MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

S 160 (M 54)

Schattingsmethoden voor overlevingskansen.

I. Schattingen op grond van drie opeenvolgende waarnemingsreeksen.

door

Gerda Klerk-Grobben.

1954

Inleiding.[1]

Onder de overlevingskans, θ, van een dier wordt in dit memorandum verstaan de kans dat een dier, dat op een bepaald moment in leven is, ook een jaar later nog in leven zal zijn.

De schattingen voor θ, die in dit memorandum gegeven worden berusten op waarnemingsmateriaal dat als volgt is opgebouwd.

Ieder jaar worden in eenzelfde gebied (b.v. eenzelfde complex grotten als wij populaties vleermuizen bekijken) een aantal dieren gevangen, gemerkt en weer vrijgelaten. Het aantal gevangen dieren en de jaren waarin een dier reeds eerder gevangen werd worden genoteerd. Na een aantal jaren bestaan dus de waarnemingen uit aantallen dieren die één, twee, enz. malen werden gevangen.

Voor deze aantallen zullen wij de volgende notatie gebruiken:

$z_{i_1, i_2, \cdots, i_k} \overset{\text{def}}{=}$ het aantal dieren, dat tot en met jaar i_k gevangen werd in de jaren $i_1, i_2, \cdots, i_k$

Hierbij zullen wij de opeenvolgende jaren, waarin vangsten werden verricht nummeren van 0 af; de rij $(i_1, i_2, \cdots, i_k)$ is dus een combinatie uit de nummers $0, 1, 2, \ldots, i_k$ met $i_1 < i_2 < \cdots < i_k$.

Bij voorbeeld $z_{1,5,7}$ stelt het aantal dieren voor van de in jaar 7 gevangen dieren die ook reeds in de jaren 1 en 5 gevangen werden en in geen van de andere jaren.

Verder is:

$u_i \overset{\text{def}}{=}$ het aantal dieren dat in jaar i voor het eerst gevangen werd.

$R_i \overset{\text{def}}{=}$ het totale aantal in jaar i gevangen dieren.

$s_i \overset{\text{def}}{=}$ het totale aantal in jaar i gevangen dieren, die ook reeds in enig vroeger jaar gevangen werden.

In dit memorandum zullen wij ons beperken tot schattingen gebaseerd op waarnemingen uit drie opeenvolgende jaren; in een volgend memorandum zullen wij twee van deze schattingsmethoden (nl. die van P.H. LESLIE en D. CHITTY) ook voor een groter aantal jaren uitwerken.

Voor drie opeenvolgende jaren geeft schema I een overzicht van het waarneminsmateriaal.

---

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

Schema I. Specificatie van de aantallen gevangen dieren in drie opeenvolgende jaren.

jaar	0	1	2
aantal eerste vangsten	u_0	u_1	u_2
aantal herhaalde vangsten		z_{01}	z_{02} z_{012} z_{12}
totale aantal herhaalde vangsten		$s_1 = z_{01}$	$s_2 = z_{02} + z_{012} + z_{12}$
totale vangst	$R_0 = u_0$	$R_1 = u_1 + s_1$	$R_2 = u_2 + s_2$

Onderstellingen.

Bij het opstellen van schattingsmethoden voor de overlevingskans van de dieren, kunnen wij nu van verschillende onderstellingen over de populatie van de dieren en de vangmethode uitgaan. In dit memorandum is gebruik gemaakt van de volgende onderstellingen:

1) de overlevingskans, θ, is ieder jaar dezelfde en voor ieder dier gelijk,

2) er verdwijnen geen vleermuizen door emigratie,

en

òf 3a) de vangkans is ieder jaar dezelfde en voor ieder dier gelijk (g),

òf 3b) de vangkans is per jaar voor ieder dier gelijk (g_i), maar mag van jaar tot jaar variëren.

