STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

S 7
(M 1)

Theil H

Over de ongelijkheid van Camp en Meidell.

1949

Overdruk uit Statistica, Jaargang 3, no. 5 en 6, 1949.

Over de ongelijkheid van Camp en Meidell

door H. Theil
Mathematisch Centrum, Amsterdam

Summary
A note on the inequality of Camp and Meidell.

This paper describes a simple and illustrative proof of the inequality of Camp and Meidell according to a line of argument originally due to Dr B. H. de Jong. The conditions to be satisfied in order that the inequality shall be applicalbe are discussed, while in a final section some practical applications are considered.

§ 1. *Inleiding.*

Bij de behandeling van statistische vraagstukken stuit men herhaaldelijk op problemen omtrent het bepalen van overschrijdingskansen. Men beschouwt dan een stochastische grootheid, waarvan aangenomen wordt dat zij gekarakteriseerd is door een zekere verdelingsfunctie, en berekent de waarschijnlijkheid, dat deze grootheid een bepaalde waarde overschrijdt.

Wanneer men op grond van zekere overwegingen de verdelingsfunctie volledig kent, is het probleem eenvoudig. Anders wordt het, indien men omtrent de verdelingsfunctie slechts weinig weet; bijv. als men een steekproef voor zich heeft, die niet voldoende groot is om te besluiten tot een bepaalde verdelingsfunctie van het universum, waaruit de steekproef is getrokken. Men heeft dan echter steeds enig houvast aan de voor elke verdelingsfunctie geldende ongelijkheid van Bienaymé — Tchebycheff[1]), die luidt

$$P\left[|\underline{x}| > x\right] \leqslant \frac{a_k}{x^k}; \tag{1}$$

in woorden: de waarschijnlijkheid, dat de stochastische grootheid $\underline{x}$ in absolute waarde de waarde x overtreft, is hoogstens gelijk aan het absolute moment van de orde k,

$$a_k = \int_{-\infty}^{\infty} |x|^k \, \mathrm{d}\, F(x),$$

gedeeld door de k-de macht van x.

De ongelijkheid van Bienaymé — Tchebycheff is voor de meest voorkomende verdelingen zeer weinig scherp. Een verscherping van deze ongelijkheid kan alleen tot stand komen, als men bepaalde eigenschappen van de verdelingsfunctie veronderstelt. Op dit terrein is het belangrijkste resultaat afkomstig van Meidell en Camp[2]), die hebben aangetoond, dat onder bepaalde voorwaarden de ongelijkheid van Bienaymé—Tchebychef verscherpt kan worden tot

$$P\,[|\underline{x}| > x] \leqslant \left(\frac{k}{k+1}\right)^k \frac{a_k}{x^k}\,. \tag{2}$$

Onlangs heeft Dr B. H. de Jongh een methode aangegeven voor een betrekkelijk eenvoudig bewijs van de ongelijkheid van Camp-Meidell; deze methode heeft hij bovendien benut voor een verdere verscherping van deze ongelijkheid. In tegenstelling tot de bewijzen van Camp en Meidell zelf bleek de methode van Dr de Jongh op eenvoudige wijze voor generalisatie vatbaar te zijn. Wanneer men omtrent de verdelingsfunctie steeds verder gaande veronderstellingen maakt, zal de coefficient C_k in de ongelijkheid

$$P\,[|\underline{x}| > x] \leqslant C_k \frac{a_k}{x^k}\,, \tag{3}$$

die bij Bienaymé — Tchebycheff gelijk is aan 1 en bij Camp-Meidell gelijk is aan $\{k/(k+1)\}^k$, verder dalen. Voor het geval $k = 2$ heeft deze coëfficiënt bij Camp-Meidell de bekende waarde $\frac{4}{9}$ en bij de eerstvolgende verscherping is deze coefficient $\frac{3}{8}$, hetgeen iets kleiner is. Over deze generalisatie zal wellicht elders een artikel verschijnen van de hand van Prof. van Dantzig. Hier zal slechts de ongelijkheid van Camp-Meidell bewezen worden volgens de methode van Dr B. H. de Jongh. Tevens zal aandacht worden geschonken aan de vereiste voorwaarden, waaronder deze ongelijkheid geldig is.

