- 1. (18 bodů) Informační kanál se šumem a zámlkou Z, který je znázorněn na obrázku, má parametry přenosu $p=0.1,\ q=0.08,\ r_0=r_1=0.03.$
 - a) Vypočtěte pravděpodobnosti $(P(B_0), P(Z), P(B_1))$ symbolů na výstupu, jestliže jsou pravděpodobnosti $(P(A_0), P(A_1))$ symbolů 0 a 1 na vstupu rovny (0.6, 0.4). (10 bodů)
 - b) Výstup kanálu (B_0, B_1) připojíme jako vstup (A_0^*, A_1^*) do dalšího informačního kanálu se stejnými parametry. Vypočtěte pravděpodobnosti symbolů $(P(B_0^*), P(Z^*), P(B_1^*))$ na jeho výstupu. $(8 \ bodů)$

Řešení:

Pro parametry kanálu je zadáno:

$$P(B_0|A_0) = 1 - p - r_0 = 0.87, \quad P(B_0|A_1) = q = 0.08;$$

 $P(B_1|A_0) = p = 0.1, \quad P(B_1|A_1) = 1 - q - r_1 = 0.89;$
 $P(Z|A_0) = r_0 = 0.03, \quad P(Z|A_1) = r_1 = 0.03.$
a pravděpodobnosti na vstupu $(P(A_0), P(A_1)) = (0.6, 0.4).$

a) Pro pravděpodobnosti na výstupu dostaneme ze vzorce pro úplnou pravděpodobnost

$$(P(B_0), P(Z), P(B_1)) = (P(A_0), P(A_1)) \begin{pmatrix} P(B_0|A_0) & P(Z|A_0) & P(B_1|A_0) \\ P(B_0|A_1) & P(Z|A_1) & P(B_1|A_1) \end{pmatrix} =$$

$$= (0.6, 0.4) \begin{pmatrix} 0.87 & 0.03 & 0.1 \\ 0.08 & 0.03 & 0.89 \end{pmatrix} =$$

$$= (0.6 \cdot 0.87 + 0.4 \cdot 0.08, 0.6 \cdot 0.03 + 0.4 \cdot 0.03, 0.6 \cdot 0.1 + 0.4 \cdot 0.89) =$$

$$= (0.554, 0.03, 0.416).$$

b) Na vstup dalšího kanálu se dostanou symboly 0 a 1 za podmínky, že se neprojeví zámlka, tedy nastane jev $B_0 \cup B_1$. Protože jsou oba jevy neslučitelné, je

$$P(B_0 \cup B_1) = P(B_0) + P(B_1) = 0.554 + 0.416 = 0.97$$
.

Pro pravděpodobnosti $(P(A_0^*), P(A_1^*))$ na vstupu do druhého kanálu dostaneme ze vzorce pro podmíněnou pravděpodobnost

$$P(A_0^*) = P(B_0|B_0 \cup B_1) = \frac{P(B_0 \cap (B_0 \cup B_1))}{P(B_0 \cup B_1)} = \frac{P(B_0)}{P(B_0 \cup B_1)} = \frac{0.554}{0.97} \doteq 0.571,$$

$$P(A_1^*) = P(B_1|B_1 \cup B_1) = \frac{P(B_1 \cap (B_0 \cup B_1))}{P(B_0 \cup B_1)} = \frac{P(B_1)}{P(B_0 \cup B_1)} = \frac{0.416}{0.97} \doteq 0.429.$$

Jako v případě a) dostaneme pro výstup pravděpodobnosti

$$(P(B_0^*), P(Z^*), P(B_1^*)) = (0.571, 0.429) \begin{pmatrix} 0.87 & 0.03 & 0.1 \\ 0.08 & 0.03 & 0.89 \end{pmatrix} =$$

$$= (0.571 \cdot 0.87 + 0.429 \cdot 0.08, 0.571 \cdot 0.03 + 0.429 \cdot 0.03, 0.571 \cdot 0.1 + 0.429 \cdot 0.89) =$$

$$= (\mathbf{0.531}, \mathbf{0.03}, \mathbf{0.439}).$$

2. (18 bodů) Na N=150 jedincích byla zjišťována přítomnost tří variant téhož genu (H, L a K) a současně sledováno, zda se projevuje jistá vlastnost V. Data jsou shrnuta v tabulce.

	H	$\mid L \mid$	K
V se projevuje	7	22	31
V se neprojevuje	23	28	39

- a) Na hladině $\alpha=5\%$ testujte hypotézu o nezávislosti výskytu variant genu a projevu vlastnosti $V.~(9~bod\mathring{u})$
- b) Pro rozdělení variant genu H, L a K v populaci předpokládejme teoretické pravděpodobnosti podle tabulky níže. Z naměřených hodnot odhadněte metodou maximální věrohodnosti parametry c a q. (9 bodů)

varianta	H	L	K
teoretická pravděpodobnost	c	cq	2cq

Řešení:

(a) Označme si X veličinu, která přiřazuje danému jedinci variantu genu a Y veličinu, která zjišťuje projev vlastnosti V. Budeme testovat hypotézu:

 $\mathbf{H}_0:$ rozdělení veličinXa Yjsou nezávislá

proti alternativní hypotéze:

 \mathbf{H}_1 : rozdělení veličin X a Y jsou závislá.

na hladině významnosti $\alpha = 5\%$.

