Praktická aplikace Galtonova – Watsonova procesu větvení v demografii

Výchozí data
Budeme vyšetřovat sled generací ženské populace Československa. Máme k dispozici údaje z roku 1961, které popisují rozdělení žen ve věkovém intervalu (45 let, 50 let) (tedy na konci reprodukčního období) podle počtu živě narozených dětí.

Tabulka 1 - výchozí data

počet dětí	označení	počet žen
0	c0	65387
1	c1	78901
2	c2	136150
3	c3	79878
4	c4	39387
5	c5	19856
6	c6	15365
7	c7	7683
8	c8	3841
9	c9	1921
10	c10	960
11	c11	480
12	c12	240
13	c13	120
14	c14	60
15	c15	30

Upozornění: skupiny c_6 až c_{15} byly vytvořeny uměle. K dispozici totiž byly jen údaje o skupinách $c_{6\text{-}10}$ (24604 žen), $c_{11,\dots}$ (2041 žen) a nezařazeno (4055 žen). Počty žen v těchto uměle vytvořených skupinách geometricky klesají tak, že počet žen v každé následující skupině je polovinou počtu žen v předešlé skupině. Taková aproximace je přípustná, protože se týká jenom 30700 žen, což je zanedbatelné vzhledem k celkovému počtu 450259 žen ve věkové skupině (45 let, 50 let).

Stanovení pravděpodobnosti narození dcery, která se dožije reprodukčního věku 25 let
Zajímají nás pouze potomci ženského pohlaví. Je známo, že v r. 1961 byl poměr počtu živě narozených chlapců k počtu živě narozených dívek 1,055 (tzv. ukazatel maskulinity). Tedy pravděpodobnost narození dívky je $\frac{1}{1+1{,}055}$. Dále je známo z úmrtnostních tabulek ČSR pro rok 1961, že pravděpodobnost dožití 25 let pro ženu je 0,96788. Tedy hodnotu $h = 0{,}96788 \cdot \frac{1}{1+1{,}055} = 0{,}470988$ lze považovat za pravděpodobnost, že živě narozený potomek je dcera, která se dožije věku 25 let.

Rozložení počtu žen podle počtu dcer, které se dožijí reprodukčního věku 25 let
Nechť c_i je počet žen s i potomky. Z vlastností binomického rozložení plyne, že z počtu c_i připadá $(1\text{-}h)^i c_i$ na ženy s žádnou 25 letou dcerou, $ih(1\text{-}h)^{i-1}c^i$ na ženy s právě jednou 25 letou dcerou atd.

Tabulka 2 – rozdělení $p_0, p_1, \ldots$ žen podle počtu 0, 1, ... 25 letých dcer konstruované na základě počtů $c_0, c_1, \ldots$ žen s 0, 1, ... potomky.

počet potomků	počet žen	rozdělení žen podle počtu 25 letých dcer			
		0	1	2	…
0	c_0	c_0	0	0	…
1	c_1	$(1-h)c_1$	hc_1	0	…
2	c_2	$(1-h)^2c_2$	$2h(1-h)c_2$	h^2c_2	…
3	c_3	$(1-h)^3c_3$	$3h(1-h)^2c_3$	$3h^2(1-h)c_3$	…
4	c_4	$(1-h)^4c_4$	$4h(1-h)^3c_4$	$6h^2(1-h)^2c_4$	…
.	.	.	.	.	…
.	.	.	.	.	…
.	.	.	.	.	…
suma	c	p_0c	p_1c	p_2c	

Numerické vyhodnocení

počet potomků	počet žen	rozdělení žen podle počtu 25 letých dcer					
		0	1	2	3	4	…
0	65387	65387	0	0	0	0	…
1	78901	41740	37161	0	0	0	…
2	136150	38102	67846	30202	0	0	…
3	79878	11826	31586	28121	8645	0	…
4	39387	3045	10985	14671	8708	1938	0
.	.	.	.		.		…
.	.	.	.		.		…
.	.	.	.		.		…
suma	450259	161420	153833	85789	31330	11549	

$$p_0 = \frac{161259}{450259} = 0{,}3985047,\ p_1 = 0{,}3416544,\ p_2 = 0{,}1905325,\ p_3 = 0{,}0695821$$