§ 2. *De ongelijkheid van Camp en Meidell* *.)

Noem de verdelingsdichtheid, waarvoor men de ongelijkheid wil afleiden, $\psi(x)$; beschouw verder de verdelingsdichtheid $f(x)$ van de absolute waarde van x. Dan geldt:

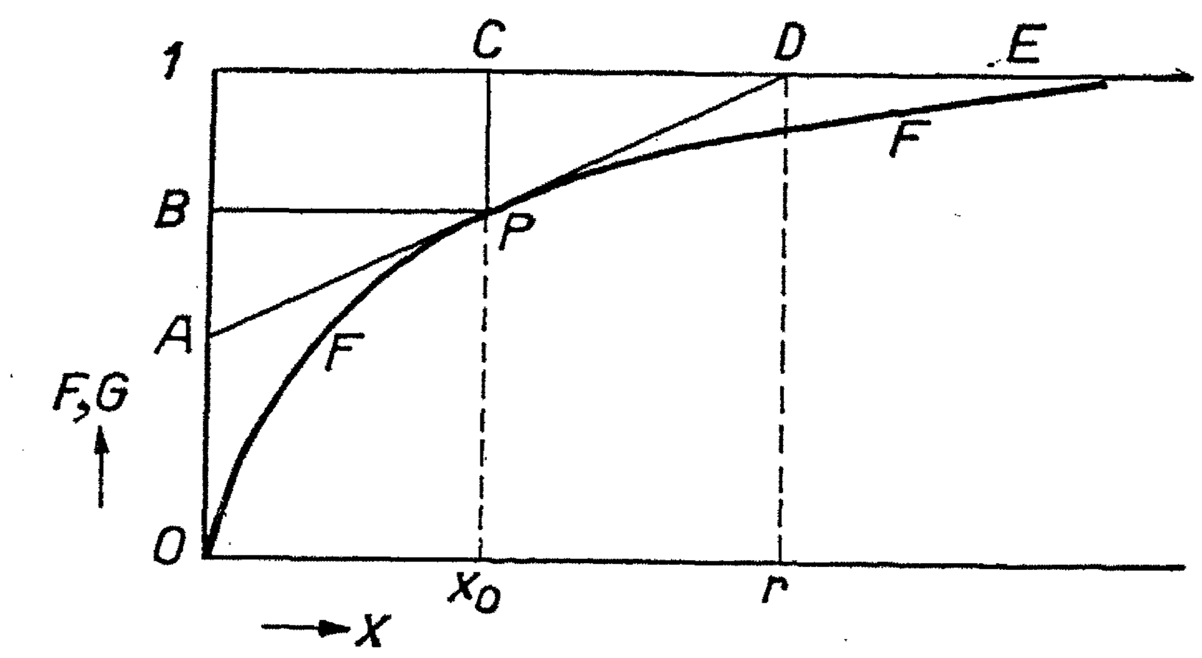
$$f(x) = \psi(x) + \psi(-x).$$

Men interpreteert op de bekende wijze $\psi(x)$ als de dichtheid van een massaverdeling op de x-as; $f(x)$ is dan de dichtheid van de massaverdeling, die

*) Meidell en Camp beschouwden algemenere gevallen dan hier plaats zal hebben. Camp veronderstelde, dat de verdelingsdichtheid $f(x)$ (zie § 2) monotoon niet-stijgend is tussen a en ∞ ($a \geqslant 0$); voor $a > 0$ krijgt de ongelijkheid door het optreden van „correctietermen" een minder eenvoudige gedaante. Meidell veronderstelde dat $f(x)$ monotoon niet-stijgend is tussen 0 en b. In dit artikel wordt verondersteld, dat $f(x)$ monotoon niet-stijgend is tussen 0 en ∞; het is echter ook mogelijk te volstaan met een beperking van de monotonieveronderstelling tot een interval (a, b), hetgeen natuurlijk eveneens correctietermen noodzakelijk maakt.

ontstaat als men de negatieve x-as met de erop gelegen massa's om O omklapt en met de positieve x-as laat samen vallen. We beschouwen nu verder *deze* verdelingsdichtheid $f(x)$. Noem de hierbij beschouwde verdelingsfunctie $F(x)$ en zijn complement, de restfunctie $R(x) = 1 - F(x)$.

