Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Radim Navrátil

Urnové modely

Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: Mgr. Martin Schindler
Studijní obor: Matemtika
Studijní program: Obecná matematika

2007

Na tomto místě bych rád poděkoval panu Mgr. Martinu Schindlerovi za veškerou pomoc a cenné rady a podněty, které mi pomohly při psaní této práce.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 12.5.2007 Radim Navrátil

Obsah

Název práce: Urnové modely
Autor: Radim Navrátil
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: Mgr. Martin Schindler
e-mail vedoucího: martin.schindler@karlin.mff.cuni.cz

Abstrakt: V této práci popisujeme nejrůznější varianty urnových modelů od jednoduchých modelů bez vracení a s vracením, přes Pólyovy - Eggenbergerovy modely až po umisťovací a přemisťovací modely. Tyto modely se budeme snažit co nejvíce zobecňovat. Ukážeme souvislosti s některými rozděleními jako například binomické, hypergeometrické nebo multinomické. Všechny tyto modely si ilustrujeme na příkladech, a to i početních. Ukážeme souvislosti s hazardními hrami a soutěžemi. V této souvislosti budeme testovat, jestli jsou míčky v různých loteriích taženy rovnoměrně, zaměříme se hlavně na nejznámější loterii u nás - Sportku. Neopomeneme ani fyzikální aplikaci některých umisťovacích schémat. Na závěr se zmíníme o některých dalších modelech využívaných i v jiných oblastech lidské činnosti než matematika.

Klíčová slova: Výběr bez vracení, výběr s vracením, Pólyovo schéma, capture-recapture model, okupační rozdělení

Title: Urn models
Author: Radim Navrátil
Department: Department of Probability and Mathematical Statistics
Supervisor: Mgr. Martin Schindler
Supervisor's e-mail address: martin.schindler@karlin.mff.cuni.cz

Abstract: In this work we describe some variants of urn models, a simple model with replacement, a simple model without replacement, Pólya - Eggenberger's models, an occupancy model and an urn transfer model. We will try to generalize these models. We will show connections with some distributions, for example binomial, hypergeometric and multinomial. We will illustrate all these models with the examples namely numerical, too. We will show conections with gambles and competitions. In this connection we will test if the balls in some lotteries are drawn equally. Mainly we will target best known lottery in the Czech republic - Sportka. We will not forget physical applications of some occupancy schemes. Finally we will mention some other models used in other fields of human activity than mathematics.

Keywords: Sample with replacement, sample without replacement, Pólya's scheme, capture-recapture model, occupancy distributions

Kapitola 1

Úvod

1.1 Základní urnové modely

Urnové modely si můžeme představovat nejrůznějšími způsoby. Nejčastěji však jako losování různých míčků z uren nebo jako umisťování míčků do uren. V prvním případě nás nejvíce zajímá rozdělení počtu míčků v urně nebo doba čekání než bude splněna určitá podmínka. Ve druhém případě se zajímáme většinou o počet různých rozmistění míčků do uren nebo pravděpodobnost, že určitý počet uren bude prázdný. V některých případech výsledky dokonce splývají.

Urnové modely mají spoustu nejrůznějších aplikací. Pomocí urnových modelů mohou být modelována některá známá rozdělení, např. binomické, hypergeometrické, multinomické a mnohá další. Ukážeme jejich různé aplikace ve statistice, konkrétně v teorii odhadu a testování hypotéz. Urnové modely však mají také široké uplatnění i v jiných oborech jako je například epidemiologie, kde se urnové modely využívají k simulaci šíření nakažlivých chorob v populaci nebo kvantová fyzika, kde jsou pomocí urnových modelů popisovány různé stavy systému. Neopomeňme ani aplikaci v hazardních hrách, kde se urnové modely používají jako losovací zařízení.

V této práci si popíšeme základní urnové modely - model s vracením a bez vracení, budeme se snažit odvodit jejich různá zobecnění, např. Póllyovo-Eggenbergerovo schéma, Ehrenfestův model a další. Zmíníme se také o tzv. umisťovacích modelech. Všechny tyto modely si budeme ilustrovat na nejrůznějších příkladech a ukážeme si i jejich některé praktické aplikace. Příklady 2,4,6,8,9 jsou převzaty z [4], příklad 7 z [5] a příklad 10 z [7].

1.2 Historie urnových modelů

Historie urnových modelů se začala psát již v dobách, kdy vznikala vlastní teorie pravděpodobnosti. Vznik urnových modelů je spjat s rozvojem hazardních her, např. házení mincí, hra v kostky a dalších. Ty sebou přinášely nové otázky a problémy. Jako dobrý prostředek k modelování těchto situací posloužily právě urnové modely. Jako ilustraci si uveďme jeden prastarý příklad, který si můžeme mimo jiné představit pomocí urnového modelu.

Příklad 1 (B.Pascal, P.de Fermat, 1654) *Základy teorie pravděpodobnosti vznikly v korespondenci mezi dvěma francouzskými matematiky B. Pascalem a P. de Fermatem v roce* 1654*. Zabývali se mimo jiné problémem, se kterým přišel Chevalier de Méré (velký milovník hazardních her). Ze zkušenosti věděl, že je výhodné sázet na to, že při* 4 *hodech šestistěnnou kostkou padne šestka. Usuzoval, že při* 24 *hodech dvěma kostkami bude opět výhodné sázet na to, že padnou na obou kostkách šestky. Spočítejme tedy pravděpodobnost, že při* 4 *hodech padne aspoň jednou šestka a pravděpodobnost, že při* 24 *hodech dvěma kostkami padnou aspoň jednou dvě šestky.*

Řešení
Předpokládejme, že kostka je pravidelná, a tedy každé číslo může padnout se stejnou pravděpodobností $\frac{1}{6}$. Místo házení kostkou si můžeme představit následující model. Pro první případ mějme 4 urny obsahující každá šest míčků s čísly $1, 2, \ldots, 6$ a z každé urny losujeme jeden míček. Pro druhý případ mějme 24 uren obsahující každá 12 míčků, přičemž se čísla $1, 2, \ldots, 6$ objevují právě na dvou míčcích v každé urně. Nyní z každé urny losujeme dva míčky.

Počítejme první pravděpodobnost, že při 4 tazích vytáhneme aspoň jeden míček s číslem 6. My ale spočítáme pravděpodobnost, že nevylosujeme ani jednou šestku a hledaná pravděpodobnost pak bude doplněk do jedné. Počet uspořádaných čtveřic z čísel $1, 2, \ldots, 6$ je 6^4 a počet uspořádaných čtveřic neobsahující šestku, tj. z čísel $1, 2, \ldots, 5$ je 5^4. Tudíž pravděpodobnost, že při 4 tazích vytáhneme aspoň jednu šestku je

$$1 - \frac{5^4}{6^4} = 1 - \left(\frac{5}{6}\right)^4 \doteq 0,518.$$

Analogickou úvahou jako v předchozím je pravděpodobnost, že při 24 tazích vytáhneme aspoň jednou dvě šestky je

$$1 - \frac{35^{24}}{36^{24}} = 1 - \left(\frac{35}{36}\right)^{24} \doteq 0,491.$$

Vidíme tedy, že je pravděpodobnější, že při 4 hodech padne aspoň jednou šestka, než že při 24 hodech dvěma kostkami padnou aspoň jednou dvě šestky. $\triangle$

První zmínku o urnových modelech můžeme nalézt v Huygensově díle *Oeuvres Complétes* z roku 1665 (viz Příklad 2). Urnové modely ve svých pracech zmiňoval J.Bernoulli, např. dílo *Ars Conjectandi* a také D. Bernoulli, který navrhl model nazývaný v současnosti jako *Ehrenfestův* model. Urnové modely se objevovaly i v dílech dalších matematematiků, např. de Moivre, Laplace, Ostrogradskii, Poisson. Celá třída urnových modelů, zvaná jako *Pólyovy - Eggenbergerovy* modely, byla uvedena již v práci Markova (1906).

Nalézt počátky tzv. *umisťovacích modelů* je však velmi obtížné. První použití těchto modelů pravděpodobně souviselo s problémy teoretické fyziky, konkrétně s teorií plynů. Tyto problémy řešili ve svých dílech Boltzmann, Maxwell, Clausius a mnozí další.

Následující příklad ilustruje jedno z prvních použití urnových modelů v souvislosti s hazardními hrami.

Příklad 2 (Hudde a Huygens, 1665) *Uvažujme následujíci hru, hráč A má urnu obsahující dva bílé a jeden černý míček. Hráč B má urnu obsahující bílé a černé míčky v poměru* $p : (1-p)$. *A a B střídavě tahají míček z vlastní urny a vrátí ho zpět do své urny. Když některý z obou hráčů vytáhne bílý míček, hra končí a daný hráč vyhrává celý bank. Pokud hráč vytáhne černý míček, musí dát do banku jeden dukát. Předpokládejme, že hru začíná hráč A a v banku nejsou žádné peníze. Jaký má být poměr míčků v urně hráče B, aby hra byla spravedlivá?*

Řešení

Aby hra byla spravedlivá, musí být střední hodnota výhry hráče A rovna střední hodnotě výhry hráče B. Vzhledem k tomu, že čistá výhra jednoho hráče je rovna prohrané částce druhého hráče, znamená výše uvedená podmínka to, že střední hodnota výhry obou hráčů má být nulová.

Pokud hráč A vytáhne v 1. tahu bílý míček vyhrává 0, to se stane s pravděpodobností $\frac{2}{3}$. S pravděpodobností $\frac{1-p}{3} \cdot \frac{2}{3}$ vyhraje 1 dukát (tj. v prvním tahu hráč A vytáhne černý míček s pravděpodobností $\frac{1}{3}$ a dá 1 dukát do banku, pak hráč B vytáhne černý míček s pravděpodobností $1-p$ a dá 1 dukát do banku. Při druhém tahu hráč A vytáhne bílý míček s pravděpodobností $\frac{2}{3}$ a hra končí). A dále analogicky s pravděpodobností $\frac{2}{3} \cdot \left(\frac{1-p}{3}\right)^k$ vyhraje hráč A přesně k dukátů. Stejnou úvahou dostáváme, že s pravděpodobností $p \cdot \frac{(1-p)^{k-1}}{3^k}$ hráč A prohraje k dukátů. Odtud již snadno plyne, že střední hodnota výhry hráče A je

$$\mathrm{E}X_1 = \sum_{k=0}^{\infty} \frac{2k}{3} \left(\frac{1-p}{3}\right)^k - \sum_{k=1}^{\infty} pk \, \frac{(1-p)^{k-1}}{3^k} = \frac{\frac{2}{9} - \frac{2}{9}p}{(\frac{2}{3} + \frac{1}{3}p)^2} - \frac{\frac{p}{3}}{(\frac{2}{3} + \frac{1}{3}p)^2}.$$

Odtud dostáváme $\frac{2}{9} - \frac{2}{9}p = \frac{p}{3}$. Čili $p = \frac{2}{5}$. To znamená, že v urně hráče B mohou být například dva bílé a tři černé míčky. $\triangle$

S urnovými modely se můžeme setkávat dennodenně na televizní obrazovce při různých hrách a loteriích, kde člověk tipuje losovaná čísla míčků tažených z urny. Asi nejznámější z nich jsou hra Sportka nebo Šťastných deset.

