STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

De asymptotische eigenschappen van de binomiale verdeling

door

J. Hemelrijk

(Statistica Neerlandica, 16(1962), p 255-258)

De asymptotische eigenschappen van de binomiale verdeling *)

In een reeks van n onafhankelijke experimenten, ieder met kans p op succes, bezit het totale aantal successen, $\underline{x}$, een binomiale verdeling:

$$(1) \qquad P(\underline{x} = x) = \binom{n}{x} p^x q^{n-x} \qquad (q = 1 - p).$$

Het is algemeen bekend, dat $\underline{x}$ asymptotisch normaal is, als n, bij constante p (> 0), onbegrensd toeneemt. In formule:

$$(2) \qquad \lim_{n \to \infty} P\left(\frac{\underline{x} - np}{\sqrt{npq}} \leq a\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}u^2}\, du.$$

Houdt men bij deze limietovergang p niet constant, maar laat men p op zodanige wijze tot 0 naderen dat

$$(3) \qquad \lim_{n \to \infty} np = \mu \qquad (\mu > 0),$$

dan bezit $\underline{x}$ in de limiet een P o i s s o n - verdeling:

$$(4) \qquad \lim_{\substack{n \to \infty \\ np \to \mu}} P(\underline{x} = x) = e^{-\mu} \frac{\mu^x}{x!}.$$

Hiermee zijn echter nog niet alle mogelijkheden opgesomd.

In het bijzonder kan (met $n \to \infty$) p zo langzaam tot 0 naderen, dat $np \to \infty$. Het ligt voor de hand te veronderstellen, dat $\underline{x}$ in dat geval weer asymptotisch normaal is. Immers ook de P o i s s o n - verdeling heeft, voor $\mu \to \infty$, deze eigenschap. Deze veronderstelling is inderdaad juist, zoals wij hieronder zullen laten zien.

Een duidelijk overzicht wordt verkregen door beschouwing van de variantie

$$(5) \qquad \sigma_n^2 = npq,$$

die het voordeel heeft symmetrisch te zijn in p en q.

Hierin stellen p en q geen constanten voor, maar functies van n. Wij veronderstellen nu, dat σ_n^2 voor $n \to \infty$ een limiet bezit, waarbij ∞ als limiet toegelaten is. Is aan deze voorwaarde niet voldaan, dan is er ook geen limietverdeling. De volgende drie gevallen kunnen zich nu voordoen.

*) Rapport S 294 (M85) van de Afdeling Mathematische Statistiek van het Mathematisch Centrum, Amsterdam.

I. *Asymptotische normaliteit.* Noodzakelijk en voldoende hiervoor is

$$\lim_{n \to \infty} \sigma_n^2 = \infty. \tag{6}$$

In dat geval geldt (2), doch ook $n - \underline{x}$ is dan asymptotisch normaal. De grootheden p en q behoeven geen limiet te bezitten, maar (6) sluit evenmin uit dat p (of q) tot 0 nadert. Het bewijs van deze stelling wordt verderop gegeven.

II. *P o i s s o n - verdeling als limiet.*

Zonder verlies van algemeenheid kunnen wij stellen, dat voor iedere n geldt: $p \leqq q$, dus $p \leqq \frac{1}{2}$. Is $p > q$, dan treedt $n - \underline{x}$ in de plaats van $\underline{x}$.
Noodzakelijk en voldoende, opdat de limietverdeling een P o i s s o n - verdeling is, is nu

$$\lim_{n \to \infty} \sigma_n^2 = \mu \qquad (0 < \mu < \infty). \tag{7}$$

III. *Ontaarding.* Stellen wij weer $p \leqq q$, dan is de voorwaarde

$$\lim_{n \to \infty} \sigma_n^2 = 0 \tag{8}$$

noodzakelijk en voldoende voor

$$\lim_{n \to \infty} P(\underline{x} = 0) = 1. \tag{9}$$

Bewijs van I.

Het bewijs van de asymptotische normaliteit kan gevoerd worden met behulp van een belangrijke moderne vorm van de centrale limietstelling, nl. de stelling van L i n d e b e r g - L é v y - F e l l e r : *)

Laat $\underline{x}_{n1}, \underline{x}_{n2}, \ldots, \underline{x}_{nn}$ onderling onafhankelijke stochastische grootheden zijn met verwachtingen en varianties die voor iedere n voldoen aan

$$\mathscr{E}\, \underline{x}_{n\nu} = 0 \quad (\nu = 1, \ldots, n) \ \textit{en} \ \sum_{\nu=1}^{n} \sigma^2(\underline{x}_{n\nu}) = 1, \tag{10}$$

dan is $\sum_{\nu=1}^{n} \underline{x}_{n\nu}$ dan en slechts dan asymptotisch normaal als

$$\lim_{n \to \infty} \sum_{\nu=1}^{n} \int_{|w| \geqq \varepsilon} w^2 dF_{n\nu}(w) = 0 \qquad \textit{voor iedere } \varepsilon > 0, \tag{11}$$

waarin $F_{n\nu}$ de verdelingsfunctie van $\underline{x}_{n\nu}$ voorstelt (en tevens is dan $\max \sigma^2(\underline{x}_{n\nu}) \to 0$).

