De Poisson benadering voor de binomiale verdeling *)

Om het gebruik van de Poisson verdeling als benadering voor een binomiale verdeling te rechtvaardigen beroept men zich meestal op de bekende limietstelling die zegt, dat de binomiale verdeling met parameters n en p naar de Poisson verdeling met parameter λ convergeert als $n \to \infty$ en tegelijkertijd $p \to 0$, zodanig dat $np \to \lambda$. Strikt genomen is een dergelijke redenering zonder meer niet voldoende, omdat deze limietstelling geen antwoord geeft op de vraag hoe groot n en hoe klein p dan wel moet zijn opdat de benadering redelijk is. Daarvoor is een approximatiestelling nodig, d.w.z. een stelling, die bij vaste n en p aangeeft in welke mate de binomiale verdeling met deze parameters afwijkt van de Poisson verdeling met parameter $\lambda = np$.

Nu zijn approximatiestellingen in het algemeen veel moeilijker te bewijzen dan limietstellingen, maar de volgende stelling van HODGES en LE CAM, [1], toont aan, dat in dit geval het omgekeerde waar is.

Stelling

Als S een binomiaal verdeelde stochastische grootheid is met parameters n en p, terwijl de stochastische grootheid T de Poisson verdeling met parameter $\lambda = np$ heeft, dan geldt de ongelijkheid

$$\left| P\{S \in A\} - P\{T \in A\} \right| \leqq np^2$$

voor elke verzameling A van reële getallen.

Bewijs

Laten $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ onderling onafhankelijke, identiek verdeelde, stochastische vectoren zijn, wier gemeenschappelijke twee-dimensionale verdeling gegeven wordt door

$$P\{X_i = Y_i = 0\} = 1 - p,$$

$$P\{X_i = 1,\ Y_i = 0\} = e^{-p} - 1 + p,$$

$$P\{X_i = 1,\ Y_i = k\} = e^{-p} \frac{p^k}{k!} \qquad (k = 1, 2, \cdots).$$

Dit houdt in, dat X_i de alternatieve verdeling met kans op succes p heeft, dat Y_i de Poisson verdeling met parameter p heeft en dat $P\{X_i = Y_i\}$ zo groot

*) Rapport S 378 van de Statistische Afdeling van het Mathematisch Centrum.

is als bij deze gegeven marginale verdelingen van X_i en Y_i maar mogelijk is. Uit deze marginalen volgt immers de ongelijkheid

$$P\{X_i = Y_i\} = P\{X_i = Y_i = 0\} + P\{X_i = Y_i = 1\} \leqq$$
$$\leqq P\{X_i = 0\} + P\{Y_i = 1\} = 1 - p + pe^{-p},$$

terwijl de gelijkheid

$$P\{X_i = Y_i\} = 1 - p + pe^{-p}$$

gemakkelijk is af te leiden uit de opgegeven simultane verdeling van X_i en Y_i.

Omdat de stochastische vectoren (X_i, Y_i), $i = 1, 2, \ldots, n$ onderling onafhankelijk zijn, hebben

$$\sum_{i=1}^{n} X_i \text{ en } \sum_{i=1}^{n} Y_i$$

dezelfde verdelingen als S resp. T, en dus hebben we

$$| P\{S \in A\} - P\{T \in A\} | = \left| P\left\{\sum_1^n X_i \in A\right\} - P\left\{\sum_1^n Y_i \in A\right\} \right| =$$
$$= \left| P\left\{\sum_1^n X_i \neq \sum_1^n Y_i, \sum_1^n X_i \in A\right\} - P\left\{\sum_1^n X_i \neq \sum_1^n Y_i, \sum_1^n Y_i \in A\right\} \right| \leqq$$
$$\leqq P\left\{\sum_1^n X_i \neq \sum_1^n Y_i\right\} \leqq P\left\{\bigcup_1^n \{X_i \neq Y_i\}\right\} \leqq$$
$$\leqq \sum_1^n P\{X_i \neq Y_i\} = np\,(1 - e^{-p}) \leqq np^2,$$

hetgeen te bewijzen was.

Veronderstelt men dat S verdeeld is als het aantal successen in n onderling onafhankelijke, doch niet noodzakelijk identieke, alternatieven met kans op succes $p_1, p_2, \ldots, p_n$, dan kan men op dezelfde manier bewijzen dat

$$| P\{S \in A\} - P\{T \in A\} | \leqq \sum_1^n p_i^2$$

voor elke verzameling $A \subset (-\infty, +\infty)$, als T wederom de Poisson verdeling met dezelfde verwachting als S heeft. Men hoeft daarvoor in het hierboven gegeven bewijs slechts de veronderstelling, dat de stochastische vectoren (X_i, Y_i) identiek verdeeld zijn, te laten vallen en in de specificatie van de verdeling van (X_i, Y_i) overal p door p_i te vervangen.

Uitgaande van dit resultaat, en gebruik makend van niet meer dan de formule van STIRLING, de ongelijkheid van CHEBYCHEV, en enig gereken, bewezen HODGES en LE CAM in [1] ook nog de ongelijkheid

$$\sup_x \left| P\{S \leqq x\} - P\{T \leqq x\} \right| \leqq 3 \sqrt[3]{\max_{1 \leqq i \leqq n} p_i}.$$

Met veel meer moeite heeft LE CAM in [2] tenslotte langs geheel andere weg kunnen aantonen dat men het rechterlid van deze laatste ongelijkheid zelfs mag vervangen door 9 max $(p_1, \cdots, p_n)$.

Men kan in deze approximatiestellingen een verklaring zien voor het feit dat vele in de statistische praktijk voorkomende stochastische grootheden (aantallen telefoongesprekken, ongelukken e.d.) verdelingen blijken te hebben, die zeer goed benaderd kunnen worden met Poisson verdelingen.

Het directe praktische nut van de stellingen is echter gering, daar numerieke vergelijking leert, dat de beschouwde benaderingen doorgaans aanzienlijk beter zijn dan de stellingen zouden doen vermoeden.

Literatuur

[1] J. L. HODGES, Jr. and LUCIEN LE CAM, The Poisson approximation to the Poisson binomial distribution, Ann. Math. Statist. **31** (1960), 737-740.

[2] LUCIEN LE CAM, An approximation theorem for the Poisson binomial distribution, Pacific J. Math. **10** (1960), 1181-1197.

J. FABIUS