De onderstellingen 1) en 2) worden steeds gebruikt, terwijl bovendien òf van onderstelling 3a) òf van 3b) wordt uitgegaan. Wij merken op, dat bij de schattingsmethoden in dit memorandum niets ondersteld wordt over de grootte van de populatie en dat nieuwe dieren (door geboorte of immigratie) aan de populatie toegevoegd mogen worden.

Op grond van de onderstellingen 1) en 2) kunnen wij in elk jaar de verwachting van de aantallen dieren met verschillende merkencombinaties in de populatie aangeven, uitgaande van de waarnemingen in vorige jaren. Wij gebruiken hierbij de volgende notatie:

$p_{i_1,\dots,i_\ell;\,i_k} \overset{\text{def}}{=}$ de verwachting van het aantal dieren in de populatie in het jaar i_k, die in de jaren $i_1, i_2, \dots, i_\ell$ werden gevangen.

Zo is bij voorbeeld $p_{0;1}$ de verwachting van het aantal dieren, dat in jaar 1 nog in de populatie aanwezig is van de dieren die in jaar 0 werden gevangen. Nu werden in jaar 0 u_0 dieren gevangen; de overlevingskans is θ , dus:

$$p_{0;1} = \theta u_0 .$$

Voor de jaren 0,1 en 2 geeft Schema II een overzicht van deze verwachtingen. Nogmaals dit zijn dus verwachtingen van aantallen in de populatie en <u>niet</u> bij de vangsten.

<u>Schema II.</u> Verwachtingen van de aantallen één of meer keren gevangen dieren in de populatie in drie opeenvolgende jaren.

jaar	0	1	2
verwachtingen		$p_{0;1} = \theta u_0$	$p_{0,1;2} = \theta t_{01}$ $p_{0;2} = \theta^2 u_0 - \theta t_{01}$ $p_{1;2} = \theta u_1$
totaal		θu_0	$\theta^2 u_0 + \theta u_1$

In de regel "totaal" staat de verwachting van het totale aantal reeds eerder gevangen dieren in de populatie in het betreffende jaar.

Is in jaar i de vangkans g_i dan is voor de in dit jaar te vangen aantallen dieren, $t_{i_1,\dots,i_\ell;i}$, de verwachting:

$$\mathcal{E}\, \underline{t}_{i_1,\dots,i_\ell,i} = g_i\, p_{i_1,\dots,i_\ell;i} \qquad 2)$$

<u>Verschillende schattingen voor θ.</u>

Van de nu volgende schattingen voor θ zijn de eerste twee afkomstig van P.H. LESLIE en D. CHITTY (zie het artikel genoemd aan het einde van dit memorandum). Zij onderstellen dat, ten tijde i , inderdaad de aantallen $p_{i_1,\dots,i_\ell;i}$ in de populatie aanwezig zijn en laten dus mogelijke fluctuaties in deze aantallen buiten beschouwing.

2)Door onderstreping van het symbool wordt aangegeven dat de grootheid een waarschijnlijkheidsverdeling bezit (een stochastische grootheid is). De verwachting of het theoretische gemiddelde van een stochastische grootheid $\underline{x}$ wordt met het symbool $\mathcal{E}\,\underline{x}$ aangegeven. Een door $\underline{x}$ aangenomen waarde geven wij aan met x , dus zonder onderstreping.

1) Om de eerste schatting voor θ af te leiden, beschouwen wij de volgende verwachtingen:

$$\begin{array}{l} \mathcal{E}(\underline{z}_{012}+\underline{z}_{12}) = g_2\,\theta(z_{01}+u_1) \\ \mathcal{E}(\underline{z}_{012}+\underline{z}_{02}) = g_2\,\theta^2 u_0 \\ \hline \mathcal{E}(2\underline{z}_{012}+\underline{z}_{02}+\underline{z}_{12}) = g_2\,(\theta^2 u_0+\theta z_{01}+\theta u_1) \end{array} +$$

Deze verwachtingen zijn dus bepaald met behulp van de in de jaren 0 en 1 gevangen aantallen dieren.

