^{\infty} e^{ixt}\, d\{F(x) - G(x)\} .$$

Het is bekend dat als dit verschil klein is in een eindig interval om $t = 0$, dat dan M klein is. Partiele integratie geeft

$$\frac{\varphi(t) - \psi(t)}{-it} = \int_{-\infty}^{\infty} \{F(x) - G(x)\}\, e^{ixt}\, dx .$$

Wij vervangen nu x door $x+a$ (de a van lemma 1)

$$\frac{\varphi(t) - \psi(t)}{-it}\, e^{-iat} = \int_{-\infty}^{\infty} \{F(x+a) - G(x+a)\}\, e^{ixt}\, dx .$$

Wij voeren in de gewichtsfunctie:

$$w(t) = \begin{cases} T - |t| & \text{voor } |t| \leqq T \\ 0 & \text{voor } |t| > T . \end{cases}$$

De Fouriergetransformeerde van $w(t)$ is, op een constante na:

$$W(x) = \int_{-T}^{T} w(t)\, e^{ixt} dt = \frac{2(1 - \cos Tx)}{x^2} .$$

Volgens de stelling van Parseval geldt nu:

$$\int_{-\infty}^{\infty} W(x) \{F(x+a) - G(x+a)\}\, dx = \int_{-T}^{T} w(t)\, \frac{\varphi(t) - \psi(t)}{-it}\, e^{-iat} dt .$$

Hieruit volgt:

$$\left|\int_{-\infty}^{\infty}\frac{1-\cos Tx}{x^2}\{F(x+a)-G(x+a)\}\,dx\right|\leqq\int_0^T (T-t)\frac{|\varphi(t)-\psi(t)|}{t}\,dt.$$

Hieruit wordt afgeleid:

(5) $$A(T\delta)\leqq\int_0^T (T-t)\frac{\varphi(t)-\psi(t)}{t}\,dt\quad ,$$

waarin

$$A(u)=(2/\pi)^{1/2}\,u\left\{3\int_0^u\frac{1-\cos x}{x^2}\,dx-\pi\right\}.$$

Als men nu voor T neemt $1{,}1/\varepsilon$, dan wordt bewezen dat het rechterlid van (5) niet groter is dan

(6) $$(1{,}1/6).(\pi/2)^{1/2}.$$

Dit bewijs verloopt als volgt: De stochastische grootheid X_k heeft de karakteristieke functie

$$\varphi_k(t)=\int_{-\infty}^{\infty}e^{ixt}\,d\,F_k(x)=1-\sigma_k^2t^2/2+\ldots\ldots.$$

Achtereenvolgens worden nu afgeleid

<u>Lemma 2</u> In het interval $0\leqq t<\sqrt{2}/\varepsilon$ geldt

$$\log\varphi_k(t)=-\sigma_k^2t^2/2+\sigma_k^2\Delta_k\quad ,$$

waarin voor iedere k $(k=1,\ldots\ldots,n)$, $|\Delta_k|\leqq t^2h(\varepsilon t)$

$$h(x)=x/6-\{\log(1-x^2/2)+x^2/2\}/x^2.$$

<u>Lemma 3</u> In het interval $0\leqq t<\sqrt{2}/\varepsilon$ geldt

$$\log\varphi(t)=-t^2/2+\Delta\quad ,$$

waarin

$$|\Delta|\leqq t^2\,h(\varepsilon t).$$

<u>Lemma 4</u> In het interval $0\leqq t<\sqrt{2}/\varepsilon$ geldt

$$|\varphi(t)-\psi(t)|\leqq\left\{e^{t^2h(\varepsilon t)}-1\right\}e^{-t^2/2}.$$

Uit lemma 4 volgt dat het rechterlid van (5) voor $T=1{,}1/\varepsilon$ gemajoreerd wordt door de integraal

$$B=\int_0^{1,1/\varepsilon}(1{,}1/\varepsilon-t)\,\frac{e^{t^2h(\varepsilon t)}-1}{t}\,e^{-t^2/2}\,dt$$