Matice pravděpodobností přechodu

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ p_0 & p_1 & p_2 & \cdots \\ p_0^2 & 2p_0p_1 & p_1^2 + 2p_0p_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

. V našem případě tedy dostáváme:

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0{,}358504 & 0{,}341655 & 0{,}190533 & 0{,}069583 & \cdots \\ 0{,}128525 & 0{,}244969 & 0{,}253342 & 0{,}180085 & \cdots \\ 0{,}046077 & 0{,}131733 & 0{,}199067 & 0{,}206734 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

Pravděpodobnostní rozložení počtu 25 letých dcer v n-té generaci

Pro vektor absolutních pravděpodobností platí: $\mathbf{p}(n) = \mathbf{p}(0).\mathbf{P}^n$. V následující tabulce jsou uvedeny složky vektoru $\mathbf{p}(n)$ pro n = 0, 1, ..., 10

Tabulka 3 – pravděpodobnostní rozložení $p_i(n)$ počtu i 25 letých dcer v n-té generaci

n \ i	0	1	2	3	4	5	6	7	...
0	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	...
1	0,358504	0,341655	0,190533	0,065830	0,025651	0,009428	0,003191	0,001013	...
2	0,509170	0,174495	0,131214	0,078222	0,046294	0,026754	0,015134	0,008447	...
3	0,593159	0,105974	0,090446	0,064523	0,045799	0,031962	0,022014	0,015040	...
4	0,646756	0,071021	0,065064	0,051054	0,039920	0,030799	0,023537	0,017881	...
5	0,683843	0,050729	0,048627	0,040485	0,033619	0,027604	0,022493	0,018246	...
6	0,710930	0,037884	0,037477	0,032484	0,028103	0,024071	0,020485	0,017367	...
7	0,731494	0,029234	0,029590	0,026409	0,023533	0,020781	0,018246	0,015970	...
8	0,747563	0,023131	0,023824	0,021736	0,019805	0,017894	0,016085	0,014416	.
9	0,760402	0,018665	0,019489	0,180087	0,016768	0,015422	0,014117	0,012887	.
10	0,770845	0,015300	0,016151	0,015194	0,014281	0,013322	0,012371	0,011460	
.	.	.	.	.	.	.	.	.	.

V prvním sloupci jsou pravděpodobnosti vyhynutí (v ženské linii). Je vidět, že v 10. generaci je pravděpodobnost vyhynutí $p_0(10) = 0,770845$ hodně vysoká. Naproti tomu pravděpodobnosti libovolného nenulového počtu dcer jsou velice malé (např. pro jednu dceru $p_1(10) = 0,015300$ a s rostoucím počtem generací se stále snižují.

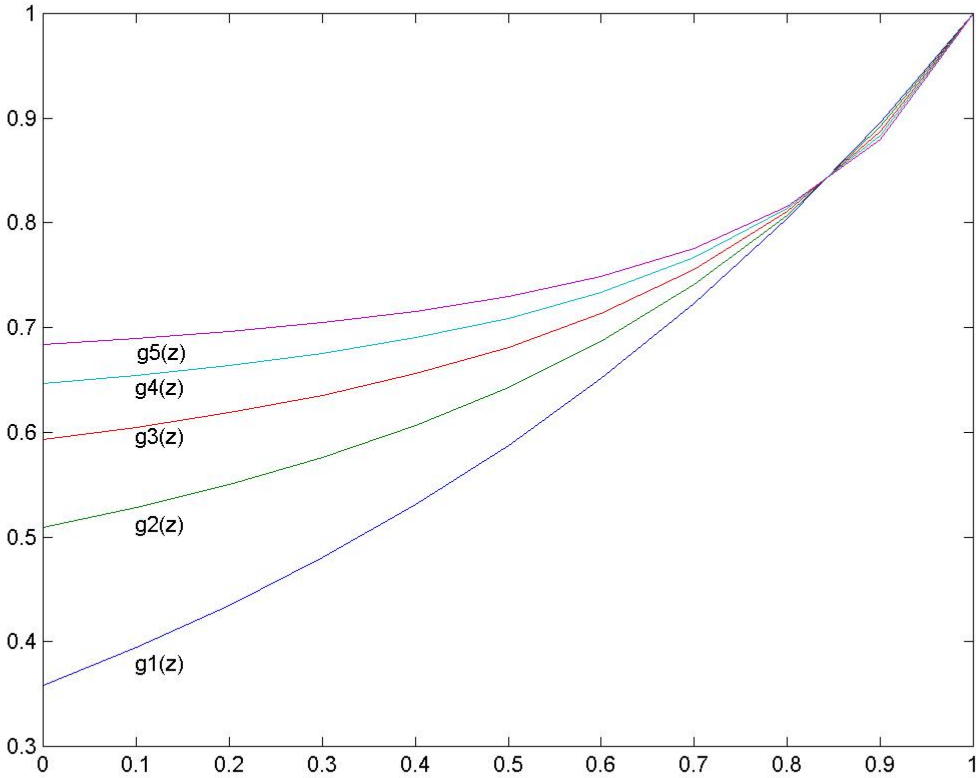
Vytvořující funkce počtu 25 letých dcer v n-té generaci