Onderstellen wij, dat in de populatie in jaar 2 inderdaad $\theta(z_{01}+u_1)$ dieren aanwezig zijn die in jaar 1 ook gevangen zijn en $\theta^2 u_0$ dieren die ook in jaar 0 gevangen zijn dan is de voorwaardelijke waarschijnlijkheidsverdeling van de grootheden $(\underline{z}_{012}+\underline{z}_{12})$ en $(\underline{z}_{012}+\underline{z}_{02})$, onder voorwaarde van het gevonden totaal, dus onder voorwaarde $(2\underline{z}_{012}+\underline{z}_{12}+\underline{z}_{02}) = (2z_{012}+z_{12}+z_{02})$ binomiaal met kansen

$$\frac{\theta(z_{01}+u_1)}{\theta^2 u_0+\theta z_{01}+\theta u_1} = \frac{z_{01}+u_1}{\theta u_0+z_{01}+u_1} \quad en \quad \frac{\theta^2 u_0}{\theta^2 u_0+\theta z_{01}+\theta u_1} = \frac{\theta u_0}{\theta u_0+z_{01}+u_1} \quad dus: \quad 3)$$

$$P = P\left[\underline{z}_{012}+\underline{z}_{12} = z_{012}+z_{12} \text{ en } \underline{z}_{012}+\underline{z}_{02} = z_{012}+z_{02} \,\middle|\, 2\underline{z}_{012}+\underline{z}_{02}+\underline{z}_{12} = 2z_{012}+z_{02}+z_{12}\right] =$$

$$\frac{(2z_{012}+z_{12}+z_{02})!}{(z_{012}+z_{12})!\,(z_{012}+z_{02})!}\left(\frac{z_{01}+u_1}{\theta u_0+z_{01}+u_1}\right)^{z_{012}+z_{12}}\left(\frac{\theta u_0}{\theta u_0+z_{01}+u_1}\right)^{z_{012}+z_{02}}.$$

Beschouwd als functie van θ is dus:

$$P = C\,\frac{\theta^{z_{012}+z_{02}}}{(\theta u_0+z_{01}+u_1)^{2z_{012}+z_{02}+z_{12}}}$$

De meest aannemelijke schatting ("maximum likelihood estimate") $\hat{\theta}$ voor θ is nu die waarde van θ welke deze kans P maximaal maakt; en waarvoor dus geldt:

$$\left[\frac{\partial P}{\partial \theta}\right]_{\theta=\hat{\theta}} = 0$$

of, wat op hetzelfde neerkomt:

$$\left[\frac{\partial \log P}{\partial \theta}\right]_{\theta=\hat{\theta}} = \frac{z_{012}+z_{02}}{\hat{\theta}} - \frac{(2z_{012}+z_{02}+z_{12})\,u_0}{\hat{\theta} u_0+z_{01}+u_1} = 0$$

----

3) $P[\underline{x}=x \text{ en } \underline{y}=y \,|\, \underline{x}+\underline{y}=x+y]$ is de voorwaardelijke waarschijnlijkheid dat de stochastische grootheid $\underline{x}$ de waarde x en de stochastische grootheid $\underline{y}$ de waarde y aanneemt, berekend onder de voorwaarde dat hun som $\underline{x}+\underline{y}$ een gegeven waarde $x+y$ bezit(zie ook voetnoot 2)).

dus:

$$\boxed{\hat{\theta}_1 = \frac{z_{01} + u_1}{u_0} \cdot \frac{z_{012} + z_{02}}{z_{012} + z_{12}}} \qquad (1)$$

2) Deze methode van LESLIE en CHITTY is eveneens toe te passen op de verwachtingen:

$$\begin{aligned}
\mathcal{E}\, \underline{z}_{012} &= g_2\, \theta\, z_{01} \\
\mathcal{E}\, \underline{z}_{02} &= g_2 (\theta^2 u_0 - \theta z_{01}) \\
\mathcal{E}\, \underline{z}_{12} &= g_2\, \theta\, u_1 \\
\hline
\mathcal{E}\, \underline{s}_2 &= g_2 (\theta^2 u_0 + \theta u_1).
\end{aligned} \quad +$$