*) Zie bijv. M. L o è v e, Probability Theory, N.Y. 1955, p. 295.

Zij nu $\underline{y}_{n\nu}$ het aantal (0 of 1) successen bij het ν^e experiment van de reeks met lengte n, zodat het totale aantal successen, $\underline{x}$, gelijk is aan $\sum_{\nu=1}^{n} \underline{y}_{n\nu}$, dan geldt

$$\mathscr{E}\underline{y}_{n\nu} = p \quad \text{en} \quad \sigma^2(\underline{y}_{n\nu}) = pq \qquad (\nu = 1, \ldots, n; \; n = 1, 2, \ldots),$$

zodat de grootheden

$$\underline{x}_{n\nu} \overset{\text{def}}{=} \frac{\underline{y}_{n\nu} - p}{\sqrt{npq}} \tag{12}$$

aan (10) voldoen. De door $\underline{x}_{n\nu}$ aangenomen waarden zijn

$$\frac{q}{\sqrt{npq}} \text{ (met kans } p\text{) en } \frac{-p}{\sqrt{npq}} \text{ (met kans } q\text{)}, \tag{13}$$

zodat de integraal in (11), als wij even afzien van de voorwaarde $|w| \geqq \varepsilon$, tot twee termen gereduceerd wordt. Deze bezitten de vorm

$$\frac{pq^2}{npq} = \frac{q}{n} \quad \text{resp.} \quad \frac{qp^2}{npq} = \frac{p}{n} \tag{14}$$

en zij tellen hoogstens mee als

$$\frac{q}{\sqrt{npq}} \quad \text{resp.} \quad \frac{p}{\sqrt{npq}} \geqq \varepsilon. \tag{15}$$

Is aan (6) voldaan, dan is voor voldoende grote n aan (15) niet meer voldaan, daar $npq \to \infty$ en p en $q \leqq 1$. De integraal wordt dan dus $= 0$ en aan (11) is voldaan, zodat de asymptotische normaliteit geldt.

Is aan (6) niet voldaan dan is er een $\delta > 0$ zodanig, dat er willekeurig grote waarden van n te vinden zijn met

$$\frac{1}{\sqrt{npq}} > \delta. \tag{16}$$

Nemen wij nu $\varepsilon = \frac{1}{2}\,\delta$ dan is, voor de beschouwde waarden van n, steeds voldaan aan minstens één van beide ongelijkheden van (15), daar de som van de linkerleden gelijk is aan $1/\sqrt{npq}$, dus $> 2\,\varepsilon$ is. Is $p \leqq q$ (de algemeenheid wordt hierdoor niet geschaad), dan is de integraal in (11) dus minstens gelijk aan q/n, dus $\geqq 1/2n$. De som in (11) bestaat uit n identieke termen en is $\geqq \frac{1}{2}$. Derhalve is (11) niet vervuld en is de verdeling niet asymptotisch normaal.

Het elementaire *bewijs van II* is voldoende bekend en behoeft hier niet herhaald te worden. Ook III is gemakkelijk te bewijzen. Voor de volledigheid vermelden wij echter een algemene limietstelling, eveneens te vinden in het boek van L o è v e, p. 296, die noodzakelijke en voldoende voorwaarden geeft voor het optreden van de verdeling van P o i s s o n als limietverdeling van een som onafhankelijke stochastische grootheden.

Laat $\underline{x}_{n1}, \underline{x}_{n2}, \ldots, \underline{x}_{nn}$ *onderling onafhankelijke stochastische grootheden zijn met*

$$\max_{\nu} \sigma^2(\underline{x}_{n\nu}) \to 0 \quad en \quad \sum_{\nu=1}^{n} \sigma^2(\underline{x}_{n\nu}) \to \mu, \tag{14}$$

dan is de limietverdeling van $\sum_{\nu=1}^{n} \underline{x}_{n\nu}$ *dan en slechts dan een P o i s s o n - verdeling als*

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} \mathscr{E} \underline{x}_{n\nu} = \mu \tag{15}$$

en

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} \int\limits_{|w-1| \geq \epsilon} w^2 \, dF_{n\nu}(w + \mathscr{E} \underline{x}_{n\nu}) = 0 \ voor\ iedere\ \varepsilon > 0, \tag{16}$$

waarin $F_{n\nu}$ *de verdelingsfunctie van* $\underline{x}_{n\nu}$ *voorstelt.*

Hoewel uit deze stelling blijkt, dat de $\underline{x}_{n\nu}$ niet dezelfde verdelingen behoeven te bezitten en zowel continu als discreet verdeeld mogen zijn, leert een nadere beschouwing van de stelling dat deze generalisatie van de limietstelling van P o i s s o n geen essentieel nieuwe mogelijkheden onthult.

In tegenstelling tot stelling I is het belang van stelling II eerder dat de laatstgenoemde aangeeft hoe beperkt de mogelijkheden zijn om als limietverdeling voor de som van onafhankelijke grootheden een P o i s s o n - verdeling te verkrijgen. Essentieel zijn alleen vrij triviale modificaties mogelijk op het oorspronkelijk door P o i s s o n beschouwde binomiale geval.

J. H e m e l r ij k.