

n	p_n	n	p_n	n	p_n
1	1.000 000 000	6	0.631 944 444	11	0.632 120 561
2	0.500 000 000	7	0.632 142 857	12	0.632 120 559
3	0.666 666 667	8	0.632 118 056	13	0.632 120 559
4	0.625 000 000	9	0.632 120 811	14	0.632 120 559
5	0.633 333 333	10	0.632 120 536	15	0.632 120 559

Tab. 3.1: K příkladu 1 — problém šatnářky

3.1 Aplikace

3.1.1 Lincolnova-Petersenova metoda odhadu počtu volně žijících živočichů

V určité lokalitě žije neznámý počet živočichů, například kaprů v rybníce, zubrů v pralese, antilop na savaně ap. Nemůžeme je všechny najednou spočítat, snažíme se jejich počet nějak „rozumně“ odhadnout. Postupujeme přitom takto: ulovíme několik jedinců, nějak je označujeme a vypustíme. Počkáme, až se všichni označení promíchají s ostatními, znovu jich několik odchytíme a spočítáme mezi nimi označované. Čekání mezi dvěma odchty musí být dostatečně dlouhé, aby se jedinci stačili dostatečně promísit, nesmí být však příliš dlouhé, neboť pak by se — např. v důsledku úhynu, narození mladých, migrace — změnil jejich počet. Navíc označení musí být dostatečně trvanlivé, aby o ně živočich do dalšího odchytu nepřišel, ale nesmí ho nějak handicapovat, aby pravděpodobnosti odchytu označeného a neoznačeného jedince byly stejné.

Označme n neznámý počet živočichů, m počet označovaných, tj. ulovených při prvním odchytu, r počet ulovených při druhém odchytu a k počet označovaných mezi nimi. Čísla n , m , r a k musí samozřejmě splňovat nerovnosti $n \geq m \geq k$, $n \geq r \geq k$. Navíc všech neoznačovaných živočichů nemůže být méně, než kolik je neoznačovaných při druhém odchytu, tj. $n - m \geq r - k$,

$$n \geq m + r - k.$$

Odchyt příslušných živočichů lze považovat za náhodný pokus „výběr r předmětů z n daných“, nalezení k označovaných mezi r vybranými za náhodný jev. Základním prostorem výsledků jsou všechny r -prvkové podmnožiny z n prvkové množiny, pravděpodobnost budeme počítat jako klasickou. Počet r -prvkových podmnožin n prvkové je podle 2.1.3 roven $c(n, r)$. Symbolem $A_n^{m, r, k}$ označme jev „bylo označováno m živočichů a mezi r ulovenými při druhém odchytu je právě k označovaných“. Příznivými výsledky pokusu jsou tedy ty výběry, v nichž je právě k označovaných živočichů a $r - k$ neoznačovaných. Označované živočichy vybíráme z m označovaných, těchto výběrů je tedy $c(m, k)$; neoznačované vybíráme z $n - m$ možných, těchto výběrů je $c(n - m, r - k)$. Libovolný takový výběr označovaných lze skombinovat s libovolným výběrem neoznačovaných, takže příznivých

výsledků je $c(m, k) c(n - m, r - k)$ a

$$\begin{aligned} P(A_n^{m,r,k}) &= \frac{c(m, k) c(n - m, r - k)}{c(n, r)} = \frac{\binom{m}{k} \binom{n - m}{r - k}}{\binom{n}{r}} = \\ &= \frac{m! (n - m)! r! (n - r)!}{k! (m - k)! (r - k)! (n - m - r + k)! n!}. \end{aligned}$$

Podle Descartesovy zásady „... nemůžeme-li poznat nejpravdivější soudy, musíme se řídit soudy nejpravděpodobnějšími“, zformulované v jeho Rozpravě o metodě, budeme za počet živočichů na dané lokalitě považovat takové číslo n , pro které je pravděpodobnost $P(A_n^{m,r,k})$ při známých m, r, k největší.

Uvedený způsob odhadu velikosti populací pohyblivých organismů bývá nazýván *metoda opakovaného odchytu* (mark-recapture nebo mark-resight).

