

Autosomální dědičnost a lineární algebra

Co je autosomální dědičnost

Živočichové obecně dědí od svých rodičů značnou část svých tělesných znaků. Tyto znaky jsou určeny geneticky a právě genetika se zabývá děděním určitých znaků mezi generacemi v závislosti na jejich genetické výbavě. Geny určují takové znaky jako barva očí či ochlupení (u zvířat a u člověka), případně barva květů a tvar listů (u rostlin).

Typů dědičnosti různých znaků máme několik. Nás zajímá autosomálně recesivní dědičnost. To znamená, že znak, který je tímto způsobem děděn, je kódován jen jedním genem. Obvykle má každý takový gen dvě formy - alely (značeno A, případně a). Tyto alely kódují genotyp jedince. Pro každý znak tak máme tři různé genotypy: AA, Aa a aa. Předpokládejme nyní, že se v nějakém stádu krav jsou rohy/bezrohost krávy ovládány autosomálně recesivní dědičností. Krávy, které mají genotyp AA, mají dlouhé rohy, stejně tak krávy s genotypem Aa. Krávy s genotypem aa rohy nemají. Pak řekneme, že alela A (kódující rohy) je dominantní nad alelou a (kódující bezrohost), případně alela a je recesivní k alele A, neboť krávy s genotypem AA a Aa mají stejný sledovaný znak. Poměr jedinců s různými genotypy v určité populaci nazýváme genetickou distribucí populace.

Zajímá nás, jak určit genotyp potomků nějakých rodičů a případně určit genotyp jakékoliv následující n-té generace jedinců.

Model situace

V případě genotypu potomků nějakých rodičů nás vlastně zajímá, jakým způsobem můžeme nakombinovat jejich geny a s jakou pravděpodobností každá možná kombinace nastane. Každému jedinci můžeme přiřadit vektor:

$$j = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad \text{kde } v_i \text{ je genetická výbava, kterou rodič předá potomku}$$

Tento vektor má vždy 2^m složek, kdy m je počet znaků, jejichž dědičnost sledujeme. Dále budeme pro zjednodušení sledovat dědičnost jen jednoho znaku.

Když nám půjde o genotyp n-té generace, půjde nám o určení genetické distribuce n-té generace. Genetickou distribuci každé generace můžeme chápat jako stav, jednoznačně určený svým vektorem:

$$s_n = \begin{pmatrix} p_{AA} \\ p_{Aa} \\ p_{aa} \end{pmatrix} \quad \text{kde } p_g \text{ je podíl jedinců s genotypem } g$$

Budeme tedy chtít vypočítat libovolný vektor s_n v závislosti na vektoru s_0 , který určuje počáteční genetickou distribuci. Případně spočítat i limitu posloupnosti vektorů s_n . Ta určuje rovnovážný stav, kterému se bude genetická distribuce populace blížit.

Řešení problému

Nejdříve zjistíme, jak určit podíl genotypů potomstva jednoho páru. Každý potomek zdědí od každého rodiče jednu alelu kódující daný znak. Pokud mají oba rodiče genotyp AA, případně aa je zřejmé, že tento genotyp budou mít i všichni jejich potomci. Budeme předpokládat, že se každá alela dědí se stejnou pravděpodobností (což je jistě rozumný předpoklad).

Vyšetříme situaci, kdy jeden z rodičů má genotyp AA a druhý aa. Alely rodičů, které mohou předat svým potomkům, můžeme zapsat do vektorů.

$$r_1 = \begin{pmatrix} A \\ A \end{pmatrix}, r_2 = \begin{pmatrix} a \\ a \end{pmatrix}$$

Abychom nyní určili genotyp potomstva stačí mezi sebou tyto vektory vynásobit.

$$P = r_1 r_2^T = \begin{pmatrix} Aa & Aa \\ Aa & Aa \end{pmatrix}$$

Vidíme, že v tomto případě budou mít všichni potomci hybridní genotyp Aa. Nyní zjistíme, co se stane v případě, že jsou oba rodiče hybridy.

$$r_1 = \begin{pmatrix} A \\ a \end{pmatrix}, r_2 = \begin{pmatrix} A \\ a \end{pmatrix} \quad P = r_1 r_2^T = \begin{pmatrix} AA & Aa \\ Aa & aa \end{pmatrix}$$

Vidíme, že 1/4 potomků bude mít dominantní genotyp AA, 1/4 potomků bude mít recesivní genotyp aa a 1/2 potomků získá hybridní genotyp Aa. Stejně můžeme postupovat pro libovolné genotypy rodičů. Dokonce takto můžeme zkoumat i více znaků najednou, pouze se nám uměrně zvětší matice.