Příklad 3 (Hra Šťastných 10) *Vypočítejme střední hodnotu výhry ve hře Šťastných* 10. *Je to sázková hra, ve které sázející tipuje jedno až deset čísel z osmdesáti. V každém slosování je vylosováno* 20 *výherních čísel. Pro* *jednoduchost předpokládejme, že sázíme* 1 *korunu a tipujeme* 10 *čísel. V následující tabulce jsou uvedeny výhry V_k podle počtu uhodnutých čísel k:*

k	0	1	2	3	4	5	6	7	8	9	10
V_k	1	0	0	0	0	3	10	20	500	10 000	200 000

Tabulka 1.1: Výhry ve hře Šťastných 10

Řešení

Spočítejme nejprve pravděpodobnost, že uhádneme právě k čísel, kde $k = 0, 1, \ldots, 10$. Počet možností jak z 80 míčků vylosovat 20 je právě $\binom{80}{20}$. To, že uhádneme právě k čísel znamená, že mezi našimi 10 tipovanými čísly je právě k čísel losovaných a mezi netipovanými 70 čísly je zbylých $20 - k$ čísel. Takových možností je $\binom{10}{k}\binom{70}{20-k}$. Tedy pro $k = 0, 1, \ldots, 10$ je hledaná pravděpodobnost

$$P_k = \frac{\binom{10}{k}\binom{70}{20-k}}{\binom{80}{20}}.$$

V následující tabulce jsou tyto pravděpodobnosti vypočteny s přesností na 3 desetinná místa.

k	0	1	2	3	4	5
P_k	0,0458	0,180	0,295	0,267	0,147	0,0514

k	6	7	8	9	10
P_k	0,0115	$1{,}61\cdot10^{-3}$	$1{,}35\cdot10^{-4}$	$6{,}12\cdot10^{-6}$	$1{,}12\cdot10^{-7}$

Tabulka 1.2: Pravděpodobnosti výhry ve hře Šťastných 10

Odtud se již jednoduchým výpočtem určí střední hodnota výhry $\sum_{k=0}^{10} V_k P_k - 1 = -0,502$. $\triangle$

Kapitola 2

Výběr s vracením

2.1 Binomické rozdělení

Předpokládejme, že máme urnu obsahující a bílých a b černých míčků. Budeme náhodně losovat míčky z urny a po každém vytažení vrátíme míček zpět do urny. Tento proces opakujeme n krát. Označme X náhodnou veličinu udávající počet vytažených bílých míčků.

Pravděpodobnost, že v jednom tahu vytáhneme bílý míček je $\frac{a}{a+b}$ a černý míček vytáhneme s pravděpodobností $\frac{b}{a+b}$. Počet možností, jak vybrat k bílých míčků z n je $\binom{n}{k}$. Pak pravděpodobnost, že vylosujeme právě k bílých míčků v n tazích je

$$P(X=k)=\binom{n}{k}\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^{n-k}, \qquad k=0,1,\ldots,n.$$

Označíme-li nyní $p=\frac{a}{a+b}$, vidíme, že X má *binomické rozdělení* s parametry n a $p=\frac{a}{a+b}$, neboť

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \qquad k=0,1,\ldots,n. \tag{2.1}$$

Existuje celá řada zobecnění binomického rozdělení, v následujícím příkladě odvodíme tzv. *kvazi-binomické rozdělení.*

Příklad 4 (Kvazi-binomické rozdělení) *Uvažujme následující hru: Hráč má* 4 *urny obsahující bílé a černé míčky:* 1. *a* 4. *obsahuje* a *bílých a* b *černých, druhá urna obsahuje* b *černých míčků a třetí urna* a *bílých. Hráč si zvolí nějaké přirozené číslo* k *mezi nulou a* n*. Poté* $(n-k)\theta$ *bílých míčků přidáme do* 2. *a* 4. *urny a* $k\theta$ *černých míčků přidáme do* 3. *a* 4. *urny, kde* θ *je známé přirozené číslo.*

Hráč nyní náhodně vylosuje míček z 1. *urny. Pokud vytáhne bílý míček, losuje dále z* 2. *urny; pokud první vytažený míček byl černý, losuje z* 3. *urny. Jestliže první dva tažené míčky jsou různé barvy, pak losuje* n *míčků ze* 4. *urny s vracením. V opačném případě hra končí. Hráč vyhrává, pokud vylosuje právě* k *bílých míčků z poslední urny. Jaká je pravděpodobnost výhry?*

Řešení
Určeme nejprve pravděpodobnost, že se vůbec hráč dostane k losování ze čtvrté urny, tj. že vytáhne bílý míček z první urny a černý míček z druhé urny nebo vytáhne černý míček z první urny a bílý míček z třetí urny. Čili hledaná pravděpodobnost je

$$\frac{a}{(a+b)}\cdot\frac{b}{(b+(n-k)\theta)}+\frac{b}{(a+b)}\cdot\frac{a}{(a+k\theta)}=\frac{ab(a+b+n\theta)}{(a+b)(a+k\theta)(b+(n-k)\theta)}.$$

Nyní určíme pravděpodobnost, že při losování z poslední urny vylosuje právě k bílých míčků. Dosadíme-li do (2.1) za $p=\frac{a+k\theta}{a+b+n\theta}$ máme hledanou pravděpodobnost

$$\binom{n}{k}(a+k\theta)^k\cdot(b+(n-k)\theta)^{n-k}\cdot(a+b+n\theta)^{-n}.$$

Odtud dostáváme, že pro $k=0,1,\dots n$ je pravděpodobnost výhry

$$P_k=\binom{n}{k}\frac{ab}{a+b}\cdot\frac{1}{a+b+n\theta}\cdot\left(\frac{a+k\theta}{a+b+n\theta}\right)^{k-1}\cdot\left(\frac{b+(n-k)\theta}{a+b+n\theta}\right)^{n-k-1}. \tag{2.2}$$

Dá se ukázat, že $P_0+\ldots+P_n=1$. A tedy (2.2) určuje rozdělení, tzv. *kvazi-binomické rozdělení.* $\triangle$

2.2 Negativně binomické rozdělení

Mějme opět urnu obsahující a bílých a b černých míčků. Losujeme náhodně míček z urny a po vytažení ho vrátíme zpět do urny. Nyní nás bude zajímat rozdělení počtu tahů X potřebných k vytažení právě k bílých míčků. To, že budeme potřebovat právě j tahů znamená, že v j-tém tahu vylosujeme bílý míček a v prvních $j-1$ tazích vytáhneme přesně $k-1$ bílých míčků. Odtud s použitím (2.1) dostáváme

$$P(X=j)=\frac{a}{a+b}\binom{j-1}{k-1}\left(\frac{a}{a+b}\right)^{k-1}\left(\frac{b}{a+b}\right)^{j-k}.$$

Označíme-li opět jako v předchozím $p=\frac{a}{a+b}$, po drobné úpravě obdržíme

$$P(X=j)=\binom{j-1}{k-1}p^k(1-p)^{j-k},\qquad j=k,k+1,\dots. \tag{2.3}$$

Rozdělení náhodné veličiny X je nazýváno *negativně binomické* s parametry k a p. Ve speciálním případě, kdy $k=1$, (2.3) přejde v

$$P(X=j)=p(1-p)^{j-1},\qquad j=1,2,\dots.$$

Toto rozdělení se označuje jako *geometrické* s parametrem p.

2.3 Multinomické rozdělení

Nyní si zobecníme základní model s vracením. Již nebudeme předpokládat, že jsou v urně míčky jen dvou barev, nýbrž předpokládejme, že v urně je m míčků, z nichž m_j je barvy β_j, $j = 1, 2, \ldots, k$, přičemž $m_1 + \cdots + m_k = m$. Losujeme n míčků z urny s vracením. Označme $X_1, \ldots, X_k$ počet vylosovaných míčků barvy $\beta_1, \ldots, \beta_k$. Bude nás zajímat rozdělení náhodného vektoru $\boldsymbol{X} = (X_1, \ldots, X_k)'$.

V každém tahu pravděpodobnost, že vytáhneme míček barvy β_j je $\frac{m_j}{m}$. Tedy pravděpodobnost, že vytáhneme n_j míčků barvy β_j je $\left(\frac{m_j}{m}\right)^{n_j}$. Počet způsobů, jak vybrat n_1 míčků barvy β_1, n_2 míčků barvy β_2,…,n_k míčků barvy β_k z n je $\binom{n}{n_1 \cdots n_k}$. Čili

$$P(X_1 = n_1, \ldots, X_k = n_k) = \binom{n}{n_1 \cdots n_k} \prod_{j=1}^{k} \left(\frac{m_j}{m}\right)^{n_j}, \qquad \sum_{j=1}^{k} n_j = n.$$

Označíme-li $p_j = \frac{m_j}{m}$ dostaneme

$$P(X_1 = n_1, \ldots, X_k = n_k) = \binom{n}{n_1 \cdots n_k} \prod_{j=1}^{k} p_j^{n_j}, \qquad \sum_{j=1}^{k} n_j = n. \tag{2.4}$$

Rozdělení náhodného vektoru $\boldsymbol{X} = (X_1, \ldots, X_k)'$ je nazýváno *multinomické* s parametry $n, p_1, \ldots, p_k$.

Nyní si uvedeme jednu aplikaci multinomického rozdělení v teorie testování hypotéz. Začněme s větou, která má pro toto testování zásadní význam.

Věta 1 *Nechť náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_k)'$ má multinomické rozdělení s parametry $n, p_1, \ldots, p_k$. Pak Pearsonova statistika*

$$\chi^2 = \sum_{i=1}^{k} \frac{X_i^2}{np_i} - n \tag{2.5}$$

má při $n \to \infty$ asymptoticky rozdělení χ^2_{k-1}.

Důkaz. Viz [1], str. 270. □

Pomocí veličiny χ^2 z (2.5) můžeme testovat hypotézu H_0, že skutečné hodnoty pravděpodobností v multinomickém rozdělení jsou rovny $p_1, \ldots, p_k$ proti alternativě H_1, že tomu tak není. Hypotézu H_0 zamítneme asymptoticky na hladině α, pokud $\chi^2 \geq \chi^2_{k-1}(\alpha)$. Kde $\chi^2_{k-1}(\alpha)$ značí kritickou hodnotu χ^2 rozdělení o $k - 1$ stupních volnosti.

Tento test se nazývá *Pearsonův χ^2 test*. Můžeme ho použít například při ověřování pravidelnosti hrací kostky nebo kontrole generátorů náhodných čísel. My si ho ukážeme na příkladu, jestli jsou čísla ve hře Euromiliony tažena rovnoměrně.

Příklad 5 (Hra Euromiliony) *Ve hře Euromiliony se losují čísla ze 2 osudí, přičemž z prvního osudí 7 čísel z 35 možných a z druhého osudí jedno číslo od 1 do 5. Bude nás zajímat, jestli čísla z druhého osudí jsou tažena rovnoměrně, nebo zda-li vykazují nějaké nenáhodné vlivy. V následující tabulce jsou uvedeny četnosti tažených čísel z druhého osudí za období 10.10.2003 – 16.03.2007.*

Tažené číslo	1	2	3	4	5
Počet tažených čísel	32	35	47	35	31

Tabulka 2.1: Četnosti tažených čísel ve hře Euromiliony

Řešení
Protože zřejmě náhodný vektor $\boldsymbol{X} = (X_1, \dots, X_5)'$, kde X_i značí počet vytažených míčků s číslem i, má multinomické rozdělení, budeme testovat hypotézu $H_0 : p_1 = \dots = p_5 = \frac{1}{5}$ proti alternativě, že tomu tak není asymptoticky na hladině $\alpha = 0,05$. K tomu využijeme Pearsonův χ^2 test. V našem případě máme $n = 180, k = 5$ a s pomocí tabulky a (2.5) určíme hodnotu statistiky

$$\chi^2 = 4,556$$

a kritická hodnota je

$$\chi^2_4(0,05) = 9,488.$$

Protože $\chi^2 < \chi^2_4(0,05)$, hypotézu H_0 nezamítáme asymptoticky na hladině 0,05. $\triangle$

2.4 Negativně multinomické rozdělení

Budeme předpokládat, že v urně je m míčků, z nichž m_j je barvy β_j, $j = 0, 1, \dots, k$, přičemž $m_0 + m_1 + \dots + m_k = m$. Losujeme n míčků z urny s vracením. Označme opět $X_0, X_1, \dots, X_k$ počet vylosovaných míčků barvy $\beta_0, \beta_1, \dots, \beta_k$. Nyní budeme zvyšovat n, dokud X_0 nedosáhne předepsané hodnoty x_0. Bude nás zajímat rozdělení náhodného vektoru $\boldsymbol{X} = (X_1, \dots, X_k)'$, když $X_0 = x_0$.