De methode van Dr B. H. de Jongh houdt nu in, dat een massaverschuiving tot stand wordt gebracht, waarbij geen enkele massa ,,naar rechts'' (d.w.z. van o af) verplaatst wordt.



Daardoor wordt de verdelingsfunctie $F(x)$ vervangen door een andere verdelingsfunctie $G(x)$, die bij eenzelfde x-waarde nooit een kleinere waarde dan $F(x)$ aanneemt (d.w.z. $G(x) \geqslant F(x)$) of — hetgeen hetzelfde is — bij dezelfde functie-waarde nooit een grotere waarde van x geeft (d.w.z. is $G(x_1) = F(x_2)$, dan is $x_1 \leqslant x_2$).

Door $G(x)$ doelmatig te kiezen, krijgt men dan verschillende ongelijkheden. Zo ligt bij ieder punt P met coödinaten $(x_0, F(x_0))$ de gebroken lijn O B P C E (zie figuur) nergens beneden de grafiek van $F(x)$; immers iedere verdelingsfunctie, waarvan alle massa's $\geqslant$ o zijn, is monotoon niet-dalend. Deze gebroken lijn ontstaat door alle massa's rechts van x_0 gelegen naar x_0, en alle links ervan gelegen massa's naar O te schuiven. Nu geldt steeds, dat het moment van de orde k van een dergelijke naar links verschoven verdeling (α_k') hoogstens gelijk is aan het k-de moment van $F(x)$; immers

$$\alpha_k' = \int_0^\infty x_1^k \, d\, G(x_1) \leqslant \int_0^\infty x_2^k \, d\, F(x_2) = \alpha_k, \tag{4}$$

omdat — zoals reeds gezien — voor dezelfde infinitesimale intervallen $d\, G(x_1) = d\, F(x_2)$ geldt: $x_1 \leqslant x_2$. In het geval van de gebroken lijn O B P C E krijgen we dan:

$$\alpha_k \geqslant \alpha_k' = x_0^k \{1 - F(x_0)\},$$

of

$$P[|\underline{x}| > x_0] = \{1 - F(x_0)\} \leqslant \frac{\alpha_k}{x_0^k}\,. \tag{5}$$

Dit is de ongelijkheid van Bienaymé — Tchebycheff. Daar het mogelijk is, dat $F(x)$ zelf reeds met $G(x)$ overeenstemde (of er willekeurig weinig van verschilde), kan deze ongelijkheid niet verscherpt worden, tenzij men iets meer omtrent $F(x)$ weet dan monotonie alleen.

Indien $F(x)$ niet-positief-convex („concaaf'') is (dus indien $f(x)$ monotoon niet-stijgend is), kan men de massa's zodanig verschuiven dat $F(x)$ vervangen wordt door zijn raaklijn in het punt $(x_0, F(x_0))$; precieser: door de gebroken lijn O A D E; hierbij ontstaat, zoals we beneden zullen zien, de ongelijkheid van Camp-Meidell. De generalisatie, waarop in § 1 gewezen werd houdt in, dat de massaverschuivingen volgens de gebroken lijnen O B P C E en O A D E beschouwd moeten worden als bijzondere gevallen van een bepaalde wijze van verschuiving. Deze wijze van verschuiving moet zodanig plaats hebben, dat het k-de moment van de nieuw ontstane verdelingsfunctie eenvoudig berekend kan worden. In het bijzonder kunnen we — voor het geval, dat $F(x)$ niet-positief-convex is — voor $\bar{G}(x)$ de gebroken lijn O A D E nemen en daarvan het moment van de orde k bepalen. Aangezien de verdelingsdichtheid behorende bij deze verdelingsfunctie (de verdeling is „homogeen'' tussen o en r en heeft een afzonderlijke massa in O) de constante waarde

