Demographische Modelle (Teil 2)

SoSe 2011

LS Sozialwissenschaftliche Methodenlehre und Sozialstatistik

C. Dudel

Bisher...

Leslie-Modell

- $\mathbf{n}_{t+1} = \mathbf{A}\mathbf{n}_t$
- Deterministisches Modell
- Diverse Annahmen (keine Migration etc.)
- Bei konstanten Raten: stabile Bevölkerung
- Stabile Bevölkerung und empirische Bevölkerung?
- Echte Bevölkerungsvorausberechnungen besser?

Worum geht's?

Stochastisches Bevölkerungsmodell, anknüpfend an relativ alte Fragestellung:

$\rightarrow$ Galton-Watson-Prozess
$\rightarrow$ Verzweigungsprozess (engl.: branching process)

Fragestellung

Galton & Watson (1874):
„A large nation, of whom we will concern ourselves with the adult males, N in number, and who each bear separate surnames, collonise a district. Their law of population is such that, in each generation, a_0 per cent of the adult males have no male children who reach adult life; a_1 have one such male child; a_2 have two; and so on up to a_5 who have five. Find (1) what proportion of the surnames will have become extinct after r generations; and (2) how many instances there will be of the same surname being held by m persons. “

Fragestellung

Also:

- „Vererbung von Namen“: Aussterbewahrscheinlichkeit
- Betrachtet wird eine Bevölkerung prinzipiell beliebiger Größe
- Die Wahrscheinlichkeit, dass ein Mitglied dieser Bevölkerung genau k Nachfahren zeugt, a_k, ist bekannt
- Beschränkung auf maximal 5 Nachfahren kann beliebig verändert werden
- Das betrachtete Geschlecht ist unerheblich (aber immer nur ein Geschlecht)
- Auch Eingrenzung auf Vererbung von Namen ist unerheblich

Fragestellung

Also:

- „Vererbung von Namen“: Aussterbewahrscheinlichkeit
- Betrachtet wird eine Bevölkerung prinzipiell beliebiger Größe
- Die Wahrscheinlichkeit, dass ein Mitglied dieser Bevölkerung genau k Nachfahren zeugt, a_k, ist bekannt
- Beschränkung auf maximal 5 Nachfahren kann beliebig verändert werden
- Das betrachtete Geschlecht ist unerheblich (aber immer nur ein Geschlecht)
- Auch Eingrenzung auf Vererbung von Namen ist unerheblich

Ergänzende Überlegungen

Eine Vereinfachung lässt sich dadurch erzielen, dass man „nicht überlappende“ Generationen betrachtet

Ergänzende Überlegungen

- Man beginnt mit einer Generation n_0
- Die Mitglieder dieser Generation bekommen entsprechend der Wahrscheinlichkeiten a_k Nachfahren
- Diese bilden die Generation n_1
- Die Zahl der Mitglieder der Generation n_0 ist nun nicht weiter interessant – man muss sich bspw. nicht darum kümmern, wie lange die Mitglieder von n_0 noch leben
- n_0 kann also für weitere Berechnungen vernachlässigt werden
- Nun wird n_1 in Kombination mit den Wahrscheinlichkeiten a_k verwendet, um n_2 zu berechnen

Ergänzende Überlegungen

- Man beginnt mit einer Generation n_0
- Die Mitglieder dieser Generation bekommen entsprechend der Wahrscheinlichkeiten a_k Nachfahren
- Diese bilden die Generation n_1
- Die Zahl der Mitglieder der Generation n_0 ist nun nicht weiter interessant – man muss sich bspw. nicht darum kümmern, wie lange die Mitglieder von n_0 noch leben
- n_0 kann also für weitere Berechnungen vernachlässigt werden
- Nun wird n_1 in Kombination mit den Wahrscheinlichkeiten a_k verwendet, um n_2 zu berechnen

Ergänzende Überlegungen

- Man beginnt mit einer Generation n_0
- Die Mitglieder dieser Generation bekommen entsprechend der Wahrscheinlichkeiten a_k Nachfahren
- Diese bilden die Generation n_1
- Die Zahl der Mitglieder der Generation n_0 ist nun nicht weiter interessant – man muss sich bspw. nicht darum kümmern, wie lange die Mitglieder von n_0 noch leben
- n_0 kann also für weitere Berechnungen vernachlässigt werden
- Nun wird n_1 in Kombination mit den Wahrscheinlichkeiten a_k verwendet, um n_2 zu berechnen