To, že $X_1 = x_1, \dots, X_k = x_k$ v naší situaci znamená, že jako poslední vylosujeme míček barvy β_0 s pravděpodobností $\frac{m_0}{m}$ a v prvních $x_0 - 1 + \sum_{j=1}^k x_j$ tazích vylosujeme n_1 míčků barvy β_1, n_2 míčků barvy $\beta_2, \dots, n_k$ míčků barvy β_k. Odtud s použitím (2.4) plyne

$$P(X_1 = x_1, \dots, X_k = x_k) = \binom{x_0 - 1 + \sum_{j=1}^k x_j}{x_0 - 1 \; x_1 \cdots x_k} \left(\frac{m_0}{m}\right)^{x_0} \prod_{j=1}^{k} \left(\frac{m_j}{m}\right)^{x_j}.$$

Položíme-li $p_j = \frac{m_j}{m}$, po drobné úpravě dostaneme

$$P(X_1 = x_1, \ldots, X_k = x_k) = \binom{\sum_{j=0}^{k} x_j - 1}{x_0 - 1 \; x_1 \cdots x_k} \prod_{j=0}^{k} p_j^{x_j}.$$

Náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_k)'$ má *negativně multinomické rozdělení.*

Kapitola 3

Výběr bez vracení

3.1 Hypergeometrické rozdělení

Předpokládejme, že máme urnu obsahující a bílých a b černých míčků. Budeme náhodně losovat míčky z urny, vytažené míčky již nebudeme vracet zpět do urny. Tento proces opakujeme n krát ($n \leq a + b$). Označme X náhodnou veličinu, udávající počet vytažených bílých míčků.

Počet možností, jak vybrat n míčků z $a+b$ je $\binom{a+b}{n}$. Dále počet způsobů, jak zvolit k bílých míčků z a je $\binom{a}{k}$ a $(n-k)$ černých míčků z b je $\binom{b}{n-k}$. Odtud již plyne, že

$$P(X = k) = \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}, \qquad max(0, n-b) \leq k \leq min(a, n). \tag{3.1}$$

Náhodná veličina X má *hypergeometrické rozdělení.*

Nyní si uveďme jednu aplikaci hypergeometrického rozdělení.

Příklad 6 (Murty, 1975) *Urna obsahuje n míčků, přičemž každý míček má na sobě jedno číslo. Označme n_r počet čísel, která se objeví právě na r míčcích. Například pro $n = 8$ míčků s čísly $1, 1, 1, 2, 2, 3, 3, 4$ je $n_1 = 1, n_2 = 2, n_3 = 1$. Předpokládáme, že žádné číslo se nemůže objevit na více než k míčcích, tj. $n_r = 0$ pro $r > k$. Na základě výběru z urny chceme nalézt nestranné odhady $n_1, n_2, \ldots, n_k$.*

Řešení

Uvědomme si neprve, že platí

$$\sum_{j=1}^{k} j n_j = n. \tag{3.2}$$

Vezměme si náhodný vzorek m míčků (bez vracení) a pro $t = 1, 2, \ldots, k$ označme M_t počet čísel, která jsou právě na t míčcích z vylosovaného vzorku.

Analogicky jako v (3.2) musí pro náš vzorek platit

$$\sum_{t=1}^{k} tM_t = m.$$

Dále podmíněná pravděpodobnost, že se dané číslo objeví na t míčcích v našem vzorku, když víme, že se v urně objevuje na r míčcích je podle (3.1) rovna

$$P_{r,t} = \begin{cases} 0 & t > r, \\ \frac{\binom{r}{t}\binom{n-r}{m-t}}{\binom{n}{m}} & t \le r. \end{cases}$$

Protože hledáme nestranný odhad $n_1, n_2, \ldots, n_k$ spočítejme $\mathrm{E}M_t$ a položme ji rovnu M_t pro každé $t = 1, 2, \ldots, k$.

$$\mathrm{E}M_t = \sum_{r=1}^{k} n_r P_{r,t}.$$

Proto řešení $\hat{N}_1, \hat{N}_2, \ldots, \hat{N}_k$ soustavy k lineárních rovnic

$$\sum_{r=1}^{k} N_r P_{r,t} = M_t, \qquad t = 1, 2, \ldots, k$$

jsou nestranné odhady $n_1, n_2, \ldots, n_k$. $\triangle$

3.2 Negativně hypergeometrické rozdělení

Mějme v urně a bílých a b černých míčků. Losujeme míčky z urny bez vracení. Označme X počet tahů potřebných k vytáhnutí právě k bílých míčků. To, že budeme potřebovat právě j tahů znamená, že v j-tém tahu vylosujeme bílý míček (losujeme z urny obsahující $a+b-j+1$ míčků, z nichž $a-k+1$ je bílých) a v prvních $j-1$ tazích vytáhneme přesně $k-1$ bílých míčků. Odtud s použitím (3.1) dostáváme pro $j = k, k+1, \ldots$

$$\begin{aligned} P(X=j) &= \binom{j-1}{k-1}\frac{a(a-1)\cdots(a-k+2)b(b-1)\cdots(b-j+k-1)}{(a+b)(a+b-1)\cdots(a+b-j+2)} \\ &\cdot \frac{a-k+1}{a+b-j+1} = \\ &= \binom{j-1}{k-1}\frac{a(a-1)\cdots(a-k+1)b(b-1)\cdots(b-j+k-1)}{(a+b)(a+b-1)\cdots(a+b-j+1)}. \end{aligned}$$

Rozdělení náhodné veličiny X se nazývá *negativně hypergeometrické.*

3.3 Mnohorozměrné hypergeometrické rozdělení

Analogicky jako jsme zobecnili výběr s vracením můžeme zobecnit výběr bez vracení. Předpokládejme, že v urně je m míčků, z nichž m_j je barvy β_j, $j = 1, 2, \dots, k$, přičemž $m_1 + \cdots + m_k = m$. Losujeme n míčků z urny bez vracení. Označme $X_1, \dots, X_k$ počet vylosovaných míčků barvy $\beta_1, \dots, \beta_k$. Bude nás zajímat rozdělení náhodného vektoru $\boldsymbol{X} = (X_1, \dots, X_k)'$.

Počet způsobů, jak vybrat n míčků z m je $\binom{m}{n}$. Dále x_j míčků lze z m_j vybrat $\binom{m_j}{x_j}$ způsoby. Tedy

$$P(X_1 = x_1, \dots, X_k = x_k) = \frac{\prod_{j=1}^{k} \binom{m_j}{x_j}}{\binom{m}{n}}, \qquad \sum_{j=1}^{k} x_j = n,\ 0 \le x_j \le m_j. \tag{3.3}$$

Toto rozdělení se nazývá *mnohorozměrné hypergeometrické rozdělení* nebo též *multihypergeometrické*.

3.4 Mnohorozměrné negativně hypergeometrické rozdělení

Předpokládejme, že v urně je m míčků, z nichž m_j je barvy β_j, $j = 0, 1, \dots, k$, přičemž $m_0 + m_1 + \cdots + m_k = m$. Losujeme n míčků z urny bez vracení. Označme opět $X_0, X_1, \dots, X_k$ počet vylosovaných míčků barvy $\beta_0, \beta_1, \dots, \beta_k$. Provedeme stejné úvahy jako při odvození negativně multinomického rozdělení. Čili budeme zvyšovat n, dokud X_0 nedosáhne předepsané hodnoty x_0. Bude nás zajímat rozdělení náhodného vektoru $\boldsymbol{X} = (X_1, \dots, X_k)'$, když $X_0 = x_0$.

Analogickým postupem jako v předchozím a s použitím (3.3) pro $n = \sum_{j=1}^{k} x_j + x_0 - 1$ dostáváme

$$P(X_1 = x_1, \dots, X_k = x_k) = \frac{\binom{m_0}{x_0} \prod_{j=1}^{k} \binom{m_j}{x_j}}{\binom{m}{\sum_{j=1}^{k} x_j + x_0 - 1}} = \frac{\prod_{j=0}^{k} \binom{m_j}{x_j}}{\binom{m}{\sum_{j=0}^{k} x_j - 1}},$$

$$\text{kde } \sum_{j=1}^{k} x_j \le m - x_0 + 1.$$

Náhodný vektor $\boldsymbol{X} = (X_1, \dots, X_k)'$ má *mnohorozměrné negativně hypergeometrické rozdělení*.

Kapitola 4

Modely s náhodným vracením

Exisuje celá řada zobecnění urnových modelů, jde o zcela přirozená zobecnění, např. po každém losování přidáme do urny (odebereme z urny) míčky opačné (stejné) barvy; jejich počet může záviset na barvě nebo na pořadí losování; nebo můžeme uvažovat, že v urně máme míčky více než dvou barev.

4.1 Pólyovo - Eggenbergerovo schéma

Mějme urnu obsahující a bílých a b černých míčků. Losujeme míček z urny, po vylosování míček vrátíme zpět do urny a navíc do ní přidáme s míčků stejné barvy. Tento proces opakujeme n krát. Označme X náhodnou veličinu popisující počet vytažených bílých míčků. Budeme hledat rozdělení náhodné veličiny X.

Vezměme si nějaké $k = 0, 1, \ldots, n$. Předpokládejme nejprve, že při losování bylo nejprve vylosováno k bílých míčků a až poté $n - k$ černých míčků. Pravděpodobnost, že v prvních k tazích bylo vylosováno k bílých míčků je

$$\frac{a(a+s)\cdots[a+(k-1)s]}{(a+b)(a+b+s)\cdots[a+b+(k-1)s]}$$

a pravděpodobnost, že v dalších $n-k$ tazích bylo vylosováno $n-k$ černých míčků je

$$\frac{b(b+s)\cdots[b+(n-k-1)s]}{(a+b+ks)[a+b+(k+1)s]\cdots[a+b+(n-1)s]}.$$

Čili pravděpodobnost, že bylo nejprve vylosováno k bílých míčků a až poté $n-k$ černých míčků je

$$\frac{a(a+s)\cdots[a+(k-1)s]b(b+s)\cdots[b+(n-k-1)s]}{(a+b)(a+b+s)\cdots[a+b+(n-1)s]}. \tag{4.1}$$

Bude-li nyní pořadí barev vylosovaných míčků jiné, zůstane výsledná pravděpodobnost stejná, neboť se pouze změní pořadí činitelů v čitateli a jmenovateli v (4.1). Počet možností, jak vylosovat k bílých míčků z n je $\binom{n}{k}$, tedy pro $k = 0, 1, \ldots, n$ platí

$$P(X = k) = \binom{n}{k} \frac{a(a+s)\cdots[a+(k-1)s]b(b+s)\cdots[b+(n-k-1)s]}{(a+b)(a+b+s)\cdots[a+b+(n-1)s]}. \tag{4.2}$$

Náhodná veličina X má tzv. *Pólyovo-Eggenbergerovo rozdělení.*

Uveďme si dva speciální případy: *Bernoulliho schéma* ($s = 0$), které odpovídá výběru s vracením a *Pearsonovo schéma* ($s = -1$), což je výběr bez vracení.

V dalším příkladě si ukážeme aplikaci Pólyova-Eggenbergerova schématu v teorii martingalů. Připomeňme proto základní pojmy a značení.

Definice 1 *Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. Buď $T \subset \mathbb{R}$ neprázdná indexová množina. Když pro každé $t \in T$ je $\mathcal{F}_t \subset \mathcal{A}$ σ-algebra a $\mathcal{F}_t \subset \mathcal{F}_s$, pokud $t < s, s \in T$, pak řekneme, že $(\mathcal{F}_t, t \in T)$ je filtrace.*

Definice 2 *Nechť $(\mathcal{F}_t, t \in T)$ je filtrace. Když X_t jsou $\mathcal{F}_t$-měřitelné reálné náhodné veličiny pro každé $t \in T$. Pak řekneme, že $(X_t, t \in T)$ je $(\mathcal{F}_t, t \in T)$-adaptovaný proces.*

Definice 3 (Martingal) *Když $(X_t, t \in T)$ je $(\mathcal{F}_t, t \in T)$-adaptovaný proces a $X_t \in L_1$ pro každé $t \in T$, pak říkáme, že $(X_t, t \in T)$ je $(\mathcal{F}_t, t \in T)$-martingal, jestliže $E[X_t|\mathcal{F}_s] = X_s$ s.j. pro každé $s < t, s, t \in T$.*
Řekneme, že $(X_t, t \in T)$ je martingal, když $(X_t, t \in T)$ je $(\mathcal{S}_t, t \in T)$-martingal pro přirozenou filtraci $\mathcal{S}_t = \sigma(X_s, s \leq t, s, t \in T)$.

