Zápočtová úloha č.1 — Hardyho–Weinbergův zákon

Uvažujme experiment pářících se králíčků. Zajímá nás vývoj genu, který se může objevit ve dvou typech, označíme je G a g. Každý králíček má dvojici těchto genů buď GG (dominantní), Gg (hybridní - v tomto případě nezáleží na pořadí, tj. Gg a gG je to samé) a nebo gg (recesivní). Páří-li se králíček s králičicí, jejich potomek zdědí jeden gen od každého rodiče se stejnou pravděpodobností. Například, když spáříme dominantního jedince (GG) s hybridem (Gg), tak jejich potomek bude dominantní s pravděpodobností $\frac{1}{2}$ nebo hybridní s pravděpodobností $\frac{1}{2}$, neboť od dominantního může zdědit pouze gen G a od hybrida G nebo g se stejnou pravděpodobností.

Začneme s králíčkem daného charakteru (tj. buď GG, Gg nebo gg) a spáříme jej s hybridem. Jejich potomka pak s dalším hybridem a tento proces opakujeme po n generací. Vždy potomka pářím s hybridem.

Úkoly:

1. Formulujte model a dokažte, že se skutečně jedná o homogenní markovský řetězec. Napište matici pravděpodobností přechodu.
2. Předpokládejme, že začneme s hybridem. Buď μ_n rozdělení chrakteru králíčka v n-té generaci. Tedy $\mu_n(GG)$, $\mu_n(Gg)$ a $\mu_n(gg)$ jsou pravděpodobnosti, že králíček v n-té generaci je typu GG, Gg, resp. gg. Spočítejte μ_n.
3. Proveďte simulační studii. Simulujte vhodný počet generací a vytvořte grafický výstup. Porovnejte teoretickou hodnotu μ_n s empirickými μ_n spočtenými z vaší simulace.
4. Souvisí tento příklad s Hardyho-Weinbergovým zákonem (z genetiky), který je zmíněn v nadpise? A jak?
5. Předpokládejme dále, že stavová množina $S = \{1, 2, \ldots, m\}$ je konečná, tj. $|S| = m \in \mathbb{N}$. Odvoďte odhad pravděpodobností přechodu $(p_{i,j})_{i,j \in S}$ jako parametrů modelu pomocí metody maximální věrohodnosti. Jako věrohodnost použijte pravděpodobnost jedné dlouhé trajektorie podmíněné výchozím stavem (aby se ve vzorci vyskytovaly jen parametry $p_{i,j}$, nikoli pravděpodobnosti z počátečního rozdělení). Uvědomte si, že parametrů je jen $m(m-1)$, neboť mimo podmínek

   $$p_{i,j} \geq 0, \quad \forall i, j \in S$$

   musí platit i

   $$p_{i,m} = 1 - \sum_{j=1}^{m-1} p_{i,j}, \quad \forall i \in S.$$

   Vyzkoušejte vámi odvozené odhady použít na data — např. na nějakou dlouhou trajektorii z vaší simulace výše. Jak dobře fungují?