Ergänzende Überlegungen

- Man beginnt mit einer Generation n_0
- Die Mitglieder dieser Generation bekommen entsprechend der Wahrscheinlichkeiten a_k Nachfahren
- Diese bilden die Generation n_1
- Die Zahl der Mitglieder der Generation n_0 ist nun nicht weiter interessant – man muss sich bspw. nicht darum kümmern, wie lange die Mitglieder von n_0 noch leben
- n_0 kann also für weitere Berechnungen vernachlässigt werden
- Nun wird n_1 in Kombination mit den Wahrscheinlichkeiten a_k verwendet, um n_2 zu berechnen

Ergänzende Überlegungen

- Man beginnt mit einer Generation n_0
- Die Mitglieder dieser Generation bekommen entsprechend der Wahrscheinlichkeiten a_k Nachfahren
- Diese bilden die Generation n_1
- Die Zahl der Mitglieder der Generation n_0 ist nun nicht weiter interessant – man muss sich bspw. nicht darum kümmern, wie lange die Mitglieder von n_0 noch leben
- n_0 kann also für weitere Berechnungen vernachlässigt werden
- Nun wird n_1 in Kombination mit den Wahrscheinlichkeiten a_k verwendet, um n_2 zu berechnen

Ergänzende Überlegungen

- Man beginnt mit einer Generation n_0
- Die Mitglieder dieser Generation bekommen entsprechend der Wahrscheinlichkeiten a_k Nachfahren
- Diese bilden die Generation n_1
- Die Zahl der Mitglieder der Generation n_0 ist nun nicht weiter interessant – man muss sich bspw. nicht darum kümmern, wie lange die Mitglieder von n_0 noch leben
- n_0 kann also für weitere Berechnungen vernachlässigt werden
- Nun wird n_1 in Kombination mit den Wahrscheinlichkeiten a_k verwendet, um n_2 zu berechnen

Ergänzende Überlegungen

$\rightarrow$ Bevölkerung hat keine Altersstruktur, nur Größe

Einschränkungen

Wir gehen davon aus, dass sich weder die Generationen noch die Mitglieder einer jeden Generation unterscheiden: Für alle gelten die Wahrscheinlichkeiten a_k

Überlegungen für die Simulation

- n_0 kann im Prinzip beliebig gewählt werden (bspw. 1 oder 100)
- Gegeben der Wahrscheinlichkeiten a_k wird für jedes Mitglied der gerade betrachteten Generation ein zufälliger Wert aus dieser Nachfahrenverteilung gezogen
- Die Summe dieser zufälligen Werte ist dann die Größe der nächsten Generation

Überlegungen für die Simulation

Wie bestimmt man die Wahrscheinlichkeiten a_k? Einfache Möglichkeit: eine bekannte, diskrete Zufallsverteilung mit den natürlichen Zahlen als Wertebereich

Poisson-Verteilung I

Für eine Poisson-verteilte Zufallsvariable K gilt:

$$P(K = k) = \frac{\lambda^k}{k!} exp(-\lambda) \qquad \text{für } k \in \mathbb{N}_0 \text{ und } \lambda \in \mathbb{R}_{>0}$$

$$\begin{aligned} E(K) &= \lambda \\ Var(K) &= \lambda \end{aligned}$$

RUB

Poisson-Verteilung II: Beispiel

$$P(K = k) = \frac{\lambda^k}{k!} exp(-\lambda)$$

$\lambda = 1$ und $k = 2$

$$\begin{aligned} P(K = 2) &= \frac{1^2}{2!} exp(-1) \\ &= 0.184 \end{aligned}$$

Poisson-Verteilung II: Beispiel

Beispiel $\lambda = 1$ ($k > 5$ weggelassen):

k	0	1	2	3	4	5
$P(K = k)$	0.368	0.368	0.184	0.061	0.015	0.003

Poisson-Verteilung III

Wichtig: Verteilung hängt nur von λ ab

Poisson-Verteilung IV: Reproduktivität

Seien $X_1, X_2, \ldots, X_n$ Poisson-verteilte Zufallsvariablen mit Parametern $\lambda_1, \lambda_2, \ldots, \lambda_n$

Dann ist $\sum_{i=1}^{n} X_i \sim Pois(\sum_{i=1}^{n} \lambda_i)$

Wenn $\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$ dann $\sum_{i=1}^{n} X_i \sim Pois(n\lambda)$