Je-li indexová množina T spočetná, dá se martingalová vlastnost ověřovat snadněji než přímo z definice.

Tvrzení 1 *Nechť $(X_n, n \in \mathbb{N})$ je $\mathcal{F}_n$-adaptovaná posloupnost pro nějakou filtraci $(\mathcal{F}_n, n \in \mathbb{N})$. Pokud pro každé $n, n+1 \in \mathbb{N}$ platí $E[X_{n+1}|\mathcal{F}_n] = X_n$ s.j., pak $(X_n, n \in \mathbb{N})$ je $\mathcal{F}_n$-martingal.*

Důkaz. Musíme ukázat martingalovou vlastnost pro libovolné $n, n+k \in \mathbb{N}$. Budeme postupovat matematickou indukcí podle k.
Pro $k = 1$ platí podle předpokladu. Předpokládejme,že to platí pro k a chceme to ukázat pro $k+1$. Počítejme

$$\mathrm{E}[X_{n+1}|\mathcal{F}_{n-k-1}] = \mathrm{E}[\mathrm{E}[X_n|\mathcal{F}_{n-k}]|\mathcal{F}_{n-k-1}] = \mathrm{E}[X_{n-k}|\mathcal{F}_{n+k-1}] = X_{n-k-1} \quad \text{s.j.}$$

Přičemž ve druhé rovnosti jsme použili indukční předpoklad a ve třetí předpoklad tvrzení. $\square$

Příklad 7 (Martingal) *Uvažujme Pólyovo-Eggenbergerovo schéma s $s = 1$, tj. po vytažení vrátíme míček do urny spolu s dalším míčkem stejné barvy. Označme S_n relativní četnost bílých míčků v urně po n-tém tahu. Dokažme, že $(S_n, n \in \mathbb{N})$ je martingal.*

Řešení

Zaveďme pomocné n.v. X_n tak, že

$$X_n = \begin{cases} 1 & \text{je-li v } n\text{-tém tahu vylosován bílý míček,} \\ 0 & \text{je-li v } n\text{-tém tahu vylosován černý míček.} \end{cases}$$

Pak platí:

$$\mathrm{E}[X_{n+1}|X_1, \dots, X_n] = P(X_{n+1} = 1|X_1, \dots, X_n) = \frac{a + \sum_{j=1}^n X_j}{a+b+n} = S_n.$$

Pro podmíněnou střední hodnotu hodnotu S_n tedy platí:

$$\begin{aligned} \mathrm{E}[S_{n+1}|X_1, \dots, X_n] &= \mathrm{E}\left[\frac{a + \sum_{j=1}^n X_j + X_{n+1}}{a+b+n+1}\middle| X_1, \dots, X_n\right] = \\ &= \frac{a + \sum_{j=1}^n X_j}{a+b+n+1} + \frac{\mathrm{E}[X_{n+1}|X_1, \dots, X_n]}{a+b+n+1} = \\ &= \frac{a + \sum_{j=1}^n X_j + \frac{a+\sum_{j=1}^n X_j}{a+b+n}}{a+b+n+1} = S_n. \end{aligned}$$

Ukázali jsme, že $(S_n, n \in \mathbb{N})$ je $\sigma(X_1, \dots, X_n)$-martingal. Ale protože platí $\sigma(X_1, \dots, X_n) = \sigma(S_1, \dots, S_n)$, je $(S_n, n \in \mathbb{N})$ martingal. $\triangle$

Zůstaňme ještě na chvíli u modelu s $s = 1$. Pro tento případ se dá vzorec (4.2) zapsat velice elegantním způsobem

$$P(X = k) = \binom{n}{k} \frac{\Gamma(a+b)\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a)\Gamma(b)\Gamma(a+b+n)} = \binom{n}{k} \frac{B(a+k, b+n-k)}{B(a,b)}.$$

Pro s obecné, lze vydělením čitatele a jmenovatele s v (4.2) odvodit obdobný vzorec.

$$P(X = k) = \binom{n}{k} \frac{B(\frac{a}{s} + k, \frac{b}{s} + n - k)}{B(\frac{a}{s}, \frac{b}{s})}.$$

Dále označme B_i jev, že v i-tém tahu byl vytažen bílý míček. Pro jakékoliv s platí následující tvrzení.

Tvrzení 2 *Je-li $a + b + (n-1)s > 0$, pak platí*

$$P(B_i) = \frac{a}{a+b}, \qquad i = 1, 2, \dots, n. \tag{4.3}$$

Důkaz. Označme si $B_{i,j}, i = 1, 2, \dots, n, j = 0, 1, \dots, n-1$ jev, že v i-tém tahu byl vylosován bílý míček a v ostatních $n-1$ tazích bylo vytaženo právě j bílých míčků. Pro každé $i = 1, 2, \dots, n$ jsou jevy $B_{i,0}, B_{i,1}, \dots, B_{i,n-1}$ po dvou disjunktní a platí $B_i = \bigcup_{j=0}^{n-1} B_{i,j}$ pro každé $i = 1, 2, \dots, n$. S využitím vzorce (4.2) určíme pravděpodobnost jevu $B_{i,j}$

$$P(B_{i,j}) = \binom{n-1}{j} \frac{a(a+s)\cdots(a+js)b(b+s)\cdots[b+(n-j-2)s]}{(a+b)(a+b+s)\cdots[a+b+(n-1)s]}.$$

Pro zjednodušení zápisu si zaveďme nové značení pro $x \in \mathbb{N}, s \in \mathbb{Z}$

$$x^{[k]} := x(x+s) \cdot \ldots \cdot [x+(k-1)s] \qquad k = 1, 2, \ldots .$$

Přepišme si v novém značení pravděpodobnost jevu $B_{i,j}$

$$P(B_{i,j}) = \binom{n-1}{j} \frac{a^{[j+1]} b^{[n-j-1]}}{(a+b)^{[n]}}.$$

Odtud již

$$\begin{aligned} P(B_i) &= \sum_{j=0}^{n-1} \binom{n-1}{j} \frac{a^{[j+1]} b^{[n-j-1]}}{(a+b)^{[n]}} \\ &= \frac{a}{a+b} \sum_{j=0}^{n-1} \binom{n-1}{j} \frac{(a+s)^{[j]} b^{[n-j-1]}}{(a+b+s)^{[n-1]}} \\ &= \frac{a}{a+b}. \end{aligned}$$

Poslední rovnost plyne z faktu, že

$$\sum_{j=0}^{n-1} \binom{n-1}{j} \frac{(a+s)^{[j]} b^{[n-j-1]}}{(a+b+s)^{[n-1]}} = 1,$$

neboť odpovídá situaci v (4.2) s tím, že na začátku bylo v urně $(a+s)$ bílých a b černých míčků. $\square$

4.2 Mnohorozměrné Pólyovo-Eggenbergerovo schéma

Uvažujme následující zobecnění Pólyova-Eggenbergerova schématu. Nechť urna obsahuje c_j míčků barvy β_j, $j = 1, 2, \ldots, k$. Po každém vylosování je míček vrácen zpět do urny spolu s dalšími s míčky téže barvy. Označme X_j počet vylosováných míčků barvy β_j v prvních n tazích.

Rozdělení náhodného vektoru $\boldsymbol{X} = (X_1, \ldots, X_k)'$ určíme analogickou úvahou jako při odvození vztahu (4.2). Tedy

$$P(X_1 = x_1, \ldots, X_k = x_k) = \binom{n}{x_1 \ldots x_k} \frac{\prod_{i=1}^k c_i(c_i+s) \cdots [c_i + (x_i-1)s]}{c(c+s) \cdots [c+(n-1)s]},$$

$$\text{kde} \quad \sum_{i=1}^k c_i = c \quad \text{a} \quad \sum_{i=1}^k x_i = n.$$

Náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_k)'$ má *mnohorozměrné Pólyovo - Eggenberger-gerovo rozdělení.*

4.3 Zobecnění Pólyova - Eggenbergerova schématu

Nyní si zobecníme základní Pólyovo-Eggenbergerovo schéma. Předpokládejme, že platí vše, co platilo pro základní model s tou výjimkou, že pokud byl vylosován bílý míček přidáme do urny α_a bílých a β_a černých míčků, resp. α_b bílých a β_b černých míčků, byl-li vylosován černý míček. Připomeňme, že tažený míček vratíme zpět do urny. Pro přehlednost je tento proces znázorněn v následující tabulce

		Vylosovaný bílá	míček černá
Počet míčků	bílá	α_a	α_b
přidaných do urny	černá	β_a	β_b

Tabulka 4.1: Znázornění zobecněného Pólyova-Eggenbergerova schématu

Je vidět, že základní Pólyovo-Eggenbergerovo schéma odpovídá zobecněnému schématu s $\alpha_a = \beta_b = s$ a $\alpha_b = \beta_a = 0$. Uveďme si dva speciální případy nazvané podle jejich autorů.

Styve(1965) $\alpha_a = -1; \alpha_b = 0; \beta_a = 1; \beta_b = 0$

Tj. bílý míček je nahrazen černým a černý je vrácen zpět do urny. Tento model může sloužit k vyšetřování vad přístrojů, tj. zvolíme libovolně jeden ze zkoumaných přístrojů, pokud je vadný, nahradíme ho funkčním, v opačném případě jej ponecháme. Jiný model spočívá například ve značkování ptáků (nebo obecněji nějaké populace), kdy neoznačkovaní jedinci jsou postupně značkováni.

Budeme se zajímat o pravděpodobnost, že v prvních n tazích vylosujeme k bílých míčků. Poznamenejme nejprve, že celkový počet míčků v urně zůstává konstantní $a + b$. Dále si všimněme, že obsah urny zůstává stejný, dokud losujeme černé míčky. Pokud již bylo vylosováno k bílých míčků, pak pravděpodobnost vytažení bílého míčku v dalším tahu je $\frac{a-k}{a+b}$. Tudíž hledaná pravděpodobnost je rovna součtu pravděpodobností přes všechna $\binom{n}{k}$ možná pořadí, jak z n míčků vylosovat k bílých. Tyto dílčí pravděpodobnosti jsou tvaru

$$\frac{a(a-1)\cdot\ldots\cdot(a-k+1)b^{m_0}(b+1)^{m_1}\cdot\ldots\cdot(b+k)^{m_k}}{(a+b)^n}.$$

Tato pravděpodobnost odpovídá situaci, že v prvním, druhém, ..., m_0-tém tahu byl vylosován černý míček, v $(m_0 + 1)$-ním tahu bílý míček a dále byly taženy černé míčky až do tahu $(m_0 + m_1 + 1)$ včetně a v $(m_0 + m_1 + 2)$-hém tahu byl vytažen bílý míček atd. Přičemž platí $\sum_{j=0}^{k} m_j = n - k$. Tedy

$$P(X = k) = \frac{a(a-1)\cdot \ldots \cdot (a-k+1)}{(a+b)^n} \sum_{m_0,\ldots,m_k} b^{m_0}(b+1)^{m_1} \cdot \ldots \cdot (b+k)^{m_k}, \tag{4.4}$$

kde sčítáme přes všechny $m_0, \ldots, m_k$ takové, že $\sum_{j=0}^{k} m_j = n-k$. Všimněme si, že suma ve vzorci (4.4) je rovna koeficientu u t^{n-k} v rozvoji

$$\prod_{j=0}^{k} \frac{1}{1-(b+j)t}. \tag{4.5}$$

Provedeme rozklad (4.5) na parciální zlomky

$$\prod_{j=0}^{k} \frac{1}{1-(b+j)t} = \frac{A_0}{1-bt} + \frac{A_1}{1-(b+1)t} + \cdots + \frac{A_k}{1-(b+k)t}.$$

Metodami matematické analýzy se dá ukázat (viz [3]), že

$$A_j = \binom{k}{j} \frac{(-1)^{k-j}(b+j)^k}{k!}, \qquad j = 0, 1, \ldots, k.$$

Tedy máme rozklad (4.5)

$$\prod_{j=0}^{k} \frac{1}{1-(b+j)t} = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \frac{(b+j)^k}{1-(b+j)t}.$$

Tedy hledaný koeficient je

$$\frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} (b+j)^n.$$

Z (4.4) máme pro $k = 0, 1, \ldots, n$

$$P(X = k) = \frac{a(a-1)\cdot \ldots \cdot (a-k+1)}{(a+b)^n k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} (b+j)^n. \tag{4.6}$$

Poznámka

Často se užívá následující značení: $a^{(k)} = a(a-1)\cdot \cdots \cdot (a-k+1)$. V tomto značení (4.6) přejde v

$$P(X = k) = \frac{a^{(k)}}{(a+b)^n k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} (b+j)^n.$$

Kaminsky(1975) $\alpha_b = \beta_a = \beta_b = 0; \alpha_a = -1$

Tj. bílé míčky nejsou vraceny do urny, zato černé jsou. Tento model odpovídá odebrání vadných přístrojů (ty už však nejsou nahrazovány funkčními).

