

Bachelorproject wiskunde:

Het schatten van quantielen van de negatief binomiale verdeling

Begeleiders:

- dr. A. Di Bucchianico, HG 8.15, a.d.bucchianico@tue.nl
- prof.dr. E.R. van den Heuvel, Rijksuniversiteit Groningen, e.r.van.den.heuvel@epi.umcg.nl

Probleembeschrijving De normale verdeling wordt gebruikt voor continue variabelen en speelt een dominante rol in inleidende colleges statistiek. Dit komt enerzijds door de centrale limietstelling en anderzijds doordat de normale verdeling een goed hanteerbare kansverdeling is. Er zijn echter veel meer verdelingen in de statistiek die in de colleges enigszins onderbelicht zijn gebleven. Zo zijn de Poisson en negatief binomiale verdeling belangrijke verdelingen voor het beschrijven van discrete variabelen zoals tellingen.

In een productieomgeving voor het produceren van medicijnen wordt het aantal micro-organismen op verschillende plaatsen dagelijks gemeten (tellingen). Op sommige plekken mogen (bijna) geen bacteriën voorkomen (steriele omgeving), maar op andere plekken is het belangrijk dat het aantal bacteriën niet te hoog is en niet toeneemt in de tijd. Voor het detecteren van trends worden de gevonden aantallen in een trendgrafiek geplaatst samen met een bovengrens voor dit aantal. Door de data te beschrijven met een Poisson of een negatief binomiale verdeling kan deze bovengrens worden geschat met een quantiel van de verdeling.

De Poisson verdeling is vaak ontoereikend voor het beschrijven van microbiologische data omdat de standaardafwijking tussen tellingen vaak groter is dan de standaardafwijking $\sqrt{\lambda}$ van de Poissonverdeling met verwachting λ . De negatief binomiale verdeling heeft twee parameters (η en δ) en is daardoor flexibeler. Merkwaardig genoeg is er maar vrij weinig bekend over het schatten van quantielen van de negatief binomiale verdeling.

Opdracht Een stochast X met mogelijke uitkomsten in $\{0, 1, 2, \dots\}$ volgt een negatief binomiale verdeling met parameters $\eta > 0$ en $\delta \geq 0$ als het de volgende kansverdeling heeft

$$P(X = x) = \frac{\Gamma(x + 1/\delta)}{x! \Gamma(1/\delta)} \frac{(\delta\eta)^x}{(1 + \delta\eta)^{x+1/\delta}},$$

waarbij Γ de gammafunctie is. De meest voor de hand liggende en gebruikte methode voor het schatten van quantielen is eerst de parameters van de negatief binomiale verdeling te schatten en dan het quantiel te schatten door de verdelingsfunctie te inverteren. Dit leidt echter niet tot expliciete formules. Een alternatief is de staart van de negatief binomiale verdeling direct te benaderen met een kansverdeling waarvoor wel expliciete schatters voor de quantielen bekend zijn. Een ander alternatief is de data eerst te transformeren en dan te beschrijven met een eenvoudige en bekende verdeling. Er is zeer weinig literatuur waarin deze aanpakken goed met elkaar worden vergeleken. De referentielijst geeft een aantal relevante artikelen.

De doelen van dit bachelorproject zijn:

1. Het maken van een (compleet) literatuuroverzicht waarin specifiek gekeken wordt naar

- (a) Schattingsmethoden voor de parameters en de quantielen van de negatief binomiale verdeling
 - (b) Benaderingsformules voor de staart van de negatief binomiale verdeling, direct of via transformaties.
 - (c) Eenvoudige verdelingsvrije (niet-parametrische) methoden voor het schatten van quantielen voor de negatief binomiale verdeling.
2. Het selecteren van de meest geschikte methoden op basis van
- (a) Theoretische overwegingen en expliciete berekeningen.
 - (b) Simulaties van de methoden.
 - (c) Het vergelijken en afwegen van alle resultaten.

Benodigde voorkennis

1. Basiskennis kansrekening.
2. Basiskennis statistische schattingstheorie.

Referenties

1. ANSCOMBE FJ, The transformation of Poisson, Binomial and Negative-Binomial Data, *Biometrika Trust*, 1948, **35**:246-254.
2. BLOM G, Transformations of the binomial, negative binomial, Poisson and χ^2 distributions, *Biometrika Trust*, 1954, **41**:302-316.
3. CHRISTENSEN A, MELGAARD H, IWERSEN J, and THYREGOD P, Environmental monitoring based on a hierarchical Poisson-gamma model, *Journal of Quality Technology*, 2003, **35**:275-285.
4. CLARK SJ and PERRY JN, Estimation of the negative binomial parameters κ by maximum quasi-likelihood, *Biometrics*, 1989, **45**:309-316.
5. GUENTHER WC, A simple approximation to the negative binomial (and regular binomial), *Technometrics*, 1972, **14**:385-389.
6. HOFFMAN D, Negative binomial control limits for count data with extra-Poisson variation, *Pharmaceutical Statistics*, 2003, **2**:127-132.
7. PIEGORSCH WW, Maximum likelihood estimation for the negative binomial dispersion parameter, *Biometrics*, 1990, **46**:863-867.
8. PIETERS EP, GATES CE, MATIS JH, and STERLING WL, Small sample comparison of different estimators of negative binomial distribution, *Biometrics*, 1977, **33**:718-723.