

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 294
(m ps)

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, x , een binomiale verdeling:

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

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

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

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

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

dan bezit x in de limiet een P o i s s o n - verdeling:

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

Hiermee zijn echter nog niet alle mogelijkheden opgesomd.

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

Een duidelijk overzicht wordt verkregen door beschouwing van de variantie

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

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 $n \rightarrow \infty$ 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

$$(6) \quad \lim_{n \rightarrow \infty} \sigma_n^2 = \infty.$$

In dat geval geldt (2), doch ook $n - \bar{x}$ 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: $p \leq q$, dus $p \leq \frac{1}{2}$. Is $p > q$, dan treedt $n - \bar{x}$ in de plaats van $\bar{x}$.

Noodzakelijk en voldoende, opdat de limietverdeling een P o i s s o n - verdeling is, is nu

$$(7) \quad \lim_{n \rightarrow \infty} \sigma_n^2 = \mu \quad (0 < \mu < \infty).$$

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

$$(8) \quad \lim_{n \rightarrow \infty} \sigma_n^2 = 0$$

noodzakelijk en voldoende voor

$$(9) \quad \lim_{n \rightarrow \infty} P(\bar{x} = 0) = 1.$$

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

$$(10) \quad E \bar{x}_{nv} = 0 \quad (v = 1, \dots, n) \text{ en } \sum_{v=1}^n \sigma^2(\bar{x}_{nv}) = 1,$$

dan is $\sum_{v=1}^n \bar{x}_{nv}$ dan en slechts dan asymptotisch normaal als

$$(11) \quad \lim_{n \rightarrow \infty} \sum_{v=1}^n \int_{|w| \geq \epsilon} w^2 dF_{nv}(w) = 0 \quad \text{voor iedere } \epsilon > 0,$$

waarin F_{nv} de verdelingsfunctie van $\bar{x}_{nv}$ voorstelt (en tevens is dan $\max \sigma^2(\bar{x}_{nv}) \rightarrow 0$).

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

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

$$\mathcal{E} y_{nv} = p \text{ en } \sigma^2(y_{nv}) = pq \quad (v = 1, \dots, n; n = 1, 2, \dots),$$

zodat de grootheden

$$(12) \quad x_{nv} \stackrel{\text{def}}{=} \frac{y_{nv} - p}{\sqrt{npq}}$$

aan (10) voldoen. De door x_{nv} aangenomen waarden zijn

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

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

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

en zij tellen hoogstens mee als

$$(15) \quad \frac{q}{\sqrt{npq}} \text{ resp. } \frac{p}{\sqrt{npq}} \geq \varepsilon.$$

Is aan (6) voldaan, dan is voor voldoende grote n aan (15) niet meer voldaan, daar $npq \rightarrow \infty$ en p en $q \leq 1$. 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

$$(16) \quad \frac{1}{\sqrt{npq}} > \delta.$$

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 $p \leq q$ (de algemeenheid wordt hierdoor niet geschaad), dan is de integraal in (11) dus minstens gelijk aan q/n , dus $\geq 1/2n$. De som in (11) bestaat uit n identieke termen en is $\geq \frac{1}{2}$. 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 $x_{n1}, x_{n2}, \dots, x_{nn}$ onderling onafhankelijke stochastische grootheden zijn met

$$(14) \quad \max_{\nu} \sigma^2(x_{n\nu}) \rightarrow 0 \quad \text{en} \quad \sum_{\nu=1}^n \sigma^2(x_{n\nu}) \rightarrow \mu,$$

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

$$(15) \quad \lim_{n \rightarrow \infty} \sum_{\nu=1}^n \mathcal{E} x_{n\nu} = \mu$$

en

$$(16) \quad \lim_{n \rightarrow \infty} \sum_{\nu=1}^n \int_{|w-1| \geq \epsilon} w^2 dF_{n\nu}(w + \mathcal{E} x_{n\nu}) = 0 \quad \text{voor iedere } \epsilon > 0,$$

waarin $F_{n\nu}$ de verdelingsfunctie van $x_{n\nu}$ voorstelt.

Hoewel uit deze stelling blijkt, dat de $x_{n\nu}$ 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 i j k.