Nyní se budeme zajímat o počet vytažených černých míčků než bylo vylosováno g míčků bílých. Označme $Y =$ (počet tahů potřebných v vytažení g bílých míčků) - g. To, že $P(Y + g = k)$ znamená, že jsme v k-tém tahu vytáhli bílý míček a v ostatních $(k-1)$ tazích jsme vytáhli $(g-1)$ bílých míčků. Provedeme-li stejnou úvahu jako při odvození předchozího případu a využijeme značení z poznámky, dostáváme

$$P(Y + g = k) = \frac{a^{(g)}}{(a+b)^{(g)}} \sum_{m_0,\dots,m_{g-1}} \prod_{j=0}^{g-1} \left(\frac{b}{a+b-j}\right)^{m_j}, \qquad (4.7)$$

kde sčítáme přes všechny $m_0, \dots, m_{g-1}$ takové, že $\sum_{j=0}^{g-1} m_j = k - g$. Suma ve vzorci (4.7) je rovna koeficientu u t^{k-g} v rozvoji

$$\prod_{j=0}^{g-1} \left[1 - \frac{bt}{a+b-j}\right]^{-1}. \qquad (4.8)$$

Stejným postupem jako při minulém odvození provedeme rozklad (4.8) na parciální zlomky (viz [3])

$$\prod_{j=0}^{g-1} \left[1 - \frac{bt}{a+b-j}\right]^{-1} = \sum_{j=0}^{g-1} A_j \left(1 - \frac{bt}{a+b-j}\right)^{-1},$$

kde

$$A_j = \frac{(a+b)^{(g)}}{(g-1)!} (-1)^{g-j-1} \binom{g-1}{j} \frac{1}{a+b-j}.$$

Pak z (4.7) dostaneme hledanou pravděpodobnost

$$\begin{aligned} P(Y + g = k) &= \frac{a^{(g)}}{(g-1)!} \sum_{j=0}^{g-1} (-1)^{g-j-1} \binom{g-1}{j} \frac{1}{a+b-j} \left(\frac{b}{a+b-j}\right)^{k-g} \\ &= b^{k-g} \frac{a^{(g)}}{(g-1)!} \sum_{j=0}^{g-1} (-1)^{g-j-1} \binom{g-1}{j} (a+b-j)^{-(k-g+1)}. \end{aligned}$$

4.4 Ehrenfestův model

Začněme tuto kapitolu netradičně příkladem.

Příklad 8 (D.Bernoulli, 1768) *Uvažujme dvě urny A,B. Urna A obsahuje k bílých míčků a urna B obsahuje k černých míčků. Nejprve náhodně vylosujeme jeden míček z urny A a dáme ho do urny B. Poté náhodně losujeme míček z urny B a umístíme ho do urny A. Tento proces dále opakujeme. Určeme střední hodnotu E_r počtu bílých míčků v první urně poté, co bylo vylosováno r míčků z každé urny.*

Řešení
Označme G_r počet bílých míčků v urně A poté, co bylo vylosováno r míčků z každé urny. Po dalších dvou tazích (jednom z urny A a jednom z urny B) se počet bílých míčků v urně A

- Zvýší o 1 s pravděpodobností $\frac{k-G_r}{k} \cdot \frac{k-G_r}{k+1}$.
- Sníží o 1 s pravděpodobností $\frac{G_r}{k} \cdot \frac{G_r}{k+1}$.
- Nezmění s pravděpodobností $1 - \frac{(k-G_r)^2+G_r^2}{k(k+1)}$.

Odtud střední hodnota zvýšení počtu bílých míčků v urně A je

$$\frac{(k-G_r)^2}{k(k+1)} - \frac{G_r^2}{k(k+1)} = \frac{k-2G_r}{k+1}.$$

Tedy s využitím toho, že střední hodnota rozdílu je rozdíl středních hodnot je

$$E_{r+1} - E_r = \frac{k-2E_r}{k+1}.$$

Po úpravě dostáváme lineární diferenční rovnici

$$E_{r+1} - \frac{k-1}{k+1}E_r = \frac{k}{k+1}. \tag{4.9}$$

Řešení homogenní soustavy

$$E_{r+1} - \frac{k-1}{k+1}E_r = 0$$

je $E_r = C \cdot \left(\frac{k-1}{k+1}\right)^r$.
Hledejme partikulární řešení (4.9) ve tvaru $E_r = c$. Pak z

$$c - \frac{k-1}{k+1} \cdot c = \frac{k}{k+1}$$

dostáváme partikulární řešení $E_r = \frac{k}{2}$.
Obecné řešení (4.9) je tedy tvaru

$$E_r = C \cdot \left(\frac{k-1}{k+1}\right)^r + \frac{k}{2}.$$

Dosadíme-li pro $r = 0$ $E_r = k$ máme $C = \frac{k}{2}$. Čili

$$E_r = \frac{k}{2} \cdot \left[\left(\frac{k-1}{k+1}\right)^r + 1\right]. \qquad \triangle$$

V tomto modelu zůstával počet míčků v každé urně konstantní, Ehrenfest uvažoval následující model, v němž se počet míčků v urnách může měnit.

Mějme opět dvě urny A,B, přičemž v urně A je k_1 míčků a v urně B k_2 míčků. Náhodně zvolíme míček, ať už je v urně A či B, vylosujeme jej a umístíme zpět do opačné urny. Tento proces dále opakujeme. Tento model je ekvivalentní Pólyovu-Eggenbergerovu schématu popsaného v kapitole 4.3. Neboť si můžeme představit, že máme jednu urnu a v ní k_1 bílých a k_2 černých míčků, přičemž po vylosování míčku z urny je nahrazen míčkem opačné barvy.

Budeme se zabývat rozdělením počtu K_n míčků v urně A po n tazích. Položme $P_n(k) = P(K_n = k)$ a $k_0 = k_1 + k_2$. To, že po n tazích bude k míčků v urně A znamená, že po $n-1$ tazích bylo v urně A $k+1$ míčků a v n-tém tahu jsme vytáhli nějaký míček z urny A, nebo po $n-1$ tazích bylo v urně A $k-1$ míčků a v n-tém tahu jsme vytáhli nějaký míček z urny B. Neboli

$$P_n(k) = \frac{k+1}{k_0} P_{n-1}(k+1) + \frac{k_0 - k + 1}{k_0} P_{n-1}(k-1).$$

Můžeme ještě doplnit počáteční podmínky $P_0(k_1) = 1, P_0(k) = 0$.

Předpokládejme, že pro $n \to \infty$ $P_n(k) \to P(k)$ pro každé $k = 0, 1, \ldots, k_0$. Pak máme

$$P(k) = \frac{k+1}{k_0} P(k+1) + \frac{k_0 - k + 1}{k_0} P(k-1).$$

Odtud (zpětným dosazením můžeme ověřit)

$$\frac{P(k+1)}{P(k)} = \frac{k_0 - k}{k+1}. \tag{4.10}$$

Zvolme $P(0) = K$, kde K je nějaká konstanta. Pak $P(1) = K \cdot k_0$ a $P(2) = K \cdot \frac{k_0(k_0-1)}{2}$,atd. Konečnou indukcí nyní ukážeme, že pro $k = 0, 1, \ldots, k_0$ platí

$$P(k) = K \cdot \binom{k_0}{k}. \tag{4.11}$$

Již jsme ukázali, že pro $k = 0, 1, 2$ to platí. Předpokládejme, že to platí pro nějaké $0 < k < k_0$. Pak podle (4.10) a (4.11) je

$$P(k+1) = \frac{k_0 - k}{k+1} \cdot P(k) = \frac{k_0 - k}{k+1} \cdot K \cdot \binom{k_0}{k} = K \cdot \binom{k_0}{k+1}.$$

Nyní zbývá nalézt konstantu K. Neboť $\sum_{k=0}^{k_0} P(k) = 1$ máme

$$1 = K \cdot \sum_{k=0}^{k_0} \binom{k_0}{k} = K \cdot 2^{k_0}.$$

Výsledná pravděpodobnost je

$$P(k) = 2^{-k_0} \cdot \binom{k_0}{k} \qquad k = 0, 1, \ldots, k_0.$$

Vidíme, že limitním rozdělením K_n je *binomické* s parametry k_0 a $\frac{1}{2}$.

Kapitola 5

Umisťovací schémata

Až doposud jsme předpokládali, že máme urnu a z ní losujeme míčky. Nyní budeme mít několik uren a míčky do nich budeme umisťovat. Tyto modely mají široké uplatnění ve fyzice. Z tohoto důvodu se zde užívá označení částice a přihrádka místo míčku a urny. I my se proto budeme tohoto označení držet.

5.1 Základní umisťovací model

Mějme m přihrádek a předpokládejme, že částice přilétají jedna po druhé a každá z n částic náhodně skončí v jedné přihrádce. Pravděpodobnost, že skončí v j-té přihrádce je p_j ($\sum_{j=1}^m p_j = 1$).

Nechť náhodná veličina X_j udává počet částic v j-té přihrádce. Pak pro náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_m)'$ platí

$$P(X_1 = x_1, \ldots, X_m = x_m) = \binom{n}{x_1 \cdots x_m} \prod_{j=1}^{m} p_j^{x_j}, \qquad \sum_{j=1}^{n} x_j = n.$$

Tedy náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_m)'$ má *multinomické rozdělení* s parametry $n, p_1, \ldots, p_m$. Zabývejme se nyní určením pravděpodobnosti, že určitý počet přihrádek je prázdný.

Pravděpodobnost, že j-tá přihrádka bude prázdná je $(1 - p_j)^n$. Podobně pravděpodobnost, že i-tá a j-tá ($i \neq j$) budou prázdné je $(1 - p_i - p_j)^n$. Odtud již snadno plyne, že pravděpodobnost, že právě jedna z přihrádek i nebo j bude prázdná je $(1 - p_i)^n + (1 - p_j)^n - (1 - p_i - p_j)^n$.

Označme nyní Y náhodnou veličinu popisující počet prázdných přihrádek. Určeme nejprve, pravděpodobnost, že žádná přihrádka není prázdná (tj. v každé přihrádce je nějaká částice).