Poisson-Verteilung IV: Reproduktivität

Seien $X_1, X_2, \ldots, X_n$ Poisson-verteilte Zufallsvariablen mit Parametern $\lambda_1, \lambda_2, \ldots, \lambda_n$

Dann ist $\sum_{i=1}^{n} X_i \sim Pois(\sum_{i=1}^{n} \lambda_i)$

Wenn $\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$ dann $\sum_{i=1}^{n} X_i \sim Pois(n\lambda)$

Poisson-Verteilung IV: Reproduktivität

Seien $X_1, X_2, \ldots, X_n$ Poisson-verteilte Zufallsvariablen mit Parametern $\lambda_1, \lambda_2, \ldots, \lambda_n$

Dann ist $\sum_{i=1}^{n} X_i \sim Pois(\sum_{i=1}^{n} \lambda_i)$

Wenn $\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$ dann $\sum_{i=1}^{n} X_i \sim Pois(n\lambda)$

Was machen wir jetzt?

Bisher:

- Der zu beschreibende Prozess ist angegeben
- Ebenso, wie sich dieser Prozess entwickelt

Nun folgt:

- Parameter festlegen
- Modell programmieren
- Berechnen & auswerten

Parameter

Wie verhält sich das Modell in Abhängigkeit von λ?

Werte nahe 1

Kurze Einordnung

- Modell an sich Mikro-Ebene, aber für Simulation relativ egal (Reproduktivität)
- Einfache Mikrosimulation? Numerische Simulation? Grenzfall!
- Analytische Lösung der Ausgangsfragestellung möglich
- Stochastisch!

Stochastisch!

Hat Konsequenzen für die Simulation! Ergebnisse eines Durchlaufs sind nun „zufällig“! Um einen Eindruck des Verhaltens des Modells zu bekommen, müssen etliche Durchläufe erzeugt werden!

Vergleich mit Leslie-Modell

Verzweigungsprozesse und Leslie-Modell sehr unterschiedliche Annahmen und Ergebnisse

Vergleich mit Leslie-Modell

Einerseits:

- Wenn $\lambda = 1$ stirbt Bevölkerung im Verzweigungsprozess aus, im Leslie-Modell nicht!
- Auch wenn $\lambda > 1$ aussterbende Bevölkerung beim VP, bei LM hingegen stetiges Wachstum!

Andererseits:
Theoretische Verknüpfung möglich;
Multityp-Galton-Watson-Prozess hat Ergebnisse des Leslie-Modells als Erwartungswert!

Vergleich mit Leslie-Modell

Einerseits:

- Wenn $\lambda = 1$ stirbt Bevölkerung im Verzweigungsprozess aus, im Leslie-Modell nicht!
- Auch wenn $\lambda > 1$ aussterbende Bevölkerung beim VP, bei LM hingegen stetiges Wachstum!

Andererseits:
Theoretische Verknüpfung möglich;
Multityp-Galton-Watson-Prozess hat Ergebnisse des Leslie-Modells als Erwartungswert!

Vergleich mit Leslie-Modell

Einerseits:

- Wenn $\lambda = 1$ stirbt Bevölkerung im Verzweigungsprozess aus, im Leslie-Modell nicht!
- Auch wenn $\lambda > 1$ aussterbende Bevölkerung beim VP, bei LM hingegen stetiges Wachstum!

Andererseits:
Theoretische Verknüpfung möglich;
Multityp-Galton-Watson-Prozess hat Ergebnisse des Leslie-Modells als Erwartungswert!

Fazit (Demographische Modelle)

- Sehr stark vereinfachende Annahmen unrealistisch, aber durchaus nützlich (stabile Bevölkerung)
- Auch bei realistischeren Modellen bleiben nicht unerhebliche Probleme wie z.B. die Wahl von Parametern (LM)
- Modellverhalten kann stark von Annahmen abhängen (Vergleich VP und LM)

Anwendung Verzweigungsprozesse

Verzweigungsprozesse sehr flexibel, da prinzipiell beliebige Verteilungen für a_k

- Erklärung der Verteilung von Nachnamen?
- Aussterbewahrscheinlichkeiten (Ökologie)
- Ausbreitung von Epidemien (Medizin)
- Verbreitung von Mutationen (Genetik)
- Gemeinsame Vorfahren (Genetik/Demographie)
- Verwandtschaftsstrukturen (Demographie)
- Kettenbriefe