$$g_{X_n}(z) = \sum_{k=0}^{\infty} p_{nk} z^k$$

Tabulka 4 – hodnoty vytvořující funkce $g_n(z)$ počtu 25 letých dcer v n-té generaci

z	g1(z)	g2(z)	g3(z)	g4(z)	g5(z)
0,0	0,358504	0,509170	0,593159	0,646756	0,683843
0,1	0,394647	0,528015	0,604730	0,654564	0,689446
0,2	0,435057	0,550027	0,618573	0,664047	0,696322
0,3	0,480260	0,575893	0,635302	0,675717	0,704892
0,4	0,530872	0,606507	0,655773	0,690318	0,715785
0,5	0,587619	0,643056	0,681205	0,708964	0,729979
0,6	0,651360	0,687140	0,713398	0,733401	0,749073
0,7	0,723115	0,740967	0,755101	0,766513	0,775872
0,8	0,804109	0,807566	0,810729	0,813400	0,815726
0,9	0,895811	0,891733	0,887770	0,883881	0,880022
1,0	1,000000	1,000000	1,000000	1,000000	1,000000

Průběhy vytvořujících funkcí



Je vidět, že s postupem generací funkce $g_n(z)$ začínají být téměř konstantní pro $0 \le z < 1$ se skokem na hodnotu 1 v bodě z=1. To je ovšem ve shodě s teorií, neboť pro $\mu > 1$ je $\lim_{n\to\infty} g_n(z) = \begin{cases} \xi \text{ pro } 0 \le z < 1 \\ 1 \text{ pro } z = 1 \end{cases}$. V našem případě $\mu = p_1 + 2p_2 + 3p_3 + \ldots = 0{,}341655 + 2.0{,}190533 + 3.0{,}069583 + \ldots = 1{,}111166 > 1$.

Pravděpodobnost vyhynutí

Limitní hodnotu pravděpodobnosti vyhynutí lze získat jako nejmenší kladný kořen rovnice $z = g(z)$, kde $g(z) = g_1(z) = p_0 + p_1 z + p_2 z^2 + \ldots$
V našem případě řešíme rovnici $z = 0{,}358504 + 0{,}341655z + 0{,}190533z^2 + 0{,}069583z^3 + \ldots$
Výsledek: $\xi = 0{,}834043$.
Je zajímavé posoudit rychlost konvergence posloupnosti q_n pravděpodobnosti vyhynutí potomků v ženské linii v jednotlivých generacích k příslušné limitní hodnotě $\xi = 0{,}834043$ - viz 1. sloupec v tabulce 3.
Pro zajímavost ještě uvedeme limitní hodnoty pravděpodobnosti vyhynutí potomků v ženské linii pro různé země (údaje jsou z roku 1960).

země	ξ
Peru	0,2620
Japonsko	0,3242
Mexiko	0,4066
Maďarsko	0,7130
USA	0,8209