Pokud $n < m$, tj. částic je méně než přihrádek, pak zřejmě $P(Y = 0) = 0$, budeme proto dále předpokládat, že $n \geq m$. Označme A_j jev, že j-tá přihrádka je prázdná a $A = \bigcup_{j=1}^m A_j$ jev, že aspoň jedna přihrádka je prázdná. Podle principu inkluze a exkluze s přihlédnutím k tomu, že $P\left(\bigcap_{j=1}^m A_j\right) = 0$

dostáváme

$$
\begin{aligned}
P(Y=0) = P(A^C) = 1 - P(A) = 1 - P\left(\bigcup_{j=1}^{m} A_j\right) = \\
= \; 1 - \sum_{j=1}^{m} P(A_j) + \sum_{i<j}^{m} P(A_i \cap A_j) - \ldots + (-1)^{m-1} \sum_{j=1}^{m} P\left(\bigcap_{i \neq j}^{m} A_i\right) = \\
= \; 1 - \sum_{j=1}^{m} (1-p_j)^n + \sum_{i<j}^{m} (1-p_i-p_j)^n - \ldots + (-1)^{m-1} \sum_{j=1}^{m} p_j^n. \qquad (5.1)
\end{aligned}
$$

Označme dále $P_{1,2,\ldots,k}$ pravděpodobnost, že právě první, druhá,..., k-tá přihrádka jsou prázdné. Tato pravděpodobnost je rovna součinu pravděpodobnosti, že první, druhá,...,k-tá přihrádka jsou prázdné a pravděpodobnosti, že mezi $(k+1)$-ní,...,m-tou přihrádkou není žádná přihrádka prázdná za podmínky, že prvních k přihrádek prázdných je. První pravděpodobnost je rovna $\left(1 - \sum_{i=1}^{k} p_i\right)^n$. Druhá odpovídá situaci v (5.1) s $(m-k)$ přihrádkami s pravděpodobnostmi $p_j(1 - \sum_{i=1}^{k} p_i)^{-1}$ pro $j = k+1, \ldots, m$. Odtud již

$$
\begin{aligned}
P_{1,2,\ldots,k} \; = \; & \left(1 - \sum_{i=1}^{k} p_i\right)^n - \sum_{j=k+1}^{m} \left(1 - p_j - \sum_{i=1}^{k} p_i\right)^n + \\
+ \; & \sum_{j \neq l > k}^{m} \left(1 - p_j - p_l - \sum_{i=1}^{k} p_i\right)^n - \ldots.
\end{aligned}
$$

Pokud nyní sečteme $P_{a_1,a_2,\ldots,a_k}$ přes všechny možné k-prvkové podmnožiny jevů $A_1, A_2, \ldots, A_m$ dostaneme pravěpodobnosti, že počet prázdných přihrádek je k.

$$
\begin{aligned}
P(Y=k) \; = \; & \sum_{\binom{a_1 \ldots a_m}{k}} P_{a_1,a_2,\ldots,a_k} = \\
= \; & \sum_{\binom{a_1 \ldots a_m}{k}} \left(1 - \sum_{i=1}^{k} p_{a_i}\right)^n - \binom{k+1}{k} \sum_{\binom{a_1 \ldots a_m}{k+1}} \left(1 - \sum_{i=1}^{k+1} p_{a_i}\right)^n + \\
+ \; & \binom{k+2}{k} \sum_{\binom{a_1 \ldots a_m}{k+2}} \left(1 - \sum_{i=1}^{k+2} p_{a_i}\right)^n - \ldots + (-1)^{m-k} \binom{m}{k} \sum_{i=1}^{m} p_i^n.
\end{aligned} \qquad (5.2)
$$

Kde $\sum_{\binom{a_1 \ldots a_m}{k}}$ značí součet přes všechny k-prvkové podmnožiny množiny $(1, 2, \ldots, m)$.
Snadno se lze přesvědčit, že pro $k = 0$ dostaneme vzorec (5.1).

Ve speciálním případě kdy $p_1 = p_2 = \cdots = p_m = m^{-1}$ (5.2) přejde s využitím $\binom{m}{k+r}\binom{k+r}{k} = \binom{m}{k}\binom{m-k}{r}$ v

$$\begin{aligned} P(Y=k) &= \binom{m}{k}\left[\left(1-\frac{k}{m}\right)^n - \binom{m-k}{1}\left(1-\frac{k+1}{m}\right)^n + \cdots \right. \\ &+ \left.(-1)^{m-k}\binom{m-k}{m-k-1}\left(1-\frac{m-1}{m}\right)^n\right], \qquad k=0,1,\ldots,m. \end{aligned}$$

Rozdělení náhodné veličiny Y je nazýváno *klasické okupační rozdělení.* Někdy se však jako *okupační rozdělení* uvádí rozdělení počtu neprázdných (okupovaných) uren $X = m - Y$

$$P(X=k) = m^{-n}\binom{m}{k}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^n, \qquad k=0,1,\ldots,m. \tag{5.3}$$

Spočítejme ještě střední hodnotu a rozptyl náhodné veličiny Y. Tu si rozepišme pomocí indikátorových náhodných veličin $Y = I_1 + \cdots + I_m$, kde

$$I_j = \begin{cases} 1 & \text{je-li } j\text{-tá přihrádka prázdná,} \\ 0 & \text{jinak.} \end{cases}$$

Již jsme ukázali, že $P(I_j = 1) = (1-p_j)^n$ a $P(I_i = 1, I_j = 1) = (1-p_i-p_j)^n$. Jednoduchým výpočtem dostaneme

$$\begin{aligned} &\mathrm{E}I_j = (1-p_j)^n, \\ &\mathrm{E}I_iI_j = (1-p_i-p_j)^n, \qquad i \neq j, \\ &\mathrm{var}(I_j) = (1-p_j)^n[1-(1-p_j)^n], \\ &\mathrm{cov}(I_i, I_j) = (1-p_i-p_j)^n - (1-p_i)^n(1-p_j)^n, \qquad i \neq j. \end{aligned}$$

Odtud již okamžitě plyne, že

$$\mathrm{E}Y = \sum_{j=1}^{m}\mathrm{E}I_j = \sum_{j=1}^{m}(1-p_j)^n,$$

$$\begin{aligned} \mathrm{var}(Y) &= \sum_{j=1}^{m}\mathrm{var}(I_j) + 2\sum_{i<j}^{m}\mathrm{cov}(I_i, I_j) = \\ &= \sum_{j=1}^{m}(1-p_j)^n\left[1-\sum_{j=1}^{m}(1-p_j)^n\right] + 2\sum_{i<j}^{m}(1-p_i-p_j)^n. \end{aligned}$$

Uveďme si jeden příklad, který má uplatnění v chemii, i když to tak na první pohled nemusí vypadat.

Příklad 9 (P.Hagis, C.Schmidt) *V místnosti je n mužů, n žen postupně vstoupí do místnosti a vybere si náhodně jednoho muže a chytne ho za ruku, pokud muž ještě nějakou volnou ruku má, tj. každý muž může být vybrán maximálně dvěma ženami. Určete rozdělení počtu mužů, kteří mají (i) obě ruce volné (ii) jednu ruku volnou (iii) žádnou ruku volnou.*

Řešení
Nechť N_j značí počet mužů s j volnýma rukama ($j = 0, 1, 2$). Pak

$$\sum_{j=0}^{2} jN_j = n = \sum_{j=0}^{2} N_j.$$

Odtud je vidět, že $N_2 = N_0$ a $N_1 = n - 2N_2$. Takže stačí, když určíme rozdělení N_2.

Počet možností, jak si mohou ženy vybrat mužovy ruce je $\binom{2n}{n}$. To, že $N_2 = r$ znamená, že r mužů nedrží žádnou ženu za ruku, tedy n žen si vybírá jen z rukou $n - r$ mužů, tudíž

$$P(N_2 = r) = \frac{\binom{2(n-r)}{n}}{\binom{2n}{n}}, \qquad r = 0, 1, \ldots, \left\lfloor \frac{n}{2} \right\rfloor. \qquad \triangle$$

Chemická aplikace vyplyne ihned na povrch, pokud si místo mužů a žen představíme atomy a místo mužových volných rukou volné vazby.

Na závěr této kapitoly si uveďme některé speciální modely, které mají uplatnění ve statistické fyzice při popisu stavů systému různých částic.

Maxwellův-Boltzmannův model

Předpokládejme, že máme n částic a m přihrádek. Nechť je dále splněno:

- Částice jsou vzájemně rozlišitelné.
- Pro každou částici umíme určit přihrádku, ve které se částice nachází.
- Stav systému je popsán polohou částic v přihrádkách.
- Všechny stavy jsou stejně pravděpodobné.

Počet stavů tohoto systému, tj. počet rozmístění n částic do m přihrádek, je zřejmě m^n. Dále pravděpodobnost, že v j-té přihrádce je právě k částic je rovna

$$p_{n,m}(k) = \frac{\binom{n}{k}(m-1)^{n-k}}{m^n}. \tag{5.4}$$

Zkoumejme jaké je limitní chování této pravděpodobnosti, jestliže $n \to \infty$, $m \to \infty$, přičemž $\frac{n}{m} \to \lambda$, kde $\lambda > 0$.

$$\lim p_{n,m}(k) = \frac{1}{k!} \lim \left(1 - \frac{1}{m}\right)^{n-k} \frac{n(n-1) \cdot \cdots \cdot (n-k+1)}{m^k} = \frac{\lambda^k}{k!} e^{-\lambda},$$

kde lim značí limitu při $n \to \infty$, $m \to \infty$, $\frac{n}{m} \to \lambda$. Vidíme, že limitním rozdělením je *Poisonovo rozdělení* s parametrem λ.

Zkoumejme dále, jaká je pravděpodobnost, že žádná přihrádka není prázdná. Mohli bychom provést úplně stejnou úvahu jako v úvodu této kapitoly nebo můžeme rovnou dosadit do vzorce (5.1)

$$\begin{aligned} P(Y=0) &= 1 - \binom{m}{1} \cdot \left(1 - \frac{1}{m}\right)^n + \binom{m}{2} \cdot \left(1 - \frac{2}{m}\right)^n - \dots \\ &+ (-1)^n \binom{m}{m-1} \cdot \left(1 - \frac{m-1}{m}\right)^n. \end{aligned}$$

Tento model se používá ve statistické mechanice, např. pro soustavy molekul plynů.

Boseův-Einsteinův model

Opět uvažujme n částic a m přihrádek a nechť je navíc splněno

- Částice jsou vzájemně nerozlišitelné.
- Pro každou přihrádku umíme určit počet částic, které se v ní nachází.
- Stav systému je popsán počtem částic v přihrádkách.
- Všechny stavy jsou stejně pravděpodobné.

Ukážeme, že počet stavů takto popsaného systému je

$$\binom{n+m-1}{m-1}. \tag{5.5}$$

Určení tohoto počtu není již tak snadné jako u Maxwellova-Boltzmannova modelu. Použijeme proto drobného triku. Následující obrázek znázorňuje jeden z možných stavů pro $n = 7$ a $m = 5$ (symbol o znázorňuje částici a symboly || přihrádku).

|o||oo|o|ooo|

Obrázek 5.1: Znázornění jednoho možného stavu v B-E modelu

Každý stav systému můžeme jednoznačně znázornit tímto obrázkem. Náš úkol se tedy redukuje na zjištění počtu možných rozložení symbolů o a | v našem diagramu, kde počet o je n a počet | je $m+1$. Ale protože na konci, resp. na začátku, musí být |, je počet možných rozložení o a | právě $\binom{n+m-1}{m-1}$, neboť umisťuje právě $m-1$ symbolů | na $n+m-1$ možných pozic.

Dále pravděpodobnost, že v j-té přihrádce je právě k částic je rovna

$$p_{n,m}(k) = \frac{\binom{n+m-2-k}{m-2}}{\binom{n+m-1}{m-1}}.$$

Jako v předchozím zkoumejme limitu, když $n \to \infty, m \to \infty$ a $\frac{n}{m} \to \lambda$, kde $\lambda > 0$.

$$
\begin{aligned}
\lim p_{n,m}(k) &= \lim \frac{\binom{n+m-2-k}{m-2}}{\binom{n+m-1}{m-1}} \\
&= \lim \frac{(m-1)n(n-1)\cdot \dots \cdot (n-k+1)}{(n+m-1)(n+m-2)\cdot \ldots \cdot (n+m+k-1)} \\
&= \frac{\lambda^k}{(1+\lambda)^{k+1}} = \frac{1}{1+\lambda} \cdot \left(\frac{\lambda}{1+\lambda}\right)^k,
\end{aligned}
$$

kde opět lim značí limitu při $n \to \infty$, $m \to \infty$, $\frac{n}{m} \to \lambda$. Limitním rozdělením je *geometrické rozdělení* s parametrem $\frac{1}{1+\lambda}$.

Spočítejme pro tento model pravděpodobnost, že žádná přihrádka není prázdná. To, že žádná přihrádka není prázdná znamená, že m částic z n je každá v jedné z m přihrádek. Ostatních $n - m$ částic pak může být rozmístěno do m přihrádek zcela libovolně. Protože všechny stavy jsou stejně pravděpodobné, je podle (5.5)

$$P(Y = 0) = \frac{\binom{n-1}{m-1}}{\binom{n+m-1}{m-1}}.$$

Tímto modelem bývá popisováno chování soustav fotonů.