Rozdělení veličin \boldsymbol{X} ani \boldsymbol{Y} neznáme a proto je odhadneme jako

$$p_X(i) = \frac{n_{i,*}}{N} ,$$

$$p_Y(j) = \frac{n_{*,j}}{N} ,$$

kde

$$n_{i,*} = \sum_{j} n_{i,j}$$
 a $n_{*,j} = \sum_{i} n_{i,j}$

jsou marginální četnosti.

$ \begin{array}{c c} n_{i,j} & & & \\ (X =) i & & & \\ \end{array} $	Н	L	K	$n_{i,*}$
V se projevuje	7	22	31	60
V se neprojevuje	23	28	39	90
$n_{*,j}$	30	50	70	150

Za předpokladu \mathbf{H}_0 pro očekávané četnosti pro jednotlivé hodnoty (i,j) náhodného vektoru (X,Y) pak bude platit, že

$$N \cdot p_{X,Y}(i,j) = N \cdot p_X(i) \cdot p_Y(j) = \frac{n_{i,*} \cdot n_{*,j}}{N}$$
.

Tabulka pro pro tyto četnosti bude:

$ \begin{array}{c} N \cdot p_{X,Y}(i,j) \\ (X=) i \end{array} $	Н	L	K	$n_{i,*}$
V se projevuje	$\frac{30.60}{150} = 12$	$\frac{50.60}{150} = 20$	$\frac{70.60}{150} = 28$	60
V se neprojevuje	$\frac{30.90}{150} = 18$	$\frac{50.90}{150} = 30$	$\frac{70.90}{150} = 42$	90
$n_{*,j}$	30	50	70	150

Podmínka na teoretické (tj. očekávané) četnosti ≥ 5 je splněna. Pro realizaci testovací statistiky dostaneme

$$t = \sum_{i,j} \frac{\left(n_{i,j} - N \cdot p_{X,Y}(i,j)\right)^2}{N \cdot p_{X,Y}(i,j)} =$$

$$= \frac{(7-12)^2}{12} + \frac{(22-20)^2}{20} + \frac{(31-28)^2}{28} + \frac{(23-18)^2}{18} + \frac{(28-30)^2}{30} + \frac{(39-42)^2}{42} \doteq$$

$$= 2.083 + 0.2 + 0.321 + 1.389 + 0.133 + 0.214 = 4.34$$

a porovnáme ji s hodnotou kvantilu χ^2 pro $(3-1)\cdot(2-1)=2$ stupňů volnosti

$$q_{\chi^2(2)}(1-\alpha) = q_{\chi^2(2)}(0.95) \doteq 5.992$$
.

Protože

$$t \doteq 4.34 < 5.992 \doteq q_{\chi^2(2)}(0.95)$$
,

hypotézu o nezávislosti proto **NEZAMÍTÁME**.

(b) Použijeme metodu maximální věrohodnosti (nic jiného ani nelze).

hodnota $Y = j$	H	L	K
pravděpodobnost $p_Y(j)$	c	cq	2cq
četnost n_j	30	50	70

Protože součet pravděpodobností všech hodnot je 1, musí být

$$1 = c + cq + 2cq = c(1+3q)$$

tedy $c=\frac{1}{1+3q}$. Současně musí být pravděpodobnosti nezáporné, tj. $0\leq c$ a $0\leq q$. Zbývá tedy odhadnout parametr q.

Hledáme hodnotu q, která maximalizuje funkci věrohodnosti

$$L(q) = P(Y_1 = y_1, \dots, Y_n = y_n) = \prod_{i=1}^n P(Y_i = y_i) = c^{30} (cq)^{50} (2cq)^{70} = 2^{70} c^{150} q^{120} =$$

$$= 2^{70} \frac{q^{120}}{(1+3q)^{150}}$$

kde Y_i jsou jednotlivé nezávislé veličiny (pokusy) a y_i naměřené hodnoty. Budeme ekvivalentně hledat maximum logaritmu:

$$\ell(q) = \ln(L(q)) = 120 \cdot \ln q - 150 \cdot \ln(1 + 3q) + konst$$

pomocí derivace

$$0 = \ell'(q) = \frac{120}{q} - \frac{150 \cdot 3}{1 + 3q} = \frac{120 - 90q}{q(1 + 3q)}$$

Ze znamének derivace je hned jasné, že maximum je nabyto (a tedy i odhad parametru q) je v

$$q = \frac{4}{3} \doteq 1.33$$

a

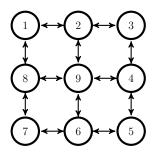
$$c = \frac{1}{1+3a} = \frac{1}{5} = 0.2$$

Odhady pravděpodobností pro hodnoty H, K a L jsou tedy

$$p_Y(H) = \frac{1}{5} = 0.2$$
 $p_Y(K) = \frac{4}{15} \doteq 0.267$ $p_Y(L) = \frac{8}{15} \doteq 0.533$

což vyhovuje zadání.