Příklad 2. Uvažujme situaci, že k odhadu počtu kaprů v jezírku bylo chyceno a označováno 10 kusů, při druhém výlovu bylo chyceno 12 kaprů a mezi nimi byli dva označovaní. Dovedete odhadnout maximálně pravděpodobný počet kaprů v jezírku ještě před provedením výpočtu?

V tomto případě je $m = 10$, $r = 12$ a $k = 2$. Počet kaprů v rybníce určitě není menší, než $10 + 12 - 2 = 20$. V tabulce 3.2 jsou uvedeny pravděpodobnosti $P(A_n^{10,12,2})$ pro $n = 20, 21, \dots, 99$. Vidíme, že stejnou největší pravděpodobnost mají jevy $A_n^{10,12,2}$ pro $n = 59$ a $n = 60$. Můžeme tedy uzavřít, že nejpravděpodobnější počet kaprů v jezírku je 59 nebo 60.

Výsledek 60 kusů není překvapivý. Bylo totiž možné očekávat, že relativní četnost označovaných kaprů mezi všemi v jezírku je zhruba stejná, jako relativní četnost označovaných v náhodném vzorku ulovených při druhém odchytu, tj.

$$\frac{m}{n} \approx \frac{k}{r}, \quad \text{neboli} \quad n \approx \frac{mr}{k}.$$

□

3.1.2 Hardyho-Weinbergův zákon

K abstraktnímu modelu dědičné informace diploidních organismů (sr. 2.4.1) přidejme abstraktní model reprodukce: dospělý jedinec vyprodukuje pohlavní buňku *gametu*, která je *haploidní*, tj. každý její lokus obsahuje pouze jeden gen. Spojením dvou gamet vznikne diploidní jedinec *zygot*. Uvažujme pro jednoduchost jeden gen A , který má dvě alely a_1 a a_2 . Dále budeme předpokládat:

- (i) Plodnost nezávisí na genotypu A , tj. geneticky rozdílní jedinci produkují stejný počet gamet.
- (ii) Alely přechází do gamety nezávisle na genotypu, náhodně a se stejnou pravděpodobností, tj. pravděpodobnost, že jedinec předá do gamety jednu svou alelu je $\frac{1}{2}$, pravděpodobnost, že předá druhou je rovněž $\frac{1}{2}$.

n	$P(A_n^{10,12,2})$	n	$P(A_n^{10,12,2})$	n	$P(A_n^{10,12,2})$	n	$P(A_n^{10,12,2})$
20	0.0004	40	0.2420	60	0.3303	80	0.2963
21	0.0017	41	0.2527	61	0.3301	81	0.2938
22	0.0046	42	0.2625	62	0.3296	82	0.2912
23	0.0095	43	0.2716	63	0.3288	83	0.2887
24	0.0167	44	0.2798	64	0.3279	84	0.2861
25	0.0260	45	0.2872	65	0.3268	85	0.2835
26	0.0373	46	0.2940	66	0.3255	86	0.2809
27	0.0503	47	0.3000	67	0.3240	87	0.2783
28	0.0647	48	0.3053	68	0.3224	88	0.2757
29	0.0801	49	0.3101	69	0.3207	89	0.2731
30	0.0961	50	0.3142	70	0.3189	90	0.2705
31	0.1125	51	0.3178	71	0.3170	91	0.2679
32	0.1289	52	0.3208	72	0.3149	92	0.2653
33	0.1451	53	0.3234	73	0.3128	93	0.2627
34	0.1609	54	0.3255	74	0.3106	94	0.2602
35	0.1763	55	0.3272	75	0.3084	95	0.2576
36	0.1910	56	0.3285	76	0.3060	96	0.2551
37	0.2049	57	0.3294	77	0.3037	97	0.2526
38	0.2181	58	0.3300	78	0.3013	98	0.2500
39	0.2305	59	0.3303	79	0.2988	99	0.2475

Tab. 3.2: K příkladu 2 — odhad počtu kaprů v jezírku

- (iii) Populace je *panmiktická*, což znamená, že křížení je náhodné, k náhodně vybrané gametě je náhodně a nezávisle přidána gameta druhá.