Pro úplnost se ještě zmiňme o Fermiově - Diracově modelu, který vznikne z Boseova - Einsteinova modelu použitím Pauliho vylučovacího principu, podle něhož může být v každé přihrádce maximálně jedna částice.

Fermiův-Diracův model

Máme n částic a m přihrádek a nechť je navíc splněno

- Částice jsou vzájemně rozlišitelné.
- V každé přihrádce může být maximálně jedna částice.
- Stav systému je popsán seznamem obsazených přihrádek.
- Všechny stavy jsou stejně pravděpodobné.

Počet stavů tohoto systému je $\binom{m}{n}$. A pravděpodobnost, že v j-té přihrádce je právě jedna částice(více jich tam ani být nemůže) je

$$\frac{\binom{m-1}{n-1}}{\binom{m}{n}} = \frac{n}{m}.$$

Fermiův-Diracův model je vhodný k popisu chování elektronů, protonů, či neutronů.

Na závěr této kapitoly uveďme ještě jeden příklad.

Příklad 10 (Problém šatnářky) *Ctihodní pánové v počtu n přijdou na shromáždění, všichni v kloboucích, a odloží si své klobouky do šatny. Při odchodu šatnářka, možná ten den velmi roztržitá, možná dokonce z mizerného osvětlení oslepla, vydá každému z pánů náhodně jeden z klobouků. Jaká je pravděpodobnost, že žádný pán nedostane od šatnářky zpět svůj klobouk?*

Řešení

Budeme postupovat stejně jako při odvození vzorce (5.1) v úvodu této kapitoly. Označme

$$\begin{aligned} A_i &\ \ldots\ i\text{-tý pán dostane svůj klobouk}, \qquad i = 1, \ldots, n, \\ A &\ \ldots\ \text{aspoň jeden pán dostane svůj klobouk.} \end{aligned}$$

Pak zřejmě $A = \bigcup_{i=1}^{n} A_i$. Podle principu inkluze a exkluze je

$$\begin{aligned} P(A^C) &= 1 - P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \sum_{i=1}^{n} P(A_i) + \\ &+ \sum_{1 \le i < j \le n} P(A_i \cap A_j) - \ldots + (-1)^n P\left(\bigcap_{i=1}^{n} A_i\right). \end{aligned} \tag{5.6}$$

Spočítejme nyní pravděpodobnosti vyskytující se ve vzorci (5.6)

$$\begin{aligned} P(A_i) &= \frac{(n-1)!}{n!}, \qquad i = 1, \ldots, n, \\ P(A_i \cap A_j) &= \frac{(n-2)!}{n!}, \qquad i \neq j, \\ &\vdots \\ P\left(\bigcap_{i=1}^{n} A_i\right) &= \frac{1}{n!}. \end{aligned}$$

Tedy z (5.6) plyne

$$\begin{aligned} P(A^C) &= 1 - \binom{n}{1}\frac{(n-1)!}{n!} + \binom{n}{2}\frac{(n-2)!}{n!} - \ldots + (-1)^n \binom{n}{n}\frac{1}{n!} = \\ &= \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n \frac{1}{n!} = \sum_{i=2}^{n} (-1)^i \frac{1}{n!}. \end{aligned}$$

Poznamenejme ještě, že

$$\lim_{n \to \infty} P(A^C) = \sum_{n=2}^{\infty} (-1)^n \frac{1}{n!} = \frac{1}{e}. \qquad \triangle$$

Poznámka

Tato úloha bývá formulována více způsoby, například máme n obálek a n

dopisů, každý dopis vložíme náhodně do jedné obálky a zajímá nás pravděpodobnost, že se žádný dopis nedostane do správné obálky; nebo uvažujme večírek, na který každý z n hostů donese jeden dárek, dárky se pak náhodně rozdělí a hledáme pravděpodnost, že si žádný z hostů neodnese domů svůj dárek.

5.2 Znáhodněný model

V této kapitole si zobecníme model uvažovaný v minulé kapitole. Pro připomenutí: Máme m přihrádek a n částic do nich přilétajících. Předpokládejme, že si částice přihrádku zvolí náhodně, tj. s pravděpodobností m^{-1} skončí v jedné z m přihrádek. Každá částice může s pravděpodobností p zůstat v přihrádce, nebo s pravděpodobností $1-p$ z přihrádky vypadne. Označme X počet přihrádek, ve kterých zůstala alespoň jedna částice. Případ $p = 1$ odpovídá základnímu modelu (kapitola 5.1). Proto předpokládejme, že $p < 1$.

V minulé kapitole jsme odvodili, že pro případ $p = 1$ je podle (5.3) pravděpodobnost, že právě k přihrádek obsahuje nějakou částici

$$H_1(k,m,n) = m^{-n}\binom{m}{k}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^n.$$

Počet částic N, které nevypadnou z přihrádky, má binomické rozdělení s parametry n, p. Hledaná pravděpodobnost $H_p(k,m,n)$ je rovna střední hodnotě $H_1(k,m,N)$, kde N má binomické rozdělení s parametry n, p. Odtud již

$$\begin{aligned} H_p(k,m,n) &= \sum_{j=0}^{n}\binom{n}{j}p^j(1-p)^{n-j}H_1(k,m,j) = \\ &= \binom{m}{k}\sum_{i=0}^{k}(-1)^i\binom{k}{i}\sum_{j=0}^{n}\binom{n}{j}\left[\frac{(k-i)p}{m}\right]^j(1-p)^{n-j} = \\ &= \binom{m}{k}\sum_{i=0}^{k}(-1)^i\binom{k}{i}\left[1-p+\frac{p}{m}(k-i)\right]^n. \end{aligned}$$

Uveďme si nyní armádní aplikaci okupačního a znáhodněného okupačního rozdělení. Tato rozdělení se používají v situacích, kdy se zajímáme o zasažení co nejvíce z m terčů pomocí n střel. Ztotožněme si terče s urnami a střely s míčky a dále předpokládejme, že každý terč může být zasažen se stejnou pravděpodobností. Pak počet zasažených, resp. nezasažených terčů má klasické okupační rozdělení. Budeme-li předpokládat, že některé střely mohou minout svůj cíl, řekněme s pravděpodobností $1-p$, pak se jedná o zobecněné okupační rozdělení.

Nebo uvažujme následují situaci. V letecké bitvě uvažujme m bombardérů (urny), nebo obecně ofenzivní síly, a n raket (míčky), nebo obecně defenzivní síly. V této situaci se předpokládá, že každá raketa může napadnout pouze jeden bombardér. Nechť p je parametr zasažení, tj. pravděpodobnost, že bombardér bude zasažen. Pak počet obsazených uren odpovídá počtu zničených bombardérů.

Kapitola 6

Některé další aplikace

V této kapitole si uvedeme některé další aplikace urnových modelů, které nelze zařadit do žádné z předchozích kapitol.

6.1 Capture - recapture model

Tyto modely slouží k odhadům demografických parametrů jako jsou například velikost polulace, emigrace, imigrace. My se budeme zabývat odhadem počtu ryb f v jezeře následující metodou:

1. Chytíme n_0 ryb, označíme je a vrátíme je zpět do jezera.

2. Vylovíme nový vzorek n_1 ryb a zjistíme počet označených ryb T_1.

3. Určíme podíl označených ryb v jezeře před druhým lovením a v druhém vzorku
$$\frac{n_0}{f} = \frac{T_1}{n_1}.$$

To vede k odhadu f
$$\hat{f} = \frac{n_0 n_1}{T_1}. \tag{6.1}$$

Protože se může stát, že $T_1 = 0$, tj. při druhém lovení nechytíme žádnou označenou rybu, vzorec (6.1) nemusí mít smysl. Proto budeme hledat nějakou modifikaci odhadu $\hat{f}$. Předpokládejme, že

1. Oba výlovy jsou náhodné.

2. Označení ryby nemá vliv na pravděpodobnost, aby ryba byla znovu vylovena.

3. Mezi prvním a druhým výlovem nejsou žádné změny v rybí populaci.

Pak T_1 má hypergeometrické rozdělení:

$$P(T_1 = t_1) = \frac{\binom{n_0}{t_1}\binom{f-n_0}{n_1-t_1}}{\binom{f}{n_1}}.$$

Dá se ukázat (viz [2], str.145), že

$$\mathrm{E}[(T_1+1)^{-1}] = \begin{cases} \frac{f+1}{(n_0+1)(n_1+1)}\left[1 - \frac{(f-n_0)!(f-n_1)!}{(f+1)!(f-n_0-n_1-1)!}\right] & \text{pro} f \geq n_0+n_1+1, \\ \frac{f+1}{(n_0+1)(n_1+1)} & \text{pro} f < n_0+n_1+1. \end{cases}$$

To nás vede k odhadu $\hat{f}^* = \frac{(n_0+1)(n_1+1)}{T_1+1} - 1$, což je jen málo vychýlený odhad f.

6.2 Křivky osvojování znalostí

Uveďme si aplikaci urnových modelů v oblasti matemetické psychologie, konkrétně v teorii učení.

Budeme se zabývat dvěma modely:

- jednoduchý model s vracením,
- jednoduchý akumulovaný model.

Oba tyto modely mohou být reprezentovány urnovými modely. Další reprezentace je pomocí Markovových řetězců, avšak urnové modely mají svou výhodu v jednoduchosti a lepší názornosti.

Jednoduchý model s vracením (lineární model)

V tomto modelu se předpokládá, že některé dispozice jsou nahrazovány jinými. V nejjednodušším případě je jejich celkový počet konstantní, přičemž jsou nahrazovány jedna za jinou.

Představme si náš zkoumaný objekt jako urnu S obsahující černé a bílé míčky. Přičemž černý míček symbolizuje správnou odezvu, naproti tomu bílý chybu. Označme jejich celkový počet m. Protože se jedná o model s vracením, číslo m zůstává pevné a učení je realizováno jako proces nahrazování bílých míčků černými.

Dále předpokládejme, že máme jinou urnu E obsahující nekonečný počet míčků (představující nevyčerpatelný zdroj nových poznatků). Nechť a značí poměr černých míčků v urně E (a je pevné). Je-li $a < 1$, znamená to, že proces učení není dokonalý. Jednoduchý model s vracením spočívá ve zvolení nějakého pevného počtu k míčků, jejich vytažení z urny S a nahrazení stejným počtem míčků z urny E v každém pokusu. Předpokládáme, že výběr z obou uren je náhodný.

Označme B_n počet černých míčků v urně S po n-tém pokusu. Pak

$$\mathrm{E}[B_{n+1}|B_n] = B_n \left(1 - \frac{k}{m}\right) + ka. \tag{6.2}$$

Neboť střední hodnota počtu odebraných černých míčků v $(n+1)$-ním tahu je rovna kB_n/m za podmínky známe-li B_n. Stejně tak střední hodnota přidaných černých míčků je ka.

Přejdeme-li nyní v (6.2) ke střední hodnotě (nepodmíněné)

$$\mathrm{E}B_{n+1} = \mathrm{E}B_n \left(1 - \frac{k}{m}\right) + ka. \tag{6.3}$$

Označme $P_n = \frac{B_n}{m}$ poměr černých míčků v urně S. To bývá interpretováno jako pravděpodobnost správné odezvy v n-tém pokusu. Dále označme $\theta = \frac{k}{m}$. Vydělíme-li nyní obě strany rovnice (6.2) počtem míčků m, dostáváme lineární diferenční rovnici

$$\mathrm{E}P_{n+1} = \mathrm{E}P_n(1-\theta) + \theta a. \tag{6.4}$$

Doplňme si ještě počáteční předpoklady; uvažujme, že na začátku byl v urně S poměr černých míčků b, tj. $P_1 = \mathrm{E}P_1 = b$. Obecné řešení homogenní rovnice

$$\mathrm{E}P_{n+1} = \mathrm{E}P_n(1-\theta)$$

je $\mathrm{E}P_n = K(1-\theta)^n$. Hledáme-li partikulární řešení (6.4) ve tvaru $\mathrm{E}P_n = c$, dostáváme $\mathrm{E}P_n = a$. Řešení (6.4) je tedy tvaru $\mathrm{E}P_n = K(1-\theta)^n + a$. Dosazením počáteční podmínky $\mathrm{E}P_1 = b$ dostáváme

$$\mathrm{E}P_n = a - (a-b)(1-\theta)^{n-1}.$$

Poznamenejme, že při $n \to \infty$, $\mathrm{E}P_n$ konverguje k hodnotě a.