Při zjišťování genetické distribuce n-té generace chceme spočítat vektor s_n . Vlastně chceme vyřešit diferenční rovnici $s_n = M s_{n-1}$ kdy M je matice přechodu mezi generacemi. Genetickou distribuci 1. generace tedy spočítáme takto: $s_1 = M s_0$. Genetická druhé generace tedy bude vypadat takto:

$s_2 = M s_1 = M (M s_0) = M^2 s_0$. Obecně bude genetická distribuce n-té generace dána vzorcem:

$$s_n = M^n s_0$$

Ukážeme, jak se bude populace vyvíjet, pokud budeme vždy znát genotyp jednoho z rodičů (například šlechtitelský program). Nejprve rozebereme případ, kdy jeden z rodičů má genotyp AA. Máme tedy tři možnosti křížení: AA s každým genotypem. Zjistíme tedy, jaké jsou pravděpodobnosti různých genotypů. Pomocí výše zmíněné metody zjistíme, že když zkřížíme dva jedince s genotypem AA, všichni potomci budou mít také genotyp AA, to odpovídá vektoru:

$$s_{AA \times AA} = \begin{pmatrix} p_{AA} \\ p_{Aa} \\ p_{aa} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Pokud bude druhým rodičem jedinec s genotypem Aa, pak stejným postupem zjistíme, že polovina potomků bude mít genotyp AA a polovina jedinců bude mít genotyp Aa. To odpovídá vektoru:

$$s_{AA \times Aa} = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$$

Stejně vyšetříme i zbývající případ, kdy druhý rodič má genotyp aa. Dostaneme vektor:

$$s_{AA \times aa} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Z těchto vektorů můžeme sestavit matici pro přechod mezi generacemi:

$$M_1 = \begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Díky tomu můžeme určit genetickou distribuci libovolné generace pomocí vzorce $s_n = M^n s_0$. Musíme ovšem mezi sebou vynásobit n matic, což je pro velká n značně nepraktický až nepoužitelný postup. Proto se pro velká n vyplatí matici M zdiagonalizovat a využít faktu, že:

$M^n = P D^n P^{-1}$ kde D je diagonální matice, která má na diagonále vlastní čísla matice M a kterou umíme snadno spočítat a matice P je pak tvořena vlastními vektory M. Víme totiž, že pro libovolnou diagonální matici platí následující:

$$D^n = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}^n = \begin{pmatrix} \lambda_1^n & 0 & 0 \\ 0 & \lambda_2^n & 0 \\ 0 & 0 & \lambda_3^n \end{pmatrix} \quad (\text{kde } D \text{ je čtvercová matice libovolného řádu})$$

Protože tedy chceme tuto matici zdiagonalizovat, musíme zjistit, zda to lze, musíme tedy zjistit její vlastní čísla a vektory odpovídající těmto vlastním číslům. Vlastní čísla spočítáme jako kořeny charakteristického polynomu matice, tedy čísla λ splňující rovnost $\det(A - \lambda I) = 0$ kde A je libovolná čtvercová matice a I je identická matice stejného řádu jako A. V našem případě rovnost $\det(M_1 - \lambda I) = 0$ splňují čísla: 1, 1/2 a 0. Protože máme matici o rozměrech 3x3 máme tři vlastní čísla, tak tato matice určitě půjde zdiagonalizovat. K těmto vlastním číslům matice M_1 tedy najdeme vlastní vektory jim příslušné jako ty vektory v , které splňují rovnost: $(M_1 - \lambda I)v = 0$, kde o je nulový vektor. Zjistíme, že vlastní vektory příslušné těmto vlastním číslům jsou vektory:

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad v_{1/2} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \quad v_0 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

Ty jsou určeny jednoznačně až na skalární násobek.