Nyní můžeme položit otázku: Známe-li relativní zastoupení genotypů v *parentální* (rodičovské) generaci, jaké bude její zastoupení v první *filiální* generaci (v generaci jejích potomků)?

Označme relativní četnosti genotypů a_1a_1 , a_1a_2 a a_2a_2 v parentální generaci symboly x , y a z (v tomto pořadí). Hodnotu x lze chápat jako pravděpodobnost, že náhodně vybraný jedinec bude genotypu a_1a_1 , analogicky i pro hodnoty y a z . Relativní četnosti x , y , z jsou samozřejmě nezáporná čísla menší než 1, která splňují rovnost

$$x + y + z = 1. \quad (3.1)$$

Dále označme relativní četnosti alel a_1 a a_2 v parentální generaci symboly p a q (v tomto pořadí). Hodnotu p lze opět chápat, jako pravděpodobnost jevu, že náhodně vybraná alela genu A bude a_1 , analogicky i hodnotu q . Relativní četnosti alel vyjádříme pomocí relativních četností genotypů. Skládá-li se celá populace z N diploidních jedinců, nachází se v ní $2N$ alel. Alela a_1 se vyskytuje dvakrát v každém homozygotním jedinci genotypu a_1a_1 a jednou v každém jedinci heterozygotního genotypu a_1a_2 . Jedinců genotypu a_1a_1 je xN , jedinců genotypu a_1a_2 je yN , takže alel a_1 je v celé populaci celkem $2xN + yN$. To znamená, že relativní četnost alely

a_1 v parentální generaci je

$$p = \frac{2xN + yN}{2N} = x + \frac{1}{2}y. \quad (3.2)$$

Analogicky lze odvodit

$$q = \frac{1}{2}y + z. \quad (3.3)$$

Z rovností (3.1), (3.2) a (3.3) plyne rovnost

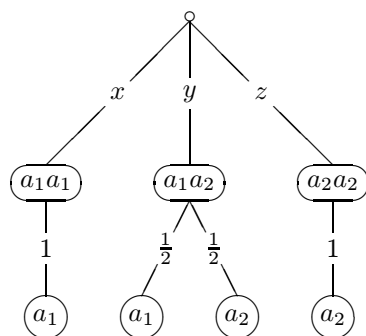
$$p + q = 1. \quad (3.4)$$

Budeme uvažovat jevy

$(a_i a_j)$... jedinec z parentální generace je genotypu $a_i a_j$,

(a_i) ... gameta obsahuje alelu a_i ;

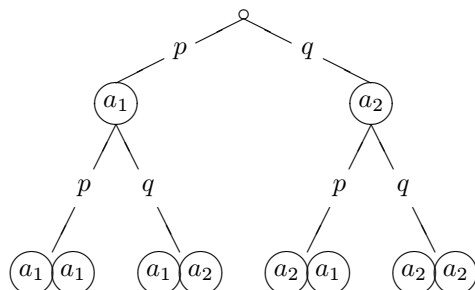
indexy i, j nabývají kterékoliv z hodnot 1, 2, j není menší než i . Proces reprodukce rozdělíme do dvou etap: „vytvoření gamety“ a „spojení gamet“. Etapu „vytvoření gamety“ rozdělíme na dvě podetapy: „výběr rodičovského jedince“ a „předání alely do gamety“. Podle předpokladu (ii) homozygotní rodič předá svou alelu do gamety jistě, heterozygotní rodič předá alelu a_1 s pravděpodobností $\frac{1}{2}$, alelu a_2 s touž pravděpodobností $\frac{1}{2}$. Toto předání nezávisí na genotypu, výběr alely je nezávislý na výběru rodičovského jedince. Ohodnocený strom průběhu a výsledků náhodného pokusu „vytvoření gamety“ tedy je takový, jak je znázorněno na obr. 3.1. Tento