$$R(x_0)/r(x_0)$$

heeft (waarbij r de absciswaarde is van het punt D), kunnen we voor dit moment schrijven:

$$a_k' = \int_0^r x^k \frac{R(x_0)}{r - x_0}\, dx = \frac{R(x_0)}{r - x_0} \frac{r^{k+1}}{k + 1} \,. \tag{6}$$

Uit (4) en (6) volgt

$$R(x_0) \leqslant (k + 1)\, a_k \frac{r - x_0}{r^{k+1}} \,. \tag{7}$$

In deze ongelijkheid heeft r een bepaalde vaste waarde, afhankelijk van x_0 en $F'(x_0) = f(x_0)$. De redenering is nu als volgt: als deze ongelijkheid geldt voor die bepaalde waarde van r, dan geldt zij zeker ook voor de waarde van r, die de uitdrukking in het rechterlid van (7) maximaal maakt. We zien dus af van de betekenis, die r in de figuur heeft en beschouwen r nu als onafhankelijke variabele. Het voordeel van deze procedure is gelegen in het feit, dat daardoor het aantal grootheden in het rechterlid van (7), dat van de verdelingsfunctie afhankelijk is (nl. a_k en r), met één vermindert.

We beschouwen dus dat deel van het rechterlid van (7), waarin r voorkomt, en schrijven

$$w(r) = \frac{r - x_0}{r^{k+1}}\,. \tag{8}$$

Opdat $w(r)$ maximaal zij, is noodzakelijk, dat zijn logarithmische afgeleide gelijk is aan nul:

$$\frac{d(\log w)}{d\,r} = \frac{1}{r - x_0} - \frac{k+1}{r} = 0,$$

of:

$$r = \frac{k+1}{k}\,x_0. \tag{9}$$

Aangezien de logarithmische afgeleide van $w(r)$ rondom het stationaire punt een dalend verloop heeft (d.w.z. $\frac{d(\log w)}{d\,r} > 0$, resp. $= 0$, resp. < 0 voor $r <$, resp. $=$, resp. $> \frac{k+1}{k}\,x_0$) *), is dit extremum inderdaad een maximum.

Invulling van deze waarde van r in de ongelijkheid (7) geeft

$$R(x_0) \leqslant \left(\frac{k}{k+1}\right)^k \frac{a_k}{x_0^{\,k}}\,, \tag{10}$$

d.w.z. de ongelijkheid van Camp en Meidell voor de verdelingsdichtheid $f(x)$; de grootheid x_0 kan uiteraard willekeurige positieve waarden aannemen.

De ongelijkheid van de verdelingsdichtheid $\psi(x)$ verschilt hiervan slechts in zoverre, dat de enkelzijdige overschrijdingskans $R(x)$ door een tweezijdige moet worden vervangen; aangezien verder a_k voor $f(x)$ en $\psi(x)$ gelijk is, geldt dus:

$$P\,[|\underline{x}| > x] \leqslant \left(\frac{k}{k+1}\right)^k \frac{a_k}{x^k}\,. \tag{11}$$

§ 3. *Over de vereiste voorwaarden voor toepassing van de ongelijkheid.*

Voor toepassing van de ongelijkheid van Camp en Meidell is noodzakelijk dat $f(x)$ monotoon niet-stijgend is. Als $\psi(x)$ unimodaal is en zijn maximum bereikt in $\hat{x}$ kan men de in § 2 besproken „omklapping'' laten plaats hebben in een punt (a, O), zodanig dat a met $\hat{x}$ samenvalt of in een niet te grote omgeving van $\hat{x}$ ligt. Dit laatste is van belang omdat de modus van de steekproef-

---

*) Korter geschreven $\frac{d(\log w)}{d\,r} \gtreqless 0$ voor $r \lesseqgtr \frac{k+1}{k}\,x_0$. Het berekenen van de tweede afgeleide is niet noodzakelijk om de aard van het extremum te onderzoeken.

verdeling in het algemeen niet zal samenvallen met de modus van de oorspronkelijke verdeling. In hoeverre men hiervan gebruik kan maken, moet nu nader worden gespecificeerd.