Onder voorwaarde van het gevonden totaal $\underline{s}_2 = s_2$ ontstaat hier een multinomiale verdeling met kansen:

$$\frac{\theta z_{01}}{\theta^2 u_0 + \theta u_1} = \frac{z_{01}}{\theta u_0 + u_1} \; ; \quad \frac{\theta u_0 - z_{01}}{\theta u_0 + u_1} \quad \text{en} \quad \frac{u_1}{\theta u_0 + u_1}.$$

De meest aannemelijke schatting voor   wordt in dit geval:

$$\boxed{\hat{\theta}_2 = \frac{z_{02} \quad u_1 + z_{01} + s_2}{u_0 (z_{012} + z_{12})}} \; . \qquad (2)$$

Bij de volgende schattingen onderstellen wij dat de vangkans ieder jaar gelijk is, dus: $g_0 = g_1 = g_2 = g$.

3) Gaan wij nu uit van de verwachtingen:

$$\mathcal{E}(\underline{z}_{01} + \underline{z}_{012} + \underline{z}_{12}) = g\,\theta (R_0 + R_1) \quad \text{en}$$

$$\mathcal{E}(\underline{z}_{012} + \underline{z}_{02}) = g\,\theta^2 R_0$$

welke uit de schema's I en II kunnen worden afgelezen, dan volgt hieruit, door deling en door substitutie van de gevonden aantallen voor hun verwachtingen, als schatting voor θ :

$$\boxed{\hat{\theta}_3 = \frac{R_0 + R_1}{R_0} \cdot \frac{z_{02} + z_{012}}{z_{01} + z_{12} + z_{012}}} \; . \qquad (3)$$

4) Op dezelfde manier vinden wij, uitgaande van

$$\mathcal{E}(\underline{z}_{12} + \underline{z}_{01}) = g\,\theta (u_1 + u_0) \quad \text{en}$$

$$\mathcal{E}(\underline{z}_{012} + \underline{z}_{02}) = g\,\theta^2 u_0 \, ,$$

als schatting voor θ:

$$\boxed{\hat{\theta}_4 = \frac{u_0 + u_1}{u_0} \cdot \frac{z_{02} + z_{012}}{z_{01} + z_{12}}} \; . \qquad (4)$$

De schatting $\hat{\theta}_3$ werd voorgesteld door Ir J. BEZEM; schatting $\hat{\theta}_4$ door A. BENARD.

Over de vergelijking van de verschillende schattingen.

Over de waarde van de verschillende schattingen is niet veel bekend (zuiverheid, spreiding); deze zal onderzocht worden aan een steekproefexperiment. LESLIE en CHITTY verkiezen schatting $\hat{\theta}_2$ boven $\hat{\theta}_1$ omdat voor $\hat{\theta}_2$ een gedetailleerder gebruik van de waarnemingen is gemaakt (dus "meer informatie" is gebruikt). Voor de schattingen $\hat{\theta}_3$ en $\hat{\theta}_4$ is van een zwaardere onderstelling over de vangkans dan bij $\hat{\theta}_1$ en $\hat{\theta}_2$ uitgegaan.

Waarnemingen uit meer dan drie jaren.

Voor de methode van LESLIE en CHITTY zullen in een volgend memorandum schattingen voor θ gebaseerd op meerdere jaren worden gegeven, terwijl in een derde memorandum een schattingsmethode voor meerdere jaren gebaseerd op regressie zal worden uitgewerkt.

Overigens kan voor meer dan drie jaren als schatting voor θ het gemiddelde genomen worden van de schattingen berekend uit telkens drie opeenvolgende jaren.

Literatuur:

P.H. LESLIE en D.CHITTY, The estimation of population parameters from data obtained by means of the capture-recapture method. I Maximum likelihood equations for estimating the death-rate. Biometrika, 38 (1951), p. 269-292.