B wordt in 3 delen gesplitst:

$$B_1 = \varepsilon/6 \int_0^{0,75/\varepsilon} (1,1/\varepsilon - t)\, t^2 e^{-t^2/2}\, dt\,,$$
$$B_2 = \int_0^{0,75/\varepsilon} (1,1/\varepsilon - t)\, \frac{e^{t^2 h(\varepsilon t)} - 1 - \varepsilon t^3/6}{t}\, e^{-t^2/2}\, dt\,,$$
$$B_3 = \int_{0,75/\varepsilon}^{1,1/\varepsilon} (1,1/\varepsilon - t)\, \frac{e^{t^2 h(\varepsilon t)} - 1}{t}\, e^{-t^2/2}\, dt\,.$$

Voor deze drie delen worden de volgende schattingen afgeleid:

$$B_1 \leqq (1,1/6)(\pi/2)^{1/2} - \varepsilon/3 \quad, \text{ als } \varepsilon \leqq 1/1,952 \text{ 2)},$$
$$B_2 \leqq (0,134)\,\varepsilon\,,$$
$$B_3 \leqq (0,195)\,\varepsilon\,,$$

zodat $B < (1,1/6)(\pi/2)^{1/2}$, waarmee de ongelijkheid (6) dus bewezen is.
Dus (volgens (5))

$$A(1,1\,\delta/\varepsilon) < 0,229\,\delta\,.$$

Met behulp van de ongelijkheid

$$\cos x \leqq 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!}$$

vindt men

$$A(2,59) > 0,229\,\delta\,.$$

Nu is $A(u)$ een niet afnemende functie, daaruit volgt dat

$$M = (2/\pi)^{1/2}\,\delta \leqq (2/\pi)^{1/2}\, 2,59\,\varepsilon/1,1 < (1,88)\,\varepsilon\,,$$

onder de voorwaarde

$$\varepsilon \leqq (1,952)^{-1}\,.$$

Voor $\varepsilon > (1,88)^{-1}$ is de stelling triviaal, want $M \leqq 1$.

Verhoogt men de constante 1,88 tot 1,952, dan is steeds aan de voorwaarde $\varepsilon \leqq (1,952)^{-1}$ voldaan, voor zover de stelling niet triviaal is. Men kan dus onvoorwaardelijk stellen:

$$M \leqq 1,952\,\varepsilon\,.$$

(zie voetnoten 1) en 2))

Hiermee is stelling 1(R) en dus stelling 1 bewezen.

Stel nu $C = \sup M/\varepsilon$.
Volgens de tweede stelling geldt dan

Stelling 2 $\boxed{C \geqq (2\pi)^{-1/2} \approx 0,3989 \approx 0,40}$.

Om dit te bewijzen beschouwen wij de grootheden $\underline{X}_1, \ldots, \underline{X}_n$ voor een speciaal geval, nl. dat ze allen dezelfde verdeling bezitten en wel dat ze alleen de waarden -1 en +1 aan kunnen nemen.

---

2) De noodzakelijkheid van deze extra voorwaarde, afkomstig van D.J. Stoker, wordt aan het einde afzonderlijk behandeld.

met waarschijnlijkheid $\frac{1}{2}$ voor iedere waarde. Als wij aannemen dat $C < (2\pi)^{-1/2}$, komen wij tot een tegenspraak.

De drie volgende stellingen geven schattingen voor M als de voorwaarden van resp. Feller, Lindeberg en Liapounoff vervuld zijn.

a) Het geval van Feller.

$X_1, \ldots\ldots, X_n$ zijn onderling onafhankelijke stochastische grootheden met verdelingsfuncties resp. $F_1(x), \ldots\ldots, F_n(x)$. $s > 0$, $a_1, \ldots, a_n$ zijn reële getallen. $F(x)$ is de verdelingsfunctie van $X_1, \ldots\ldots, X_n$.