Neem $a > \hat{x}$ en veronderstel dat $\psi(x)$ differentieerbaar is. Opdat $f(x)$ monotoon niet-stijgend is, moet voor iedere waarde van x

$$f(x) = \psi(a + x) + \psi(a - x)$$

minstens gelijk zijn aan

$$f(x + \mathrm{d}x) = \psi(a + x + \mathrm{d}x) + \psi(a - x - \mathrm{d}x),$$

waarbij $\mathrm{d}x > 0$ is. D.w.z.

$$\begin{aligned} f(x+\mathrm{d}x) - f(x) &= \{\psi(a+x+\mathrm{d}x) - \psi(a+x)\} - \{\psi(a-x) - \psi(a-x-\mathrm{d}x)\} \\ &= \psi'(a + x + \theta_1 \mathrm{d}x)\,\mathrm{d}x - \psi'(a - x - \theta_2 \mathrm{d}x)\,\mathrm{d}x \leqslant 0, \end{aligned} \tag{12}$$

waarbij $0 \leqslant \theta_i \leqslant 1$ $(i = 1,2)$. Voldoende is dus: $\psi'(a + x) \leqslant \psi'(a - x)$ voor iedere x. Deze voorwaarde behoeven we echter niet te stellen voor grotere waarden van x dan $(a - \hat{x})$, want links van de modus daalt $\psi(x)$ naar links. Dit houdt in, dat we vanaf de corresponderende waarde ter rechter zijde van a,

$$a + (a - \hat{x}) = 2a - \hat{x},$$

slechts behoeven te eisen, dat $\psi(x)$ monotoon niet-stijgend is.

Hieruit en uit de voorwaarde $\psi'(a + x) \leqslant \psi'(a - x)$ (waarbij $x \leqslant a - \hat{x}$) volgt dat de volgende aan a te stellen eis voldoende (niet noodzakelijk) is voor de geldigheid van de ongelijkheid van Camp—Meidell: $\psi'(x)$ is in het interval $(\hat{x},\ 2a - \hat{x})$ monotoon niet-stijgend; of ook: dit interval bevat geen buigpunt van $\psi(x)$. Geven we de abscis van het naastbijgelegen rechterbuigpunt aan met b_2, dan is dus voldoende dat

$$b_2 > 2a - \hat{x},$$

d.w.z.

$$a < \frac{\hat{x} + b_2}{2}\,. \tag{13}$$

Aangezien analoge beschouwingen gehouden worden voor het geval, dat a links van de modus ligt, kunnen we zeggen, dat voldoende is, dat a gekozen is in het open interval

$$\left(\frac{b_1 + \hat{x}}{2},\ \frac{b_2 + \hat{x}}{2}\right),$$

waarbij b_1, de abscis van het naastbijgelegen linker-buigpunt voorstelt.

§ 4. *Een toepassing.*

Waargenomen waren de volgende steekproeven uit twee onbekende universa:

Steekproef I		Steekproef II	
x	frequentie	y	frequentie
0,21	1	0	5
0,27	1	1	6
0,63	1	2	4
0,80	1	3	4
0,85	1	4	1
1,58	1	5	2
1,59	1	6	1
1,80	1	7	—
1,84	1	8	1
1,99	1	9	1
2,25	1	13	1
2,28	1	14	1
2,51	1	16	1
3,28	1	21	1
3,86	1	47	1
4,64	1		
8,47	1		
11,43	1		
23,27	1		
34,54	1		

Gevraagd was, voor beide verdelingen overschrijdingskansen te berekenen. Gezien het feit, dat de verdelingen voor grote x en y slechts langzaam naar nul naderen, leek het niet raadzaam tot aanpassing door middel van een standaardverdeling over te gaan. Aangezien aangenomen mocht worden, dat de verdelingsdichtheid een dalende functie is van x en y, was het dus zinvol te onderzoeken of de ongelijkheid van Camp en Meidell een bruikbaar resultaat kan geven.