3. (14 bodů) Markovův řetězec je dán obrázkem; pro každý stav platí, že všechny hrany z něj vycházející mají stejnou pravděpodobnost.



A. Vypočtěte pravděpodobnost přechodu ze stavu 1 do stavu 4 v právě třech krocích. (3 body)

- B. Oklasifikujte všechny stavy. (3 body)
- C. Stanovte všechny komponenty. (2 body)
- D. Konverguje rozdělení stavů ke stacionárnímu? Za jakých podmínek? Odůvodněte. (3 body)
- E. Stanovte rozdělení pravděpodobností stavů po 10^5 krocích, pokud jsme vyšli se stavu 1. (3 body)

(Můžete kreslit do zadání.)

Řešení:

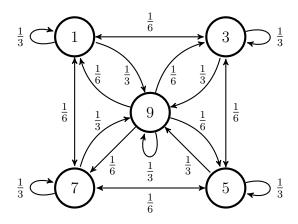
A. Např. cesta 1234 má pravděpodobnost

$$p_{12} p_{23} p_{34} = \frac{1}{2} \frac{1}{3} \frac{1}{2} = \frac{1}{12}.$$

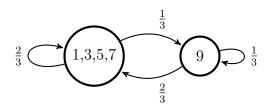
Cesty 1234, 1294, 1894 dávají celkovou pravděpodobnost

$$\frac{1}{12} + \frac{1}{24} + \frac{1}{24} = \frac{1}{6}.$$

- B. Všechny stavy jsou trvalé s periodou 2.
- C. Všechny stavy tvoří jedinou komponentu.
- D. Nějaké stacionární rozdělení existuje, ale kvůli netriviální periodě k němu tento Markovův řetězec nekonverguje (pokud se v něm právě nenachází).
- E. Po sudém počtu kroků musíme být v některém z lichých stavů 1, 3, 5, 7, 9. Po přechodu ke krokům délky 2 tvoří tyto stavy ergodickou komponentu na obrázku.



Přibližným výsledkem je její stacionární rozdělení. Můžeme je najít jako levý vlastní vektor matice přechodu (řádu 5×5), odpovídající vlastnímu číslu 1. Pravděpodobnost stavu 9 označme c. Díky symetrii bude pravděpodobnost stavů 1, 3, 5, 7 stejná, $\frac{1-c}{4}$. Pro přechod mezi stavy 1, 3, 5, 7 a stavem 9 dostáváme jednodušší graf na obrázku.



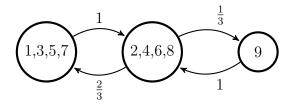
Z něj nebo z rovnice

$$(1-c,c)\cdot \begin{pmatrix} 2/3 & 1/3\\ 2/3 & 1/3 \end{pmatrix} = (1-c,c)$$

vyplývá $c=\frac{1}{3}$ a rozdělení pravděpodobností stavů po 10^5 krocích je přibližně

$$\left(\frac{1}{6}, 0, \frac{1}{6}, 0, \frac{1}{6}, 0, \frac{1}{6}, 0, \frac{1}{3}\right)$$
.

Alternativní řešení: Vzhledem k symetrii jsou po velkém počtu kroků stavy 2, 4, 6, 8 stejně pravděpodobné, stejně tak stavy 1, 3, 5, 7. Sloučíme je a řešíme zjednodušený Markovův řetězec s následujícím přechodovým grafem:



Jednodušší alternativní řešení: O jeden krok dříve jsou možné jen stavy 2, 4, 6, 8, díky symetrii stejně pravděpodobné. Na rozdělení pravděpodobností $(0, \frac{1}{4}, 0, \frac{1}{4}, 0, \frac{1}{4}, 0, \frac{1}{4}, 0)$ aplikujeme jeden krok (např. vynásobíme maticí přechodu) a dostaneme výše uvedený výsledek.

(Jediné stacionární rozdělení původního řetězce, o němž se hovořilo v části D, je $(\frac{1}{12}, \frac{1}{8}, \frac{1}{12}, \frac{1}{8}, \frac{1}{12}, \frac{1}{8}, \frac{1}{12}, \frac{1}{8}, \frac{1}{6})$, neboť poloviční pravděpodobnost sudých stavů 2, 4, 6, 8 se mezi ně dělí rovnoměrně.)