Stel $$M = \sup_{-\infty < x < \infty} \left| F(x) - G\left(\frac{x-a}{s}\right) \right| .$$

Voor een gegeven $\varepsilon > 0$ voeren wij in:

$$\varepsilon_0 = \sum_{k=1}^{n} P\left[|X_k - a_k| > \varepsilon s\right] ,$$

$$\varepsilon_1 = s^{-1} \sum_{k=1}^{n} \left| \int_{a_k - \varepsilon s}^{a_k + \varepsilon s} (x_k - a_k)\, d F_k(x) \right| ,$$

$$\varepsilon_2 = \left| 1 - \frac{1}{s^2} \sum_{k=1}^{n} \int_{a_k - \varepsilon s}^{a_k + \varepsilon s} (x_k - a_k)^2\, d F_k(x) \right| .$$

Berry bewijst nu

Stelling 3 Als $\varepsilon_0 \leqq \varepsilon$, $\varepsilon_1 \leqq \varepsilon$ en $\varepsilon_2 \leqq \varepsilon$, dan is $M \leqq (5,8)\,\varepsilon$.

b) Het geval van Lindeberg.

Stel dat de spreiding van iedere X_k eindig is. Stel $s = \sigma$, $a_1 = \alpha_1, \ldots\ldots, a_n = \alpha_n$ nu gaan ε_1 en ε_2 over in:

$$\varepsilon_1 = \sigma^{-1} \sum_{k=1}^{n} \left| \left(\int_{-\infty}^{\alpha_k - \varepsilon\sigma} + \int_{\alpha_k + \varepsilon\sigma}^{\infty} \right) (x_k - \alpha_k)\, d F_k(x) \right|$$

$$\varepsilon_2 = \sigma^{-2} \sum_{k=1}^{n} \left(\int_{-\infty}^{\alpha_k - \varepsilon\sigma} + \int_{\alpha_k + \varepsilon\sigma}^{\infty} \right) (x_k - \alpha_k)^2\, d F_k(x) .$$

Dan geldt:

Stelling 4: Als $\varepsilon_2 \leqq \varepsilon^3$, dan is $M \leqq (3,6)\,\varepsilon$.

c) Het geval van Liapounoff. Hierbij wordt ondersteld het besta[an] van de momenten van de orde $m > 2$ (m hoeft niet geheel te zijn). Stel $\eta = \sigma^{-m} \sum_{k=1}^{n} \mu_m(X_k)$. Uit stelling 4 volgt dan

Stelling 5
$$M \leqq (3,6)\, \eta^{1/(m+1)} .$$

Aanvulling van D.J. Stoker (zie voetnoten 1) en 2)).

Volgens P.L. Hsu (Ann.Math.Stat. 14 (1945), p. 3) is,

$$B_1 = (1,1/6)\sqrt{\pi/2} - \varepsilon/3 - 1/6 \int_{c/\varepsilon}^{\infty} \left\{ (1,1-c)t^2 + c - 2\varepsilon t \right\} e^{-t^2/2}\, dt ,$$

waarbij in dit geval $c = 0,75$. Hieruit volgt:

$$B_1 \leqq (1,1/6)\sqrt{\pi/2} - \varepsilon/3 ,$$

indien:

$$\int_{c/\varepsilon}^{\infty}\left\{(1{,}1-c)t^2+c-2\varepsilon t\right\}e^{-t^2/2}\,dt \geqq 0 .$$

Deze voorwaarde geldt zeker indien:

$$(1{,}1-c)t^2+c-2\varepsilon t \geqq 0 \qquad \text{voor } c/\varepsilon \leqq t \leqq \infty ,$$

en hieraan is weer voldaan indien:

$$\varepsilon^2 \leqq 0{,}2625$$

of indien

$$\varepsilon < 1/1{,}952 .$$

Door deze wijziging in het bewijs **van Berry** worden de bovenste grenzen van de integralen B_2 en B_3 niet beïnvloed.