Aan deze mogelijkheid is het nadeel verbonden, dat het k-de absolute moment van het universum in tegenstelling met dat van de steekproef onbekend is. Aangezien echter x en y alleen niet-negatieve waarden kunnen aannemen, zodat het k-de absolute moment gelijk is aan het k-de gewone moment, kan de volgende redenering toepassing vinden. De mathematische verwachting van het k-de steekproef-moment is gelijk aan het k-de moment van het universum; immers

$$\mathcal{E}\, \underline{m}_k = \mathcal{E}\, \frac{1}{n} \Sigma\, \underline{x}^k = \frac{1}{n}\, \mathcal{E}\, \Sigma\, \underline{x}^k = \mu_k. \qquad (14)$$

Dan is ook

$$\mathcal{E}\left\{\left(\frac{k}{k+1}\right)^k \frac{\underline{m}_k}{x^k}\right\} = \left(\frac{k}{k+1}\right)^k \frac{\mu_k}{x_k}. \tag{15}$$

We kunnen dus voor ε_0, in de ongelijkheid

$$R(x) \leqq \left(\frac{k}{k+1}\right)^k \frac{\mu_k}{x^k} \equiv \varepsilon_0, \tag{16}$$

uit de steekproef een „schatting" („estimate") $\underline{\varepsilon}$ vinden:

$$\underline{\varepsilon} = \left(\frac{k}{k+1}\right)^k \frac{\underline{m}_k}{x^k}. \tag{17}$$

$\underline{\varepsilon}$ is dus een „zuivere schatting" („unbiased estimate") van ε_0 (immers $\mathcal{E}\,\underline{\varepsilon} = \varepsilon_0$). Meer dan een *schatting* van de overschrijdingskans kan men op deze wijze niet verkrijgen. De kans, dat deze schatting te klein is, dus dat $\varepsilon_0 > \underline{\varepsilon}$ is, is in het algemeen ongeveer $\frac{1}{2}$ (evenals de kans op $\varepsilon_0 < \underline{\varepsilon}$). Toepassingen zijn slechts onder dit voorbehoud mogelijk.

Voor ε vindt men, indien men de ongelijkheid toepast met het eerste resp. het tweede steekproef-moment:

Verdeling	*Eerste moment*	*Tweede moment*
I	$\varepsilon = \dfrac{2{,}7}{x}$	$\varepsilon = \dfrac{31{,}9}{x^2}$
II	$\varepsilon = \dfrac{2{,}9}{y}$	$\varepsilon = \dfrac{38}{y^2}$.

De grootte van x resp. y bepaalt of het eerste of het tweede moment het scherpste resultaat geeft. Men vindt gemakkelijk, dat (voor verdeling I) als $x < 11{,}8$ is, het gebruik van het eerste moment het voordeligst is, en dat voor grotere waarden van x het tweede moment te prefereren is. Voor verdeling II is deze grens 13,1.

LITERATUUR

1. I. J. Bienaymé, C. R. Acad. Sci. Paris, **37**, 1853; P. L. Tchebycheff, Des valeurs moyennes, J. de Math. **12**, 1867.

2. B. Meidell, Sur un problème du calcul des probabilités et les statistiques mathématiques, C. R. Acad. Sci. Paris, **175**, 806—808, 1922.

   B. H. Camp, A new generalization of Tchebycheff's statistical inequality, Bull. Am. math. Soc, **28**, 427-432, 1922.