Jednoduchý akumulovaný model

Mějme opět dvě urny S a E se stejnými počátečnímí poměry černých míčků jako v předchozím případu. V tomtu modelu nebudou již žádné míčky z urny S odebírány. Opět máme nějaké pevně zvolené číslo k a při každém pokusu je k míčků z urny E přidáno do urny S.

Označme B_n počet černých míčků v urně S po n-tém pokusu a analogicky A_n počet bílých míčků v urně S po n-tém pokusu. Poznamenejme, že celkový počet míčků v urně S po n-tém pokusu je $B_n + A_n = B_1 + A_1 + (n-1)k$. Stejnou úvahou jako při odvození (6.2) a (6.3)máme

$$\mathrm{E}B_{n+1} = \mathrm{E}B_n + ka.$$

a analogicky

$$\mathrm{E}A_{n+1} = \mathrm{E}A_n + k(1-a).$$

Tedy

$$\mathrm{E}B_{n+1} = \mathrm{E}B_n + (n-1)ka.$$

Označíme-li nyní $P_n = \frac{B_n}{B_n + A_n}$ poměr černých míčků v urně S po n-tém pokusu, pak

$$\mathrm{E}P_n = \frac{B_1 + (n-1)ka}{B_1 + A_1 + (n-1)k}. \tag{6.5}$$

Nyní si uvědomme, že $P_1 = \mathrm{E}P_1 = b = \frac{B_1}{B_1+A_1}$ a označme opět $\theta = \frac{k}{B_1+A_1}$ a po vydělení čitatele a jmenovatele (6.5) výrazem $B_1 + A_1$ máme

$$\mathrm{E}P_n = \frac{b + \theta a(n-1)}{1 + \theta(n-1)} = \frac{b(n-1)^{-1} + \theta a}{(n-1)^{-1} + \theta}.$$

Stejně jako v prvním modelu při $n \to \infty$ je $\mathrm{E}P_n \to a$.

6.3 Hra Sportka

Již v úvodu jsme se zmiňovali o tom, že vznik urnových modelů souvisel s rozvojem hazardních her. Proto v této oblasti mají spoustu aplikací. My se budeme zabývat asi nejznámější hrou, kde se losují míčky s čísly z urny - Sportkou. Nebudeme však počítat pravděpodobnost hlavní výhry, protože tuto pravděpodobnost zná v dnešní době už skoro každý. Budeme se zajímat o to, jestli jsou čísla z osudí tažena skutečně náhodně.

V tomto případě již však nemůžeme použít Pearsonův χ^2 test popsaný v kapitole 2.3, neboť vektor četností tažených čísel nemá multinomické rozdělení. Odvodíme proto test jiný založený na centrální limitní větě. Nejdříve si však uveďme několik důležitých vět, které budeme při odvozování potřebovat.

Věta 2 *Nechť náhodný vektor $\boldsymbol{X}$ má mnohorozměrné normální rozdělení se střední hodnotou $\boldsymbol{\mu}$ a varianční maticí $\boldsymbol{V}$, kde $h(\boldsymbol{V}) = r \geq 1$. Pak náhodná veličina $Y = (\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{V}^{-}(\boldsymbol{X} - \boldsymbol{\mu})$ má rozdělení χ^2_r při libovolné volbě pseudoinverzní matice $\boldsymbol{V}^{-}$.*

Důkaz. Viz [1], str. 68. □

Věta 3 *Je-li $\boldsymbol{A}$ idempotentní matice, pak jednotková matice $\boldsymbol{I}$ je jednou z jejích pseudoinverzních matic.*

Důkaz. Připomeňme, že pseudoinverzní matice $\boldsymbol{A}$ musí podle definice splňovat rovnost $\boldsymbol{A}\boldsymbol{A}\ \ \boldsymbol{A} = \boldsymbol{A}$. V našem případě je $\boldsymbol{A}\ \ = \boldsymbol{I}$. A tedy

$$\boldsymbol{A}\boldsymbol{A}\ \ \boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}\boldsymbol{A} = \boldsymbol{A}^2 = \boldsymbol{A}. \qquad \square$$

Věta 4 *Hodnost idempotentní matice je rovna její stopě.*

Důkaz. Viz [1], str. 321. □

Věta 5 (Lindebergova mnohorozměrná CLV) *Nechť $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ jsou nezávislé stejně rozdělené náhodné vektory se střední hodnotou $\boldsymbol{\mu}$ a varianční maticí $\boldsymbol{V}$ s konečnými prvky. Pak pro $n \to \infty$ platí*

$$\frac{\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n - n\boldsymbol{\mu}}{\sqrt{n}} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V})$$

kde $N(\boldsymbol{0}, \boldsymbol{V})$ značí mnohorozměrné normální rozdělení s nulovou střední hodnotou a varianční maticí $\boldsymbol{V}$.

Důkaz. Viz [6]. □

Věta 6 *Buďte $\boldsymbol{X}, \boldsymbol{X}_n, n \in \mathbb{N}$ reálné náhodné vektory, nechť g je spojitá funkce. Pokud $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$, pak $g(\boldsymbol{X}_n) \xrightarrow{d} g(\boldsymbol{X})$.*

Důkaz. Viz [6]. □

Zformulujme nyní přesně naši úlohu. Budeme chtít otestovat hypotézu H_0, jestli jsou čísla ve hře Sportka tažena rovnoměrně, tj. že při každém jednotlivém slosování (7 čísel ze 49) je pravděpodobnost vylosování každého čísla $1/7$, proti alternativě, že tomu tak není.

Využijeme k tomu archiv losovaných čísel. V následující tabulce jsou uvedeny četnosti vylosovaných čísel ve středečních tazích Sportky v období 29.03.1995 - 28.03.2007. Označme $\boldsymbol{S} = (S_1, \ldots, S_{49})'$ vektor těchto četností.

i	1	2	3	4	5	6	7	8	9	10	11	12
S_i	162	176	167	148	178	200	177	192	178	179	175	195
i	13	14	15	16	17	18	19	20	21	22	23	24
S_i	171	182	176	171	180	209	207	204	173	175	155	172
i	25	26	27	28	29	30	31	32	33	34	35	36
S_i	180	169	176	170	183	175	172	167	180	195	161	190
i	37	38	39	40	41	42	43	44	45	46	47	48
S_i	182	191	191	177	187	154	173	185	156	194	187	187
i	49											
S_i	180											

Tabulka 6.1: Četnosti tažených čísel ve hře Sportka

Nyní již můžeme přikročit ke konstrukci našeho testu. Označme $\boldsymbol{X}^{(n)} = (X_1^{(n)}, \ldots, X_{49}^{(n)})'$, kde

$$X_i^{(n)} = \begin{cases} 1 & \text{pokud bylo v } n\text{-tém slosování taženo číslo } i, \\ 0 & \text{jinak.} \end{cases}$$

Pak zřejmě pro každé n platí

$$P(X_i^{(n)} = 1) = \frac{\binom{48}{6}}{\binom{49}{7}} = \frac{1}{7},$$

$$P(X_i^{(n)} = 1, X_j^{(n)} = 1) = \frac{\binom{47}{5}}{\binom{49}{7}} = \frac{1}{7 \cdot 8}, \qquad i \neq j.$$

Odtud jednoduchým výpočtem dostaneme

$$\mathrm{E}X_i^{(n)} = \frac{1}{7},$$

$$\mathrm{var}X_i^{(n)} = \frac{6}{49},$$

$$\mathrm{cov}(X_i^{(n)}, X_j^{(n)}) = -\frac{1}{49 \cdot 8}, \qquad i \neq j.$$

A tedy

$$\boldsymbol{\mu} = (\frac{1}{7}, \frac{1}{7}, \dots, \frac{1}{7})',$$

$$\mathrm{var}\boldsymbol{X}^{(n)} = \boldsymbol{V} = \frac{1}{49 \cdot 8} \cdot \begin{pmatrix} 48 & -1 & \dots & -1 \\ -1 & 48 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & 48 \end{pmatrix}.$$

Snadno nahlédneme, že matice

$$\boldsymbol{A} = \frac{1}{49} \cdot \begin{pmatrix} 48 & -1 & \dots & -1 \\ -1 & 48 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & 48 \end{pmatrix}$$

je idempotentní. Podle věty 4 je její hodnost je $h(\boldsymbol{A}) = 48$. Z věty 3 vyplývá, že jednotková matice je jednou z jejích pseudoinverzí. Odtud je hned partné, že pseudoinverzí matice $\boldsymbol{V}$ může být například $\boldsymbol{V}^- = 8 \cdot \boldsymbol{I}$. A hodnost matice $\boldsymbol{V}$ je také $h(\boldsymbol{V}) = 48$. Označme dále

$$\tilde{\boldsymbol{X}}_n = \frac{\boldsymbol{X}^{(1)} + \dots + \boldsymbol{X}^{(n)} - n\boldsymbol{\mu}}{\sqrt{n}}.$$

Podle věty 5 při $n \to \infty$ je $\tilde{\boldsymbol{X}}_n \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V})$. Pak ale z vět 2 a 6 plyne že při $n \to \infty$

$$8 \cdot \tilde{\boldsymbol{X}}'_n \tilde{\boldsymbol{X}}_n \xrightarrow{d} \chi^2_{48}.$$

Test založíme na statistice

$$\chi^2 = 8 \cdot \tilde{\boldsymbol{X}}'_n \tilde{\boldsymbol{X}}_n = \frac{8}{n} \sum_{i=1}^{49} \left(S_i - \frac{n}{7}\right)^2.$$

Pokud bude statistika χ^2 větší než kritická hodnota $\chi^2_{48}(\alpha)$, zamítneme hypotézu H_0 asymptoticky na hladině α.

Spočítejme nyní tuto statistiku pro naše hodnoty

$$n = \frac{1}{7} \sum_{i=1}^{49} S_i = 1252,$$

$$\chi^2 = 54,2236.$$

A kritická hodnota $\chi^2_{48}(0,05) = 65,1707$. Čili hypotézu H_0 nemůžeme zamítnout asymptoticky na hladině $\alpha = 0,05$. Pro zajímavost uveďme p-hodnotu tohoto testu, která činí $0,2492$.

Poznámka

Poznamenejme, proč jsme testovali právě data ze středečního tahu Sportky a ne rovnou všechna čísla tažená v historii hry Sportka. Vysvětlení je jednoduché, protože dříve, od roku 1957, kdy se Sportka začala losovat, se v každém slosování losovalo jen 6 čísel. Sedmé losované číslo, číslo dodatkové, bylo zavedeno až v roce 1962.

Proto bychom na tato data nemohli použít výše odvozený test. Avšak použitím mnohorozměrné centrální limitní věty pro nestejně rozdělené náhodné vektroy bychom mohli analogickým postupem odvodit test podobný.

Literatura

[1] Anděl J.: *Základy matematické statistiky*, Matfyzpress, Praha, 2005.

[2] Johnson N. L., Kotz S.: *Distributions in Statistics:Discrete Distributions*, John Wiley & Sons, New York, 1969.

[3] Johnson N. L., Kotz S.: *Two variants of Pólya's urn models*, The american statistician, November 1976, Vol. 30, No.4.

[4] Johnson N. L., Kotz S.: *Urn models and their application*, John Wiley & Sons, New York, 1977.

[5] Lachout P.: *Diskrétní martingaly*(skripta), MFF UK, Praha, 2007.

[6] Lachout P.: *Teorie pravděpodobnosti*(skripta), MFF UK, Praha, 1998.

[7] Matoušek J., Nešetřil J.: *Kapitoly z diskrétní matematiky*, Karolinum, Praha, 2002.

[8] Zvára K., Štěpán J.: *Pravděpodobnost a matematická statistika*, Matfyzpress, Praha, 2002.