Pro naši matici tedy platí:

$$M_1^n = \begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1^n & 0 & 0 \\ 0 & 1/2^n & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}^{-1}$$

Pokud tedy budeme předpokládat, že původní genetická distribuce bylo $s_0 = \begin{pmatrix} 1/4 \\ 1/2 \\ 1/4 \end{pmatrix}$, pak díky

tomuto postupu pak můžeme snadno spočítat, že pro velmi velká n se bude vektor genetické

distribuce populace blížit vektoru $s_{inf} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, vidíme tedy, že postupně z populace úplně vymizí

alela a. Každý jedinec v populaci bude mít genotyp AA.

Nyní předpokládejme, že mezi sebou budeme vždy křížit jedince se stejným genotypem (což může být součástí dalšího šlechtitelského programu, např. alela a může kódovat velké plody u rajčat a alela A malé). V tomto případě tedy pomocí výše zmíněné metody zjistíme, jaké je genetická distribuce potomstva v každém ze tří případů. Dostaneme tyto vektory:

$$s_{AAxAA} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad s_{AaxAa} = \begin{pmatrix} 1/4 \\ 1/2 \\ 1/4 \end{pmatrix} \quad s_{aaxaa} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Stejně jako v minulém případě vzniklou matici zdiagonalizujeme a dostaneme:

$$M_2 = \begin{pmatrix} 1 & 1/4 & 0 \\ 0 & 1/2 & 0 \\ 0 & 1/4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & -2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & -2 \\ 0 & 1 & 1 \end{pmatrix}^{-1}$$

Pokud budeme znovu předpokládat počáteční genetickou distribuci $s_0 = \begin{pmatrix} 1/4 \\ 1/2 \\ 1/4 \end{pmatrix}$, pak zjistíme, že

pro velmi velká n se bude genetická distribuce blížit $s_0 = \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \end{pmatrix}$, tedy všichni jedinci s hybridním genotypem vymizí a v populaci budou rovnoměrně zastoupeni jedinci s genotypy AA a aa.

Závěr

Vidíme, že pomocí matic, vektorů a jejich násobení můžeme snadno určit jak poměr různých genotypů potomstva jednoho páru, tak můžeme i simulovat výsledky šlechtitelských programů v různých generacích. Také jsme ukázali, že pomocí diagonalizace matic, můžeme tyto úlohy snadno řešit i pro velmi velká n . Dále vidíme, proč je je příbuzenské křížení nebezpečné – naprostá většina genetických poruch je přenášena recesivními alelami, a vidíme, že pokud budeme soustavně křížit jedince s recesivním genotypem, pak se jejich podíl v rámci populace stabilizuje na vysokém procentu. To se stalo například vyšší šlechtě v průběhu středověku, protože se snažila udržet "čistou krev", proto se v pozdější době mezi šlechtou vyskytovaly různé choroby (např. Hemofilie). Ze stejného důvodu je při ochraně ohrožených druhů důležité udržovat jejich počet nad určitou hodnotou, jinak mohou začít degenerovat. Ve šlechtitelských programech se tomuto nyní již úspěšně zabraňuje.

Seznam použitých zdrojů:

Kirkham, Jared. Math 308 Project

<http://www.math.washington.edu/~king/coursedir/m308a01/Projects/m308a01-pdf/kirkham.pdf>

Lohrenz, Caitlin. Linear Algebra Final Project

http://www.unc.edu/~marzuola/Math547_S13/Math547_S13_Projects/C_Lohrenz_Section001_Genetics.pdf

Daddel, Ali A. Applications of Linear Algebra

https://www.math.ucdavis.edu/~daddel/linear_algebra_appl/Applications/Genetics/genetics/genetics.html

Khoury, Joseph a Jessup, Barry. Linear algebra close to Earth

<http://aix1.uottawa.ca/~jkhoury/genetics.htm>

Jemelková, Simona a Sojková, Radana Co je to autosomálně recesivní dědičnost

<http://www.ttc-cz.com/news/co-je-to-autosomalne-recesivni-dedicnost/>