

Navorsingsbriewe

Foutbespeuring tydens proeflees

UITTREKSEL

Twee metodes wat gebruik kan word om die totale getal setfoute in 'n teks te beraam op grond van die getalle foute wat deur twee onafhanklike proeflesers gevind is, word bespreek. Een van hierdie metodes is nuut en is van toepassing wanneer twee proeflesers dieselfde kopie lees.

ABSTRACT

Error detection in proofreading

Two methods that can be used to estimate the total number of typographic errors in a text from the numbers of errors detected by two independent proofreaders are discussed. One of these methods is new and can be applied when two proofreaders read the same copy.

INLEIDING

'n Aantal jare gelede het Polya¹ aangedui hoe die getal foute wat deur twee onafhanklike proeflesers in verskillende kopieë van dieselfde teks gevind is, gebruik kan word om die waarskynlike totale getal foute in die teks te beraam. Gestel die twee lesers het waarskynlikhede p en q om foute te vind, en gestel verder dat hulle respektiewelik n en m foute opspoor, terwyl x foute deur beide gevind word. Onder hierdie voorwaardes kan n , m en x as beraamings van Np , Nq en Npq gebruik word, waar N die werklike (maar onbekende) getal foute in die teks is. Aangesien $N = Np.Nq/Npq$, kan die totale getal foute deur nm/x beraam word.

DIE VANG-HERVANGTEGNIK

Polya se metode is eintlik 'n toepassing van die vang-hervangtegniek wat veral in ekologiese navorsing gebruik word om die grootte van populasies, soos die getal visse in 'n dam, te beraam. Eers neem mens 'n steekproef van grootte n ('n arbitrêre getal visse). Dié word gemerk en in die dam teruggeplaas. Nadat voldoende tyd toegelaat is vir vermenging, word 'n tweede steekproef van grootte m geneem en die getal gemerkte visse in die steekproef, x , getel. As die tweede steekproef ewekansig is, is die proporsie gemerkte visse in hierdie steekproef, x/m , 'n beraaming van die proporsie gemerkte visse in die dam, n/N . Gevolglik kan nm/x as 'n beraaming van N gebruik word.

Sulke beraamings word dikwels Petersonberamings genoem, na aanleiding van C.G.J. Peterson se beskrywing van die metode in 1893. Die tegniek is in besonderhede deur Seber² behandel. Hy verskaf verwysings na 'n uitgebreide literatuur, asook 'n deeglike statistiese bespreking van alle variasies in die metode. Indien die populasie nie verander tussen die eerste en die tweede steekproefnemings nie en indien die tweede steekproef ewekansig is, is die voorwaardelike verdeling van x , vir gegewe waardes van n en m , 'n hipergeometriese verdeling. Verder is die mak-

simumaanneemlikheidsberaming van die populasiegrootte gelyk aan nm/x en kan die eienskappe van die hipergeometriese verdeling gebruik word om die betroubaarheid van hierdie beraaming te bepaal.

'N NUWE METODE

Die vang-hervangtegniek kan nie toegepas word op die proefleesprobleem wanneer twee proeflesers agtereenvolgens dieselfde kopie van die teks lees nie, aangesien die tweede leser dan bewus is van die foute wat reeds opgespoor is. Ek wil egter aantoon hoe Polya se redenering uitgebrei kan word om ook hierdie situasie te dek. Die doel is om 'n beraaming te verkry van die totale getal foute wat aanwesig is (N), sowel as beraamings van die waarskynlikhede van die proeflesers om foute te vind (p en q). Dit kan bereik word deur die teks in twee dele, I en II, te verdeel. Deel I word eers deur leser 1 gelees, en hy bespeur 'n proporsie p van die getal foute (N_1) wat aanwesig is, en daarna deur leser 2, wat 'n proporsie q van die foute wat oorbly, bespeur. Deel II (met N_2 foute) word eers deur leser 2 gelees en daarna deur leser 1. Daar moet aanvaar word dat elke leser sy waarskynlikheid om foute te vind van die begin tot die einde van die taak behou. Die getalle foute wat gevind word, kan nou soos volg voorgestel word:

Deel	Leser	Volgorde	Foute gevind	=	Beraaming van
I	1	eerste	A	=	pN_1
I	2	tweede	a	=	$q(1-p)N_1$
II	2	eerste	B	=	qN_2
II	1	tweede	b	=	$p(1-q)N_2$

Hierdie vier vergelykings, wat vier onbekendes bevat, lewer die volgende oplossings:

$$\begin{aligned}
 p &= (AB - ab)/B(A + a) \\
 q &= (AB - ab)/A(B + b) \\
 N_1 &= AB(A + a)/(AB - ab) \\
 N_2 &= AB(B + b)/(AB - ab)
 \end{aligned}$$

Hieruit kan weer verdere inligting verkry word. So is die beraamde totale getal foute $N = N_1 + N_2 = AB(A + a + B + b)/(AB - ab)$. Verder is die totale proporsie foute wat (deur beide lesers saam) gevind is, gelyk aan $(AB - ab)/AB$.

Die nutswaarde van hierdie metode sal grotendeels afhang van hoe houdbaar die aanname oor ewekansige steekproefneming is in die geval van foutbespeuring tydens proeflees. Dit is duidelik 'n saak wat deur empiriese ondersoek opgeklar moet word. Selfs al sou sommige foute meer bespeurbaar wees as ander, kan mens die metode nog steeds gebruik deur dit

apart op verskillende tipes foute toe te pas in gevalle waar die steekproewe groot genoeg is.

VERWYSINGS

1. Polya, G. (1976). Probabilities in proofreading, *Am. Math. Monthly*, 83 (1), 42.
2. Seber, G.A.F. (1982). *The estimation of animal abundance and related parameters* (Griffin, London).

C. Plug

Universiteit van Suid-Afrika, Posbus 392, Pretoria 0001