Obr. 3.1: Strom průběhu a výsledků náhodného pokusu „vytvoření gamety“

strom je stejný pro kterýhokoliv rodičovského jedince. Z něho a z rovností (3.2), (3.3) vidíme, že

$$P\left(\left(a_1\right)\right) = x + \frac{1}{2}y = p, \quad P\left(\left(a_2\right)\right) = \frac{1}{2}y + z = q.$$



Obr. 3.2: Strom průběhu a výsledků náhodného pokusu „spojení gamet“

Etapu „spojení gamet“ také rozdělíme na dvě podetapy: „výběr první gamety“ a „připojení druhé gamety“. S využitím předchozího výsledku dostaneme ohodnocený strom průběhu a výsledků náhodného pokusu „spojení gamet“ tak, jak je znázorněn na obr. 3.2

Označme nyní relativní četnosti genotypů a_1a_1 , a_1a_2 a a_2a_2 v první filiální generaci symboly x_1 , y_1 a z_1 (v tomto pořadí) a relativní četnost alel a_1 a a_2 symboly p a q (v tomto pořadí). Hodnotu x_1 lze chápat jako pravděpodobnost jevu, že náhodně vybraný jedinec z první filiální generace je genotypu a_1a_1 a podobně; hodnotu p_1 jako pravděpodobnost, že náhodně vybraná alela z první filiální generace je a_1 , podobně pro q . Tedy

$$\begin{aligned} x_1 &= P\left(\begin{array}{c} \textcircled{a_1} \\ \textcircled{a_1} \end{array}\right) = p^2, \\ y_1 &= P\left(\begin{array}{c} \textcircled{a_1} \\ \textcircled{a_2} \end{array}\right) + P\left(\begin{array}{c} \textcircled{a_2} \\ \textcircled{a_1} \end{array}\right) = pq + qp = 2pq, \\ z_1 &= P\left(\begin{array}{c} \textcircled{a_2} \\ \textcircled{a_2} \end{array}\right) = q^2. \end{aligned}$$

Stejnou úvahou jako při odvození rovností (3.2), (3.3) a potom s využitím rovnosti (3.4) dostaneme

$$\begin{aligned} p_1 &= x_1 + \frac{1}{2}y_1 = p^2 + pq = p(p + q) = p, \\ q_1 &= \frac{1}{2}y_1 + z_1 = pq + q^2 = (p + q)q = q. \end{aligned}$$

Relativní četnosti alel jsou ve filiální generaci stejné, jako v generaci parentální.

Uvažujme nyní populaci, která se reprodukuje popsáním způsobem a jejíž jednotlivé generace se nepřekrývají, tj. ke křížení dochází jen v jedné generaci, nikoliv mezi generacemi. K předpokladům (i), (ii) a (iii) přidejme další:

(iv) Přežití jedinců nezávisí na genotypu.

- (v) Nedochází k žádným mutacím a migracím, tj. alely se nemění a populace je tvořena všemi jedinci vzniklými reprodukcí a žádnými jinými.

Pak je zřejmé, že provedené úvahy můžeme se stejnými výsledky zopakovat pro každou následující generaci. Tímto způsobem dostaneme *Hardyho-Weinbergův zákon*: Za uvedených předpokladů

- relativní četnosti p , q alel a_1 , a_2 zůstávají ve všech generacích stejné,
- počínaje první filiální generací jsou relativní četnosti genotypů a_1a_1 , a_1a_2 a a_2a_2 rovny p^2 , $2pq$ a q^2 (v tomto pořadí).

Označme x_n , y_n a z_n relativní četnosti genotypů a_1a_1 , a_1a_2 a a_2a_2 (v tomto pořadí) v n -té filiální generaci. Podle Hardyho-Weinbergova zákona je

$$x_n = p^2, \quad y_n = 2pq, \quad z_n = q^2,$$

takže

$$y_n^2 = 4p^2q^2 = 4x_nz_n, \quad \text{neboli} \quad y_n = 2\sqrt{x_nz_n}.$$

Jsou-li a , b kladná čísla, pak číslo $\sqrt{ab}$ se nazývá jejich *geometrickým průměrem*. Poslední výsledek tedy lze zformulovat: relativní četnost heterozygotů v populaci je dvojnásobkem geometrického průměru relativních četností homozygotů.